The Hidden Hazards of Job Hunting: Third-Party Services on Job Search Websites

Esko Vuorinen, Sampsa Rauti, Timi Heino, Henna Lohi, Sammani Rajapaksha, Panu Puhtila
Department of Computing
University of Turku
Turku, Finland
e-mail: {etvuor | sjprau | tdhein | hmlohi | syraja | papuht}@utu.fi

***Abstract*—This paper explores third-party services and trackers used on job search websites. We analyze what kind of data on the job search process is sent to these third parties and whether users have a fair possibility to understand what kind of data collection is taking place on the studied websites. Our results show that 87.5% of the studied websites leak data on the user's actions, such as displayed job advertisements, search terms and the intent to apply for a specific position. One job search website could leak data to as many as 9 third parties. Websites run by public sector bodies had significantly less third parties and data leaks, however. While some third parties may be necessary for targeted advertising, it would be important for website maintainers to consider the number of third parties and better inform users about the data protection activities.**

Keywords-Data leaks; third parties; web analytics; job search websites; online privacy.

I. Introduction

More and more often, recruiting and seeking jobs happen online [1][2]. Job search websites have become a tool of choice for internet users looking for vacant jobs [3][4]. With these online platforms, users can now effortlessly hunt for jobs by browsing job openings and submitting applications. However, many job search websites are also businesses trying to make profit. Therefore, website maintainers add third-party web analytics services to these websites to monitor the fulfillment of their business goals and to measure the quality of user experience. Job seekers are now faced with a privacy risk, often unknowingly.

When users navigate on the job search websites, browse through job listings, and show interest in specific positions, their actions are monitored and recorded. While the justification for this is often said to be collecting data for the company that runs the job search website, data is also gathered by the third party that hosts the web analytics service, such as Google. Tracking job search activities secretly undermines users' privacy. This is especially true if users are not transparently informed of the fact that third-party services are deployed on the job search websites. Informing users inadequately of the data processing activities that taking place prevents users from giving proper and genuine consent and making truly informed decisions about their privacy [5].

In this study, we analyze 16 Finnish job search websites by performing a network traffic analysis and study *what kind of personal data concerning the user's job searches they send to third parties*. We also gauge the transparency of the privacy policy documents on these websites and identify dark patterns in their cookie consent banners. Previous research on privacy of job search websites has usually concentrated solely on the data the user willingly inputs on the job search platforms [6][7], and we extend this research by covering third-party tracking. Our paper also provides an up-to-date overview of the privacy of Finnish job search websites.

The rest of the paper is organized as follows. Section II reviews related work. Section III describes the study setting, explains how the dataset was collected, and presents the method. Section IV discusses the results of network traffic analysis, privacy policies and dark patterns. Section V covers implications for users and offers recommendations for web developers. Finally, Section VI concludes the paper.

II. Related Work

Nickel et al. [7] study user trust in e-recruitment, focusing on how perceived privacy affects user trust. They found that an interface that conveyed a high level of privacy also increased trust users place in the web service. Users that trusted the service more also disclosed more sensitive information. It is worth noting that perceived privacy and actual privacy can be very different, especially with third-party services that are invisible to the user. In a similar vein, Wijayanto et al. [6] report that platform credibility and users' awareness have a positive correlation with the willingness to share personal information in an online job portal. While the privacy of job search websites has been discussed by several studies, third parties have not received the attention they deserve. Our study aims to fill this gap.

The privacy risks of including third party analytics in web services have been widely studied [8]–[10]. The dangers of Uniform Resource Locators (URL) and search terms leaking to third parties have also been discussed in the literature [11], and the need for analytics solutions that do not share users' personal data with external parties has been acknowledged [12].

Third-party analytics and trackers have been found to be common privacy challenge in web services covering many critical application areas. The dangers of third parties have been explored for example in health-related websites [13], online pharmacies [14], goverment websites [15], and voting advice applications [16] just to name a few. The current paper presents, to the best of our knowledge, the first study on third-party analytics services on job search websites and the possible implications of such third-party data leaks.

III. Study Setting, Data collection, and Method

In the current study, we examined 16 job search websites. As there appears to be no official list of largest Finnish job search websites, we attempted to choose the most popular and well-known job search websites in Finland, based on Google search results and several sources listing job search services. Therefore, we believe the chosen websites form a representative sample of Finnish job search websites.

The selected services include both public sector websites (Kuntarekry, TE-palvelut, Työmarkkinatori, Valtiolle) and websites run by private companies (Academic Work, Adecco, Atalent, Barona, Biisoni, Duunitori, Eezy, Eilakaisla, Jobly, Manpower, Oikotie, Staffpoint). However, it is not our intention to discuss or criticize the privacy of specific companies or organizations, but rather adress the phenomenon of data leaks from a holistic perspective. Therefore, in the current paper, the websites are referred to by pseudonyms WS1–WS16, assigned in a random order.

Our experiment involved running a short testing sequence on the selected websites. All cookies were consented to upon arriving at the selected websites. First, from the main page, a search for a specific job title was initiated. On the results page, the first job advertisement in the results was chosen and clicked. Finally, on a specific job advertisement page, the "apply for the job" button or link was pressed, indicating an intent to apply for a position.

We recorded the network traffic by employing Google Chrome's Developer Tools (devtools). This set of tools allowed us to examine network activity during the testing sequence. The analyzed network traffic was filtered so that only the web requests going to third parties were inspected. The recorded traffic was also saved as log files for later analysis. From the log files, we extracted any data sent to third parties that could be used to identify the user or contained sensitive contextual information (such as data showing an intention to apply for a job). In other words, personal data items were collected from the network traffic.

In addition to network traffic analysis, we also examined on whether dark patterns – user interface design techniques to deceive users into accepting cookies and consenting to data collection [17] – were present in the cookie consent banners of the analyzed websites. We chose to use the dark pattern definitions of the European Data Protection Board's Cookie Consent Banner Taskforce [18]. We chose to focus on following specific dark patterns, which were the most unambiguous to detect and assess, detailed in the report:

- The "reject cookies" button is not present on the first layer of the cookie banner, making it harder to reject cookies and data collection than to accept them.
- Pre-ticked consent boxes or other predefined choices are used, which is not sufficient for obtaining unambiguous and freely given consent required by the General Data Protection Regulation (GDPR).
- Deceptive use of button colors or contrast is present to psychologically manipulate the user by the deployment of specific colors. For example, the colors are used to unfairly highlight the accept button so that the user is deceived into clicking it.

Finally, we briefly define the term "personal data". In this study, we employ the definition used in the GDPR and the Finnish Office of the Data Protection Ombudsman: personal data refers to "all data related to an identified or identifiable person" [19][20]. Internet Protocol (IP) addresses, accurate location data, and device or user specific identifiers used by third-party analytics services are examples of personal data. It is also important to note that many pieces of data can be used together to identify a person, and therefore, they are also considered as personal data. For instance, when window size is sent to a third-party service, this data item does not identify a specific user alone, but it can be effectively combined with other technical details to profile users.

IV. Results

In this section we look at the results we retrieved from the prior methods.

A. Network Traffic Analysis

The third parties found in the network traffic analysis are shown in Figure 1. The usual technology giants, Google and Meta (Facebook), are the most frequent third parties, appearing on 12 out of 16 studied websites. Although the use of Google Analytics has raised legal concerns in the European Union [21], Google's services are still widely prevalent. The third one on the list is LinkedIn (8 occurrences), a professional networking platform, which is an expected third party on job search websites. It is also noteworthy that TikTok, a Chinese social media platform which has recently raised privacy concerns in western countries [22], is also present with 3 occurrences.

When it comes to leaking contextual data related to searching for a job, we found that there were three notable categories of data leaks:

1) *Job advertisement page*. The URL of the job advertisement the user is viewing leaks to a third party, thus indicating that specific user shows interest in the job opening.
2) *Search term*. The search term user has used with the search function of the website (such as "Siivoaja" – a cleaner in Finnish) leaks to a third party. This often indicates the user is interested in a specific category of jobs. The search term usually leaks as a part of the URL of the search result page listing the job opportunities.
3) *Intent to apply*. The user's intent to apply for a specific position is leaked to a third party. This is probably the most valuable piece of information that a third party can receive from job search websites, because it often indicates the user's strong interest in a job opening. It often also means the user applied for the job, although some users may click the apply button just out of curiosity or otherwise abort the process of applying to a job before completion. From a technical viewpoint, there

Copyright (c) IARIA, 2024.   ISBN: 978-1-68558-194-7

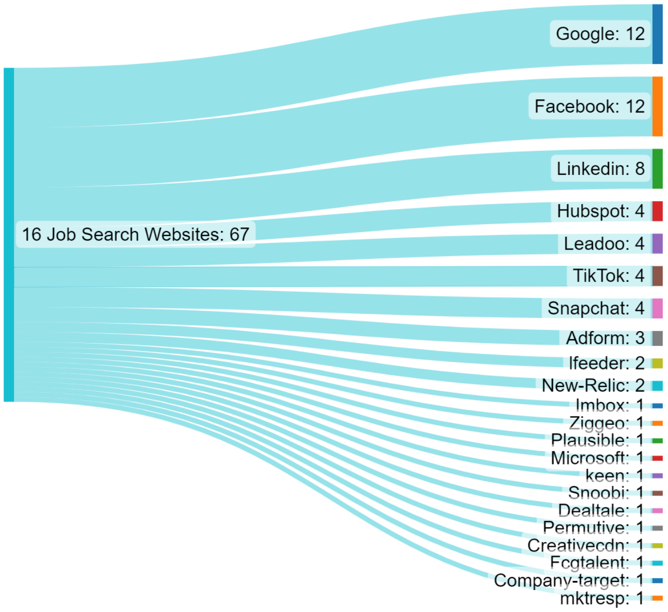

Figure 1. An alluvial diagram showing the third parties on the studied job search websites. Each third party has been counted once per website.

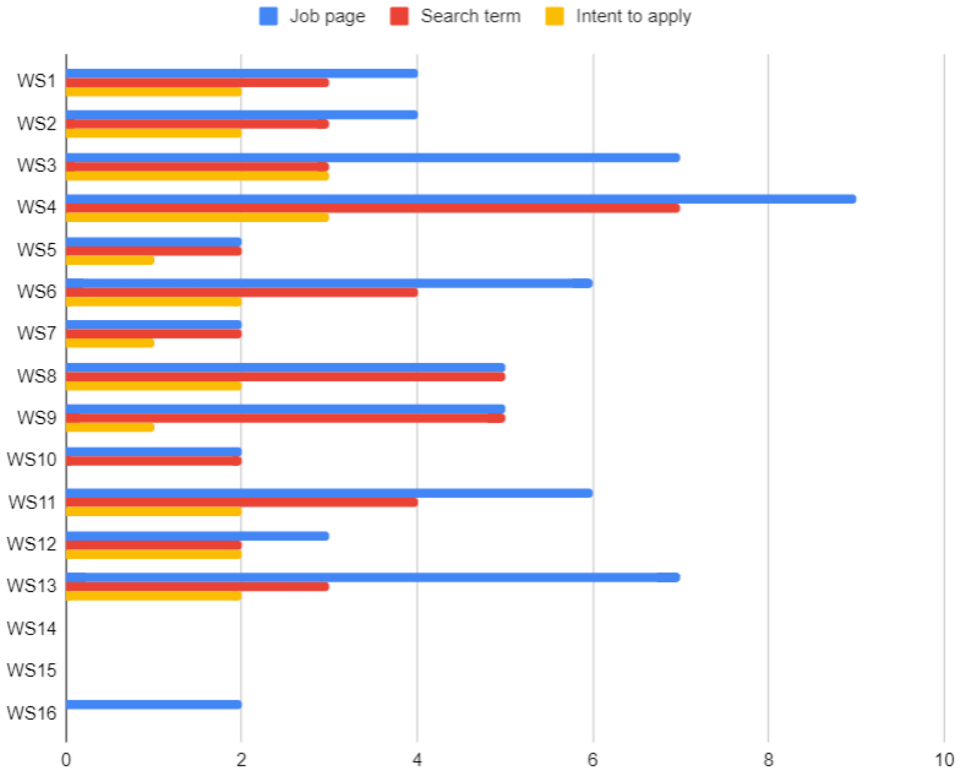

Figure 2. The leaked job search data on different job search websites and the numbers of third parties data was leaked to.

ISBN: 978-1-68558-194-7

were many ways the information about an intent to apply for a job was leaked, but usually the button click event or form submission along with explicit button or form names reveal that the user has taken an action indicating the start of the job application process.

TABLE I. Dark patterns on cookie banners of job search websites.

Website	Asks for consent	Reject button	Pre-ticked boxes	Colors & Contrast
WS1	🟦	🟦	🟦	🟦
WS2	🟦	🟦	🟦	🟦
WS3	🟦	🟦	🟦	🟥
WS4	🟦	🟦	🟦	🟥
WS5	🟦	🟦	🟦	🟥
WS6	🟦	🟥	🟦	🟥
WS7	🟦	🟦	🟥	🟦
WS8	🟦	🟦	⬛	🟥
WS9	🟦	🟥	🟥	🟥
WS10	🟦	🟥	⬛	🟥
WS11	🟦	🟦	🟦	🟥
WS12	🟦	🟥	🟦	🟥
WS13	🟦	🟦	🟦	🟥
WS14	🟦	🟦	🟦	🟥
WS15	🟦	🟦	⬛	🟥
WS16	🟦	🟦	🟦	🟥

Figure 2 shows the leaked job search data on different job search websites, sorted into three categories discussed previously. The diagram displays the number of third parties data was leaked to for each job search website and data category. The leaked job pages are shown in blue, search terms in red and intent to apply for a job in yellow.

It is quite clear that public sector websites – WS10, WS14, WS15, and WS16 – fare better in terms of privacy and third parties. WS14 and WS15 did not appear to have any third-party analytics and did not leak personal data in our experiment. The rest of the websites were maintained by private sector companies, and they had a greater number of data leaks, for example WS4 leaking the viewed job advertisement to 9 third parties, search term to 7 third parties and the intent to apply for a job to 3 third parties. The clearest divide between public and private sector job search websites, however, is in whether they leaked the intent to apply for a job. All of the private sector websites leaked this information, and none of the public sector sites did.

On average, there were 4.2 third-party services receiving personal information per job search website. Private sector websites had 5.3 such services on average, while public sector websites only had 1.0. This clearly shows that, when visiting job search websites, at least the ones run by private companies, one should expect their job search information to be leaked and monetized.

While we have focused on contextual job search data here, it is worth noting that this data would not be valuable without identifying information. Therefore, the data delivered to third parties includes identifying technical details like IP addresses, and unique device and user identifiers, as well as other technical data like screen resolution. For example, every web request contains the device's IP address, an important piece of information when trying to identify a specific user [23][24]. It is also clear that big tech companies like Google and Meta track users with cookies and also usually have the capability to know users' real names and connect them to job seeking data. When this identifying data and contextual data such as the user's intent to apply for a specific job are combined, the user's privacy is compromised.

B. Privacy Policies and Dark Patterns

While we gave consent to cookies and data collection in our study setting, it is interesting to ask whether a user can really understand that the information of the job openings they browse, the search terms they use and the job they intend to apply for are all recorded and sent to remote servers hosted by third parties. Can it be said that the user truly consents to this information collection knowingly?

When examining privacy policies, we found that only one privacy policy document mentioned all the third parties that were collecting data on the website (WS6). Also, the fact that data on the job postings user has accessed is collected was also only mentioned in one policy (WS10). The collection of search terms or recording the intent to apply for a specific job opportunity were not mentioned in any of the analyzed privacy policies. Consequently, it can be argued that the transparency of the studied job seeking websites is almost zero and the user cannot genuinely understand what kind of data collection they consent to on these websites.

Table I shows the dark patterns found on cookie banners of the studied job search websites. Positive outcomes are shown in blue, while red describes the fact that a recommended practice has not been followed. All the analyzed websites ask for consent for cookies and data collection, which is a positive result. However, 4/16 websites did not have a reject button on the first layer of the cookie banner, making it more difficult to decline cookies.

Moreover, on 2/16 occasions, cookie banners also contained pre-ticked boxes, trying to get consent without clear and affirmative user action. Three of the cookie banners, marked in black, did not have any tickable boxes. It is worth noting that the GDPR (Recital 32) explicitly addresses the fact that pre-ticked boxes should not be used: "Silence, pre-ticked boxes or inactivity should not [...] constitute consent." This makes pre-ticked boxes essentially illegal [25][26].

Finally, the most common flaw of cookie banners is using misleading colors so that the accept button is portrayed as a prominent and enticing option for users, deceiving them into accepting cookies and data collection. This dark pattern was present in 13/16 cookie consent banners.

V. Discussion

The privacy and confidentiality of job-seeking and applying for a job varies. In Finland, for example, the private sector can keep applications confidential. In the public sector, recruitment

ISBN: 978-1-68558-194-7

is public. Information on applicants can be published and anyone can request to see applications. Still, it is fair to say many applicants applying for a job generally see the application process as confidential. In this section, the implications of the data leaks found in this study are discussed from the viewpoints of users and web developers.

A. Implication for Users

Leaking data related to job searches to third parties can have various implications. The use of web analytics and tracking users' activities on job search websites are a cause for privacy concerns. Users' job search queries, job advertisement pages they display, and possibly even their intent to apply for a job are being monitored and collected by third-party services. Most users are probably not even aware that this kind data collecton is taking place [27]. The user's personal data is compromised when their job search activities are secretly monitored. Because the job seeker is often not transparently informed about third-party services being present on the job search website, it is fair to ask whether they can really give proper consent and make truly informed decisions about their privacy is greatly impeded.

Profiling and targeted advertising are another concern [28][29]. Third-party analytics services collect data on job seekers' interests and search history. It is possible to create user profiles using this information, and customize advertisements displayed to the user online based on their preferences. While this is also a positive thing to many users receiving more interesting job advertisements and opportunities, profiling can also limit the job opportunities the user sees on the web. Obviously, profiles and preferences can also be used for many other purposes besides matters related to job-seeking.

Discrimination during the recruitment process is also possible. The personal data collected by third party services could be used to treat job seekers in an unfair manner. A job seeker's job preferences or past job search activities could have an effect on recruitment decisions, if an employer were to obtain this information. An example of this would be a potential employer finding out what kinds of jobs a candidate has viewed or applied for in the past. The employer could use this data to make unjust assessments on how qualified or suitable the candidate is for a job.

A situation where the current employer finds out that the employee is looking for a job elsewhere could also sometimes become a problem. Likewise, in some cases the relatives or friends of an individual learning about what kinds of job openings the individual has been interested in could be highly undesirable. Applying for jobs in controversial companies or politically inclined organizations are some examples of this. Some lines of work, such as front-line service jobs, are also often stigmatized [30]. While it may not be very probable that the data collected by third parties becomes public, an individual's job search preferences can also leak through the advertisements regularly displayed to them, that may also be visible to their employer and the people close to them.

B. Implication for Developers

The findings of this study raise some concerns about job-seekers' privacy online. It is obvious that the job search websites need some third parties to do well against their competitors. Targeted advertising on social media is important, as job-seeking often takes place on social media [31][32] and big tech giants also have a great coverage when it comes to advertising on other websites.

Even if we accept some of the third-party tracking taking place on the studied websites, the problem of excessive use of third-party services remains. For instance, having 9 third-party services on one website and leaking personal data to them is simply too much, not to mention this is done without informing users appropriately. Dark patterns on cookie consent banners further exacerbate this problem, as users are surreptitiously coaxed into accepting the data collection.

It is evident that users should be better informed what data related to the job-seeking process is handed over to third parties. Even if job-seeking activities and applying for a job are not usually highly sensitive personal data comparable to data concerning health, for example [9][33], the users should definitely have a possibility to make an informed decision on this data collection. While some users may find targeted advertisements useful and beneficial, there should be a possibility to control them easily.

Considering and choosing the used third-party services carefully is important. The use of each service should be well justified. It is also important to understand where the services send the user's personal data. The network traffic analysis approach discussed in this paper can be used to get a good understanding of what kinds of data the website transmits to external entities. If possible, third-party services should be replaced by locally hosted services such as Matomo [12] that do not leak the users' data to third parties. This way, the privacy-by-design approach is better followed and the user's privacy is respected.

VI. Conclusion

It is clear that competition and the drive to fulfill business goals cause the job search websites to use several data-collecting third-party services on their websites. At the same time, the website owners are in many cases responsible for placing the users' privacy in jeopardy without adequately informing them about their data processing activities. We have observed that the visited job advertisements, search terms and intent to apply are regularly leaked to third parties without the user being explicitly informed. At the same time, dark patterns – especially deceptive color choices – are used to persuade users to accept cookies and data collection. Our findings call for increased attention to the use of third parties on job search websites and careful consideration on how these services handle users' personal data.

Acknowledgments

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

References

[1] A. E. Green, Y. Li, D. Owen, and M. De Hoyos, "Inequalities in use of the internet for job search: Similarities and contrasts by economic status in great britain," *Environment and Planning A*, vol. 44, no. 10, pp. 2344–2358, 2012.

[2] A. Gandini and I. Pais, "Social recruiting: Control and surveillance in a digitised job market," *Humans and machines at work: Monitoring, surveillance and automation in contemporary capitalism*, pp. 125–149, 2018.

[3] Jobvite, *Job seeker nation survey 2020: When change is the only constant*, https://www.jobvite.com/lp/2020-job-seeker-nation-report/thank-you/?submissionGuid=dcf1bc49-5922-4c08-9a49-ecf6c01ab09b, 2020.

[4] R. J. Faberman and M. Kudlyak, "What does online job search tell us about the labor market," *Economic perspectives*, vol. 40, no. 1, pp. 1–15, 2016.

[5] S. Breen, K. Ouazzane, and P. Patel, "Gdpr: Is your consent valid?" *Business Information Review*, vol. 37, no. 1, pp. 19–24, 2020.

[6] R. Wijayanto, Y. Ruldeviyani, and I. Sulistyowati, "Analysis factors of personal information sharing in online job portal application," in *AIP Conference Proceedings*, AIP Publishing, vol. 2482, 2023.

[7] J. Nickel and H. Schaumburg, "Electronic privacy, trust and self-disclosure in e-recruitment," in *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, 2004, pp. 1231–1234.

[8] R. D. Gopal, H. Hidaji, R. A. Patterson, E. Rolland, and D. Zhdanov, "How much to share with third parties? user privacy concerns and website dilemmas," *Mis Quarterly*, vol. 42, no. 1, 143–A25, 2018.

[9] M. Huo, M. Bland, and K. Levchenko, "All eyes on me: Inside third party trackers' exfiltration of phi from healthcare providers' online systems," in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, 2022, pp. 197–211.

[10] C. Utz *et al.*, "Privacy rarely considered: Exploring considerations in the adoption of third-party services by websites," *arXiv preprint arXiv:2203.11387*, 2022.

[11] T. Libert, "Privacy implications of health information seeking on the web," *Communications of the ACM*, vol. 58, no. 3, pp. 68–77, 2015.

[12] D. Quintel and R. Wilson, "Analytics and privacy," *Information Technology and Libraries*, vol. 39, no. 3, 2020.

[13] X. Yu, N. Samarasinghe, M. Mannan, and A. Youssef, "Got sick and tracked: Privacy analysis of hospital websites," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, 2022, pp. 278–286.

[14] A. R. Zheutlin, J. D. Niforatos, and J. B. Sussman, "Data-tracking among digital pharmacies," *Annals of Pharmacotherapy*, vol. 56, no. 8, pp. 958–962, 2022.

[15] N. Samarasinghe, A. Adhikari, M. Mannan, and A. Youssef, "Et tu, brute? privacy analysis of government websites and mobile apps," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 564–575.

[16] T. Heino, S. Rauti, S. Laato, R. Carlsson, and V. Leppänen, "Leaky democracy: Third parties in voting advice applications," in *International Conference on Smart Computing and Communication*, Springer, 2024, pp. 351–360.

[17] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(un) informed consent: Studying gdpr consent notices in the field," in *Proceedings of the 2019 acm sigsac conference on computer and communications security*, 2019, pp. 973–990.

[18] European Data Protection Board's Cookie Consent Banner Taskforce, *Report of the work undertaken by the cookie banner taskforce*, https://edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-cookie-banner-taskforce_en, Accessed: 2024-08-20, 2023.

[19] gdpr.eu, *What is considered personal data under the eu gdpr?* https://gdpr.eu/eu-gdpr-personal-data/, Accessed: 2024-08-20, 2024.

[20] Office of the Data Protection Ombudsman, *What is personal data?* https://tietosuoja.fi/en/what-is-personal-data, Accessed: 2024-08-20, 2024.

[21] S. Winklbauer and R. Horner, "Austrian DPA Decides EU-US Data Transfer through the use of Google Analytics to Be Unlawful," *European Data Protection Law Review*, vol. 8, p. 78, 2022.

[22] S. Patnaik and R. E. Litan, "TikTok shows why social media companies need more regulation," *The Brookings Institution, Policy Brief*, 2023.

[23] V. Mishra *et al.*, "Don't count me out: On the relevance of ip address in the tracking ecosystem," in *Proceedings of The Web Conference 2020*, 2020, pp. 808–815.

[24] T. Heino, R. Carlsson, S. Rauti, and V. Leppänen, "Assessing discrepancies between network traffic and privacy policies of public sector web services," in *Proceedings of the 17th International Conference on Availability, Reliability and Security*, 2022, pp. 1–6.

[25] T. Jakobi and M. von Grafenstein, "What HCI Can Do for (Data Protection) Law—Beyond Design," *Human Factors in Privacy Research*, pp. 115–136, 2023.

[26] P. Puhtila, R. Carlsson, and S. Rauti, "Privacy risks of third-party services on women's shelter websites," in *2023 16th International Conference on Security of Information and Networks (SIN)*, IEEE, 2023, pp. 1–6.

[27] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, communication & society*, vol. 15, no. 5, pp. 662–679, 2012.

[28] M. Gjoreski, M. Laporte, and M. Langheinrich, "Toward privacy-aware federated analytics of cohorts for smart mobility," *Frontiers in Computer Science*, vol. 4, p. 891 206, 2022.

[29] S. J. De and A. Imine, "Consent for targeted advertising: The case of facebook," *AI & SOCIETY*, vol. 35, pp. 1055–1064, 2020.

[30] C. Benoit, B. McCarthy, and M. Jansson, "Occupational stigma and mental health: Discrimination and depression among front-line service workers," *Canadian Public Policy*, vol. 41, no. Supplement 2, S61–S69, 2015.

[31] G. Karaoglu, E. Hargittai, and M. H. Nguyen, "Inequality in online job searching in the age of social media," *Information, Communication & Society*, vol. 25, no. 12, pp. 1826–1844, 2022.

[32] C. W. Piercy and S. K. Lee, "A typology of job search sources: Exploring the changing nature of job search networks," *New Media & Society*, vol. 21, no. 6, pp. 1173–1191, 2019.

[33] A. R. Zheutlin, J. D. Niforatos, and J. B. Sussman, "Data-tracking on government, non-profit, and commercial health-related websites," *Journal of general internal medicine*, pp. 1–3, 2021.