# Is It Possible to Preserve Privacy in the Age of AI?

Vijayanta Jain
University of Maine
Orono, Maine, USA
vijayanta.jain@maine.edu

Sepideh Ghanavati
University of Maine
Orono, Maine, USA
sepideh.ghanavati@maine.edu

## ABSTRACT

Artificial Intelligence (AI) hopes to provide a positive paradigm shift in technology by providing new features and personalized experience to our digital and physical world. In the future, almost all our digital services and physical devices will be enhanced by AI to provide us with better features. However, as training artificially intelligent models require a large amount of data, it poses a threat to user privacy. The increasing prevalence of AI promotes data collection and consequently poses a threat to privacy. To address these concerns, some research efforts have been directed towards developing techniques to train AI systems while preserving privacy and help users preserve their privacy. In this paper, we survey the literature and identify these privacy-preserving approaches that can be employed to preserve privacy. We also suggest some future directions based on our analysis. We find that privacy-preserving research, specifically for AI, is in its early stage and requires more effort to address the current challenges and research gaps.

## CCS CONCEPTS

• **Privacy → Privacy protections**.

## KEYWORDS

Artificial Intelligence, Privacy, Machine Learning, Survey

## 1 INTRODUCTION

Artificial Intelligence (AI) is increasingly becoming ubiquitous in our lives through its growing presence in the digital services we use and the physical devices we own. AI already powers our most commonly used digital services, such as search (Google, Bing), music (Spotify, YouTube Music), entertainment (Netflix, YouTube), and social media (Facebook, Instagram, Twitter). These services heavily rely on AI or Machine Learning (ML)[1] to provide users with personalized content and better features, such as relevant search results, the content the users would like, and the people they might know. AI/ML also enhances several physical devices that we own (or can own), for example - smart speakers, such as Google Hub and Amazon Echo, that rely on natural language processing to detect voice, understand, and execute commands such as to control lights, change the temperature, or add groceries to shopping list. Using AI to provide highly personalized experience is beneficial for the users as well as the providers; users get positive engagement with these platforms and providers get engaged users who spend more time on their services. The number of applications and devices that use AI will also increase in the near future. This is evident by the increasing number of smartphones with dedicated chips for machine learning (ML) [1–3, 27] and devices that come integrated with personal assistants .[2,3]

The proliferation of AI poses direct and indirect threats to user privacy. The direct threat is the inference of personal information and the indirect threat is the promotion of data collection. Movies such as *Her*, accurately portray the Utopian-AI future some companies hope to provide users as they increase the ubiquity of ML in their digital and physical products. However, as training AI systems, such as deep neural networks, requires a large amount of data, companies collect usage data from users whenever they interact with any of their services. There are two major problems with this collection: first, the usage data collected is used to infer information such as personal interests, habits, and behavior patterns thus invading privacy; and second, to improve the personalization, intelligent features, and AI-capabilities of the services, companies will continuously collect and increase the data collected from users, thus leading to an endless-loop of collecting data which threatens user privacy (see Figure 2). Moreover, the collected data is often used for ad-personalization or shared with third-party which does not meet user's expectations and thus, violates user privacy [23]. For example, when you interact with Google's Home Mini, the text from these recordings may be used for ad-personalization (see Figure 1) which does not meet the privacy expectations of the users [23].

Privacy violations in recent times have motivated research efforts to develop techniques and methodologies to preserve privacy. Previous research work has developed tools that provide users with more effective notice and choice [9, 18, 19, 31]. With increasing concerns about privacy because of AI, some efforts have also been directed towards training machine learning models while preserving privacy [4, 29]. User-focused techniques provide users with the necessary tools to preserve privacy whereas privacy-preserving machine learning helps companies use machine learning for their services while still preserving user privacy. In this work, we survey these methods to understand the methodologies that can be employed when users are surrounded by digital services and physical devices that use AI. The contributions of this paper are two-fold:

- We survey the machine learning based methodologies and techniques.
- Identify research gaps and suggest future directions.

The rest of the paper is organized as follows: in Section 2, we report the result of our survey. In Section 3, we discuss some related work whereas Section 4 identifies the challenges and suggests future directions. Finally in Section 5, we conclude our work.

[1] AI and ML are used interchangeably in this paper.

[2] https://www.amazon.com/Amazon-Smart-Oven/dp/B07PB21SRV

[3] https://www.amazon.com/Echo-Frames/dp/B01G62GWS4



*WSDM '20, February 3–7, 2020, Houston, TX, USA*

https://doi.org/10.1145/1122445.1122456

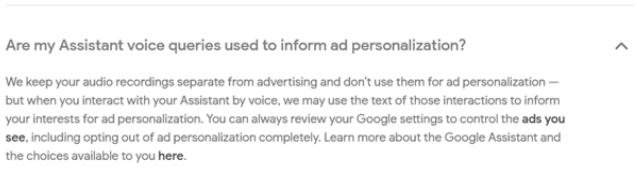


**Figure 1: The Text of Voice Recordings Can be Used for Ad-personalization**

## 2 ANALYSIS OF THE CURRENT LITERATURE

In this section, we report on our survey of machine-learning based techniques that have been developed to preserve user privacy. We divide this section into two groups: i) privacy preserving machine learning approaches and ii) techniques to provide users with notice and give them choices.

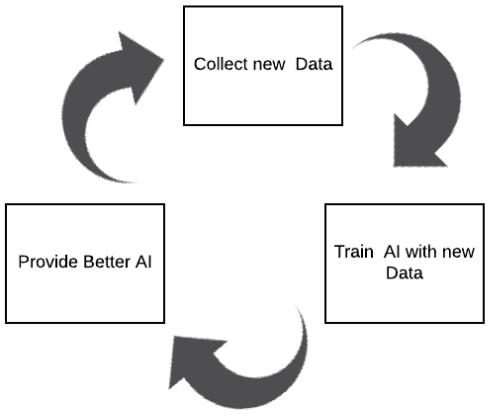


**Figure 2: Cycle of Eternal Increase in Data Collection**

### 2.1 Privacy Preserving Machine Learning Approaches

Recent research efforts have been directed to develop privacy-preserving machine learning techniques [4, 24]. Prior to machine learning, differential privacy provided a strong standard to preserve privacy for statistical analysis on public datasets. In this technique, whenever a statistical query is made to a database containing sensitive information, a randomized function $k$ adds noise to the resulting query which preserves privacy while also ensuring the usability of the database [13]. Some work has used differential privacy for training machine learning models [4, 7]. Chaudhri and Monteleoni [7] use this technique to develop a privacy-preserving algorithm for logistic regression. Abadi et al. [4] also use this technique to train deep neural networks by developing a noisy Stochastic Gradient Descent (SGD) algorithm. However, a key problem with differential privacy is that having repeated queries to the database can average out the noise and thus revealing the underlying sensitive information of the database [13]. To solve this, Dwork proposes privacy budget that considers each query to the database as a privacy cost and for each session there is a privacy budget [11, 13]. After the privacy budget has been used for the session, no query results are returned.

Other work in this area has been to develop methods to train neural networks on the device itself without sending the data back to the servers [24, 25, 29]. Shokri and Shmatikov [29] present a system to jointly train models without sharing the input dataset of each individual. In their work, they develop a system that allows several participants to train similar neural networks on their input data without sharing the data but selectively sharing the parameters with each other to avoid local minima. Similarly, in line with Shokri and Shmatikov to not share data, McMahan et al. [24] propose *Federated Learning* which allows developers to train neural networks in a decentralized and privacy-preserving manner. The ideology behind their work is that neural network models to be trained are sent to the mobile devices which contain the user sensitive data and use SGD locally to update the parameters. The models are then sent back to a central server which "averages" the update from all the models to achieve a better model. They term this algorithm FederatedAveraging. Similarly, Papernot et al. [25] propose Private Aggregation of Teacher Ensemble (PATE) - a method to train machine learning models while preserving privacy. In their approach, several "teacher" models are trained on disjoint subsets of the dataset, then the "student" model is trained by the aggregation of the "teachers" to accurately "mimic the ensemble". The goal of this work is to address the information leakage problem [15].

The goal of the work outlined above is to develop new algorithms and methods to train neural networks on a device or use differentially private algorithms. However, information leakage still provides a threat to the user's privacy. Information leakage is the concept in which the neural network implicitly contains sensitive information it was trained on. This is demonstrated in [15, 30]. This is an active research topic and new methods, such as PATE, aim to resolve this issue by not exposing the dataset to the machine learning model.

### 2.2 Mechanisms to Control User's Data

The primary goal in this field of research has been to provide users with better notice, give them choices and provide them with the means to control their personal information. Notice and Choice is one of the fundamental methods to preserve privacy and is based on the Openness principle of the OECD Fair Information Principle [16]. In Notice and Choice mechanism, the primary goal has been to improve and extract relevant information from privacy policies for the users. This is because privacy policies are lengthy and it is infeasible for users to read the privacy policies for all the digital and physical services they use/own [10]. Therefore, research has focused on providing them with better notice and choice such as in [20, 22, 28]. Other work have achieved similar results by applying machine learning techniques. Harkous et al. [18] develop *PriBot* a Q&A chatbot that analyzes a privacy policy and then provides users with sections of the privacy policy that answers their question.

Some work has focused on identifying the quality of the privacy policy. For example, Constane et al. [8] use text categorization and machine learning to categorize paragraphs of privacy policies and assess their completeness with a grade. The grade is calculated by the weight assigned by the user to each category and the coverage of the category in a selected section. This method helps users inspect a privacy policy in a structured way and read only the paragraphs that interest them. Zimmeck et al. introduce Privee [36] which integrates Constane's classification method with Sadeh's crowdsourcing. In Privee, if a privacy analysis results are available in the repository, the result is returned to the user. Otherwise, the privacy policy is automatically classified and then, it is returned. PrivacyGuide [31] uses classification techniques, such as Naïve Bayes and Support Vector Machines (SVM), to categorize privacy policies based on the EU GDPR [14], summarize them and then allocate risk factors. These above work certainly improve the previous "state-of-the-art" method of notice & choice - a privacy policy by giving users a succinct form of information. However, privacy policies often contain ambiguities that are difficult for technology to answer, for example, the number of third parties the data is shared with or how long the data will be stored by the companies.

Another active topic of research in providing control of their privacy to users is to model privacy preferences. The goal of this topic of research is to provide users with more control over what information can mobile applications or other users access. Lin et al. [21] create a small number of profiles for user's privacy preference using clustering and then based on those profiles analyze whether the user from a profile allows certain permissions or not. Similar to their work, Wijesekera et al. [32] develop a contextually-aware permission system that dynamically permits access to private data of Android applications based on user's preferences. They argue that their permission system is better than the default Android permission system of Ask-On-First-Use (AOFU) as context, "what [users] were doing on their mobile devices at the time that data was requested" [32] affect user's privacy preferences. In their system, they use SVM classifier, trained over contextual information and user's behavior, to make permission decisions. They also conduct a usability study to model the preferences of 37 users and test their system [33]. Similarly, other work to use contextual information to model privacy preferences has been done for applications in web-based services as well. Yuan et al. [34] propose a model that uses contextual information to share images, with different granularity with other users. In their work, based on the semantic image features and contextual features of a requester, they train logistic regression, SVM and Random Forest to predict whether the user would share, would not share, or partially share the image requested. Similarly, Bilogrevic et al. [6] develop Smart Privacy-aware Information Sharing Mechanism, a system that shares personal information with users, third-party, online services, or mobile apps based on the user's privacy preferences and the contextual information. They use Naïve Bayesian, SVM, and Logistic Regression to model preferences. They also conduct a user study to understand their preferences and the factors influencing their decision. Using contextual information and providing different levels of information access is a great step towards providing the user with greater control of their data but certain challenges still remain. Primarily, most of these systems have not conducted usability studies to examine the user's view. This inhibits implementing such research into real-world.

Overall we find that this line of work has focused on giving users the mechanisms to understand the privacy practices and control their data. Giving users the control of their data is important, however, this approach puts the burden on the users to preserve their privacy which might be difficult for less tech-savvy users as often the privacy settings for websites are hidden under layers of settings to control.

## 3 RELATED WORK

Papernot et al. [26] provide a Systematization of Knowledge (SoK) of security and privacy challenges in machine learning. This work surveys the existing literature to identify the security and privacy threats as well as defenses that have been developed to mitigate the threats. The research work also argues based on the analysis, to develop a framework for understanding the sensitivity of ML algorithms to its training data to foster security and privacy implications of ML algorithms. Our analysis is similar as it evaluates privacy implications of these machine learning algorithms, but our work provides a more detailed discussion on the privacy challenges as compared to [26]. Zhu et al. [35] survey different methods developed to publish and analyze differentially private data. The work analyzes differentially private data published based on the type of input data, the number of queries, accuracy, and efficiency and evaluate differentially private data analysis based on Laplace/Exponential Framework, such as [7] and Private Learning Framework, such as [4]. The paper also presents with some future directions for differential privacy, such as executing more local differential privacy. This work is the closest to our work as it surveys a privacy-preserving analysis technique and suggests future work. However, in our analysis, we also incorporate the technologies that help users preserve their privacy. Overall, our work differs from [26, 35] as we look at the big picture of privacy-preserving technologies specifically with the increase in use of AI.

## 4 DISCUSSION

In this paper, we discussed techniques and methodologies developed to preserve user privacy. Primarily, we identified two groups of work: (1) privacy-preserving machine learning, such as noisy SGD and federated learning, and (2) techniques to provide users with the tool to protect their own privacy. In this section, we discuss the advantages of each category of approaches, their existing challenges, the research gaps, and suggest some potential future work to address the challenges and gaps identified here. We summarize our analysis in Table 1.

**Differential Privacy and Machine Learning Approaches:** Differential privacy provides a strong state-of-the-art for data analysis by introducing noise to query results [12] and this method has also been used to train deep neural networks [4]. One of the biggest advantages of these approaches is the simplicity and efficiency of the methodology. Some companies have even started to use differential privacy in some of their applications .[4] Using differential privacy for deep learning provides great potential for researchers and developers. However, understanding the trade-offs between

[4]https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

**Table 1: Summary of Privacy-Preserving Approaches**

| Privacy-Preserving Approach | Advantages | Disadvantages |
|---|---|---|
| Differential Privacy and Machine Learning | • Simple and efficient<br>• Easy to Employ | • Requires large noise for effective privacy at the cost of utility. |
| Federated Learning | • Prevents sharing and profiling and thus better privacy | • More suitable for large-scale applications |
| User-Focused Privacy Tools | • Gives user the control of their privacy | • Puts the burden on user to preserve privacy<br>• Limited tools for controlling privacy |

privacy and utility for specific tasks, models, optimizers, and similar other factors can further help developers in using differentially-private machine learning. Some initial work has been done in this area [5] but future work can explore this in detail.

**Federated Learning:** Federated learning provides a unique approach to machine learning by training models on device instead of on a central server [24]. By keeping the data on a device, it will prevent sharing with third-party and even profiling user-data for ad-personalization. A key challenge with federated learning is the complexity of using Federated Learning; small-scale companies and developers might find differential privacy easier to optimize and employ on a smaller scale. Another challenge with this approach is information leakage from the gradients of the neural network [15, 30]. There has been some effort to address this issue by developing different privacy-preserving machine learning methodologies [25]. However, a critical gap in this area of research is that few research efforts have looked into providing users with mechanisms that control the data being used for federated-learning. Future work can address this gap. Another future direction for federated learning is to combine differentially-private data with federated learning. Initial work has been done in this direction, such as [17], but future work could expand the analysis by evaluating different differential privacy algorithms for privatizing data.

**User-Focused Privacy Preserving:** Several methods have been proposed that uses machine learning to preserve user-privacy [6, 18, 32] to provide users with the necessary notices and control mechanisms to have control over their data. Some of these methods [18] employ Natural Language Processing (NLP) to understand privacy text to preserve user privacy. Future work in this direction can employ more advanced architectures for this task to improve accuracy and relevance. Another future direction can be to help companies and developers create applications and systems that preserve user's privacy.

Based on our analysis of the current data practices and research development, we believe that it will be difficult to preserve privacy in the age of AI. As the ubiquity of AI and economic incentives to use AI will increase, it will passively promote data collection and thus pose a threat to user privacy. The techniques developed to preserve user privacy are not as effective as the current data practices that violates them. Increased research effort along with legal actions will be required to preserve privacy in the age of AI.

## 5 CONCLUSION

In this work, we provide a brief survey of machine learning based techniques to preserve user privacy, identify the challenges with these techniques and suggest some future work to address the challenges. We argue that the privacy-preserving technologies specifically for AI are in their early stages and it will be difficult to preserve privacy in the age of AI. We identify research gaps and suggest future work that can address some of the gaps and result in more effective privacy-preserving technologies for AI. In future, we plan on expanding this work for a more critical analysis of different algorithms and evaluate their efficacy for different use cases.

## REFERENCES


[1] [n.d.]. iPhone 11 Pro. https://www.apple.com/iphone-11-pro/.
[2] [n.d.]. OnePlus 7 Pro. https://www.oneplus.com/7pro#/specs.
[3] [n.d.]. Samsung Galaxy S10 Intelligence - Virtual Assistant & AR Photo. https://www.samsung.com/us/mobile/galaxy-s10/intelligence/.
[4] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 308–318.
[5] Brendan Avent, Javier Gonzalez, Tom Diethe, Andrei Paleyes, and Borja Balle. 2019. Automatic Discovery of Privacy-Utility Pareto Fronts. *arXiv preprint arXiv:1905.10862* (2019).
[6] Igor Bilogrevic, Kévin Huguenin, Berker Agir, Murtuza Jadliwala, Maria Gazaki, and Jean-Pierre Hubaux. 2016. A machine-learning based approach to privacy-aware information-sharing in mobile social networks. *Pervasive and Mobile Computing* 25 (2016), 125–142.
[7] Kamalika Chaudhuri and Claire Monteleoni. 2009. Privacy-preserving logistic regression. In *Advances in neural information processing systems*. 289–296.
[8] Elisa Costante, Yuanhao Sun, Milan Petković, and Jerry den Hartog. 2012. A machine learning solution to assess privacy policy completeness:(short paper). In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society*. ACM, 91–96.
[9] Lorrie Faith Cranor. 2003. P3P: Making privacy policies more useful. *IEEE Security & Privacy* 1, 6 (2003), 50–55.
[10] Lorrie Faith Cranor. 2012. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.* 10 (2012), 273.
[11] Cynthia Dwork. 2011. Differential privacy. *Encyclopedia of Cryptography and Security* (2011), 338–340.
[12] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.
[13] Aaruran Elamurugaiyan. 2018. A Brief Introduction to Differential Privacy. https://medium.com/georgian-impact-blog/a-brief-introduction-to-differential-privacy-eacf8722283b
[14] EU GDPR [n.d.]. *"The EU General Data Protection Regulation (GDPR)"*. EU GDPR. https://eugdpr.org.
[15] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. 2015. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1322–1333.

[16] Ben Gerber. [n.d.]. OECDprivacy.org. http://www.oecdprivacy.org/.

[17] Robin C Geyer, Tassilo Klein, and Moin Nabi. 2017. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557* (2017).

[18] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. 2018. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *USENIX Security Symposium*.

[19] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. 2009. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*. ACM, 4.

[20] Marc Langheinrich. 2002. A privacy awareness system for ubiquitous computing environments. In *international conference on Ubiquitous Computing*. Springer, 237–245.

[21] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I Hong. 2014. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 199–212.

[22] Fei Liu, Rohan Ramanath, Norman Sadeh, and Noah A Smith. 2014. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 884–894.

[23] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. Privacy Attitudes of Smart Speaker Users. *Proceedings on Privacy Enhancing Technologies* 2019, 4 (2019), 250–271.

[24] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. 2016. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629* (2016).

[25] Nicolas Papernot, Martín Abadi, Ulfar Erlingsson, Ian Goodfellow, and Kunal Talwar. 2016. Semi-supervised knowledge transfer for deep learning from private training data. *arXiv preprint arXiv:1610.05755* (2016).

[26] Nicolas Papernot, Patrick McDaniel, Arunesh Sinha, and Michael P Wellman. 2018. SoK: Security and privacy in machine learning. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 399–414.

[27] Brian Rakowski. 2019. Pixel 4 is here to help. https://blog.google/products/pixel/pixel-4/.

[28] Joel R Reidenberg, N Cameron Russell, Alexander J Callen, Sophia Qasir, and Thomas B Norton. 2015. Privacy harms and the effectiveness of the notice and choice framework. *ISJLP* 11 (2015), 485.

[29] Reza Shokri and Vitaly Shmatikov. 2015. Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*. ACM, 1310–1321.

[30] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 3–18.

[31] Welderufael B. Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, and Jetzabel Serna. 2018. PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA '18)*. ACM, New York, NY, USA, 15–21. https://doi.org/10.1145/3180445.3180447

[32] Lynn Tsai, Primal Wijesekera, Joel Reardon, Irwin Reyes, Serge Egelman, David Wagner, Nathan Good, and Jung-Wei Chen. 2017. Turtle guard: Helping android users apply contextual privacy preferences. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. 145–162.

[33] Primal Wijesekera, Joel Reardon, Irwin Reyes, Lynn Tsai, Jung-Wei Chen, Nathan Good, David Wagner, Konstantin Beznosov, and Serge Egelman. 2018. Contextualizing privacy decisions for better prediction (and protection). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 268.

[34] Lin Yuan, Joël Theytaz, and Touradj Ebrahimi. 2017. Context-dependent privacy-aware photo sharing based on machine learning. In *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, 93–107.

[35] Tianqing Zhu, Gang Li, Wanlei Zhou, and S Yu Philip. 2017. Differentially private data publishing and analysis: A survey. *IEEE Transactions on Knowledge and Data Engineering* 29, 8 (2017), 1619–1638.

[36] Sebastian Zimmeck and Steven M Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies.. In *USENIX Security*, Vol. 14.