Enabling Machine Learning Prediction of Consumer Stalling in Checkout Process

Social Group Influence in the California Consumer Privacy Act and California Privacy Rights Act

A Thesis Prospectus

In STS 4500

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

By

Rishika Nayak

October 27, 2023

ADVISORS

MC Forelle, Department of Engineering and Society

Madhur Behl, Department of Computer Science

Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have become pervasive technologies in the realm of eCommerce. The adoption of AI in businesses has shown exponential growth, surging by 270% in just the last four years (Makarenko 2023). Currently, more than half (51%) of e-commerce businesses have incorporated AI into their operations and 84% of eCommerce businesses are either actively working AI solutions into their business (BusinessDIT 2023). As the use of AI is expanding in eCommerce, as is concerns regarding consumer data privacy and security. While AI/ML technology infiltrates various aspects of online shopping, the vast amount of consumer data collected and how it is being used has come under scrutiny. Consumers have a large stake in their own data privacy. According to the Cisco 2022 Consumer Privacy Survey, in a group of adults across 12 countries, including countries the Americas, Europe, and Asia Pacific, "62% of survey respondents expressed concern about how organizations are using their personal data for AI, with 65% saying that they have already lost some trust in organizations as a result of their AI use" (Waitman 2023). A notable instance of consumer data privacy concern occurred in 2018 when Facebook's unauthorized use of millions of users' personal data for political campaign came to light (Kozlowska 2018). While Facebook's misuse of data does not occur within the scope of eCommerce, it displays the unethical implications of enterprises exploiting or mishandling consumer information. The rapid growth of AI/ML and large-scale data collection in ecommerce calls for attention to ethical means of handling consumer information. There are several groups, including businesses and the government, that have the means to establish change in the way consumer data is collected and used. In this paper, I will be focusing on government-level regulation of consumer data privacy. Currently, in the US there is no federal legislation that regulates how data is used in AI/ML in

eCommerce; however, some states have adopted their own laws. In the United States, regulation regarding consumer data privacy began with the California Consumer Privacy Act (CCPA) in 2018, which "gives consumers more control over the personal information that businesses collect about them," according to the California Department of Justice (*California Consumer Privacy Act* 2023). Since then, thirteen other states have followed the lead of CCPA by adopting some form of similar data privacy legislation (Bea & VandenBerk Attorneys at Law 2020). As of January 1, 2023, the California Privacy Rights Act (CPRA) has amended CCPA to expand consumer rights with regards to how their data is used by businesses' AI/ML technologies. Per the State of California Department of Justice, CPRA outlines further legislation that "grants consumers rights regarding opting out of the processing of their personal information for purposes of profiling and create requirements that impact automated decision-making" (Yavorsky 2022). The government, both state and federal, have the ability to regulate data privacy in eCommerce. As the use of AI continues to grow in online commerce, the need for proper regulation grows as well. To better understand how data privacy and AI regulation will be implemented in the US in the future, it is important to examine the players that have a vested interest in such regulation. For my STS research project, I will research the following question: How have different stakeholders, such as consumer groups and companies, influenced the passing of CCPA and CPRA? In my technical research project, I will discuss a project I developed that enabled machine learning to deliver predictive analytics in an eCommerce platform.

Technical Research

At the eCommerce company I interned for in 2022, an issue persisted where customers would sometimes stall at certain points in the process of checking out. The company viewed

these "incomplete" checkouts as missed opportunities for potential sale of goods, meaning lower potential profit. An analyst team determined that around 50% of total customers entering checkout would drop the process. The solution my team proposed was to develop a machine learning model to predict how likely a customer is to stall at a phase of the checkout process. This customer stalling prediction integration adds onto the company's existing online purchasing process. The frontend team responsible for checkout will use this prediction to deliver a personalized nudge to the customer urging them to perform certain actions depending on what phase they are likely to stall at. For example, if the ML model predicts that the customer will pause at the "Finances" stage in the process, the user interface will prompt the user to ensure that the selected vehicle meets their budget. In a more holistic view of the implementation, the process would go as follows: as the customer selects "Proceed to Checkout" on the eCommerce website, the prediction from the ML model will be retrieved in real time, and the customer may be nudged with an appropriate prompt. The team I worked with foresees that this new feature will help facilitate the checkout process for the high rate of stalling that occurs when the checkout process is entered.

My role in this project was to develop the microservice, or backend architecture, that links the frontend to the customer stalling ML model. I used the framework Flask and the language Python in order to develop the microservice. In the development of this project, I implemented the following mechanisms: processing of the front-end request, API calls to retrieve relevant data, feature engineering of retrieved data into model input form, an API call to the ML model, and processing prediction output data for delivery to the front-end. When the front-end requires a model prediction, it makes a request to my service. The first functionality I implemented was the service's ability to accept a request from the front-end and store the

customer data that the front-end sends. I then parsed the customer data to construct appropriate inputs to the required data provider APIs. The machine learning model required additional data from these data providers in order to make a prediction. In making the API calls to data providers, I had to use proper authentication credentials in order to access private customer data from the company's internal databases. After retrieving the response from the APIs, I serialized the data to prepare for the feature engineering stage. Feature engineering involved the extraction of features, or appropriate model input data, from the raw data provided by the APIs. Finally, in the last stage of development, I input the extracted features in a call to the ML model to retrieve the prediction. I transformed the prediction data into a readable format for the front-end and returned the prediction to the front-end for use.

Upon completion of my project and internship, my team integrated the microservice I developed into the checkout process nudging system. In follow up with this project, the microservice could be improved to be able to handle large volumes of front-end requests or work in faster time as it is meant to deliver predictive analytics in real-time. My technical project relates to the STS problem of data privacy in eCommerce as the company I interned for would be an example of a stakeholder with interest in AI regulation legislation. AI allows eCommerce businesses, such as mine, to predict consumer buying behavior, facilitate the shopping process for customers, increase the percentage of sales, and more (Hassan & Abdulkhaleq 2022). Therefore, due to capabilities of AI in enhancing online business, AI-driven eCommerce companies have reason to hold stake in consumer privacy regulations.

STS Topic

The focus of my research lies in addressing the growing issue of data privacy concerns in AI-driven eCommerce. The sociotechnical dimension delves into how various social groups and

stakeholders influence AI technology in eCommerce through the shaping of legislation. My research will investigate how different stakeholders, such as consumers and companies, influence data privacy and AI regulation, specifically regarding the California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA). I will be performing my research through the lens of the social determinism framework and the Social Construction of Technology (SCOT) approach. Social determinism is the idea that technical artifacts are interwoven with political qualities, in that social circumstances determine how a technology is developed, deployed, and used (Winner 1980). My research aligns with the social deterministic viewpoint because I am examining the influence of social groups upon the way data is used and AI technology is developed in eCommerce. Furthermore, the SCOT approach outlines how the development and stabilization of technical artifacts is not a linear process, but rather a multi-dimensional process influenced by the competing needs of various groups (Pinch & Bijker 1984). Considering AI/ML in eCommerce as the technical artifact, my research will follow the SCOT approach by evaluating the needs of various social groups and how those needs shape the technical artifact. Social groups, such as consumer interest groups and companies, have the power to shape legislation through various avenues such as lobbying, media exposure, and donating to certain candidates' campaigns (Cornish & Torres-Spelliscy 2021).

The California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA) offer valuable context for examining the societal pressures upon data privacy in AI-powered eCommerce. The CCPA outlines detailed data privacy regulations, granting consumers rights over their personal data, including knowing what's collected and shared, opting out of data sales, and requesting data deletion. The CPRA builds on the CCPA with additional rights, such as the right to correct inaccurate personal information and the right to limit the use and disclosure

of sensitive personal information, strengthening consumer data privacy (California Department of Justice 2023). The CCPA and CPRA significantly impact AI in eCommerce by requiring businesses to adapt their data collection and processing practices. For example, as stated by the State of California Department of Justice, since CCPA took effect on July 1, 2020, several online retailers were found sharing consumers' personal data with third parties for services like advertising and analytics. They did so without offering an opt-out choice or verifying the third parties' CCPA compliance. Upon being alerted to their noncompliance, these companies introduced technology to signal "restricted use" to third-party recipients of personal data and blocked infringing data transfers (*CCPA Enforcement Case Examples 2021*). Therefore, through shaping legislation, social groups have the power to influence the development and deployment of technologies in eCommerce.

Examining how social groups have shaped the CCPA and CPRA is important in understanding how similar legislation may be shaped in the future. In order to explore this topic, I will research the specific stakeholders who had vested interest in CCPA/CPRA and identify their needs. To understand the influence these stakeholders had, I will investigate the actions that were taken to shape the mentioned legislations. I will make observations through the perspectives of social determinism and SCOT in order to understand the influential factors behind CCPA/CPRA and future data privacy legislation.

Conclusion

I successfully implemented my technical project: a microservice, enabling an ML model to deliver real-time, customer-centric predictive analytics during the checkout process. This project enables the eCommerce company to nudge consumers to complete relevant actions before entering checkout in order to streamline the sales process. My STS research project aims

to offer insight into how certain social groups influence data privacy legislation in the United States. Understanding how specific social groups influence data privacy legislation provides a basis for achieving better AI regulation in eCommerce in the United States. These groups often represent diverse interests and perspectives, including those of consumers, businesses, privacy advocates, and industry stakeholders. By understanding their influence, policymakers can design regulations that attempt to balance consumer data protection, business innovation, and ethical AI practices.

References

California and Europe Lead the Way on Data Privacy – More States to Follow – Bea & VandenBerk Attorneys at Law. (2020). Bea & VandenBerk Attorneys at Law. Retrieved October 27, 2023, from https://www.beavandenberk.com/ip/computer-internet/california-and-europe-lead-the-way-on-data-privacy-more-states-to-follow/

California Consumer Privacy Act (CCPA) | State of California - Department of Justice - Office of the Attorney General. (2023, May 10). California Department of Justice. Retrieved October 27, 2023, from https://oag.ca.gov/privacy/ccpa

CCPA Enforcement Case Examples | State of California - Department of Justice - Office of the Attorney General. (2021, July 19). California Department of Justice. Retrieved October 27, 2023, from https://oag.ca.gov/privacy/ccpa/enforcement

Cornish, A., & Torres-Spelliscy, C. (2021, April 2). *The Power Corporations Have In Changing Laws*. NPR. Retrieved October 27, 2023, from https://www.npr.org/2021/04/02/983925056/the-power-corporations-have-in-changing-laws

Hassan, A., Abdulkhaleq, N.M.S. (2022). *The Use of Artificial Intelligence in the Field of Electronic Commerce.* In: Musleh Al-Sartawi, A.M.A. (eds) Artificial Intelligence for Sustainable Finance and Sustainable Technology. ICGER 2021. Lecture Notes in Networks and Systems, vol 423. Springer, Cham. https://doi.org/10.1007/978-3-030-93464-4_21

Kozlowska, I. (2018, April 30). *Facebook and Data Privacy in the Age of Cambridge Analytica*. Jackson School of International Studies. Retrieved October 27, 2023, from https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/

M., Y. (2023, May 16). *AI in eCommerce Statistics, Benefits & Trends, Example 2023*. BusinessDIT. Retrieved October 27, 2023, from https://www.businessdit.com/ai-in-ecommerce/

Makarenko, E. (2023, September 26). *AI in eCommerce: Benefits, Statistics & Facts, Use Cases and Deployment*. Master of Code Global. Retrieved October 27, 2023, from https://masterofcode.com/blog/state-of-artificial-intelligence-ai-in-ecommerce-statistics-and-deployment

Pinch, T. J., & Bijker, W. E. (1984, August). The Social Construction of Facts and Artefacts: On How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. Social Studies of Science, 14(3), 399–441. http://www.jstor.org/stable/285355

Waitman, R. (2023). *Cisco Consumer Privacy Survey*. Cisco. Retrieved October 27, 2023, from https://www.cisco.com/c/en/us/about/trust-center/consumer-privacy-survey.html?CCID=cc000742

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.
http://www.jstor.org/stable/20024652

Yavorsky, S. (2022, August 11). *New U.S. State Privacy Laws Zero in on Artificial Intelligence*.
Orrick, Herrington & Sutcliffe LLP. Retrieved October 27, 2023, from
https://www.orrick.com/en/Insights/2022/08/New-State-Privacy-Laws-Zero-in-on-AI