Toward Transparent Web Browsing: A Design for a Privacy and Data Fairness Assessment Tool

Sampsa Rauti[1,*], Sini Salmi[1], Panu Puhtila[1], Shashika Harshani[1] and Sammani Rajapaksha[1]

[1]University of Turku, Vesilinnantie 5, 20014 Turku, Finland

Abstract

In this study, we present a conceptual design for a web privacy assessment tool, which is a browser extension promoting privacy and fair data practices. The tool gives the user a near real-time view on network traffic and data processing of a website. The solution makes use of large language models (LLMs) to analyze network traffic sent to third parties like analytics services and advertisers in the web environment. It detects leaks of identifying and contextual personal data, such as details related to the user's health. The tool assesses the transparency of the privacy policy document provided on the website in order to see whether the user is adequately informed about the data shared with third parties. The tool also has the ability to detect dark patterns on cookie consent banners. Our solution is targeted for end users, and it is meant to be an intuitive and usable tool providing clear and timely information about the occurring data leaks. The novelty of the proposed solution lies in combining several features. It aims to offer near real-time notifications for users, puts emphasis on the sensitive contextual data leaks, has the ability to detect discrepancies between the actual network traffic and the privacy policy using AI, and provides a concise summary of data leaks detected on the analyzed website, as well as an assessment of the privacy and fairness of data processing practices. Our solution is informational rather than preventive by design. It increases the transparency of data processing and supports the user's own decision-making when it comes to data protection.

Keywords

Privacy tool, data leaks, transparency, web privacy, fair data, third-party services

1. Introduction

The data collection taking place in the web environment is increasingly common and invisible [1, 2]. Many websites employ a wide array of third-party services that monitor the user's browsing behavior and share this data with third parties, such as big tech companies and advertising networks. Such covert data collection can also include gathering highly sensitive personal data, for example, in the form of visited websites or search terms. At the same time, users do not usually get a transparent understanding of what kinds of personal data is collected on them, where does it end up, and how well privacy policies on websites correspond to actual data collection practices [3, 4].

In this paper, we present a design for a browser extension that aims to improve the transparency and fairness of data processing on websites. The tool analyzes the network traffic in real time and notifies users if their personal data is leaking to third parties. In addition, the proposed tool examines the privacy policy of the studied website and compares its contents to the actual network traffic. It also analyzes cookie consent banners and detects dark patterns, which are deceptive design choices in the user interfaces [5, 6].

The proposed solution differs from earlier solutions due to the unique set of features it offers. The key idea is to combine the semantic analysis of LLMs with a user-friendly, easily installable browser extension for assessing online privacy. It especially focuses on sensitive contextual data leaks, such as a website leaking health data to third parties. The ability to detect discrepancies between actual network

SQAMIA 2025: 12th Workshop on Software Quality Analysis, Monitoring, Improvement, and Applications, September 10–12, 2025, Maribor, Slovenia

*Corresponding author.

sjprau@utu.fi (S. Rauti); siosal@utu.fi (S. Salmi); papuht@utu.fi (P. Puhtila); shhaku@utu.fi (S. Harshani); syraja@utu.fi (S. Rajapaksha)

0000-0002-1891-2353 (S. Rauti); 0009-0008-0691-5988 (S. Salmi); 0009-0004-6418-1063 (P. Puhtila); 0009-0003-1479-0170 (S. Harshani); 0000-0003-4647-3885 (S. Rajapaksha)

traffic and data collection reported in the privacy policy and provide real-time data leak notifications as well as digestible privacy and transparency summaries and scores for the user make the proposed solution novel. Instead of preventing data leaks, our proposed solution is designed to adequately inform the user about the leaks, improving transparency and making more informed decisions possible.

The remainder of the current study is structured as follows. First, we discuss third-party data leaks and the motivations for proposing a novel design for the tool. Next, we look at the previous tools and solutions with similar features and goals. Then, we present the tool's architecture and its essential components. Finally, we explore the technical challenges related to the tool and the ways to address them in the design and implementation of the browser extension.

2. Data Collection by Third-Party Services

In the last few decades, there has been a clear shift towards digital platforms and online business models. As this trend gains momentum, businesses and organizations have begun to make use of web analytics to collect data on their customers to optimize their websites to meet business goals and make data-driven decisions. Therefore, on modern websites, there are many kinds of third-party services that analyze users' browsing behavior, measure the website's performance, and collect demographic information about visitors [7, 8].

An important reason for using web analytics is tracking conversions [9, 10]. In online marketing, a conversion is a desirable action taken by a web visitor, meaning that positive engagement with the website or a successful outcome of marketing strategy is reached through a user's actions. Therefore, a conversion is something that produces value for the company. By making use of web analytics, it is possible to set conversion goals for websites and monitor users' behavior that leads to successful conversions. Examples include submitting a web form or placing an order in an online store.

Many non-commercial websites also make use of the same analytics services as online businesses [3, 11]. The organization behind these websites might not be interested in business goals, but they still track conversions. As Bekos et al. [12] put it, "actions the website has configured to be tracked are defined as conversions". Viewed in this light, using a search function can also be defined as a conversion and searches are often very interesting to web analytics. The visited pages are also often tracked by the third-party services. They can be used in a process called funneling, which involves tracking and analyzing the steps a user takes towards a specific goal or conversion [13]. A user reaching an important page on a website can also be seen as a conversion.

While monitoring browsing behavior can be beneficial for the website maintainer, third parties tracking users and collecting data also cause serious privacy concerns. Sensitive search terms, visits to delicate pages, and many other private actions (for instance, purchasing a prescription medicine in an online pharmacy) can leak to third parties. Figure 1 depicts such a situation. Third-party analytics services, shown as red boxes, collect identifying data (like IP addresses and device identifiers) and contextual data about the user's page visits and behavior on the website. The collected data is then sent to the external servers of third-party analytics companies, like Google and Meta. Often, organizations and the developers of the websites do not seem to fully understand that serious data leaks are happening on their web services. Consequently, users are not adequately informed about these data leaks.

Contextual data, such as URL addresses and search terms, leaking to third parties would not be a privacy problem by itself if it could not be linked to identifying personal data. However, the fact that these two categories are combined in HTTP requests makes data leaks dangerous. For example, an individual user can usually be identified by their IP address. When this identifying personal data is sent to a third party along with sensitive contextual data, like the user's medical conditions, third parties may be able to infer the state of the user's health and, with enough time, their patient history.

All third parties cannot link identifying personal data, such as IP addresses, to a specific user's identity (real name), but for large technology companies, such as Google and Meta, this is often possible. The users often use the same device to login to the services run by these companies, which makes it possible to connect the IP address to the user's real name. Moreover, these companies often store

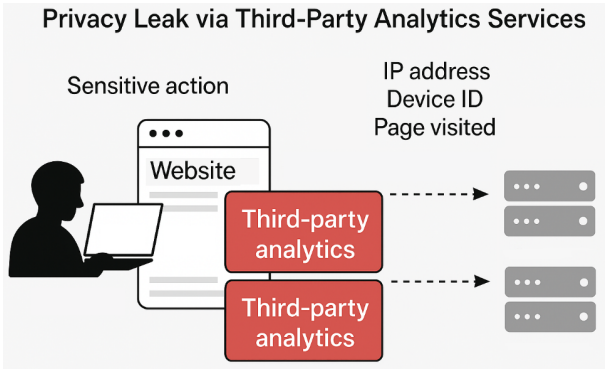

Figure 1: Third-party analytics services, shown as red boxes, collect identifying data (like IP addresses and device identifiers) combined with contextual data (such as the URL addresses of the pages the user has visited on the website).

cookies containing unique identifiers on the user's computer for extended periods of time, which helps them to keep track of the user's identity.

An example of a sensitive third-party data leak would be an online pharmacy leaking names of prescription medicines the user has viewed or ordered, combined with the user's IP address and possible real name. This would make it possible for an external party to infer the user's private medical conditions or treatments, which is a serious privacy violation. The users are exposed to potential medical profiling and other misuse of their health-related personal data.

The tool design we propose in this paper makes invisible, stealthy data leaks visible to the user. Performing network traffic analysis using the browser's developer tools and going through HTTP requests manually is laborious and not practical for an ordinary user. To get an adequate understanding of data processing activities, an ordinary user simply needs a tool that delivers real-time, transparent information about privacy matters while browsing the web.

3. Existing Tools and the Motivations for a Novel Privacy Extension

A significant move can be seen in recent research in the context of automatically evaluating privacy practices of web services. One central approach is comparing the privacy policy analysis with actual network traffic monitoring. As an example, the OVRSEEN tool by Trimananda et al. [14] captures and inspects the network traffic and compares this empirical evidence against privacy policies declared in the website. Our proposed tool's methodology closely aligns with the strategy that OVRSEEN follows, and it underscores the need for a more user-friendly and accessible web browser extension implementation. Andow et al.'s [15] work on "Poli-check" introduces an "entity-sensitive" model that identifies the discrepancies between data collected by the first party and data shared with third parties, such as advertisers and analytics providers.

The technological motivation behind the proposed tool is the innovative use of LLMs to conduct advanced analysis automatically. Previously, such use of LLMs have been reported in literature in the domain of memory leak detection, such as the LAMED system by Shemetova et al. [16]. Their study demonstrates that LLMs can efficiently generate "context-aware annotations" for static analysis, which traditionally requires a large amount of manual effort.

The third component of our tool design is partly influenced by the VeraCookie tool by Jo et al. [17] with its primary focus on identifying dark patterns in cookie consent banners. Although VeraCookie's technological implementation lies on machine learning, including computer vision and natural language processing, its successful demonstration of dark pattern recognition gives a possible basis for our design.

The novelty of our tool stems from the following set of features:

- *Semantic analysis of network traffic using LLMs.* The design of the tool enables the use of LLMs that can semantically analyze third-party HTTP requests in real time. Therefore, the analysis does

not only rely on heuristics or static rule sets, but the tool can analyze contextual data, for example, in URLs and search terms. Therefore, by analyzing natural language, it can determine whether sensitive personal data, such as health symptoms or political views, is leaking to third-party domains. This kind of context-aware leak detection can greatly aid in adequately informing users about data processing and data leaks.

- *Near real-time data leak detection.* The designed tool monitors network traffic and data flows to third parties. The third parties receiving personal data and the types of data leaked are identified – with special emphasis on sensitive data categories, such as data concerning health or religious beliefs. The user immediately receives notifications when data leaks happen, clearly and transparently explaining what kind of data is leaking and to where.
- *Browser extension for end-users.* There are many academic tool prototypes that analyze network traffic, but our tool is targeted at normal web users (and can also be used by developers), designed as a browser extension with a user-friendly interface. The goal is to provide clear and actionable information for ordinary users without any prior technical knowledge. By building a practical and deployable tool, its real-world impact is increased.
- *Privacy policy transparency analysis.* The tool automatically finds and parses a website's privacy policy document. The contents of the document are analyzed with an LLM. The information given about third parties and personal data leaked to them in the privacy policy document is then compared with the actual observed network traffic. The discrepancies between the privacy policy and the actual data flows (for example, missing disclosures about third-party trackers) are detected and the information is conveyed to the user. The website is also given a transparency score based on how honest the policy is about data sharing practices, which is then displayed to the user.
- *Privacy and fairness scores.* The observations the tool has made are summarized into simple scores representing a website's privacy and fairness of data processing. Metrics like the number of third-party data transfers, the nature of the leaked personal data, and the transparency of the analyzed privacy policy documents are taken into account in these scores. This helps users make more informed decisions when browsing websites.
- *Dark pattern detection.* By using both computer vision and LLMs, the tool finds deceptive interface elements in cookie consent banners. For example, hidden or absent "Reject All" buttons, misleading color contrasts, and pre-ticked checkboxes are detected. Warning users about these dark patterns makes them visible, and users become aware that the website tries to manipulate them.
- *Support for multiple LLMs.* The modular design and use of interfaces make it possible to replace and compare different LLMs as necessary. Therefore, LLMs ranging from cloud-hosted solutions to on-device models can be used. Studying challenges like LLM performance, possible hallucinations, costs, and response time is important when choosing a model to be used with the actual tool implementation. Adaptability like this supports changing the LLM and experimenting with different options as they keep evolving.
- *Focus on informing users.* Our tool design does not involve removing third parties from a website or blocking network traffic to them. Instead, the idea is just to inform users and make personal data sharing more visible. The users get a detailed, near real-time overview of data leaks, inconsistencies in privacy policies, and deceptive design practices.

4. Tool Design

This section presents a design for a browser extension designed to provide the users (and developers) with transparent insights into the privacy and data fairness of the website that is being browsed. The main goal of the extension is to provide a transparent view into data processing practices and potential third-party data leaks that take place on the website the user is browsing. The extension aims to present user-friendly interfaces and provide an easily understandable assessment of the privacy and transparency of the website.

An example of how the tool works could be the following. The user visits a mental health website and is informed of the dark pattern on the website's cookie banner, as the "Reject all" button does not exist. The user uses the search option on the website and searches for "anxiety". The tool detects the search term and the user's IP address are collected by Google Analytics. The tool also notes that the privacy policy on the website does not mention sharing this data to any third party. The user is alerted in real time about this data leak. This sensitive data leak also affects the privacy score that is shown to the user.

4.1. Core Components

The tool has three core components: 1) Third-Party Data Leak Detector, 2) Privacy Policy Transparency Analyzer, and 3) Dark Pattern Detector for Cookie Consent Banners. These components are discussed next.

Third-Party Data Leak Detector

The first component monitors network traffic in real time and uses an LLM to analyze third-party HTTP requests of the website the user is browsing. Particularly, it singles out the third parties receiving sensitive personal data and personal data types being sent to third-party services.

The user of the tool gets near real-time notifications of leaked personal data and the third parties receiving it, especially in the case of sensitive data (e.g. when data concerning health or political opinions leak). As we have seen previously, these leaks often manifest in the form of URLs or sensitive inputs, such as search terms, leaking. Therefore, the user may receive a notification that the medical symptom they inputted as a search term was just leaked to a specific third party, for example.

In addition to alerting the user, a view listing the detected third parties and personal data items sent to them is provided. The tool also gives a short, easily understandable summary of the leaked personal data and its destinations. A simple privacy score is given to the website based on the number of third parties personal data leaks to and the number of leaked sensitive data items.

For a tool analyzing network traffic and looking for personal data, it is important to also determine what kinds of leaks and personal data types the tool needs to detect when analyzing HTTP requests. When it comes to data leak types (the different ways personal data leaks to third parties), previous research has shown that the URL addresses visited by the user and user inputs (especially search terms) in particular are worth analyzing [18, 19]. They often contain hints of sensitive actions the user is taking on the website. There are also specific actions, such as adding a product to a shopping cart and clicking a specific button, that are worth analyzing.

Regarding the types of personal data leaking, we focus on two main types of personal data, as discussed in Section 2. First, identifying personal data, such as IP addresses and device identifiers, should be considered. There are also more traditional data items directly connected to the identity of the user, such as the name, street address, email address, and credit card number. Second, there are many kinds of contextual data, which cause problems when leaked together with identifying personal data. Special categories of personal data, as outlined in Article 9 of the GDPR, are especially sensitive [20]. These include, for example, data concerning health, such as symptoms, mental health status, clinics, medical conditions or diagnoses, medication or treatment details, and laboratory results. In addition, political or religious beliefs and sexual orientation or behavior are considered particularly sensitive personal data according to the GDPR.

Figure 2 shows a UI mockup of the tool, mainly concentrating on the data leaks. It shows a general privacy score based on the number of detected third parties and leaked personal data items, as well as their sensitivity. It also lists the third parties and leaked data items.

Privacy Policy Transparency Analyzer

The second component, the privacy policy transparency analyzer, analyzes privacy policies with an LLM and compares their content to the actual network traffic and third-party communications detected by

the previous component. More specifically, the transparency of a privacy policy is analyzed by testing whether it mentions 1) all the third parties actually receiving personal data, and 2) the categories of personal data leaked to third parties (in particular, sensitive categories such as data concerning health).

Therefore, the user can use this tool to complement lacking privacy policies and to see how transparent they are, but it can also be useful for developers to see what kinds of third parties there are on their website and the flaws their privacy policies have when it comes to adequately informing the user. In addition to evaluating transparency, the tool gives a score for the clarity of the privacy policy. For example, the Flesch-Kincaid readability test [21] is a possible option for assessing the clarity of the analyzed privacy policies.

A simplified summary can be generated of the privacy policy of the analyzed website. This summary can increase transparency and make it easier for an ordinary user to digest the data processing practices of the website, especially if the policy is complicated, unclear, or vague. It allows the users to make more informed decisions about their personal data and whether they want to use a specific website.

Dark Pattern Detector for Cookie Consent Banners

The third main component of the tool automatically analyzes the cookie consent banner on the target website to find dark patterns. In this context, dark patterns refer to deceptive UI elements that mislead the user into giving their consent to cookies and data collection [22] and supplant user value in favor of shareholder value [23]. The topic of automatically detecting the use of dark patterns has been studied in a number of different publications [24, 25, 26, 27, 28, 29, 30, 31] presenting several possible ways to solve this problem. The presented methods include, for example, text based-detection [24] and combining computer vision and natural language processing [31].

When implementing a solution for detecting dark patterns, methods based on this earlier research can be used. While our design intentionally leaves the exact implementation of this component open, current research indicates combining computer vision and LLMs is a promising approach [17]. The dark pattern detection functionality will then be combined with a front-end functionality that presents the user with this information through visual cues, such as highlighting the parts of the cookie consent banner that contain dark patterns. The idea here is to provide the user with a clear and easily accessible

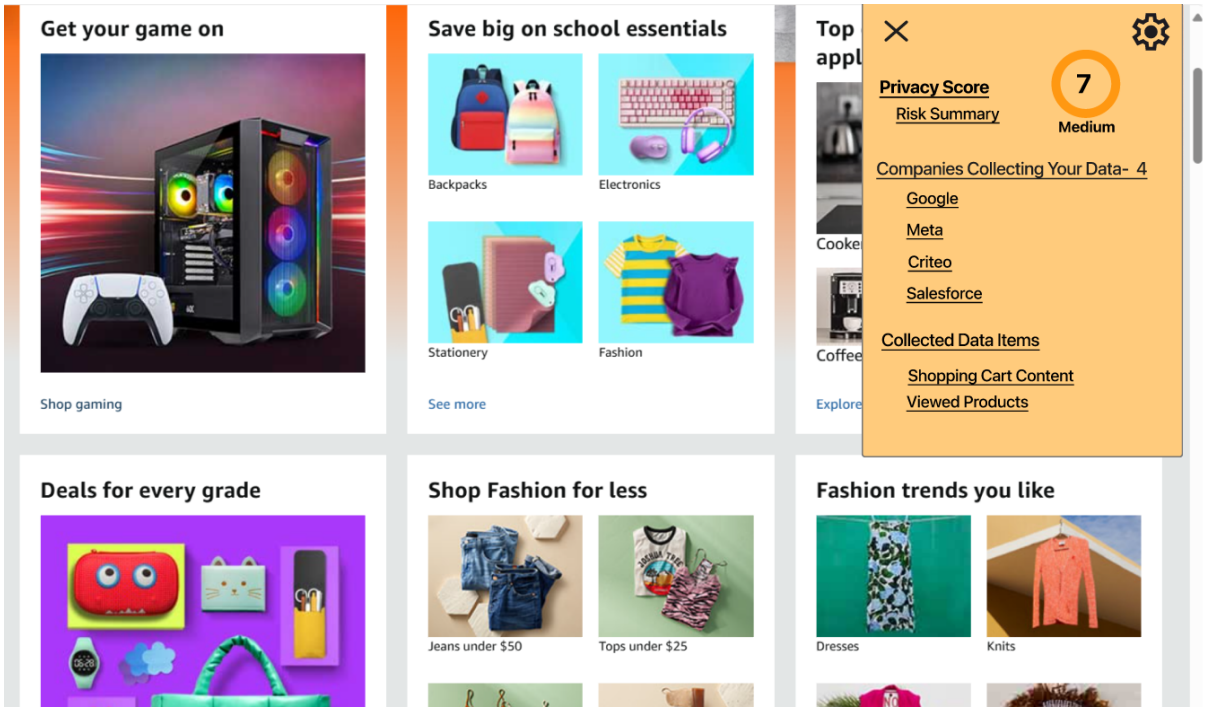

Figure 2: A UI mockup of the tool, displayed on the top of website the user is browsing.

way to recognize dark patterns, and, based on this recognition, make a better decision in regard to giving consent to data collection.

The most common and important dark patterns to detect are the following:

1. *Hidden rejection buttons*: The situation where the users ability to reject data collection and/or cookies is actively hidden by using a cookie consent banner that consists of several stacked layers or tabs. This also includes the situations where the rejection button does not exist at all.
2. *Pre-selected options*: These are situations where the user is presented with the ability to accept/reject specific categories of cookies or data collection, but where the selector elements for these choices are by default set to accepting the data collection.
3. *Deceptive color schemes or button sizes*: This means the use of psychologically appealing colors and contrasts and other visual elements that can be used to manipulate user attention and decision making. For example, situations where the acceptance button is of bright and positive appearance, while the rejection button is gray and dull looking.

Figure 3 shows the architecture of the designed tool as a component diagram, with the three main components in the center, and the supporting components surrounding them. For example, all main components feed information to the popup that displays the results to the user. Network traffic analysis, privacy policy analysis and dark pattern analysis all make use of the LLMs through an adapter interface.

4.2. LLMs and Prompting

The tool design implements an adapter layer to easily manage changes of the used language model. This is important as new language models keep surfacing and the existing models continue to develop. The design of this layer is based mainly on the LangChain framework to integrate LLMs. LangChain offers an API, which makes it easy to switch between different LLMs without changing the code of

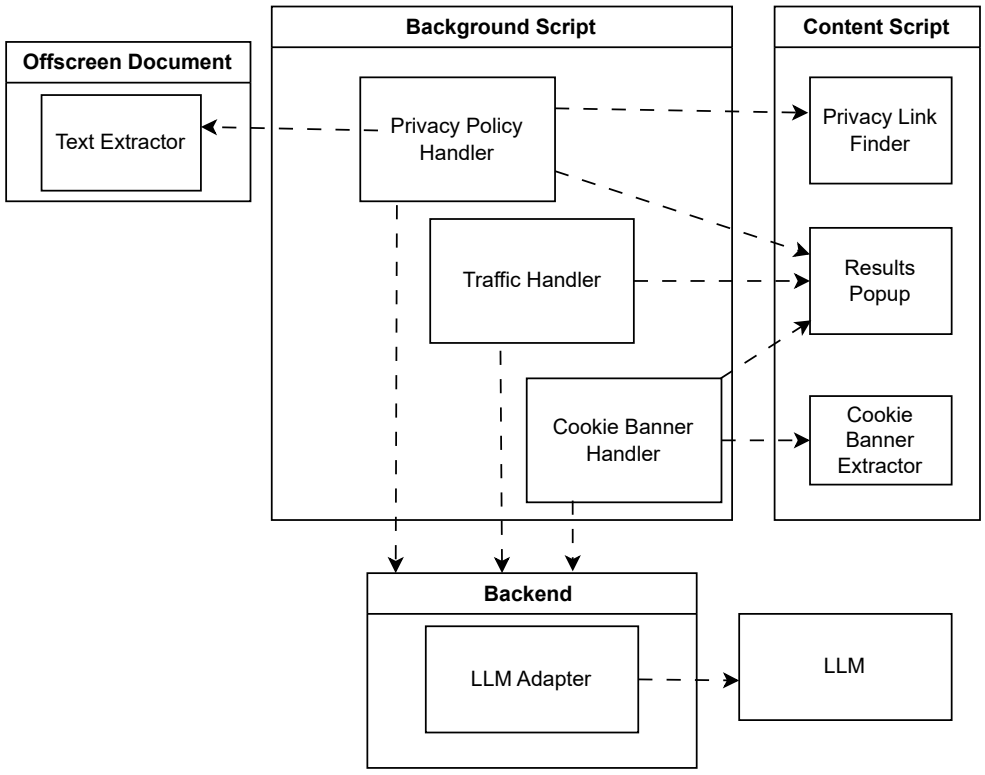

Figure 3: The component structure of the designed tool.

the three core components discussed previously. LangChain supports the majority of commercially available LLMs and several locally deployable open-source models, but leaves much to be desired for certain popular models. For these cases a custom integration component must be developed.

The adapter layer offers several potential and desirable attributes. Firstly, it allows for the quick addition of new models to the application as they are released, making the life-cycle of this tool much more sustainable. Secondly, it enables easy comparison between models in terms of performance, costs, response times, and hallucinations, which provides important perspectives for the stakeholders. Lastly, it allows for the end-users and other stakeholders the ability to use the application with the LLM of their choice, giving the tool a much wider potential base of adopters. This allows for the integration of general-purpose tools as well as models that are fine-tuned specifically for privacy research purposes as the user sees fit.

Prompting in the designed tool will be based on ready-made prompt templates, which will receive contextual information important for task completion through data injection. The data to be injected into these prompts will be extracted from network traffic as well as the privacy policies and cookie consent banners used in the websites. The prompt structure will be designed based on the principles explained in the scientific literature about prompt engineering [32, 33, 34], chief among them the idea that the prompt must be seen in terms of four fundamental elements which are 1) Instruction, 2) Input Data, 3) Context, and 4) Output indicator.

4.3. Operational Flow

Figure 4 shows the operational flow of the designed tool. The actions and components included in the tool are as follows:

- **User visits a website:** When the user opens the website, the front page is loaded and many HTTP requests containing information about the user and their actions are sent to third parties. The user should also be presented with a cookie consent banner.
- **Near real-time third-party data leak detector:** This component monitors the HTTP requests and collects the ones going to third parties. These requests are further inspected for search terms, location data, product names, and click events. Preliminary results are given to the user based on these findings. The results are also sent to the LLM for further analysis.
- **Preliminary results:** The user is notified about which third parties information is being sent to. Updates are given when information, such as search terms or data on product orders, is sent to third parties.
- **LLM analysis of data leaks:** The LLM analyzes whether the data being sent to third parties includes specific sensitive data items. This covers, for instance, the user's name or email address, but also especially sensitive contextual personal data on, for example, the user's health, sexual orientation, or political views.
- **Get privacy policy of the site:** The contents of the front page are searched for links to the website's privacy policy. Once it is found, it is sent to the LLM for analysis.
- **Privacy policy transparency analyzer:** The LLM is used to compare the website's privacy policy to the actual data being sent to third parties according to the third-party data leak detector.
- **Summary and score:** The user is given a summary of what personal data is being sent to third parties and scores for the website on transparency, clarity, and legal compliance.
- **Dark pattern detector for cookie consent banners:** The contents of the cookie consent banner are inspected for dark patterns. If dark patterns are found, the user is informed.
- **Display warning:** The user is informed about the dark patterns detected in the cookie consent banner.

5. Challenges and Future Considerations

Using LLMs in analyzing third-party requests is a significant challenge. Analyzing all requests in their entirety is likely to be too computationally expensive and time-consuming (cf. [35]). To keep the notifications given to the user in near real-time, the use of LLMs has to be lightweight and cost-efficient. One clear solution to this challenge is the preprocessing of requests so that important parts are searched for, for example, using regular expressions and then given to the LLM. Besides only giving the model parts of the requests to analyze, some third parties can be ignored altogether. For example, large analytics companies are generally more interesting as third parties than low-risk content delivery networks. Moreover, the most popular third-party requests, such as requests to Google Analytics, have well-known structures, which makes it easier to choose critical parts from the request. The response-time of the current language models is also an issue [36, 37]. This challenge can be alleviated by not requiring immediate notifications and combining the analysis of several data leaks into one

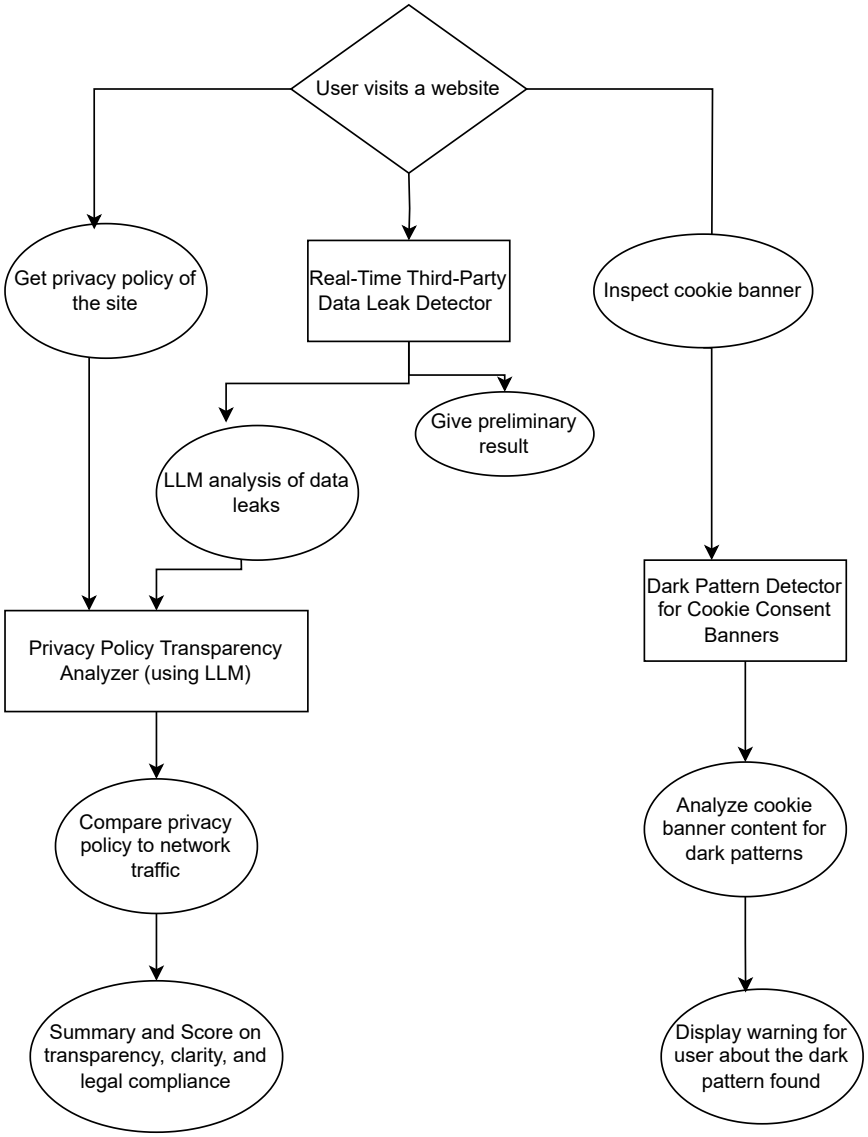

Figure 4: The operational flow of the designed tool.

prompt.

Combining the analysis of several requests and leaks into one prompt is also necessary because of repeating data leaks, in which the same personal data item repeatedly leaks to the same or different third parties. These leaks, when occurring within a short time window, should be handled in one prompt and in a single notification to the user. Not flooding the user with notifications improves user experience and decreases the amount of unnecessary information.

Although LLMs run by some external party are very efficient and accurate in many ways, they also obviously present a privacy problem [38]. After all, the purpose of the tool is to educate the user about privacy issues and make the browsing experience more transparent, not to leak the user's search terms to a third party (LLM maintainer), for example. A solution to this is to use LLMs run on trustworthy servers or completely local models. Currently, however, local models can be too slow to be practical, especially on resource-limited devices [39].

There is also the practical challenge of finding websites' privacy policies automatically before they can be analyzed with an LLM. There may also be several privacy policies or the necessary information can be given in other documents, like terms of service. To successfully compare these documents with actual network traffic, it is also important to be very specific about the categories of personal data that the model should search from the document. Discrepancies between the policy and network traffic should be presented to the user in an easily understandable form, possibly as a visual summary.

Dark pattern detection can also be fraught with many difficulties. Not all dark patterns and deceptive designs can be detected, because they appear in many forms and there are too many different types of cookie consent banners [40]. Ready-made components for dark pattern detection can also be used as a basis for the tool. These solutions need to be investigated and compared.

6. Conclusion

In this study, we have presented a conceptual design for a browser extension that aims to increase the transparency and fairness of data processing on websites. By combining near real-time web traffic analysis, comparing the traffic to information given in privacy policies, and detection of dark patterns, the tool provides the users with easily digestible information about data processing and data leaks by using notifications. The solution does not aim to prevent the data leaks but concentrates on informing the user and supporting their decision-making. Although the design shows promise, many technical implementation challenges were also identified. These include LLM response times, the coverage of network traffic analysis, and privacy concerns when using external LLMs. The future work involves implementing and testing the designed tool.

Acknowledgments

This research has been funded by Academy of Finland project 327397, IDA – Intimacy in Data-Driven Culture.

Declaration on Generative AI

No generative AI tools have been used in the preparation of this work.

References

[1] B. J. Jansen, Understanding user-web interactions via web analytics, Springer Nature, 2022.
[2] I. Önder, A. Berbekova, Web analytics: more than website performance evaluation?, International Journal of Tourism Cities 8 (2022) 603–615.

[3] S. Rauti, R. Carlsson, P. Puhtila, T. Heino, T. Mäkilä, V. Leppänen, Third-party data leaks on websites of medical condition support associations, Journal of Surveillance, Security and Safety 6 (2025) 1–16.
[4] R. Tucker, C. Tucker, J. Zheng, Privacy pal: improving permission safety awareness of third party applications in online social networks, in: 2015 IEEE 17th international conference on high performance computing and communications, 2015 IEEE 7th international symposium on cyberspace safety and security, and 2015 IEEE 12th international conference on embedded software and systems, IEEE, 2015, pp. 1268–1273.
[5] A. Mathur, M. Kshirsagar, J. Mayer, What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods, in: Proceedings of the 2021 CHI conference on human factors in computing systems, 2021, pp. 1–18.
[6] Ş. Özdemir, Digital nudges and dark patterns: The angels and the archfiends of digital communication, Digital Scholarship in the Humanities 35 (2020) 417–428.
[7] E. L. Fundingsland Jr, J. Fike, J. Calvano, J. Beach, D. Lai, S. He, Methodological guidelines for systematic assessments of health care websites using web analytics: tutorial, Journal of Medical Internet Research 24 (2022) e28291.
[8] F. Palomino, F. Paz, A. Moquillaza, Web Analytics for User Experience: A Systematic Literature Review, in: International Conference on Human-Computer Interaction, Springer, 2021, pp. 312–326.
[9] A. S. Shaheen, Maximizing website performance with Google Analytics, Turkish Journal of Computer and Mathematics Education 14 (2023) 1273–1278.
[10] A. Huidobro, R. Monroy, M. A. Godoy, B. Cervantes, A Contrast-Pattern Characterization of Web Site Visitors in Terms of Conversions, in: Technology-Enabled Innovations in Education: Select Proceedings of CIIE 2020, Springer, 2022, pp. 31–51.
[11] A. R. Zheutlin, J. D. Niforatos, J. B. Sussman, Data-tracking on government, non-profit, and commercial health-related websites, Journal of general internal medicine (2021) 1–3.
[12] P. Bekos, P. Papadopoulos, E. P. Markatos, N. Kourtellis, The Hitchhiker's Guide to Facebook Web Tracking with Invisible Pixels and Click IDs, in: Proceedings of the ACM Web Conference 2023, 2023, pp. 2132–2143.
[13] A. Aluwala, Funnel Charts to Streamline Conversions in Monitoring and Data Analytical Tools, North American Journal of Engineering Research 3 (2022).
[14] R. Trimananda, H. Le, H. Cui, J. T. Ho, A. Shuba, A. Markopoulou, OVRSEEN: Auditing network traffic and privacy policies in Oculus VR, in: 31st USENIX Security Symposium (USENIX Security 22), USENIX Association, 2022, pp. 3789–3806.
[15] B. Andow, S. Y. Mahmud, J. Whitaker, W. Enck, B. Reaves, K. Singh, S. Egelman, Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with POLICHECK, in: 29th USENIX Security Symposium (USENIX Security 20), USENIX Association, 2020, pp. 75–92.
[16] E. Shemetova, I. Shenbin, I. Smirnov, A. Alekseev, A. Rukhovich, S. Nikolenko, V. Lomshakov, I. Piontkovskaya, LAMED: LLM-generated annotations for memory leak detection, 2025. arXiv:2505.02376.
[17] B. J. A. Jo, S. J. Olino, L. L. Figueroa, M. R. C. Solamo, R. P. Feria, Developing a web-based tool for detecting deceptive designs in cookie banners, in: Proceedings of the Workshop on Computation: Theory and Practice (WCTP 2024), volume 23 of *Atlantis Highlights in Computer Sciences*, Atlantis Press, 2025, pp. 85–107.
[18] S. Rauti, R. Carlsson, P. Puhtila, V. Leppänen, Third-party data leaks on municipal websites, in: International Congress on Information and Communication Technology, Springer, 2024, pp. 599–610.
[19] T. Libert, Privacy implications of health information seeking on the web, Communications of the ACM 58 (2015) 68–77.
[20] C. F. Mondschein, C. Monda, The EU's General Data Protection Regulation (GDPR) in a research context, Fundamentals of clinical data science 1 (2019) 55–71.
[21] R. Flesch, Flesch-Kincaid readability test, Retrieved October 26 (2007) 2007.

[22] G. Guerra, Dark patterns and the scraping consumer consent: Comparative remarks on more effective legal compliance, in: Privacy, Data Protection and Data-driven Technologies, Routledge, 2025, pp. 41–67.
[23] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, 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.
[24] T. H. Soe, C. T. Santos, M. Slavkovik, Automated detection of dark patterns in cookie banners: how to do it poorly and why it is hard to do it any other way, 2022. URL: https://arxiv.org/abs/2204.11836. arXiv:2204.11836.
[25] D. Kirkman, K. Vaniea, D. W. Woods, Darkdialogs: Automated detection of 10 dark patterns on cookie dialogs, in: 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P), 2023, pp. 847–867.
[26] S. R. Kodandaram, M. Sunkara, S. Jayarathna, V. Ashok, Detecting deceptive dark-pattern web advertisements for blind screen-reader users, Journal of Imaging 9 (2023). URL: https://www.mdpi.com/2313-433X/9/11/239.
[27] Y. Sazid, M. M. Nafis Fuad, K. Sakib, Automated detection of dark patterns using in-context learning capabilities of gpt-3, in: 2023 30th Asia-Pacific Software Engineering Conference (APSEC), 2023, pp. 569–573.
[28] S. Mills, R. Whittle, Detecting dark patterns using generative ai: Some preliminary results, Available at SSRN (2023).
[29] Y. Yada, T. Matsumoto, F. Kido, H. Yamana, Why is the user interface a dark pattern?: Explainable auto-detection and its analysis, in: 2023 IEEE International Conference on Big Data (BigData), 2023, pp. 6308–6310.
[30] S. M. Hasan Mansur, S. Salma, D. Awofisayo, K. Moran, Aidui: Toward automated recognition of dark patterns in user interfaces, in: 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE), 2023, pp. 1958–1970.
[31] J. Chen, J. Sun, S. Feng, Z. Xing, Q. Lu, X. Xu, C. Chen, Unveiling the tricks: Automated detection of dark patterns in mobile applications, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST '23, Association for Computing Machinery, 2023.
[32] G. Marvin, N. Hellen, D. Jjingo, J. Nakatumba-Nabende, Prompt engineering in large language models, in: International conference on data intelligence and cognitive informatics, Springer, 2023, pp. 387–402.
[33] L. Giray, Prompt engineering with chatgpt: a guide for academic writers, Annals of biomedical engineering 51 (2023) 2629–2633.
[34] S. Ekin, Prompt engineering for chatgpt: a quick guide to techniques, tips, and best practices, Authorea Preprints (2023).
[35] F. Bang, Gptcache: An open-source semantic cache for llm applications enabling faster answers and cost savings, in: Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023), 2023, pp. 212–218.
[36] M. Nass, E. Alégroth, R. Feldt, Improving web element localization by using a large language model, Software Testing, Verification and Reliability 34 (2024) e1893.
[37] M. Zhang, X. Shen, J. Cao, Z. Cui, S. Jiang, Edgeshard: Efficient llm inference via collaborative edge computing, IEEE Internet of Things Journal (2024).
[38] T. Li, S. Das, H.-P. Lee, D. Wang, B. Yao, Z. Zhang, Human-centered privacy research in the age of large language models, in: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–4.
[39] X. Yan, Y. Ding, Are we there yet? a measurement study of efficiency for llm applications on mobile devices, in: Proceedings of the 2nd International Workshop on Foundation Models for Cyber-Physical Systems & Internet of Things, 2025, pp. 19–24.
[40] N. Bielova, L. Litvine, A. Nguyen, M. Chammat, V. Toubiana, E. Hary, The effect of design patterns on (present and future) cookie consent decisions, in: 33rd USENIX Security Symposium (USENIX Security 24), 2024, pp. 2813–2830.