**Mitigating Security Risks in Commonly Used Alexa Skills**
(Technical Paper)

**Balancing Care and Privacy: A Competition for Security Standards Governing Electronic Medical Records**

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Jammie Wang
Fall, 2020

Technical Project Team Members

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

Signature ________________________________________ Date __________
Jammie Wang

Approved ________________________________________ Date __________
Yuan Tian, Department of Computer Science

Approved ________________________________________ Date __________
Peter Norton, Department of Engineering and Society

## General Research Problem

*How can digital privacy be improved and enforced?*

Currently, about 4,000 cybercrimes are reported to the FBI daily, nearly quadrupling the number from before the COVID-19 pandemic. Yet for every 1,000 cybercrime reports, only three arrests occur. The development of new privacy regulations is becoming increasingly difficult and controversial, with civilians and companies demanding more action to combat cybercrime, and the Trump administration failing to increase budgets for federal cybercrime enforcement agencies that focus on privacy and security (Garcia & Hindocha, 2020).

## Mitigating Security Risks in Commonly Used Alexa Skills

*"How can we improve the security of voice personal assistants (VPAs)?"*

The technical advisor is Yuan Tian in the CS department. This is an Independent Research project where I will be working with a Ph.D. student, Tu Le.

Commercially available voice assistants on the market often exhibit a variety of privacy issues ranging from recording personal conversations to downloading skills that ask for extensive personal information (Pal et al., 2020). Amazon Echo (Alexa) faced public scrutiny in 2018 after it recorded a private conversation and sent it to another person on the owner's contact list (Sacks, 2018). Such incidences led to an extensive debate on the privacy of voice assistants, especially since 15.4% of the US population already owned an Amazon Echo in 2018 (Pridmore & Mols, 2020).

Our technical research aims to identify where these sources of privacy violations and inappropriate content occur in Alexa skills.

Current state of the art research has identified the relationship between the usage of voice assistants and its role in surveillance capitalism (Pridmore & Mols, 2020). Similar research has developed privacy-preserving trust models, which emphasizes a “privacy by design” approach for enhancing end-user trust (Pal et al., 2020).

I will analyze a large dataset of critical reviews from various Alexa skills to identify the status of risky skills on the market, and what kinds of risky behaviors they exhibit. The current dataset includes URLs of skills divided by functional categories, profiles of a large set of skills, and corresponding critical reviews for each skill stored in JSON files.

To parse through critical reviews, I will set up a Jupyter notebook to read in each JSON file. To analyze them, I can assemble a list of “negative” words or phrases such as “credit card,” which would indicate the skill has a reported issue with finances. By going through each skill and recording the total number of occurrences of each negative phrase, I can transform the reviews into rows of numerical data. I plan to perform some form of cross-tabulation analysis between categories of skills (ex. Food, Games, etc.) and incidences of specific negative phrases by recording frequency of occurrences, and potentially calculating mean, median, standard deviation, and correlation. It would also be interesting to identify if there were specific skills that were particularly unsafe compared to other skills in that category.

If I succeed, the result will be a list of the most common risks found in Alexa skills, where categories of Alexa skills are associated with most common critical words. I hope these results can aid further research and development of advanced solutions by focusing on a specific set of risky attributes.

## Balancing Care and Privacy: A Competition for Security Standards Governing Electronic Medical Records

*"In the U.S., how do physicians, patients, clinics, and insurers compete to determine the privacy and security standards governing patients' electronic medical records?"*

The transition from paper to electronic records in healthcare became prominent in the 1990s once the Internet allowed faster access to patient medical information. Electronic health records provide patients with greater access to health information and communication resources, leading to increased patient safety, healthcare accessibility, and "improved continuity of care and efficiency" across fragmented healthcare systems (Pagliari et al., 2007). Doctors must strike a balance between alleviating patient distrust while providing the best possible care (Barrows & Clayton, 1996, p. 141). HIPAA was initially passed in 1996 to ensure health information privacy by requiring consent before sharing patient information. Today, there is a market for cybersecurity services in healthcare specifically to comply with HIPAA regulations and prevent cyberattacks.

To improve care, many healthcare providers have adopted healthcare analytics; such providers demand patient data. Gopalakrishna-Remani et al. (2016) attribute adoption of analytics to "mimetic pressure" or "coercive pressure" in upper management (p. 204). Angst & Agarwal (2009) attributes patient distrust of electronic records to choice architecture, while Whiddett et al. (2006) reveal that patients typically trust physicians but not other stakeholders. Gunter & Terry (2005) has discovered that physicians in particular face the most legal challenges. This research fails to evaluate the role of other players involved in the development of privacy standards, notably organizations and insurer groups.

Participants include advocacies that demand strict privacy policies to protect patients' records. Some unorganized patients also demand stricter privacy standards, including some who have personally experienced a breach of privacy.

Patient advocacy organizations such as the American Patient Rights Association often provide a guide for patients to protect the privacy and security of their health information (Hunt, 2019). The Empowered Patient Coalition (2017) advises patients on how to report medical violations, including privacy violations. Other non-profit organizations such as the Center for Democracy and Technology argue that HIPAA is not enough to ensure patient privacy since it only applies to health records that flow through a traditional healthcare system. Personal health records "should be governed by a comprehensive framework of privacy and security protections," which would protect health records kept by individuals in addition to HIPAA laws (CDT, 2009). The Confidentiality Coalition (2020) is a privacy organization that proposes "Congress should establish a single national privacy and security standard for all health information not subject to HIPAA" and the disclosure of health information should be "written in a meaningful and understandable manner" and be easily accessible, which would help prevent a majority of patients from being confused by legal jargon. These organizations call for reform of privacy laws that would better protect patient privacy.

Unorganized patients bring awareness to patient privacy issues by sharing personal stories about ethical or legal breaches of privacy. Stories have emerged of various privacy violations, including a woman suing a local hospital for sharing details about her 11-year-old son's attempted suicide, a patient care technician making a public Facebook post about her friend's HPV-positive status, and a nurse snooping through a family member's medical records without her permission (Ornstein, 2015). Recent violations in the midst of the COVID-19 pandemic include a nurse accidentally revealing the name of a dead patient on camera, and another nurse sharing confidential hospital policies and patient information with friends over Facebook (Clark, 2020). Clark concludes that a majority of these violations stem from a lack of professional training. Patients who reported an incident of privacy violation feel that they can no longer trust physicians and laws to protect them, and thus favor healthcare privacy law reform by sharing personal experiences.

Professional physician organizations are participants that teach physicians how to best preserve their patients' privacy within the bounds of the law. The Radiological Society of North America aims to share research and medical cases with fellow physicians while still maintaining patient privacy, which is accomplished through technological means such as erasing sensitive metadata from images (RSNA, 2020). As both caregivers and lifelong learners, it is particularly tricky for physicians in professional organizations to maintain patient privacy while still sharing enough information on a medical case such that other physicians will be able to learn from them. The Association for Healthcare Documentation Integrity (2020) is a professional organization whose code of ethics emphasizes conducting “business in accordance with ethical privacy and security practices and maintaining confidentiality of all patient information.” Peer pressure and moral obligations encourage physicians to adhere to privacy regulations, lest they lose their membership in an organization due to a violation. During the COVID-19 pandemic, the American Health Information Management Association stated that “providers will need complete visibility into their patient populations in order to track infection patterns,” yet this data must also be anonymized to protect individual patients, which is a particularly daunting task for both physicians and the engineers working on contact tracing technology (Cidon, 2020).

Hospitals and other private care centers ensure patient privacy to comply with legal regulations and typically do so by requiring their employees (mainly physicians and nurses) to comply with care center protocols (BMC, 2014). Major non-profit care organizations such as the American Cancer Society are more transparent about their privacy practices than many private care centers. Their privacy policy is detailed enough to state that "we incorporate industry-standard security controls (like firewalls) and protocols (like SSL/TLS)" to reduce the risk of internet-based threats (ACS, 2020). Similarly, as a religious non-profit healthcare provider, the Catholic Health Association of the United States is more bound by ethics than other organizations to protect their patients. They describe a relationship between a patient and their physician as such:  "In the physician-patient relationship there is a presumption of respect for privacy, and this presumption is one of the parameters essential to the relationship, a necessary condition for the relationship" (Schick, 1998). It appears that non-profit care providers typically go above and beyond to ensure quality patient care in comparison to private care centers, who may focus more on profit than protecting patients.

Unlike previous participants, insurance companies may comply with patient privacy laws, but the nature of insurance inherently conflicts with the interests of patients due to the amount of information needed about the patient's medical treatments and medical history. Insurance companies typically have flexible privacy policies for this reason and may even share information with third parties for data mining while legally still complying with privacy laws (Frankenmuth Insurance, 2020). Larger insurance companies such as Prudential (2017) state that "we may also use and disclose Protected Health Information for our health care operations" which typically includes hiring third-party vendors to process patient data. However, doing so passes the responsibility of patient privacy into the hands of the vendor, which can endanger patient privacy.

Some patients claim that healthcare systems have grown so complex that HIPAA is no longer sufficient. Healthcare providers must do their best to adhere to new regulations while still maintaining efficiency, while insurance companies neglect taking responsibility for patients. In light of Obama's HITECH Act in 2009 and the COVID-19 pandemic, patient privacy is now more important than ever. This can be examined through the Wicked Problem framework, given the lack of a definitive answer and the circular arguments given by the relevant parties and stakeholders. There are many combinations of solutions that can result from a compromise between parties, yet the debate over security will never be fully "solved" through the classic scientific approach (Rittel & Webber, 1973).

## References

AHDI (2020). Association for Healthcare Documentation Integrity Code of Ethics. Association for Healthcare Documentation Integrity. https://www.ahdionline.org/page/code_of_ethics

ACS (2020, March 23). American Cancer Society. Privacy Statement. https://www.cancer.org/about-us/policies/privacy-statement.html#security-protection/

Angst, C. M., & Agarwal, R. (2009). Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. MIS Quarterly, 33(2), 339. Web of Science.

Barrows, R. C., & Clayton, P. D. (1996). Privacy, Confidentiality, and Electronic Medical Records. Journal of the American Medical Informatics Association, 3(2), 139–148. Web of Science.

BMC (n.d.). Boston Medical Center. Patient Confidentiality, Privacy, and Security Awareness. Boston University Medical Center. Retrieved September 17, 2020, from http://www.bumc.bu.edu/isep/files/2014/08/HIPAA_Presentation.pdf

CDT (2009, April 21). Center for Democracy and Technology. Personal Health Records - is HIPAA the Answer? https://cdt.org/insights/personal-health-records-is-hipaa-the-answer/

Cidon, D. (2020, August 3). Patient Matching in the Era of COVID-19: Maintaining Control Over Patient Privacy and Data Governance. Journal of AHIMA. https://journal.ahima.org/maintaining-control-over-patient-privacy-and-data-governance/

Clark, M. (2020, October 21). Real-World Examples of Social Media HIPAA Violations. Etactics. https://etactics.com/blog/social-media-hipaa-violations

Confidentiality Coalition. (2020). Beyond HIPAA Principles. Confidentiality Coalition. https://www.confidentialitycoalition.org/about/beyond-hipaa-principles/

Empowered Patient Coalition. (2017, December 12). File A Privacy Complaint. https://empoweredpatientcoalition.org/report-a-medical-event/file-a-privacy-complaint/

Frankenmuth Insurance. (2020, August 4). Company Privacy Notice - Frankenmuth Insurance. https://www.fmins.com/company-privacy-notice/

Garcia, M., & Hindocha, A. (2020, June). Where Are We Now?: Examining the Trump Administration's Efforts to Combat Cybercrime. Third Way. JSTOR.

Gopalakrishna-Remani, V., Jones, R., & Wooldridge, B. (2016). Influence of Institutional Forces on Managerial Beliefs and Healthcare Analytics Adoption. Journal of Managerial Issues, 28(3/4), 191-209. JSTOR.

Gunter, T. D., & Terry, N. P. (2005). The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions. Journal of Medical Internet Research, 7(1), e3. Web of Science.

Hunt, P. (2019, March 5). Protecting the Privacy and Security of Your Health Information. APRA. https://www.americanpatient.org/protecting-your-privacy-security/

Ornstein, C. (2015, December 10). Small-Scale Violations of Medical Privacy Often Cause the Most Harm. ProPublica. https://www.propublica.org/article/small-scale-violations-of-medical-privacy-often-cause-the-most-harm

Pagliari, C., Detmer, D., & Singleton, P. (2007). Potential of electronic personal health records. BMJ: British Medical Journal, 335(7615), 330-333. JSTOR.

Pal, D., Arpnikanondt, C., Razzaque, M. A., & Funilkul, S. (2020). To Trust or Not-Trust: Privacy Issues With Voice Assistants. IT Professional, 22(5), 46-53.

Pridmore, J., & Mols, A. (2020). Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants. Big Data & Society, 7(1), 2053951719891748. Web of Science.

Prudential. (2017, November). HIPAA Notice of Privacy Practices. Prudential Financial. https://www.prudential.com/links/hipaa

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2), 155–169. https://doi.org/10.1007/bf01405730

RSNA (2020, July 17). Radiological Society of North America. Protecting Patient Information in Medical Presentations, Publications and Products. https://www.rsna.org/-/media/Files/RSNA/Practice-Tools/RemovingPHI.pdf

Sacks, E. (2018, May 26). Alexa privacy fail highlights risks of smart speakers. NBC News. https://www.nbcnews.com/tech/innovation/alexa-privacy-fail-highlights-risks-smart-speakers-n877671

Schick I. C. (1998). Protecting patients' privacy. Health information networks raise new questions. Health progress (Saint Louis, Mo.), 79(3), 26–31.

Whiddett, R., Hunter, I., Engelbrecht, J., & Handy, J. (2006). Patients' attitudes towards sharing their health information. International Journal of Medical Informatics, 75(7), 530–541.  Web of Science.