

Measuring the Prevalence and Variety of Online Age Gates

Tajveer Singh Dhesei
Department of Computer Science
Colgate University
Hamilton, New York 13346
Email: tsinghdhesi@colgate.edu

Noah Apthorpe
Department of Computer Science
Colgate University
Hamilton, New York 13346
Email: naphthorpe@colgate.edu

Abstract—The legal landscape regarding age-based restrictions (age gates) for online services is rapidly changing. In order to comply with existing and proposed regulations, online services must determine whether users are older or younger than mandated age thresholds. The implementation details of these age gates are highly relevant for consumer protection advocates given the risk of user circumvention and/or chilling effects. We therefore propose a study measuring the prevalence and variety of age gating mechanisms across the Internet. We start with a case study of the e-cigarette industry, finding that nearly all site arrival age gates merely require users to self-attest that they are older than an age threshold. We plan to expand this study to additional industries, website interaction points, and automated classification techniques to produce a comprehensive assessment of online age gating practices.

1. Introduction

Policies and policy proposals mandating age restrictions online, especially targeting social media platforms, have become more popular in recent years [1]. In the United States, California’s 2022 “Age Appropriate Design Code Act” and New York State’s 2024 “SAFE for Kids Act” seek to limit potentially harmful design practices on social media platforms for underage users. Several European nations statutorily limit social media access to children under 13 to 16 absent parental consent [2], and Australia’s 2024 social media ban seeks to completely prevent access to social media for children under 16 [3].

These and similar laws ostensibly require that online services determine whether users are older or younger than mandated age thresholds. This is a difficult technical problem [1], [4], [5]. Reliable verification of user attributes is notoriously difficult online and is made even more so by the accessibility of anonymous browsing tools and incentives for circumvention by users outside of allowed age ranges. Incorrect age determination may infringe on age-appropriate users’ speech, raising the stakes for reliable implementations. Age checking may also result in privacy violations [6] or chilling effects [7], motivating thoughtful design of age gating mechanisms.

The combination of increased policymaking and these technical challenges mean that regular, large-scale audits of

deployed age gating technologies will be essential to ensure that regulators and consumer rights advocates base their decision-making about online age restrictions on the actual state of the Internet rather than on assumptions or anecdotal examples.

We therefore propose a research study to measure the prevalence and variety of online age gating mechanisms across the Internet. Specifically, we seek to determine the extent to which deployed age gates fall into three previously defined categories with different implications for correctness, privacy, circumvention risks, and chilling effects [1]: 1) *self-attestation*, in which “the user provides their age (or birth date) without any verification by the online service,” 2) *age estimation*, in which “the online service uses some available information about the user to attempt to infer their age,” and 3) *age verification*, in which “the online service uses information about the user in conjunction with another trusted source (e.g. a government ID, credit card information, etc.) in order to confirm the user’s age.” We then plan to examine the correctness, privacy, and circumvention implications of specific versions of age gates within these broad categories.

To demonstrate the feasibility of this research, we begin with a case study on the e-cigarette industry, itself the subject of recent U.S. regulation [8]. A 2023 study found that 76.3% out of 64 analyzed online vape shops in Washington DC allowed users to reach the checkout page without verifying they were over 21, exposing gaps in current age restriction laws and their enforcement [9]. Our case study expands on this prior work, as we identify and classify 103 site arrival age gates across 243 international e-cigarette websites. The following sections describe these findings and how we intend to scale our approach to produce a comprehensive assessment of online age gating practices.

2. E-Cigarette Case Study

We filtered the top 10 million website domains from the Open PageRank initiative [10] down to the 534 domains that contain the word “vape.” We used the Selenium web scraping tool [11] with Google Chrome to visit each of these sites and take a screenshot on site arrival for offline analysis.

We manually classified these screenshots, finding that 243 were actually e-cigarette websites and 103 of these contained an age gate. We then divided these age gates into self-attestation, age estimation, and age verification categories. **An overwhelming majority (100/103, > 97%) of age gates required only self-attestation to bypass** (Figure 1). Less than 3% of age-gates required an account login, which we conservatively classified as age verification given that some verifiable information might be collected during account creation. No observed age gates used any form of age estimation.

We further divided the self-attestation age gates into subcategories based on the user actions required to pass the gate (Figure 2). The most common (81/100, 81%) required users to click a button confirming that their age is above a listed threshold (e.g., Figure 3). Far less common versions required the user to enter their date of birth in a series of dropdown fields (10/100, 10%) or required users to click a button confirming that they are of “legal age” without specifying the exact age threshold (7/100, 7%). Finally, a few listed an age threshold but did not make the user take any actions, merely assuming that the user was above the threshold if they proceeded on the site (2/100, 2%).

These results indicate that the current state of site arrival age gating in the e-cigarette industry amounts to little more than security theater. The self-attestation age gates we observe are trivial to bypass and provide no real barrier to circumvention. It is clear that these gates are in place for minimal legal compliance by e-cigarette companies rather than to meaningfully prevent access by underage users.

These initial findings also support the effectiveness of our screenshot-based web scraping approach to measuring the prevalence and variety of online age gates. We therefore intend to expand this approach into a large-scale cross-industry study as described in the following section.

3. Future Directions

The software we developed for the above case study is easily applicable to websites from any industry, as it automatically takes screenshots on specified websites of interest and stores them for offline analysis. No HTML parsing nor JavaScript analysis is needed. The screenshots also accurately reflect what a user would experience during a website visit. We plan to apply this tool to many additional industries, including social media platforms and gambling websites as motivated by proposed regulation [1], [12].

We plan to improve the comprehensiveness of the study to include age gates on website interaction points beyond site arrival. Navigation to the “store” section of a website, adding a product to a cart, account creation, account login, checkout, and content posting are all reasonable moments for sites to apply an age gate. We will update our web scraper such that it automatically navigates to and takes screenshots of these interaction points for offline analysis.

As we scale the study to additional industries and more website interaction points, we plan to use machine learning methods, such as fine-tuned large language models (LLMs)

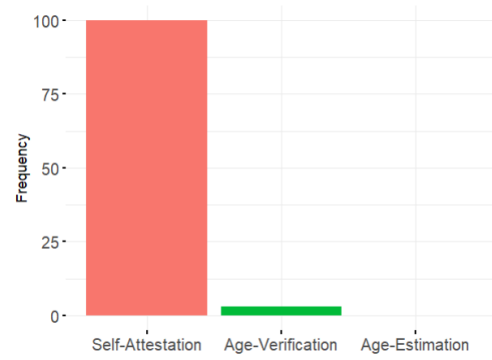


Figure 1. Distribution of site arrival age gates on e-cigarette websites by age determination method. Categories from [1].

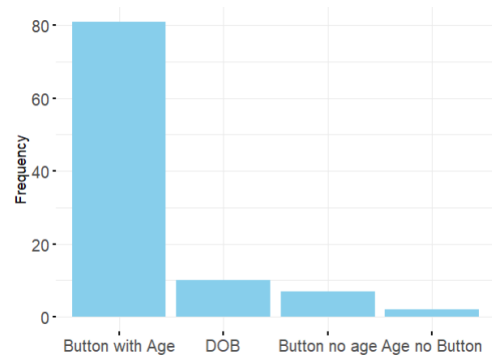


Figure 2. Variety of self-attestation site arrival age gates on e-cigarette websites. All types require no external validation of user age.

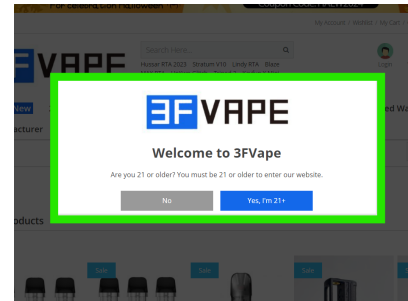


Figure 3. Example self-attestation age gate requiring a button click with the threshold age listed.

with appropriate accuracy verification, for automatic age gate classification. Preliminary tests suggest that LLMs can quickly and correctly identify the presence of age gates in website screenshots and determine whether the age gate uses self-attestation, age estimation, or age verification.

We also plan to identify the presence of user interface dark patterns [13] across age gates, such as nudges toward bypassing the gate or opportunities to retry the gate shortly after site access is denied. Finally, we hope to measure the prevalence of third-party age gating services, which may necessitate further audits.

References

- [1] N. Aphthorpe, B. M. Frischmann, and Y. Shvartzshnaider, "Online age gating: An interdisciplinary evaluation," *Available at SSRN 4937328*, 2024.
- [2] Reuters, "What countries do regulate children's social media access?" Nov. 2024, accessed: 2024-12-12. [Online]. Available: <https://www.reuters.com/technology/what-countries-do-regulate-childrens-social-media-access-2024-11-28>
- [3] G. Conroy, "Governments are banning kids from social media: will that protect them from harm?" *Nature*, Dec 2024.
- [4] C. Jarvie and K. Renaud, "Are you over 18? a snapshot of current age verification mechanisms," in *2021 Dewald Roode Workshop*, 2021.
- [5] A. Adib, W.-P. Zhu, and M. O. Ahmad, "Improvising age verification technologies in canada: Technical, regulatory and social dynamics," in *2023 IEEE International Humanitarian Technology Conference (IHTC)*. IEEE, 2023, pp. 1–6.
- [6] S. Scheffler, "Age verification systems will be a personal identifiable information nightmare," *Communications of the ACM*, vol. 67, no. 7, pp. 31–33, 2024.
- [7] J. W. Penney, "Chilling effects: Online surveillance and wikipedia use," *Berkeley Tech. LJ*, vol. 31, p. 117, 2016.
- [8] Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF), "Vapes and e-cigarettes," 2024, accessed: 2024-12-12. [Online]. Available: <https://www.atf.gov/alcohol-tobacco/vapes-and-e-cigarettes>
- [9] A. Bertrand, M. C. Diaz, E. C. Hair, and B. A. Schillo, "Easy access: identification verification and shipping methods used by online vape shops," *Tobacco Control*, 2024.
- [10] DomCop, "Top 10 million websites based on open data from common crawl & common search," 2021. [Online]. Available: <https://www.domcop.com/top-10-million-websites>
- [11] Software Freedom Conservancy, "Selenium," 2024. [Online]. Available: <https://www.selenium.dev/>
- [12] Reuters, "New zealand plans to limit online casino gambling licenses, ban ads aimed at children," *Reuters*, November 13 2024. [Online]. Available: <https://www.reuters.com/world/asia-pacific/new-zealand-plans-limit-online-casino-gambling-licenses-ban-ads-aimed-children-2024-11-13/>
- [13] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, "The dark (patterns) side of ux design," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–14.