3-2022

FERPA Self-Compliant Cloud Storage for Institutions and Professionals in Higher Education

David McCandless

FERPA Self-Compliant Cloud Storage for Institutions and Professionals in Higher Education

by

David L. McCandless

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Information Assurance

March, 2022

Thesis Committee:
Abdullah Abu Hussein, Chairperson
Lynn Collen
Erich Rice
Balasubramanian Kasi

Abstract

FERPA compliance for institutions and professionals in higher education is an important component of any institutional information security strategy. As campuses face an expanding threat environment, they are limited by aging technological infrastructure. Over the coming years, this infrastructure will require replacement, which presents an opportunity to increase FERPA compliance using technological solutions that build on new technology. Using a FERPA self-compliant cloud as a service, this paper proposes a novel approach to compliance and security which can be adapted for use in today's environment while easily growing to incorporate coming technological change.

Acknowledgements

I would like to thank Professors Abdullah Abu Hussein and Hazem Farra, without whom this study would not have been possible. Additionally, I would like to thank my friends and family for their ongoing support of my work.

Table of Contents

Chapter Page

List of Tables

Table Page

List of Figures

Figure Page

Chapter I: Introduction

Introduction

Since 1974, the Family Educational Rights and Privacy Act, or FERPA, has governed rights of post-secondary students and the responsibilities of post-secondary educational institutions as related to the security and privacy of student data (*Legislative History of Major FERPA Provisions,* 2005). Originally passed as a response to the campus unrest and civil rights movements of the 50's and 60's, FERPA signaled an end to the era of In Loco Parentis whereby institutions were seen as a guardian of their students in the absence of parents while the student was away at school. Arguably, FERPA can be traced back to the Freedom Summer efforts by the NAACP and other allied groups, when white students from prominent Northern and Western universities spent the summer helping register and organize Black voters in Mississippi.

As those students returned to campuses in the fall of 1964, they found themselves subject to restrictive speech and protest codes (Burner, 1996). Over the rest of the decade, students engaged in protest and civil disobedience around both Civil Rights and the war in Vietnam that largely dismantled In Loco Parentis and saw the restrictive codes repealed (Lombardi, 1969). An additional outcome of these protests was the objection of these students to policies around parental notification or public release of student records, both academic and disciplinary (Stone, 2002). These objections culminated in the passage of what was largely known as the Buckley Amendment in 1974 but which, once enacted, became FERPA (Lake, n.d.).

While most speech protections apply only to students at public institutions, FERPA applies to any institution which receives federal funding. As the vast majority of student grants in aid and loans comes from the Federal government (*Two Decades of Change in Federal and State Higher Education Funding*, 2019), this effectively makes FERPA a universal requirement for all institutions of Higher Education in the United States. Broadly, FERPA classifies various types of personally identifiable data that institutions retain related to their students and governs the circumstances under which that data may be released to anyone other than the student, including parents and guardians. Institutions may make “directory information” publicly available, which includes data like student’s name, major, and attendance status (though more recently many institutions have kept this data private due to non-FERPA privacy concerns around student safety) (Center, n.d.-a), However, FERPA classifies various categories of student data as private, including grades and disciplinary records, which may only be released with the written consent of the student. The penalty for violations is a potential loss of Federal funding (Hlavac & Easterly, 2015), a substantial sum when considering most institutions’ student aid budgets.

Problem Statement

FERPA was written in an era where student records were almost exclusively paper documents kept in file cabinets and was applied in that environment for the first several decades of its existence. However, starting in the 1990s, student records became increasingly digitized and their transmission increasingly occurs electronically thru email and websites. These changes significantly increase the risk of privacy

breaches which can lead to FERPA violations and represent two separate problems which pose a substantial financial and reputational risk to institutions of higher education: 1) most University employees have limited knowledge of information security practices and principles and thus are at risk of violating FERPA due to inadvertent data security breaches; and, 2) many University employees have limited knowledge of FERPA requirements and practices and thus are at risk of mistakenly releasing protected data.

Nature and Significance of the Problem

It is a truism in information security that humans are the weakest link in the proverbial security chain (Vroom & von Solms, 2004). In enterprise, this is sometimes appended to employees but there is general agreement that the human element is the most significant information security vulnerability. According to (Bissell et al., 2019), the business costs of attacks targeting human resources increased at a greater rate than those targeting systems or technology infrastructure. Even the most sophisticated phishing attacks can require far fewer computing resources than attacks based on cryptography, injection, or other technological prowess.

This dynamic is even more present in higher education. Colleges and Universities are highly distributed human networks with power dynamics that are both diffuse and hierarchical. The vast majority of employees, whether faculty or staff, are subject-matter experts within particular disciplines or service areas but who possess incomplete knowledge about other areas of the organization they nonetheless regularly interact with. Additionally, training around data security practices may be limited or non-

existent, though this dynamic is slowly changing. Students add further complications as they maintain access to many university resources with limited restrictions (McKenzie, 2020). In this environment, the vulnerability to phishing and other social engineering attacks increases (Moramarco, n.d.). From an information security perspective, these factors present significant challenges and institutions of higher education are usually more resource-limited than their corporate counterparts, which leaves them unable to respond as quickly or effectively (Vidwans, n.d.). A number of these sources document that such institutions are frequent targets for attackers and that the number of scope of those attacks is increasing.

Limited knowledge of FERPA is at the core of the second problem detailed previously. Most employees at institutions of higher education have limited knowledge of FERPA's requirements (Turnage, 2007; Maycunich, 2002), and may especially lack knowledge of recent changes to the law. These employees often know some, but not all, of the types of data covered by FERPA but may not be able to look at a document and quickly identify protected data. This problem becomes magnified when they are looking at 10 or 50 or 100 documents. Incomplete knowledge of FERPA requirements then combines with incomplete knowledge of data security risks and best practices to create vulnerabilities which have significant financial consequences for institutions. Fortunately, the same technological advances which pose risks to data security also provide potential protective measures which can simultaneously address both of these

problems and represent a significant opportunity for institutions to reduce both risk and potential harm.

Objective of the Study

In this document, I will propose and test the viability of an FERPA self-compliant cloud application which would include the ability to scan documents to identify and flag potential FERPA violations for users while also providing a means for secure transmission and storage of those sensitive documents.

Study Questions/Hypotheses

The intent is to answer these key research questions:

1) What technical solutions can help institutions and professionals in higher education environments to maintain FERPA compliance?

2) What document-level privacy solutions can help institutions and professionals in higher education environments to maintain FERPA compliance?

3) What cloud storage applications can help institutions and professionals in higher education environments to maintain FERPA compliance?

Definition of Terms

In this proposal, the included terms will be defined within the following parameters:

Cloud Storage: A model of computing where businesses or consumers keep various components of their data, applications, or other network resources stored remotely from their physical location, accessible over the internet.

Document-level privacy: The ability to secure the confidentiality of information contained in an official document which is being transmitted or shared electronically for legitimate business uses.

Encryption-based: An application or service which utilizes one or more cryptographic elements to secure, obscure, or otherwise keep confidential the contents of documents, data, or other information being stored or transmitted.

FERPA-compliant: An application which, through programming, can differentiate, detect, or otherwise denote information which is protected by, related to, or covered by the Family Educational Rights and Privacy Act of 1974 or any of its subsequent amendments.

Institutions of Higher Education: post-secondary institutions in the United State including 2- and 4-year colleges, universities, graduate schools or other public or private institutions which receive funding from the U.S. Federal government.

Professionals in higher education environments: Individuals employed or contracted by an institution of higher education who are engaged in any aspect of their official duties or otherwise acting on behalf of the institution.

Technical solutions: Applications of computing power and/or resources which can be applied to particular issues or problems in a manner that mitigates risk, increases efficiency, and/or otherwise reduces the cost of doing business as a net result of their adoption.

Summary

FERPA compliance is a key issue facing institutions and professionals in higher education environments. The adoption and advancement of technology within these environments has outpaced the general knowledge level of employees in ways that present a significant financial and reputational risk to institutions.

This is compounded by an often-diffuse institutional structure whereby many employees have broad access to various components of student records but may not always have the knowledge or training about specifics of what is or is not covered or protected by FERPA. Encrypted or otherwise secure communications have not been widely adopted by institutions of higher education and many professionals within that environment have limited knowledge of secure communications practices as related to electronic transmission of information using various elements of the Transmission Control Protocol/Internet Protocol stack.

Through the development of an encryption-based, FERPA-compliant cloud storage application for institutions and professionals in higher education, this study will address these problems in a novel way which has significant potential to mitigate risk and add value for these institutions and professionals.

Chapter II: Background and Review of Literature

Introduction

Compliance with the Family Educational Rights and Privacy Act is an important obligation for institutions of higher education. With distributed structures where individual units and personnel often operate with significant autonomy within the larger institution, it can be difficult to monitor, let alone manage, who has access to which data or to comprehensively educate those personnel about their responsibilities as data custodians.

This dynamic plays out in an environment where the storage and transmission of all types of records has transitioned from entirely paper and analog when FERPA was passed in 1974 to almost entirely digital and electronic in the contemporary environment. The same technological advances that have pushed this revolution in record-keeping have provided external actors with new avenues to compromise these systems and their data; avenues which most often rely on compromising the human element within these systems rather than directly attacking the technological element.

This section will review the history of FERPA-related breaches and their causes as well as address some of the technological background, which can lead to heightened risk. Then, it will look at potential solutions to these issues by studying similar applications from other industries, particularly the health care field which has undergone a similar transition from paper to digital record-keeping and has similar requirements for the protection of data.

Background Related to the Problem

Institutions and professionals in higher education are custodians of information that reflects not only their role as educators but often also their roles as s landlord, restaurant, medical clinic, disciplinarian, and entertainer. Often, data about students is maintained in siloes by functional area, with partial aspects of a student's footprint kept by the various functional areas that interact with that student.

This decentralization of record-keeping and decision-making is one of the key contributing factors which makes FERPA compliance uniquely challenging. The dynamic often leaves individual staff or small units with access to types and amounts of data which significantly outstrips their relative levels of responsibility.

This is compounded by a general lack of knowledge of best practices related to data security. Electronic mail is the currency of communication in the modern office environment and this is certainly true of institutions of higher education. Further, because of the decentralized nature of organizational structures within higher education, it is far more common that individual employees will communicate with others outside their immediate department or work area and that such communication will often be over e-mail. While institutions have taken great steps to increase the security of their campus e-mail systems, there are significant vulnerabilities built into e-mail communication which have been difficult to mitigate across all industries.

To recap, higher education institutions are distributed systems where individuals have access to significant amounts of data without complete (or often any) training about their responsibilities as data custodians. Communication within this environment

is necessarily electronic and thus subject to compromise by external parties. General knowledge of secure electronic data practices and risks among higher education professionals is limited. All of these factors lead to a dynamic of heightened reputational and financial risk to institutions related to their obligations under FERPA.

Literature Related to the Problem

Institutions have been somewhat protected by the courts over the course of jurisprudence related to FERPA. In Gonzaga v. Doe (2002), the United States Supreme Court held that FERPA failed to confer an enforceable individual right to non-disclosure of student data. This finding strictly limits the rights of individual students or groups of students to sue institutions for damages related to FERPA breaches. This ruling held that enforcement of FERPA could only be carried out by the appropriate office within the federal Department of Education and that sanctions within that process could include revocation of Federal funds (Rehnquist, 2002). This ruling is interpreted as providing protection against sanctions for individual violations of FERPA and rather tying financial penalties to cases that represent systemic breaches.

Further clarification was provided in that same year by the Court's opinion in Owasso v. Falvo (2002). In a case brought by the parent of an elementary student, the court ruled that peer grading did not represent a violation of FERPA. The opinion in Falvo lays out some additional protections for institutions: In order to be covered by FERPA, records must be institutionally maintained and that peer grading of homework was part of the educational process and did not represent a formal educational record in the sense that it would be covered by FERPA. This ruling is interpreted as clarifying that

FERPA applies to formal records maintained centrally by an institution rather than incidental records as granular as individual student assignments. While a student's final grade in a course would be covered by FERPA, the assignments which comprise that grade would not necessarily be covered (*Owasso Independent School Dist. No. I—011v. Falvo (Opinion of the Court)*, 2002).

The totality of the jurisprudence beyond these cases does present some interesting inconsistencies of which institutions should be aware (Center, 2020). In Board of Ed. Of ISD 92 of Pottawatomie Cty. V. Earls (2002), the Supreme Court ruled that students voluntarily participating in extracurricular activities have a limited expectation of privacy as related to their participation in those activities (*Board of Ed. Of Independent School Dist.no. 92 of Pottawatomie Cty. V. Earls (Syllabus)*, 2002). Individuals such as student athletes may expect the institution to release details such as their height, weight, and roster status as part of their participation.

In United States v. Miami Univ (2002), the 6th Circuit Court of Appeals ruled that student disciplinary records are education records and may be redacted in order to protect the privacy of individuals (*Forester*, 2002). However, in Bauer v. Kincaid (1991), the U.S. District Court for the Western District of Missouri ruled that campus security reports and the results of criminal investigations were not education records and thus not covered by FERPA (*BAUER v. KINCAID | 759 F.Supp. 575 (1991) | upp57511244 | Leagle.com*, 1991). Much of the recent public and legal discourse around FERPA has been focused on this area of student disciplinary records and the extent to which they may or may not be disclosed. Ongoing litigation involving the University of North

Carolina (Neil, 2020), the University of Kentucky (Childress, 2020b), and Syracuse University (Darnell, 2020) has the potential to provide more clarity to these ongoing questions around student disciplinary records as these cases work their way through the court system. Generally, these cases all represent the legal intersection between various states' open records laws as applied under the jurisdiction FERPA and the specifics of what institutions may and must release or protect. The North Carolina and Kentucky cases were recently argued before the respective state Supreme Courts, and the North Carolina case is expected to be appealed to the U.S. Supreme Court in the near future (Friedman, 2020).

While these court cases, both past and ongoing, have shaped and are shaping institutions' understanding of the scope and reach of FERPA, there have been a number of recent situations where institutions and professionals in higher education have been responsible for violating FERPA restrictions. These violations represent a number of different types of situations but share the common thread addressed in this paper of the joint problems of limited knowledge of FERPA and limited knowledge of data security practices among professionals and others working in higher education.

The first subset of incidents can be broadly grouped as related to the ongoing COVID-19 pandemic and the competing responsibilities of institutions to FERPA versus public health directives, as well as the additional FERPA complications arising from remote learning. Institutions may and have released generalized data about case counts but are required to protect data related to an individual student's status (Gross, 2020). However, in August the University of Kentucky was found to have left a publicly

accessible Sharepoint file that contained the names, birthdates, and negative COVID test results of students as well as employees (Childress, 2020-a). While the error was discovered quickly, the information was accessible to anyone with university credentials over the course of an entire weekend.

The questions and concerns related to remote learning are varied and range from whether professors can require cameras to be turned on for zoom course meetings (Blackman, 2020) to the complications from recording or taking screenshots of these meetings (Wan, 2020). As institutions turn to 3rd party applications to manage various aspects of remote learning, they need to be very proactive in their privacy management practices to ensure they do not inadvertently violate FERPA through the release of protected data to these 3rd parties (St. Amour, 2020). At the University of Texas at Dallas, students expressed concern about the exam proctoring software Honorlock and its data retention and sharing practices (Hidalgo, 2020).

However, the Coronavirus pandemic is far from the only driver of FERPA violations. A number of other recent incidents show the scope of compliance problems institutions face. In some cases, a lack of care in developing routine technological tools has exposed protected data. This was the case with Indiana University, which in February 2020 was notified that an online GPA calculator tool for students had inadvertently exposed the “grades of more than 100,000 current and former students” (Fernando, 2020). The University pointed to an incomplete internal review process which inadvertently caused the page to be made public.

Also in February 2020, a Teaching Assistant at Stanford University was removed from his post after using a course enrollment list to recruit for a private company (Srivastava, 2020), which represented an improper release of protected student data to an external entity. Also at Stanford, several classes have transitioned from Piazza, a 3rd party course discussion tool after becoming aware of potential concerns around outside employers using features of the tool to recruit students in violation of FERPA's guidelines (Tsai, 2020).

In October 2019, personally identifiable information of students at Southeast Missouri State University, including GPAs and Student ID numbers, was inadvertently exposed when several excel files were attached to an email that was sent to approximately 50 students at the institution (Tate & Stuerman, 2019).

These recent examples all document a similar dynamic that demonstrates the second part of the problem statement: 1) sensitive data is accessible to a wide variety of University employees from graduate students to office managers to professors, as well as adjunct instructors; and, 2) these employees have received varying degrees of training (Turnage, 2007; Maycunich, 2002) and experience inconsistent institutional expectations (Coulture et al., 2018) around FERPA compliance.

There is also significant literature describing the first part of the problem statement. Institutions of higher education are facing an uptick in spear-phishing attacks as the COVID-19 pandemic has emboldened potential attackers and provided opportunities to conduct scams with COVID as a cover (Coker, 2020). With the expansion of the "attack surface" to include home and personal devices (Jay, 2020)

there are also more endpoints to protect. According to Check Point research, “COVID-19 related phishing and malware attacks increased dramatically from under 5,000 per week in February to over 200,000 per week in late April” (Check Point Software Technologies, 2020).

This has included an uptick in phishing-enabled ransomware attacks using the NetWalker software, which occurred against Michigan State University, the University of California, San Francisco, and Columbia College Chicago early in the summer (McKenzie, 2020) as well as 89 documented attacks against “universities, colleges, and school districts” (Emisoft, 2019) in 2019. Institutions of higher education continue to be targeted. The University of Utah paid $457,000 to resolve a ransomware attack in July 2020 involving the College of Social and Behavioral Science (Sussman, 2020).

University Administrators are aware of the risk but often struggle to get buy in from their dispersed workforces as well as distracted students. Michael Tran Duff, the Chief Information Security Officer at Stanford University, is quoted as saying “We recognize phishing as the single greatest threat to privacy and security today” (Zalaznick, 2020). Institutions have gone so far as to phish their own students and faculty but are focused on automating email systems to detect phishing attacks before they make it to inboxes.

Stopping messages has not always been possible. St. Louis Community College was impacted in January by a phishing scam which compromised the personal information of more than 5,000 students and employees (Wood, 2020). In June 2020, thousands of students at Iowa State, Harvard, and Stanford received racist messages

from a compromised system at Equity Prime Mortgages (Miller, 2020b). Additionally, in September 2020 eight faculty email accounts at Sacramento State University were compromised by scammers following a phishing attack, and at least one of those emails was used to send further phishing messages to others (Robison, 2020).

Not only do these attacks represent reputational and financial risk on their own, but those risks are also compounded by potential FERPA violations. While courts have generally held that institutions must systemically violate FERPA in order to be financially penalized, it is incumbent upon institutions to avoid putting themselves and their students' data, at risk.

Literature Related to the Methodology

FERPA is unique to the education industry but challenges related to data privacy and secure data practices are experienced across industries. The most similar analogue for FERPA is likely HIPAA, the Health Insurance Portability and Accountability Act of 1996, which covers patient privacy and data practices within the Health Care industry. One of the primary differences between the two laws is that FERPA was developed prior to the widespread adoption of electronic recordkeeping and transmission while HIPAA was developed specifically to spur the adoption of electronic recordkeeping and transmission (Security, 2019). Therefore, HIPPA goes into significantly more detail about processes and procedures while FERPA is largely limited to which types of data can be disclosed, to whom, and under which circumstances. However, the laws are similar enough to allow for cross-applicability of solutions.

One area where HIPAA compliance outpaces FERPA compliance is that education of applicable staff about HIPAA standards is much more clear and consistent than with FERPA (Cannon & Caldwell, 2016). Because many health care professionals must receive professional licensing, they must demonstrate their knowledge of HIPAA as a pre-requisite to entering the profession (Agris & Spandorfer, 2016).

As far back as 2013, cloud platforms were being proposed as a solution to what was still a lagging effort across the health care industry to digitize records (Gerard et al., 2013). Early efforts by major technology companies failed to gain traction or users. Google Health was launched in 2008 but shuttered in January 2012 after failing to gain widespread adoption (Mearian, 2011), while Microsoft HealthVault lasted longer but ultimately shut down in November 2019 (Truong, 2019). These were consumer-facing products that failed to get buy-in from health care providers or consumers, with all parties concerned about privacy and desiring to keep their sensitive health information as closely held as possible. However, both of these companies, as well as Amazon and Apple, have made significant entries into the enterprise health care market by offering solutions for providers, insurers, and other major financial players in the health care market. With Healthcare comprising 17.7 percent of US GDP as of 2018 (Hartman et al., 2020), the financial stakes and rewards are quite high.

In 2019, Microsoft launched Azure API for FHIR (Fast Healthcare Interoperability Resource) which integrates the open-source FHIR platform specification into the Azure Cloud environment as a means of allowing healthcare providers and businesses to integrate their existing records systems into the cloud (Cartwright, 2020). Apple

announced last November a partnership with the Veteran's Administration to integrate the VA's Health API into Apple's Health Records platform, opening access to more than 9 million veterans (Parmar, 2019). AWS, in 2019, announced a partnership with medical records pioneer Cerner to incorporate artificial intelligence and predictive modeling as part of their larger service offerings within the healthcare cloud ecosystem (Landi, 2019). Google launched its own cloud healthcare API in April 2020 (Wiggers, 2020) and announced in October 2020 a 10-year partnership with the Mayo Clinic to develop and train machine learning tools to fight cancer (Pearson, 2020).

These developments point to a key technological advantage the Healthcare industry has versus higher education: the development of shared programming infrastructures to enable the secure transmission and storage of data; FHIR is a key component of this. Developed beginning in 2012, FHIR is a set of Application Programming Interfaces (APIs) which use the REST approach to provide easy implementation within a secure environment (*What Is FHIR?*, n.d.). While Google Cloud has developed its own healthcare API, FHIR is currently used by Apple, Microsoft, AWS, Epic, and Cerner, which comprise some of the largest entities in the medical records field.

In contrast, there is very little standardization in records management among higher education institutions. As of January 2020, an EDUCAUSE study showed that only 13 percent of colleges and universities in the U.S. were actively engaged in systematically digitizing their student records (Grajek, 2020). While data is difficult to come by, a report by EDUCAUSE identifies eight institutions with separate records

management policies, procedures, and systems and offers resources for a theoretical CISO to present on their own campus with a very earnest “Good Luck!” (*Electronic Records Management Toolkit*, 2020). A 2017 report by the AACRAO (American Association of Collegiate Registrars and Admissions Officers) surveyed respondents at 1045 institutions, including 988 within the U.S., and found broad inconsistencies in the types of records considered for retention or digitization, the institutional owner of those records, retention policies, and even the format in which records were to be retained (*Student Records Management Practice*, 2017).

This dynamic points to an opportunity for a novel approach to records management as it relates to FERPA. Institutions of Higher Education have largely not adopted cloud storage technology and largely maintain their records on legacy database systems while investment in infrastructure (technology or otherwise) has been limited (Riddell, 2016). Such limited budgets have led many institutions to continue investing in extending the life of these legacy systems rather than developing new systems (Berman, 2019), but it is likely that institutions will need to make significant investments in upgrading their Student Information Systems infrastructure sooner rather than later (Miller, 2020a) as Minnesota State is currently doing with the NextGen development process.

This points to an emerging market for technological solutions that can provide portability, access, security, privacy, and data analytics within a streamlined framework that is accessible to both students and institutions (Berman, 2019; Miller, 2020a). New applications which integrate modern software models using APIs will gain significant

market share in the coming years as more and more legacy systems reach their end of life and can no longer be patched into the modern computing environment. The IMS Global Learning Consortium is in the process of developing EDU-API as a “standard model for the integration of core education enterprise data” (*EDU-API | IMS Global Learning Consortium*, n.d.). Built on the foundation of the Learning Information Services specification and working in concert with Brigham Young University’s PERSONS API, EDU-API is an early entrant in what will likely be a crowded field of next generation software intended to modernize Student Information Systems.

The California State University System is additionally developing an API to simplify the integration of their Student Information Systems with the PeopleSoft system on their 23 campuses, with expected implementation by 2021 (Berman, 2019). BYU sought a standardized API for their campus after discovering the existing technology left them using up to 25 different APIs to integrate their SIS with various mobile apps (Raths, 2017). Other early adopters include Northwestern University, UC Berkeley, the University of Michigan, the University of Washington, and Yale, while dozens of additional institutions are in various stages of development (Lane, n.d.).

These developments are still largely FERPA neutral. While they will all incorporate robust security and privacy features, the ability to quickly flag and identify FERPA-related data within an electronic environment would be a key development with significant potential to add value to these or other efforts. A new paradigm of data management is long overdue as institutions fully adapt FERPA into the digital environment (Kellen, 2019).

A software package that incorporates a combination of text classification and/or keyword extraction methods which could examine documents while flagging and labeling potential FERPA-related information is a methodology that could be applied to these joint problems. There are a wide variety of text classification methods which perform with varying degrees of success depending on how they are deployed (Hartmann et al., 2018). These methods can be tuned in scope to classify full documents or can be adjusted to the paragraph, sentence, or sub-sentence level and generally work by applying a four-step process of 1) feature extraction, 2) dimensionality reduction, 3) classification, and 4) Evaluation (Kowsari et al., 2019).

Step 3, Classification, is the most important step (Kowsari et al., 2019) and has the widest parameters in terms of the types of decision-making that can be applied. These types include Rocchio, ensemble-based, logistic regression, Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), decision tree, random forest, and conditional random fields (CRFs) (Kowsari et al., 2019). It is likely that a model or tool which uses a composite approach may be most effective and that application of multiple approaches in a controlled testing environment will ultimately be needed to determine the most effective path forward.

An additional step to this process is clustering. Allahyari et al., (2017) point to a number of different applications of clustering including context-based retrieval systems and discuss software tools such as Lemur and BOW which implement common clustering algorithms. Other methods include hierarchical clustering, where clusters are grouped by characteristics and then slowly merged; k-means clustering, where a

partitioning algorithm groups like content; and probabilistic clustering, where topics and themes are extracted and grouped.

In addition to clustering, several other techniques providing common means to classify output data. These include tokenization (or segmentation), the process of breaking down each word into characters or subwords in order to provide more clarity in frequency counts by ensuring that grammatical factors such as word tense don't cause the classifier to miss similar words or meanings (Pai, 2020). Additional techniques include part-of-speech tagging and named entity extraction, which attempt to parse and classify words by their part of speech (noun, verb, etc.) and to classify proper nouns within text (Li, 2018). Other classification methods include chunking, the process of extracting phrases from unstructured text (Bachani, 2020), and parsing, the process of breaking a sentence down into its component parts (Gupta, 2020). Broadly, these techniques enable classification to work more effectively by providing additional context beyond only the text and allowing the algorithm to assess multiple layers of meaning.

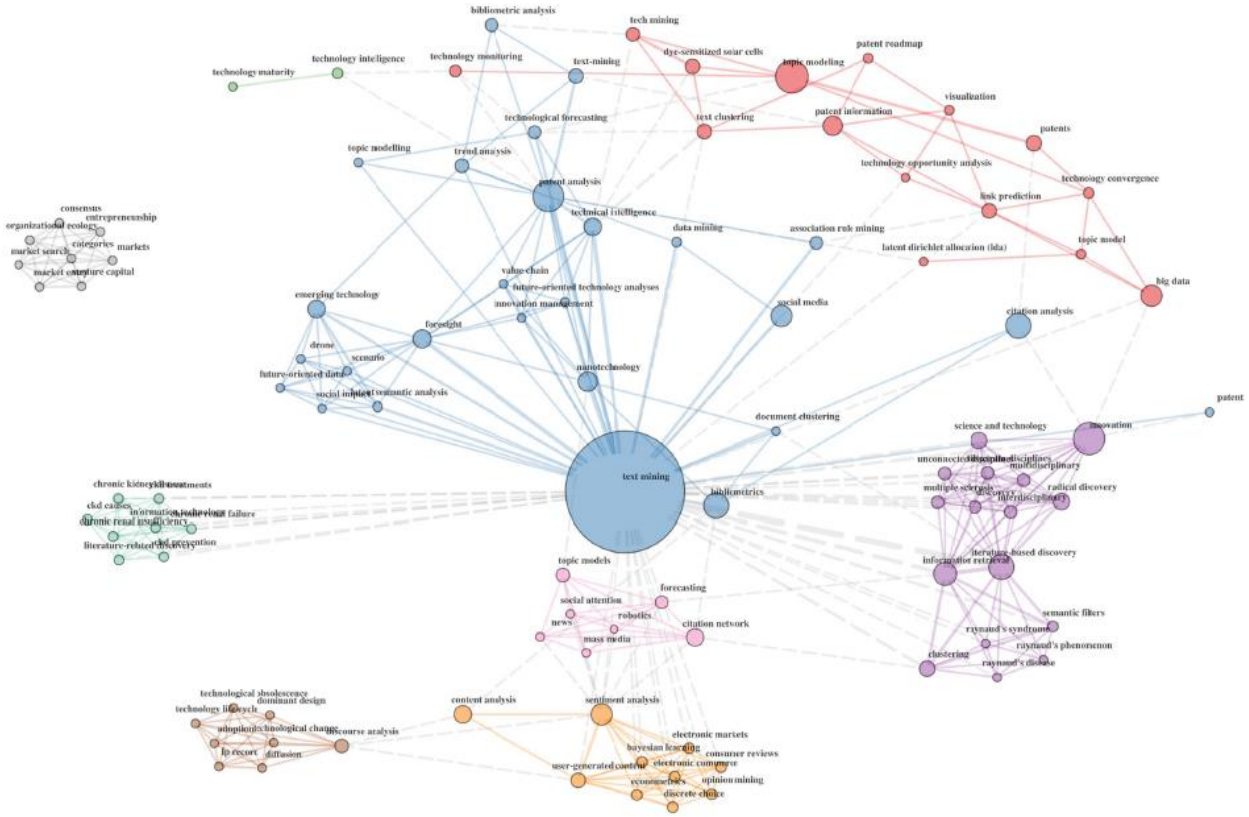
This additional context enables several different kinds of analysis to occur. One of the most common is sentiment analysis, where the algorithm classifies blocks of text such as user reviews as positive or negative or otherwise determines the general sentiment of the text snippet (Branavan et al., 2009). Keyword or entity extraction is another type of analysis where the algorithm identifies keywords or other patterns within the text that meet the specifications input by the programmer (Arras et al., 2017).

Antons et al., (2020) provide an excellent network diagram showing the various applications of text and data mining mapped into clusters. They found that while there

are a broad array of applications for text mining there are significant additional applications that need further study and development.

Figure 1

Antons, et al's Cluster Map of Text Mining Applications



There are a number of existing approaches to text classification which might be suitable to be adapted to this type of project. The most notable of these are likely the existing tools used by Amazon Web Services and Microsoft Azure. SageMaker Blazing Text algorithm is one of Amazon's in-house utilities and provides optimized implementations of the Word2vec and text classification algorithms (*BlazingText*

Algorithm - Amazon SageMaker, n.d.). According to Amazon, Blazing Text provides both unsupervised and supervised learning approaches to text classification and is one of several machine learning tools made available by Amazon Web Services (Lasseter, 2020).

Similarly, Microsoft offers the Azure Text Analytics API. Part of Azure Cognitive Services (Hill, 2020), this API offers sentiment analysis as well as key phrase extraction, language detection, and other features. More broadly, Microsoft offers more than a dozen text analytics models within the Azure Machine Learning Studio. Many of these are based on iterations of the Vowpal Wabbit Model, an open-source machine learning library developed in part by Microsoft Research.

In addition to many proprietary and open-source tools, another major player is Google Cloud NLP. This NLP include both the AutoML Natural Language as well as the Natural Language API developed by Google. AutoML is Google's more advance machine-learning model which can conduct classification, entity extraction, and sentiment analysis at a more advanced level than their Cloud Natural Language API which is most effective for sentiment analysis.

A final step in managing any machine learning application is an algorithm audit. This process includes a number of important steps to ensure the model is functioning correctly "in the wild" (Clark, 2017). These steps include making sure the model effectively understands its role in the context of the business use case, streamlining inputs to eliminate data errors, and ensuring the model is making correct decisions

based on the data. All these steps ensure that the model is still functioning effectively as it has shifted from the training and testing environments to the deployed environment.

With an effective text classification algorithm which can reliably flag FERPA-related data, the next step in expanding usability is to pair the algorithm with an encryption-secured cloud storage or transfer service. Most higher education professionals do not have easy or simple means to transfer sensitive data. As documented earlier in this report, sensitive files are often attached to unsecure email messages or otherwise exposed in ways which compromise student data.

A secure cloud storage application would allow the transfer of data between colleagues while maintaining the security of the data in a manner that is much more effective than email, especially in an era when email systems are so often targeted and easily compromised. In this way, such an application could function as a Data Loss Prevention tool to be incorporated into Student Information Systems technologies and applications, as well as effectively be incorporated into OneDrive or Google Drive for use among education industry consumers as an additional means of securing sensitive data.

Summary

FERPA was developed in a very different technological era, and institutions and professionals in higher education environments are largely underprepared to secure student data due to several factors. These include diffuse organizational structures which allow for broad access to data by employees who work largely within their own departments or units with significant autonomy; inconsistent training about FERPA

requirements; a lack of centralized decision-making and control over record-keeping; and a general lack of knowledge about data security practices.

These factors are compounded by institutions largely under-investing in data management and security, leading to significant amounts of FERPA-protected data stored in aging legacy systems which themselves were developed in a very different technological era. The health care field has seen overwhelming investment in data management platforms and practices over the last decade and it is expected that such a revolution will be coming to higher education over the next decade.

An opportunity exists to leverage advances in data management platforms and practices from health care and other industries to engage in the transition of Student Information Systems by developing a tool which can identify protected information within a document and flag it for the user while also providing a secure means of storage and transmission for that document.

Chapter III: Methodology

Introduction

An opportunity exists within the field of higher education to develop and implement a tool which can assist institutions and professionals in higher education with FERPA compliance and general data privacy by scanning documents and flagging potentially protected data. Future versions may also allow the secure transmission and storage of those documents using an encryption-based cloud storage application or using an API to transmit those documents to existing cloud applications such as Microsoft OneDrive or Google Drive. This section will discuss the development of such a tool and its potential applications.

Design of the Study

Broadly speaking, this study will involve two key steps:

(I) The development or adaptation, testing, and training of a text classification or extraction algorithm which can detect within text-based documents potential pieces of data which may be protected by or related to FERPA within a higher education environment.

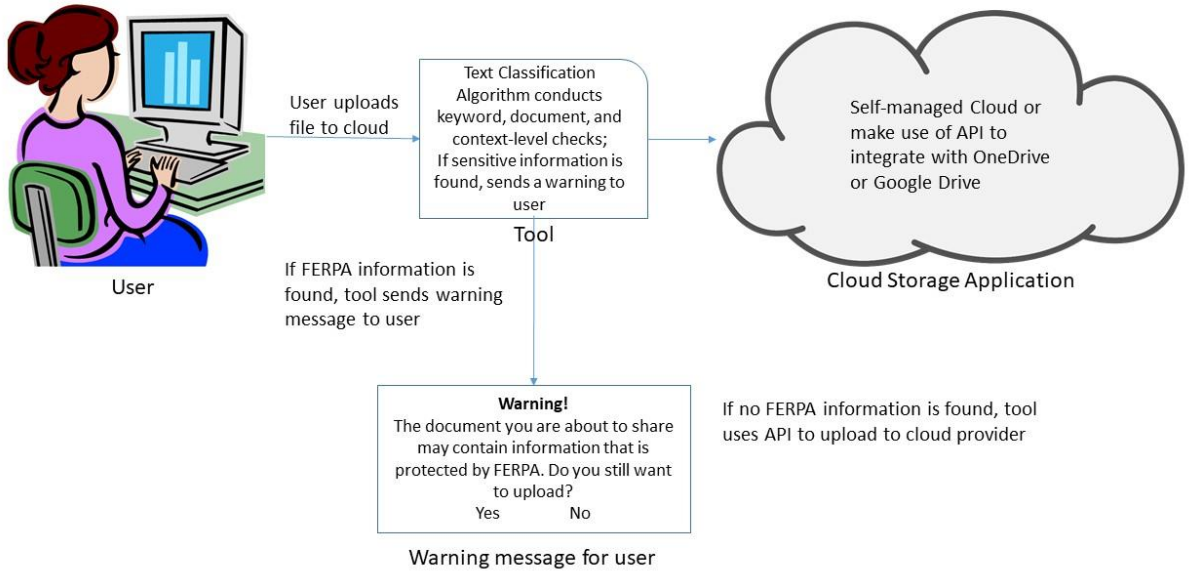
(II) The development or adaptation of an API which would allow the application of this algorithm within an encryption-based cloud storage as a service application such as OneDrive or Google Drive.

The overall goal for this initial study will be to develop an algorithm with > .75 accuracy in detecting FERPA-covered data with the eventual outcome through further development, testing, and training of > .95 or > .99 accuracy. Many algorithms initially

have lower accuracy but initial success proves the concept, and further refinement makes them operable. In order to be viable as a product, it is likely this tool would need to reach > .95 accuracy consistently, which is the basis for selecting those parameters.

Figure 2

Data Flow for Proposed Tool



In figure 1, the process is outlined visually. A user uploads a file, document, or other text-based electronic information and the tool scans the document and identifies potentially protected data. It flags that data and provides a report to the user that would be similar to using a spellchecker. Then, if they so choose, the user may upload the

document to an encrypted cloud storage application where it can be accessed by other authorized users with a key or password.

In this way, colleagues can share documents through a format that is much more secure than email, which is the standard industry tool. With most institutions using a siloed server system, individuals in different departments or units often are unable to access the server resources of other units. With this tool, not only would users have the ability to know quickly and reliably whether a document contains potentially protected data but would also be provided with a secure method of storing and sharing data with colleagues.

An important step in the development of the tool is a clear catalog and understanding of the specific types of data that are covered by FERPA. The statute (Family Educational Rights and Privacy Act, n.d.) declares that Education Records are covered by FERPA and defines those as “records that are directly related to a student and maintained by an educational agency or institution or by a party acting for the agency or institution” (pp. 5). The statute also excludes specific records such as records maintained by a law enforcement unit within the institution, records that exclusively relate to an individual in their capacity as an employee, and records maintained by health care or other “recognized professional or paraprofessional.”

Within these restrictions, FERPA allows for the limited release of “directory information” (*Directory Information | Protecting Student Privacy*, n.d.) which comprises a subset of all the student records that might be maintained by the institution. Students may elect to restrict the disclosure of their information. The statute (Family Educational

Rights and Privacy Act, n.d.) defines directory information specifically as “information contained in an education record of a student that would not generally be considered harmful or an invasion of privacy if disclosed” (pp. 4).

In the context of FERPA’s passage in 1973, it may have made sense to allow this information to remain public. However, many institutions have since decided to restrict public disclosure of much or all of this information with notable exceptions for varsity student athletes and other students who may be public figures. From a privacy standpoint, the tool will be as restrictive as possible because an individual user may not know which students have restrictions on their directory information or which information the institution has chosen not to release as a matter of policy.

The listing of directory information does provide a helpful start in designing and programming the text classification elements of the tool, however, by enumerating specific types of data for inclusion. A limitation of the tool is that there may be new types or classifications of protected data that arise after implementation. Another limitation of the tool is that non-text data such as images and biometrics will not be classifiable. However, the ability to categorize and flag text is still sufficient to be a highly useful and valuable tool. For the purposes of this study the types of information that will likely need to be included in the tool are:

Table 1

Types of Records to Be Recognized by the Tool

Text-Based Student Records Covered by FERPA	
Directory Information	**Protected Information (Non-Directory Records)**
Student Name	Social Security Number
Student Email	Student ID Number/PIN
Major/Field of Study	Grades/Academic Records (except Peer-graded)
Dates of Attendance	Disciplinary Records
Degrees, Honors, and Awards Received	Names of Parents or Family Members
Local Address	Home Address
Grade Level	Financial Aid/Financial Records
Participation in Recognized Activities	Other Records that Include Personally Identifiable Information
Most Recent Educational Institution Attended	
Telephone Number	
Date and Place of Birth	
Enrollment Status	
Height and Weight of Athletic Team Members	

Much of the rest of this data should be recognizable for Natural Language Processing or text classification algorithms using keyword level and document level checks and specifically named entity extraction and keyword/key phrase extraction. Named entity recognition is an subfield of Natural Language Processing (NLP), a field of study which uses machine learning and computational linguistics to manipulate speech and text data (Brownlee, 2017). Named entity recognition determines the “parts of a text

that can be identified categorized into present groups” such as names of people and places (Garbade, 2018) using the various processes of tokenization, stemming, and chunking (Li, 2018) discussed earlier.

Keyword and key phrase extraction is an additional application of NLP and is typically a 2 step process: first, a set of words and phrases are extracted from the document; second, those words and phrases are ranked to determine relevance (DeWilde, 2014) with the use of one or more statistical approaches or algorithms (*Keyword Extraction*, 2020). The two approaches to keyword extraction are supervised and unsupervised methods (Chaudhary, 2020). Unsupervised keyword extraction does not need a training phase and are useful for determining themes and “metadata for indexing and tagging documents (Chaudhary, 2020).” Supervised keyword extraction would be most suitable for this project due to the need to identify specific types of data. This process involves training an algorithm using test documents with pre-identified key words and phrases, which the model will then learn and apply to future documents (Raskar & Pathan, 2014).

Fortunately, there are powerful existing NLP applications which can be brought to bear on this problem. Microsoft Azure Text Analytics API 3.1 release recognizes a variety of named entity types related to Personally Identifiable Information, including names, telephone numbers, email addresses, birthdates, mailing addresses, organization names, dates, and social security numbers (Hill & Yeo, 2021a). That library is an important component which can be built into the tool.

The second key component to build into the tool is keyword and key phrase extraction. Information such as degrees, majors, enrollment status, student ID numbers, and other data can be programmed into the Microsoft API's key phrase extraction feature and then trained to recognize words and phrases related to the field of higher education and specifically FERPA.

Building the keyword/key phrase dictionary will be a key component of the project. Some information will be relatively easy to compile, such as degrees, majors, and enrollment status, and grades. A decision point will be how broadly to include possible inputs. Different institutions offer a broad range of degrees and programs of study and in some cases use different terminology to describe them. An initial version of the tool may include a limited subset of variations as a proof of concept; the likely choice of subset are those majors, programs, and degrees found in the Minnesota State University System. This is also a challenge related to student ID numbers as different institutions use different conventions and configurations to determine these identifiers.

Another category to consider are those records which may ultimately be difficult to classify. Student medical records, student disciplinary records, and financial aid records are all covered by FERPA but may be excluded from the tool initially. A rationale for this is that far fewer employees tend to have access to those records and those that do tend to have more training and awareness around privacy concerns. Additionally, height and weight will likely not be included as those records are only kept for athletics participants outside of a medical context, and athletics participants have agreed to the release of that information.

Based on these considerations, the proposed tool will address and flag the following types of records using the following methods.

Table 2

Method of Classification by Record Type

Microsoft Azure Text Analytics PII NLP	**Keyword/Key phrase Extractor Using Minnesota State University Classifications**	**Categories Excluded from the Tool**
Student Name	Major/Field of Study	Disciplinary Records
Student Email	Student ID Number/PIN	Medical Records
Social Security Number	Grades/Academic Records (except Peer-graded)	Height and Weight of Athletic Team Members
Local Address	Degrees, Honors, and Awards Received	Financial Aid Records
Home Address	Participation in Recognized Activities *	
Telephone Number	Most Recent Educational Institution Attended *	
Date of Birth	Grade Level	
Place of Birth	Enrollment Status	
Dates of Attendance		
Names of Parents or Family Members		
Participation in Recognized Activities *		
Most Recent Educational Institution Attended *		
Other Records that Include Personally Identifiable Information		

* Some records appear in multiple categories as both methods may be used

Data Collection

One of the key challenges involved in this project will be assembling a document library/test corpus to train the keyword extraction model. It should be possible to build out a library of majors, degrees, programs of study, and activities by canvassing the websites and registrar's offices of the 7 4-year institutions in the Minnesota State system to determine a complete list of degree programs, enrollment statuses, recognized activities, and other categories that are offered by those institutions. It will not be possible to build a complete list of possible honors or awards but a assembling a reasonably robust list will be an achievable outcome. One challenge will be the development of a training corpus for the keyword extractor. In order to complete the study, a collection of de-identified documents will need to be found or assembled to effectively train and test the algorithm.

Tools and Techniques

In developing a proof-of-concept tool, I utilized the Microsoft Visual Studio and Windows Forms Applications to develop a working prototype which, using several libraries as well as Regular Expression capability to detect and flag certain types of protected information in Microsoft Word documents.

The Microsoft Office Interoperability package for Word was imported to allow the tool to access and parse Word documents. Interoperability packages are a feature of C# that allow for ease of access to Office API objects to simplify programming and allow better integration between C# and Office (Wagner, 2015). This allows us to get the file content of a Word document, in this instance, and use that content as the input for tool

by using elements of Visual Studio and C# to convert the information within the Word document to a string which can be read compatibly by other programming elements within the environment.

Additionally, the tool takes advantage of Regular Expression code by importing the Regular Expressions Text package. Regular expressions are patterns that can be matched against input text (C# - Regular Expressions, 2021) and allow for the definition of specific patterns within the data to be matched against an input string or file. In the course of developing the code, various patterns can be defined such as social security numbers, dates, and other types of data in addition to string literals or specific words or phrases.

The tool functions by having a user drop a Word file into the Windows Forms application. The tool then identifies particular subsets of protected information, including Social Security Numbers, Minnesota State Tech ID numbers, and dates, and flags them as matches for the user to review. The user can then assess whether the file needs to be redacted or otherwise altered before being shared with others. This type of functionality is user-friendly and increases chances of adoption by using an informational rather than directive approach to managing sensitive or protected data.

Summary

The outcome of the study is the development of a tool which will use Regular Expression pattern matching to detect certain subsets of personally identifiable information within a Microsoft Word file as a test case for the proof of concept. The second outcome will be the development of an Application Programming Interface

which will allow the tool to interact and communicate with a cloud storage application, in this case Microsoft OneDrive.

In the process of narrowing the scope of the study, some categories of FERPA protected data will be left out. Non-text data is not suitable for detection by text classification or NLP. Additionally, financial records, disciplinary records, and medical records will be left out of the initial study as access to those types of records tends to be strictly limited already. The fully developed tool will ideally be able to delineate between directory information and fully protected information and include that in the warning to the user when potential violations are detected. However, for the proof of concept prototype this functionality will not be included.

The scope of the proof-of-concept tool will be limited to detecting Social Security Numbers, Minnesota State Tech ID numbers, and dates. Ultimately, the tool will provide institutions and professionals in higher education with a secure method to detect and classify potential FERPA violations in Word documents prior to storing or sharing those documents.

Chapter IV: Data Presentation and Analysis

Introduction

In order to develop a working prototype, the scope of the tool was further narrowed to a proof-of-concept application. Working in Visual Studio, using the Windows Forms Application, a tool was developed that can accurately identify certain subsets of personally identifiable information and flag them for review in a user-friendly format. While not a fully operational tool to identify all potential FERPA violations, this tool works as designed and works with accuracy in a way that shows the concept of detecting FERPA violations to be valid and workable in a manner that can be expanded upon in future applications.

Data Presentation

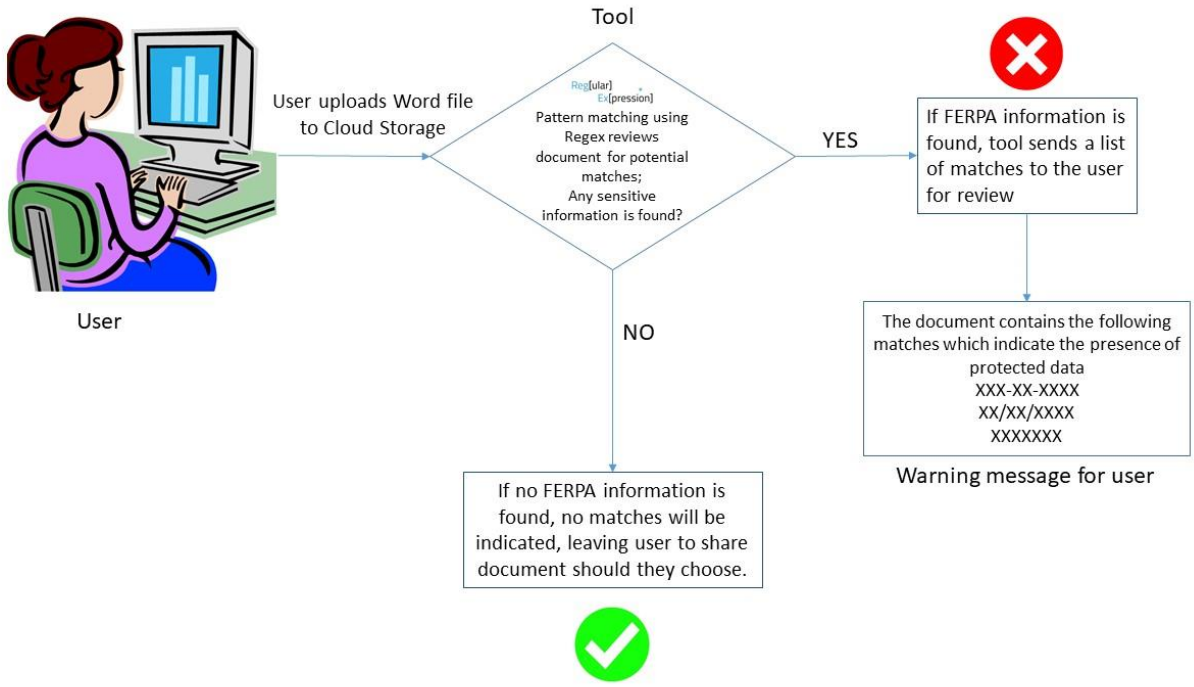
The tool is constructed in the Windows Forms Application using Visual Studio and C#. This is a widely used and user-friendly application that allows for ease of access and use by potential users. The tool functions by using the Interoperability for Office functionality to upload a Word file into an input string and then using the Regular Expression functionality to parse that input string for matches to particular defined patterns. Those patterns, for a proof-of-concept test case, include Social Security Numbers, Minnesota State Tech ID numbers, and dates to include birthdates.

The data is displayed as a list of matches to pre-defined patterns which may indicate the presence of protected data. The user can review which matches are relevant and make changes to the document or how it is shared in order to limit exposure of personal information. This type of functionality indicates an informational

approach rather than a directive approach and may lead to better adoption of recommendations by leaving decisions in the hands of the user rather than proscribing or otherwise controlling what users do with the information provided by the tool.

Figure 3

Operational Flow of Tool



As diagrammed in Figure 3, the tool takes an uploaded file and either does or does not identify any matches. If matches are identified, it provides them for the user to review so they may make changes accordingly, either to the document (such as redaction) or to potential methods of sharing the document or their intended recipients.

Microsoft Visual Studio was utilized for the proof-of-concept tool, using Windows Forms Applications for the development of a working prototype. Additionally, several

libraries as well as Regular Expression capability were used to detect and flag certain types of protected information in Microsoft Word documents.

The tool makes use of the imported Microsoft Office Interoperability package for Word (Chowdhury, 2017) to access and parse Word documents. The Interoperability package allows the tool to get the file content of a Word document and use that content as the input for tool by using elements of Visual Studio and C# to convert the information within the Word document to a string which can be read compatibly by other programming elements within the environment.

The tool also takes advantage of Regular Expression code by importing the Regular Expressions Text package (Kumar, 2014). Using this code package, various patterns can be defined such as social security numbers, dates, and other types of data in addition to string literals or specific words or phrases. The tool is coded to identify Social Security Numbers, Minnesota State Tech ID numbers, and dates using this Regular Expression code.

The tool functions by having a user drop a Word file into the Windows Forms application. The Word Interoperability package then converts the contents of the file into a string to be read by the tool. Regular expression code takes this string and searches it for patterns that match the signatures of protected information, including Social Security Numbers, Minnesota State Tech ID numbers, and dates, and flags them as matches for the user to review. If no matches are found, the tool indicates this condition and the user may share the document without addressing any privacy issues.

If matches are found, each match is presented to the user for review. This process leaves the user to decide whether the file needs to be redacted or otherwise altered before being shared with others. This type of functionality is user-friendly and increases chances of adoption by using an informational rather than directive approach to managing sensitive or protected data.

Additionally, a keyword search algorithm was tested and employed to identify and flag user-entered keywords within text. The solution was implemented in C# using MS Visual Studio 2019. The keywords are then highlighted within the text for the user to review and redact if needed.

Figure 4

Keyword Search Code

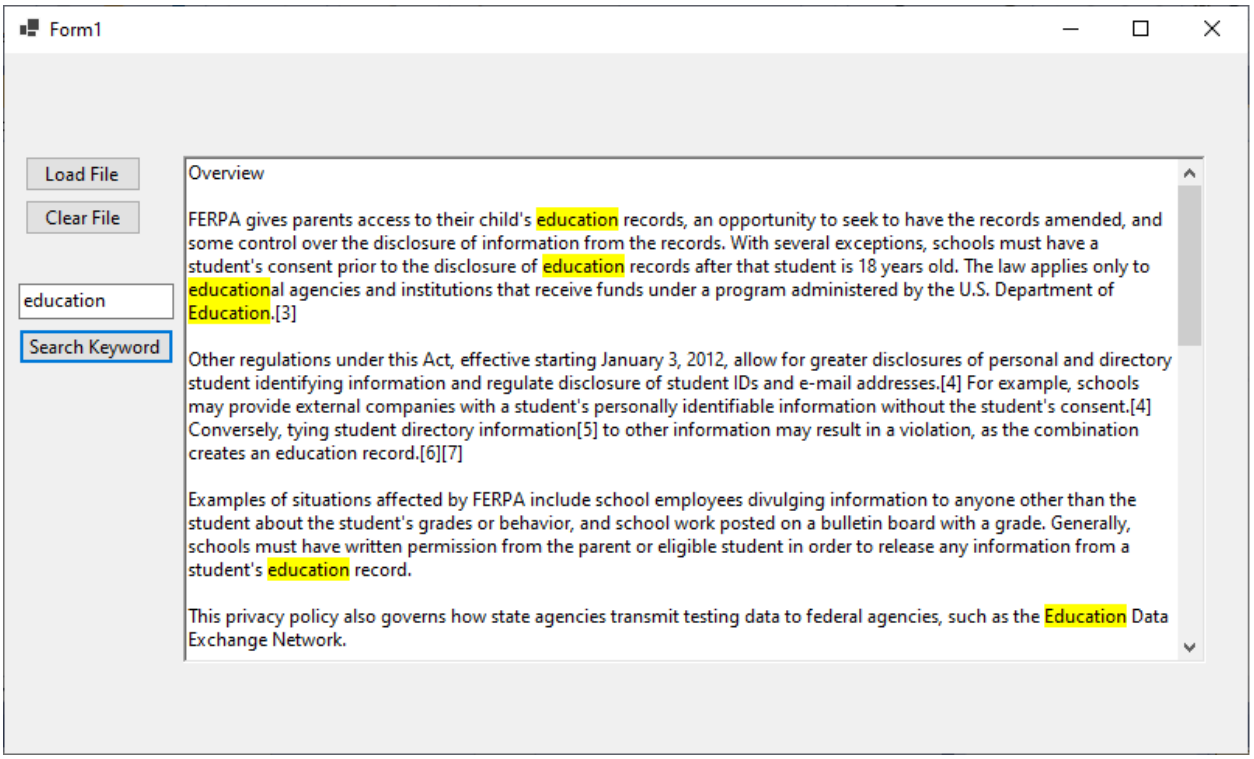
```
        1 reference
44      private void SearchBtn_Click(object sender, EventArgs e)
45      {
46          int index = 0;
47          String temp = richTextBox1.Text;
48          richTextBox1.Text = "";
49          richTextBox1.Text = temp;
50
51          while (index < richTextBox1.Text.LastIndexOf(textBox1.Text))
52          {
53              //searches text in RichTextBox control for a string from Search textbox
54              richTextBox1.Find(textBox1.Text, index, richTextBox1.TextLength, RichTextBoxFinds.None);
55              //Adds a color highlight to matched words
56              richTextBox1.SelectionBackColor = Color.Yellow;
57              //Increase the index after a match is found
58              index = richTextBox1.Text.IndexOf(textBox1.Text, index) + 1;
59          }
60      }
61  }
62  }
```

The algorithm takes a text box input search keyword and searches for instances of that keyword within an uploaded text file displayed using a Rich Text box within visual

studio. Within the Rich Text Box, all instances of the input keyword will be highlighted for the user. In this way, a user can be aware of potential violations within a document that may match with particular keywords.

Figure 5

Example Test of Keyword Search



This approach of adding a keyword search to Regular Expression pattern matching provides additional functionality to the proof-of-concept tool and shows some of the potential for helping users to monitor and limit their FERPA-related sharing. In this way, student data can be better protected both by institutions and professionals.

Data Analysis

Several test files were created to test the tool in operation. Overall, the tool found 20 out of 20 potential matches and thus performed will within its limited scope of operation. This included 12 social security numbers, 3 tech ID numbers, and 5 dates inside the various test files. Efficacy of the tool was thus determined within the limited scope of its current programming.

This level of efficacy was very positive and was well within the proscribed range to determine effectiveness. The downside of the tool is that its scope is still limited at this point to a few categories of data. Expanding the types and categories of data which the tool can detect is a key area of future research and application.

In addition to technical testing, a survey was prepared and distributed to gauge the need and potential efficacy of tools for managing FERPA compliance. While the number of responses was limited, several clear themes emerged. Respondents indicated varying levels of understanding of FERPA as well as data security practices, which matched the research in the review of literature. As shown in Figure 6, the clearest theme was that respondents wanted and needed secure methods of sharing documents as a means of protecting student data.

Figure 6

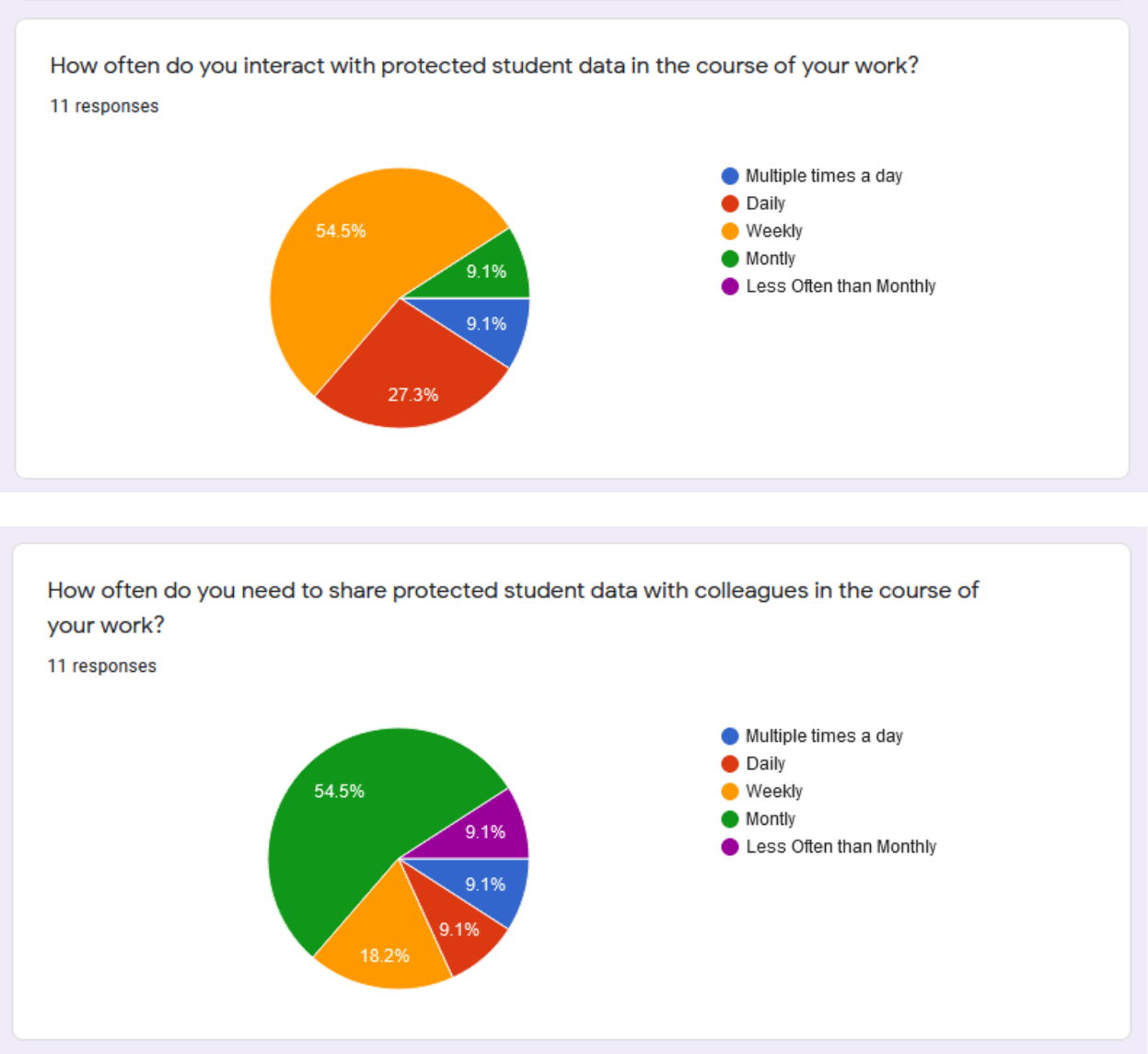
Selected Survey Responses

Additional themes were that the tool seemed relatively easy to use and that most respondents interacted with student data on at least a weekly basis along with needing

to share student data on at least a monthly basis, as shown in Figure 7. Overall, participants responded that they would regularly use a tool which allowed for secure sharing of student data.

Figure 7

How Often Roles Interact with Student Data



Summary

Using Visual Studio and the Windows Forms application and with the imported packages for Microsoft Office Interoperability and for Regular Expression text matching, the tool performed effectively within a limited scope which lends credence to its functionality as a proof-of-concept upon which future applications and expansions can be built and based.

Chapter V: Results, Conclusion, and Recommendations

Introduction

In the current information security environment, many institutions in higher education lack the resources to fully protect their assets. In the case of student data that is covered by FERPA, this under-resourced condition leaves institutions vulnerable to breaches that could have long-term financial and reputational consequences which adversely affect the institution as well as its students and employees. From ransomware to data mining, the threat environment is active and aggressive in seeking to exploit this situation.

In this environment, it is therefore incumbent upon institutions to develop new precautions that protect their resources and data from those who would seek illicit access. A key component in this strategy is taking the opportunity to limit inadvertent data breaches which may lead to FERPA violations. By developing a tool that allows institutional actors an easy method of stopping or minimizing data breaches, the results of this study show there is an opportunity for institutions and professionals in higher education to limit their exposure and vulnerability at relatively minimal cost and with minimal inconvenience.

Results

This study incorporates Regular Expression API to conduct pattern-matching within Microsoft Word documents that can identify and flag certain categories of protected student data and provide a list of potential issues to the user to allow for informed decision-making around sharing Word documents. This is a proof-of-concept

study that speaks to a much broader application which uses pattern-matching, text classification, Natural Language Processing, and similar tools to identify and flag potential violations in a variety of document types to include most or all text formats. Such an tool would ideally be linked to a cloud storage application which could then be utilized for secure sharing and storage of sensitive information.

Within the scope of the proof-of-concept application, the tool was highly effective in identifying limited types of data which suggests that its broader application would also have successful outcomes. Overall, the study attempted to answer three questions in a manner which could aid institutions and professionals in compliance with the Family Educational Rights and Privacy Act:

1) What technical solutions can help institutions and professionals in higher education environments to maintain FERPA compliance?

The technical solutions to the challenge of FERPA compliance are currently limited. While the health care field has devoted significant resources to developing privacy and compliance solutions, the field of higher education has been limited by an overall lack of investment as well as a diffuse, decentralized structure both within and among institutions. As proposed in this study, an application that can review text-based documents and files and flag potential FERPA violations for a user has great promise as a tool which can bridge this compliance gap and provide a user-friendly solution for adoption by both institutions and professions. The potential of pairing this with a cloud storage application shows additional potential to effectively limit the occurrence and scope of unintentional data breaches.

2) What document-level privacy solutions can help institutions and professionals in higher education environments to maintain FERPA compliance?

Similar to the first question, the potential for a tool which can parse text from documents and identify information which is protected by FERPA shows great potential based on the results of this study. Future applications can increase this functionality further and would give institutions and professionals an easy-to-use solution which can be widely incorporated and adopted at all levels of the organization.

3) What cloud storage applications can help institutions and professionals in higher education environments to maintain FERPA compliance?

The ability to link a document-scanning tool with a cloud storage application shows significant promise as a means of not just managing the information within documents but taking advantage of storage systems already in place, such as OneDrive or Google Drive, to provide a means of secure sharing of electronic documents within organizations between individuals who are authorized to receive such information. This would limit the use of unsecure methods such as email sharing of sensitive information and would create a more secure and effective environment for FERPA compliance.

Conclusion

Institutions and professional in higher education, like those in almost every field, face myriad challenges to maintaining data security. These challenges are significant and varied and test both the resources institutions are able to commit as well as specific

issues related to diffuse and varied organizational structures including who has access to data, their knowledge of FERPA, and their knowledge of best practices for safe and secure data sharing and management. As organizations have transitioned from paper to electronic storage of records, they have not kept up with the latest technology to keep those records secure and limit potential for breaches.

The development of a document-level tool which can identify and flag protected data shows significant potential in helping these individuals and organizations to manage compliance by limiting unintentional data breaches. Pairing this with a could storage application to allow for secure sharing of documents shows further potential by reducing the reliance on unsecure methods of sharing data. Such a tool would have great utility and high potential for adoption by individual and institutional users.

Future Work

There are several recommendations to be made for future work, both in the development of further features and applications for the tool as well as in further strengthening privacy practices within the field of higher education. The tool is a proof-of-concept application but its effectiveness in a limited scope portends great promise for a more fully developed tool with additional capabilities and features.

The first recommendation is the incorporation of additional data types into the tool, to include a broader scope of categories of protected information such as personally identifiable information, majors/programs of study, activities, and other information indicated in Tables 1 and 2. This would create a truly comprehensive tool that could reasonably detect most or all types of protected data. Another priority should

be adding additional file formats other than word such as .pdf, excel, email, .txt, and other file types. This would further increase utility and functionality for the tool.

The second recommendation is to pair the tool with a cloud storage application. This was originally part of the study but there was not an opportunity to incorporate it with the proof-of-concept application. Pairing with cloud storage would allow for secure and seamless sharing of documents among authorized individuals and within organizations and could be a key component of a final tool,

A third recommendation is to incorporate the tool into a browser extension that would provide a user-friendly option to review documents within a web browser prior to sending an email or otherwise sharing information. This would increase the utility of the tool and likely increase chances of adoption by providing additional situations in which the tool can be used.

Another feature which can be added to the tool is to program the tool to highlight the data within a document with a notation indicating what type of data it is (i.e. “Directory Information” or “Private/Protected Information – Do Not Release!”) so the user can get a quick overview of the type of data contained in the document. This will help with decision-making by providing additional granular detail about the classifications of data included in the document.

A final recommended addition to the tool is the incorporation of student disciplinary, health, financial records. These additional record classes are often stored separately within 3rd Party vendor platforms rather than on the primary institutional

database. The ability to scan these records would be an added feature which would broaden the scope of the tool and increase usability.

In addition to these recommendations for future capabilities to be added to the tool, there are several suggestions related to protecting privacy within the field of higher education. While FERPA is a national standard, there are few centralized standards for how it should be implemented between and among institutions. The development of a privacy consortium or some other group or collection of interested individuals and institutions could be a powerful force for providing more centralized recommendations and best practices for student records privacy.

Additionally, within institutions it would be valuable to transition from email sharing to cloud sharing of documents in order to better protect sensitive or personally identifiable data. Email is less secure and more subject to secondary sharing versus cloud storage where access controls can be more strictly enforced.

These recommendations can go a long way toward building out a powerful tool and developing strong best practices within and among institutions which will allow professionals in higher education to better protect student data and privacy.

References

Agris, J. L., & Spandorfer, J. M. (2016). HIPAA Compliance and Training: A Perfect Storm for Professionalism Education? *The Journal of Law, Medicine & Ethics*, *44*(4), 652–656. https://doi.org/10.1177/1073110516684812

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *ArXiv:1707.02919 [Cs]*. http://arxiv.org/abs/1707.02919

Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: Current state, evolution patterns, and development priorities. *R&D Management*, *50*(3), 329–351. https://doi.org/10.1111/radm.12408

Arras, L., Horn, F., Montavon, G., Müller, K.-R., & Samek, W. (2017). "What is relevant in a text document?": An interpretable machine learning approach. *PLOS ONE*, *12*(8), e0181142. https://doi.org/10.1371/journal.pone.0181142

Bachani, N. (2020, April 17). *Chunking in NLP :Decoded*. Medium. https://towardsdatascience.com/chunking-in-nlp-decoded-b4a71b2b4e24

BAUER v. KINCAID | 759 F.Supp. 575 (1991) | upp57511244 | Leagle.com, (March 13, 1991). https://www.leagle.com/decision/19911334759fsupp57511244

Berman, M. (2019, August 26). New Life for Legacy Systems. *EDUCAUSE Review*. https://er.educause.edu/articles/2019/8/new-life-for-legacy-systems

Bissell, K., LaSalle, R., & Dal Cin, P. (2019). *2019 Cost of Cybercrime Study | 9th Annual | Accenture*. https://www.accenture.com/us-en/insights/security/cost-cybercrime-study

Blackman, J. (2020, August 7). Should Colleges Force Students To Turn Their Cameras On? *Reason.Com*. https://reason.com/volokh/2020/08/07/should-colleges-force-students-to-turn-their-cameras-on/

BlazingText algorithm—Amazon SageMaker. (n.d.). Retrieved January 6, 2021, from https://docs.aws.amazon.com/sagemaker/latest/dg/blazingtext.html

Board of Ed. Of Independent School Dist.no. 92 of Pottawatomie Cty. V. Earls (Syllabus), 536 U.S. 822 (U.S. Supreme Court 2002), CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE TENTH CIRCUIT. https://www.law.cornell.edu/supct/html/01-332.ZS.html

Branavan, S. R. K., Chen, H., Eisenstein, J., & Barzilay, R. (2009). Learning Document-Level Semantic Properties from Free-Text Annotations. *Journal of Artificial Intelligence Research*, *34*, 569–603. https://doi.org/10.1613/jair.2633

Brownlee, J. (2017, September 21). What Is Natural Language Processing? *Machine Learning Mastery*. https://machinelearningmastery.com/natural-language-processing/

Burner, D. (1996). *Berkeley Free Speech Movement, 1963-64*. https://www.writing.upenn.edu/~afilreis/50s/berkeley.html

Cannon, A. A., & Caldwell, H. (2016). HIPAA violations among nursing students: Teachable moment or terminal mistake-A case study. *Journal of Nursing Education and Practice*, *6*(12), 41. https://doi.org/10.5430/jnep.v6n12p41

Cartwright, H. J. (n.d.). *Lighting up healthcare data with FHIR®: Announcing the Azure API for FHIR*. Retrieved November 1, 2020, from https://azure.microsoft.com/en-us/blog/lighting-up-healthcare-data-with-fhir-announcing-the-azure-api-for-fhir/

Chaudhary, A. (2020, August 30). *Unsupervised Keyphrase Extraction*. Amit Chaudhary. https://amitness.com/keyphrase-extraction/

Check Point Software Technologies. (2020, July 22). *Check Point Research: COVID-19 Pandemic Drives Criminal and Political Cyber-attacks Across Networks, Cloud and Mobile in H1 2020*. GlobeNewswire News Room. http://www.globenewswire.com/news-release/2020/07/22/2065610/0/en/Check-Point-Research-COVID-19-Pandemic-Drives-Criminal-and-Political-Cyber-attacks-Across-Networks-Cloud-and-Mobile-in-H1-2020.html

Childress, R. (2020a, August 25). *University of Kentucky Leaves COVID-19 Test Results Unguarded*. https://www.govtech.com/security/University-of-Kentucky-Leaves-COVID-19-Test-Results-Unguarded.html

Childress, R. (2020b, October 23). Should UK release sexual harassment records? Decision now rests with Ky. Supreme Court. *Kentucky*. https://www.kentucky.com/news/local/education/article246608918.html

Chowdhury, K. (2017, February 3). *How to read Microsoft Word document contents using C#/.NET?* Www.Kunal-Chowdhury.Com. https://www.kunal-chowdhury.com/2017/02/how-to-read-microsoft-word-document.html

Clark, A. (2017). *The Machine Learning Audit*. 30. Retrieved from https://www.dallasiia.org/wp-content/uploads/2017/11/The-Machine-Learning-Audit-Andrew-Clark.pdf on April 3, 2020.

Coker, J. (2020, October 29). Education Sector Facing Disproportionate Level of Spear-Phishing Attacks. *Infosecurity Magazine*. https://www.infosecurity-magazine.com:443/news/education-disproportionate-spear/

Coulture, R., Schwehm, J., & Coulture, V. (2018). FERPA Fear or FERPA Flex: Student Affairs Practitioners' Understanding of Federal Privacy Laws on Campus. *Journal of Student Affairs*, *28*, 39–50.

C#—Regular Expressions. (n.d.). Retrieved September 28, 2021, from https://www.tutorialspoint.com/csharp/csharp_regular_expressions.htm

Darnell, C. (2020, February 25). Naming students behind hate acts would violate federal law, experts say. *The Daily Orange*. http://dailyorange.com/2020/02/naming-students-hate-acts-violates-federal-privacy-law/

DeWilde, B. (2014, September 23). *Intro to Automatic Keyphrase Extraction*. https://bdewilde.github.io/blog/2014/09/23/intro-to-automatic-keyphrase-extraction/

Directory Information | Protecting Student Privacy. (n.d.). Retrieved February 9, 2021, from https://studentprivacy.ed.gov/content/directory-information

EDU-API | IMS Global Learning Consortium. (n.d.). Retrieved November 1, 2020, from https://www.imsglobal.org/edu-api

Electronic Privacy Information Center (n.d.). *EPIC - Student Privacy Case Law*. Retrieved October 31, 2020, from https://epic.org/privacy/student/cases/

Electronic Records Management Toolkit. (2020). https://www.educause.edu/focus-areas-and-initiatives/policy-and-security/cybersecurity-program/resources/information-security-guide/toolkits/electronic-records-management-toolkit

Emisoft. (2019). *The State of Ransomware in the US: Report and Statistics 2019*. https://blog.emsisoft.com/en/34822/the-state-of-ransomware-in-the-us-report-and-statistics-2019/

Family Educational Rights and Privacy Act, 34 CFR Part 99. Retrieved February 9, 2021, from https://www2.ed.gov/policy/gen/guid/fpco/pdf/ferparegs.pdf

Family Educational Rights and Privacy Act (FERPA). (2018, March 1). [Guides]. US Department of Education (ED). https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

Fernando, C. (2020, February 4). IU GPA calculator gave students access to other students' GPA, grades. *The Indianapolis Star*.

https://www.indystar.com/story/news/education/2020/02/04/indiana-university-gpa-calculator-may-violate-federal-privacy-law/4659126002/

Forester, K., U.S. v. Miami University, No. 98-00097 F. Supp. (June 27, 2002). https://www2.ed.gov/policy/gen/guid/fpco/pdf/miami-fulldecision.pdf.

Friedman, C. (2020, October 14). *Corey Friedman: Will Supreme Court Allow 'Double Secret Probation' for College Title IX Trials?* https://www.noozhawk.com/article/corey_friedman_will_supreme_court_allow_double_secret_probation_20201014

Garbade, D. M. J. (2018, October 15). *A Simple Introduction to Natural Language Processing*. Medium. https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32

Gerard, P., Kapadia, N., Chang, P. T., Acharya, J., Seiler, M., & Lefkovitz, Z. (2013). Extended Outlook: Description, Utilization, and Daily Applications of Cloud Technology in Radiology. *American Journal of Roentgenology*, *201*(6), W809–W811. https://doi.org/10.2214/AJR.12.9673

Grajek, S. (2020, January 27). *How Colleges and Universities Are Driving to Digital Transformation Today*. https://er.educause.edu/articles/2020/1/how-colleges-and-universities-are-driving-to-digital-transformation-today

Gross, N. (2020, October 5). How schools are navigating privacy concerns in COVID-19 contact tracing. *Education Dive*. https://www.educationdive.com/news/k-12-schools-covid-contact-tracing-student-privacy-coronavirus/586352/

Gupta, M. (2020, May 18). *Syntactic / Constituency Parsing using the CYK algorithm in NLP*. Medium. https://medium.com/data-science-in-your-pocket/syntactic-constituency-parsing-using-the-cyk-algorithm-in-nlp-eff9c2912b09

Hartman, M., Martin, A. B., Benson, J., Catlin, A., & The National Health Expenditure Accounts Team. (2020). National Health Care Spending In 2018: Growth Driven By Accelerations In Medicare And Private Insurance Spending: US health care spending increased 4.6 percent to reach $3.6 trillion in 2018, a faster growth rate than that of 4.2 percent in 2017 but the same rate as in 2016. *Health Affairs*, *39*(1), 8–17. https://doi.org/10.1377/hlthaff.2019.01451

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2018). Comparing automated text classification methods | Elsevier Enhanced Reader. *International Journal of Research in Marketing*, *36*, 20–38. https://doi.org/10.1016/j.ijresmar.2018.09.009

Hidalgo, E. B. (2020, September 14). What's up with Honorlock? *THE MERCURY*. https://utdmercury.com/whats-up-with-honorlock/

Hill, A. (2020, November 17). *Text mining and analysis with the Text Analytics API - Azure Cognitive Services | Microsoft Docs*. https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/overview

Hill, A., & Yeo, A. (2021, January 22). *Supported Categories for Named Entity Recognition—Azure Cognitive Services*. https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/named-entity-types

Hlavac, G. C., & Easterly, E. J. (2015). FERPA Primer: The Basics and Beyond. NACE Journal. Electronic document, https://www.naceweb.org/public-policy-and-legal/legal-issues/ferpa-primer-the-basics-and-beyond/,

Jay, W. (2020, October 16). Cybercrime increases amid COVID-19. *El Camino College The Union*. https://eccunion.com/news/2020/10/16/cybercrime-increases-amid-covid-19/

Kellen, V. (2019). 21st-Century Analytics: New Technologies and New Rules. *EDUCAUSE Review*. https://er.educause.edu/articles/2019/5/21st-century-analytics-new-technologies-and-new-rules

Keyword Extraction: A Guide to Finding Keywords in Text. (2020). MonkeyLearn. https://monkeylearn.com/keyword-extraction/

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2019). Text Classification Algorithms: A Survey. *Information*, *10*(4), 150. https://doi.org/10.3390/info10040150

Kumar. (2014, June 22). How to Validate a Social Security Number using Regex in C#? *Abundant Code*. https://abundantcode.com/how-to-validate-a-social-security-number-using-regex-in-c/

Lake, P. (n.d.). *Student-Privacy Rules Show a Renewed Trust in Colleges*. 4. Retrieved from https://www.haverford.edu/sites/default/files/Office/Safety/FERPA_Impact_2009.pdf on April 3, 2020.

Landi, H. (2019, October 9). Cerner, Amazon Web Services partner on new cloud-based cognitive health platform. *FierceHealthcare*. https://www.fiercehealthcare.com/tech/deal-amazon-web-services-cerner-unveils-new-cloud-based-health-platform

Lane, K. (n.d.). *University API Workshop The University API Guide*. Retrieved November 1, 2020, from https://kinlane.github.io/university-api-workshop/unversitiy-api/guide/

Lasseter, A. (2020, November 10). *Deploy an NLP classification model with Amazon SageMaker and Lambda*. Medium. https://austinlasseter.medium.com/deploy-an-nlp-classification-model-with-amazon-sagemaker-and-lambda-cd5ea6339781

Legislative History of Major FERPA Provisions. (2005, December 19). [Guides]. US Department of Education (ED). https://www2.ed.gov/policy/gen/guid/fpco/ferpa/leg-history.html

Li, S. (2018a, December 6). *Named Entity Recognition with NLTK and SpaCy*. Medium. https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da

Lombardi, J. (1969). Student Activism in Junior Colleges, An Administrator's Views. *American Association of Junior Colleges Monograph Series*. https://files.eric.ed.gov/fulltext/ED028767.pdf

Maycunich, A. (2002). *FERPA: An investigation of faculty knowledge levels and organization practices at three land-grant universities*. [Doctor of Philosophy, Iowa State University, Digital Repository]. https://doi.org/10.31274/rtd-180813-11714

McKenzie, L. (2020, June 11). *Colleges face evolving cyber extortion threat*. https://www.insidehighered.com/news/2020/06/11/colleges-face-evolving-cyber-extortion-threat

Mearian, L. (2011, June 28). Why Google Health failed: Too little, too soon. *Computerworld*. https://www.computerworld.com/article/2509515/why-google-health-failed--too-little--too-soon.html

Miller, M. (2020a, February 5). Do Student Information Systems Need a Tuneup? *Technology Solutions That Drive Education*. https://edtechmagazine.com/higher/article/2020/02/do-student-information-systems-need-tuneup

Miller, M. (2020b, June 23). *FBI probing racist emails sent to thousands affiliated with major universities: Reports* [Text]. TheHill. https://thehill.com/policy/cybersecurity/504199-fbi-investigating-racist-emails-sent-to-thousands-affiliated-with-major

Moramarco, S. (n.d.). *Phishing Attacks in the Education Industry*. Infosec Resources. Retrieved October 19, 2020, from https://resources.infosecinstitute.com/category/enterprise/phishing/the-phishing-landscape/phishing-attacks-by-demographic/phishing-attacks-in-the-education-industry/

Neil, A. (2020, October 1). UNC petitions U.S. Supreme Court to review ruling in sexual assault records case. *The Daily Tar Heel*. https://www.dailytarheel.com/article/2020/10/university-records-court-ruling-review

Owasso Independent School Dist. No. I—011v. Falvo (Opinion of the Court), 534 U.S. 426 (U.S. Supreme Court 2002), ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE TENTH CIRCUIT. https://www.law.cornell.edu/supct/html/00-1073.ZO.html

Pai, A. (2020, May 25). What is Tokenization | Tokenization In NLP. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/

Parmar, A. (2019, November 6). Apple adds VA to its list of organizations making data available on its Health Records. *MedCity News*. https://medcitynews.com/2019/11/apple-adds-va-to-its-list-of-organizations-making-data-available-on-its-health-records/

Pearson, D. (2020, October 29). *Up first for the Mayo Clinic–Google Health alliance: Better radiation therapy through AI*. https://www.aiin.healthcare/topics/connected-care/mayo-clinic-google-health-radiation-therapy-ai

Raskar, S. S., & Pathan, S. H. (2014). Keyphrase Extraction using supervise learning. *IJARCCE*, 8510–8512. https://doi.org/10.17148/IJARCCE.2014.31135

Raths, D. (2017, April 13). Building APIs for the University and the Student -. *Campus Technology*. https://campustechnology.com/articles/2017/04/13/building-apis-for-the-university-and-the-student.aspx

Rehnquist, W. in Gonzaga Univ. V. Doe (Opinion of the Court), 536 U.S. 273 (U.S. Supreme Court 2002), ON WRIT OF CERTIORARI TO THE SUPREME COURT OF WASHINGTON. https://www.law.cornell.edu/supct/html/01-679.ZO.html

Riddell, R. (2016, September 29). Legacy upended: Higher ed CIOs look to modernize campus systems. *Education Dive*. https://www.educationdive.com/news/legacy-upended-higher-ed-cios-look-to-modernize-campus-systems/427214/

Robison, K.. (2020, September 4). 8 Sac State faculty accounts compromised in phishing attack. *The State Hornet*. https://statehornet.com/2020/09/sac-state-faculty-email-scam/

RSI Security. (2019, November 27). What Is the Difference Between HIPAA vs. FERPA? *RSI Security*. https://blog.rsisecurity.com/what-is-the-difference-between-hipaa-vs-ferpa/

Srivastava, U. (2020, February 21). CS TA removed from staff following potential FERPA violation. *The Stanford Daily*. https://www.stanforddaily.com/2020/02/20/computer-science-ta-removed-from-staff-following-potential-ferpa-violation/

St. Amour, M. (2020, March 25). *Pivot to online raises concerns for FERPA, surveillance*. https://www.insidehighered.com/news/2020/03/25/pivot-online-raises-concerns-ferpa-surveillance

Stone, K. J. (2002). *STETSON UNIVERSITY COLLEGE OF LAW 23RD ANNUAL NATIONAL CONFERENCE ON LAW AND HIGHER EDUCATION*. 18. Retrieved from https://www.stetson.edu/law/academics/highered/home/media/2002/Revisiting_the_Purpose_of_FERPA.pdf on April 3, 2020.

Student Records Management Practice. (2017). AACRAO - American Association of Collegiate Registrars and Admissions Officers. https://www.aacrao.org/docs/default-source/research-docs/student-records-management-practice-nbsp---january-60-second-surveycab5ed2c9c694f02b241b39153251a7a.pdf?Status=Temp&sfvrsn=7385eb1e_6

Sussman, B. (2020, August 21). *Higher Ed Ransomware Attack: University Pays $457K Despite Having Backups*. https://www.secureworldexpo.com/industry-news/university-of-utah-ransomware-attack

Tate, Z., & Stuerman, M. (2019, October 8). Southeast department sends email violating student privacy. *Southeast Arrow*. https://www.southeastarrow.com/story/2639662.html

Truong, K. (2019, April 8). Microsoft HealthVault is officially shutting down in November. *MedCity News*. https://medcitynews.com/2019/04/microsoft-healthvault-is-officially-shutting-down-in-november/

Tsai, A. (2020, October 5). Piazza privacy concerns push some professors to other discussion forums. *The Stanford Daily*. https://www.stanforddaily.com/2020/10/04/concerned-with-piazzas-data-privacy-management-some-professors-look-to-alternative-discussion-forums/

Turnage, C. C. (2007). School Officials' Knowledge of the Family Educational Rights and Privacy Act of 1974 at the University of Southern Mississippi. Retrieved from https://aquila.usm.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com/&httpsredir=1&article=2361&context=dissertations on Oct. 31, 2020.

Two Decades of Change in Federal and State Higher Education Funding. (2019). https://pew.org/2M7okiZ

Vidwans, R. (n.d.). *WHY PHISHING EMAILS TARGET UNIVERSITIES & HOW TO STAY PROTECTED*. Retrieved October 19, 2020, from https://www.clearedin.com/blog/phishing-emails-target-universities

Vroom, C., & von Solms, R. (2004). Towards information security behavioural compliance. *Computers & Security*, *23*(3), 191–198. https://doi.org/10.1016/j.cose.2004.01.012

Wagner, B. (2015, July 20). *How to access Office interop objects—C# Programming Guide*. https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide/interop/how-to-access-office-onterop-objects

Wan, T. (2020, March 27). Holding Class on Zoom? Beware of These Hacks, Hijinks and Hazards - EdSurge News. *EdSurge*. https://www.edsurge.com/news/2020-03-27-holding-class-on-zoom-beware-of-these-hacks-hijinks-and-hazards

What is FHIR? (n.d.). The Office of the National Coordinator for Health Information Technology. Retrieved November 1, 2020, from https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf

Wiggers, K. (2020, April 20). Google launches Cloud Healthcare API in general availability. *VentureBeat*. https://venturebeat.com/2020/04/20/google-launches-cloud-healthcare-api-in-general-availability/

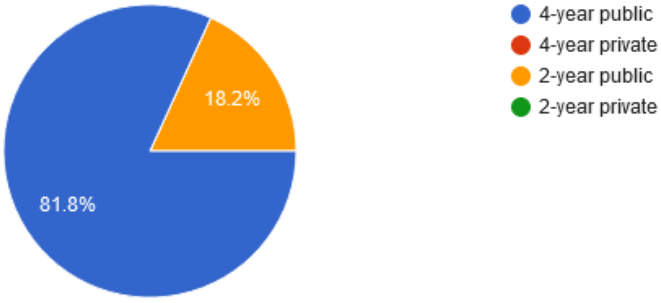
Wood, C. (2020, February 5). Phishing attack exposes personal information of 5,000 at community college. *EdScoop*. https://edscoop.com/phishing-attack-exposes-personal-information-of-5000-at-community-college/

Zalaznick, M. (2020, August 20). *4 COVID-era cybersecurity threats CISOs are confronting* |. https://universitybusiness.com/college-university-covid-cybersecurity-cisos-chief-information-officer-phishing/

Appendix A: Survey Questions and Responses

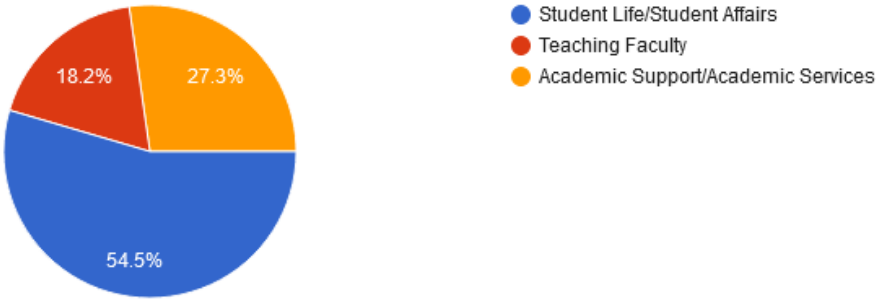
Which best describes your institution type?

11 responses



Which best describes your role?

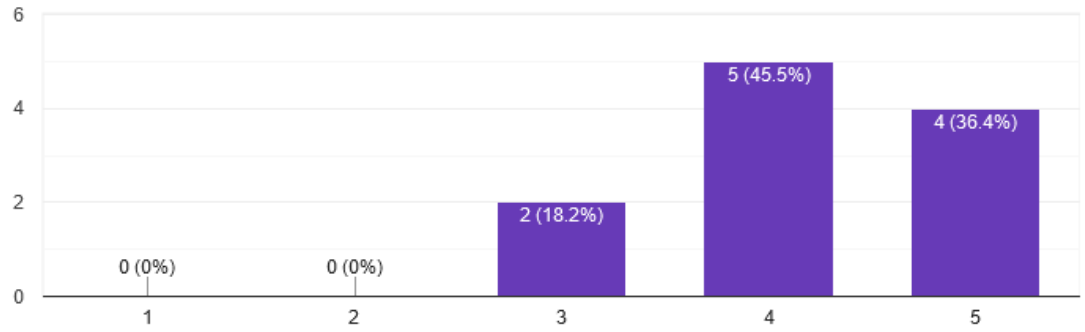
11 responses



For all Likert responses, 1 indicates Strongly Disagree and 5 indicates Strongly Agree

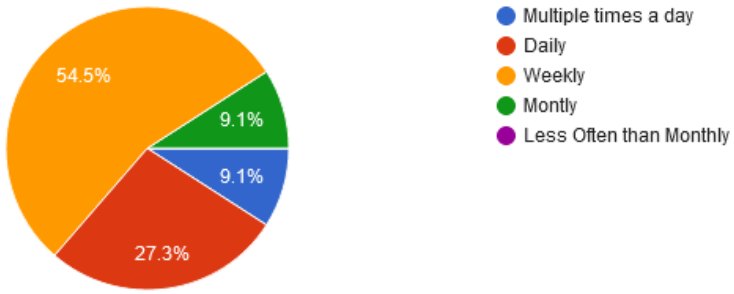
How familiar are you with FERPA, the Family Educational Rights and Privacy Act?

11 responses



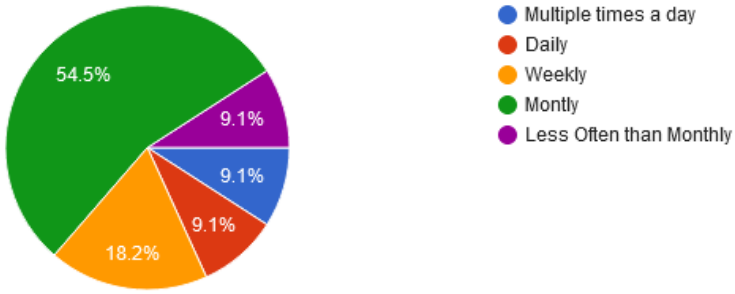
How often do you interact with protected student data in the course of your work?

11 responses



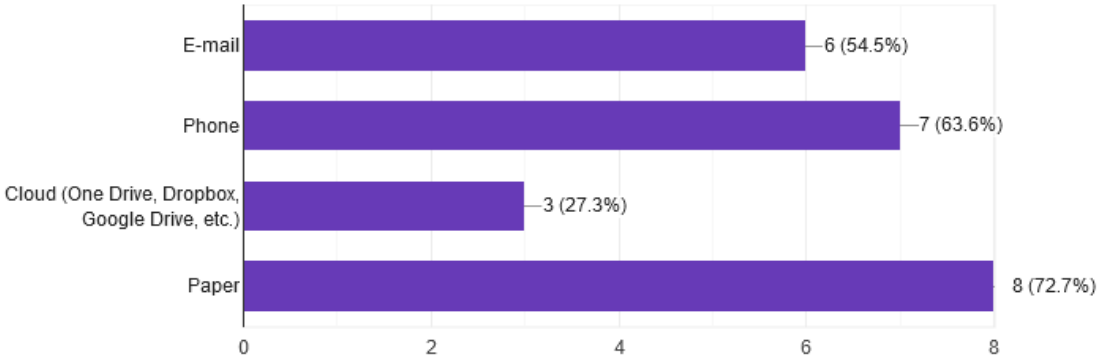
How often do you need to share protected student data with colleagues in the course of your work?

11 responses



Which methods do you regularly use to share protected data? (check all that apply)

11 responses

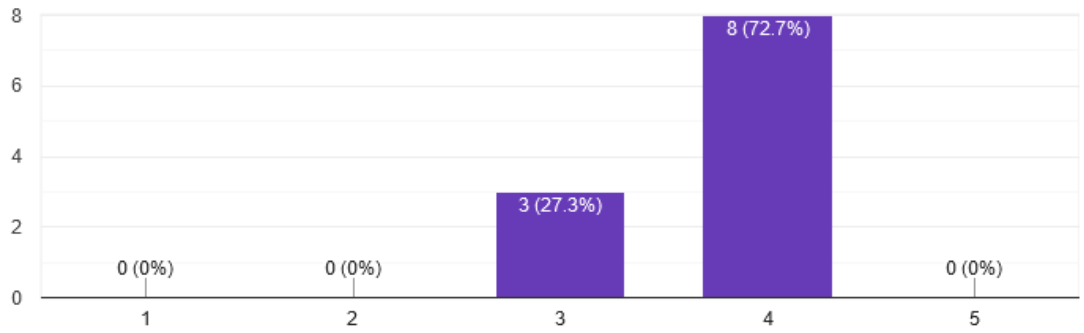


Descriptive Text for Section 2: My research is focused on developing a tool that can scan documents and identify potential FERPA violations for review by the user. The tool could be connected to a Cloud Storage application such as One Drive or DropBox or could be a stand-alone tool which could review documents and emails and flag potential data prior to sharing. It would not block a user's ability to share but would alert the user to potentially protected data. By connecting to a cloud storage service such as One Drive, it would allow for secure sharing of documents within an organization compared to e-mail, which is not a secure method of sharing. Based on this description, please answer the questions below.

Section 2: Tools for FERPA Compliance

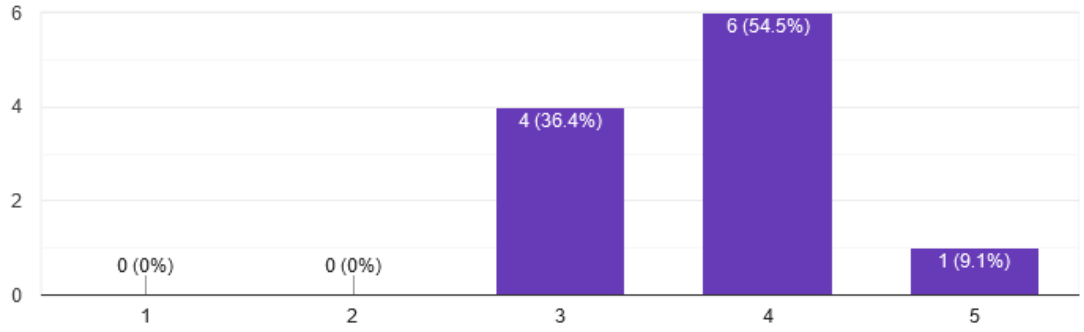
This tool seems easy to use.

11 responses



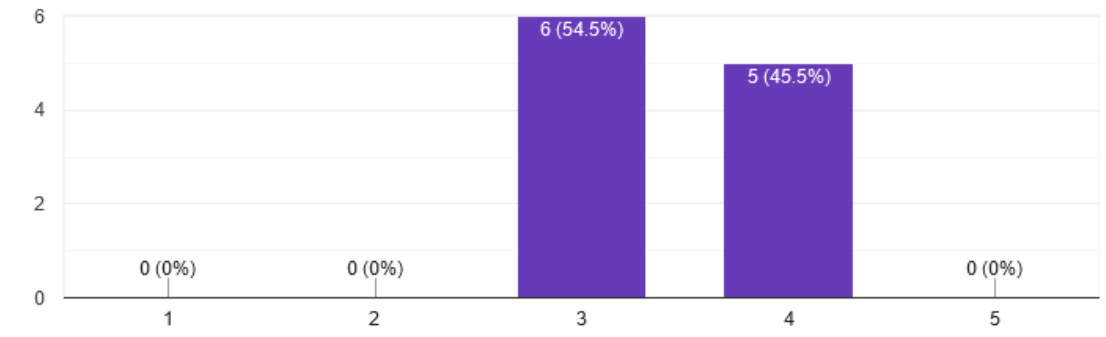
This tool would be useful to me in my job.

11 responses



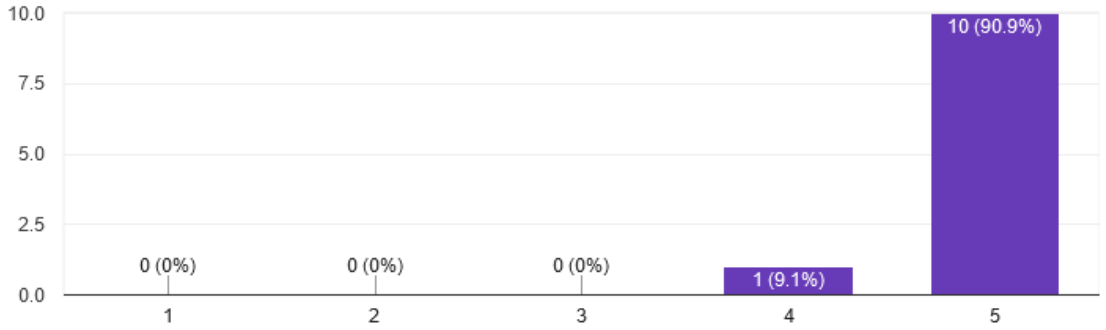
I would use a tool like this on a regular basis.

11 responses



The ability to share documents securely would help protect student data on my campus.

11 responses



I feel I have a solid understanding of data privacy issues related to technology.

11 responses