**Journal of Computing and Electronic Information Management**
ISSN: 2413-1660 | Vol. 13, No. 3, 2024

# Algorithmic Intrusion: The Erosion of Personal Privacy in Digital Age

**Shuqi Chen**

School of Humanities and Social Sciences, The University of Melbourne, Melbourne, Australia
shuqchen1@student.unimelb.edu.au

**Abstract:** In the era of global electronicization, the digital transformation of communication technology has led to the widespread collection and storage of personal information in the databases of online services, raising concerns about personal privacy. Companies use algorithms to collect and analyze this data and effectively predict user behavior to optimize services and drive profit growth. However, with the rapid development of the data economy and the increase in the scale of data collection, it has become increasingly difficult for the media to manage privacy, the boundaries of personal privacy have gradually blurred, and the public's concerns about privacy have also increased. This paper will take algorithms, the primary driving force of the data economy, as the starting point to analyze how algorithms invade personal privacy through recommender systems, advertising target, and social media functions, and how they use this private data to influence users' perceptions and behaviors.

**Keywords:** Data economy; Algorithms; Privacy erosion; Data collection.

## 1. Introduction

Privacy is dead – get over it'. The statement made by Scott McNealy, former CEO of Sun Microsystems in 1999, was controversial at the time, but it has become a reality in today's data-driven society. In the era of global digitization, the digital transformation of communication technology has led to the generation and storage of personal information, including whereabouts, preferences, and social circles, in the database of the provision of online services, which has led to the concept of integrated privacy, which is the right to privacy of individuals in the digital age (Liu, Y et al., 2022). Businesses use algorithms to collect personal data and make effective predictions about user behavior to optimize services and drive revenue (Bartlett, 2021). However, with the rapid development of the data economy and the exponential growth of data collection scale, it is increasingly difficult for media to manage privacy (Miao & Li, 2022). The infinite expansion of data has blurred the boundaries of individual privacy, exacerbating overall public concerns about privacy (Westin, 2003). Also, because of the nature of data networks, even if the users realize that their information is a data node and chooses not to provide it anymore, algorithms can still generate insights about those users, which can be a trigger for questioning the non-existence of privacy at all (Bucher, 2017).

The causes and effects of privacy breaches have been extensively discussed and analyzed by categorical analysis of users' reactions, but the analysis of how converged media collects and uses user data is scattered. This paper will take the algorithms, which are the main driving forces of the data economy, as the starting point, to demonstrate how algorithms, which are the core forces of digital media development, erode personal privacy from three aspects: recommendation system, advertising target and social media function.

## 2. Background Information

The digital age, as the fourth industrial revolution, has revolutionized people's lives through the development of information and communication technologies (Bartlett, 2021). Among them, the wide application of the Internet and artificial intelligence has brought about a change in the economic model, resulting in a data economy with data as the core asset and driving force. In the big data environment, people leave a huge amount of information on the Internet, including browsing, generating, shopping, and other behaviors that can be searched, aggregated, and measured (Klinger & Svensson, 2018). Enterprises have robust application programming interfaces (APIs) for collecting and publishing user data (Y. Liu et al., 2022). These data nodes were initially stored only in Internet service providers and online service providers, but with the development of artificial intelligence, the introduction of algorithms has provided new uses for this data (Rauhofer, 2008).

Algorithms are etymically derived from the Greek word "arithmos" and the Arabic word "al-jabr", which stands for numbers and computation (Klinger & Svensson, 2018). According to Khoo (2023), an algorithm in the electronic age is defined as a set of instructions or calculations that solve a problem. Algorithms compute data based on expressions given by programmers, which can be seen as materialization or socialization processes (Klinger & Svensson, 2018). Algorithms are built into systems that continuously configure the creation, distribution, and consumption of information (Matamoros-Fernandez et al., 2021). It predicts user behavior through operations such as calculating, sorting, and filtering data. These predictions, based on historical user behavior, are accurate and detailed, maximizing understanding of users' preferences and needs (Bartlett, 2021). Therefore, the processing of user data with algorithms has become the technical basis for enterprises to attract customers and achieve long-term sustainable development (Miao & Li, 2022). Enterprises obtain business benefits through the insight and prediction of user behavior, thus forming an economic model that uses data as a resource to create value. Therefore, in the context of data becoming capital, privacy leakage has also become an inevitable problem.

# 3. Literature Review

According to Balleys and Coll (2017), privacy is a relationship which means that people with intimacy are connected through it. The concept of privacy in the law dates back to the 80s of the 19th century, and it was described as the right not to be disturbed. Although scholars in different fields have different definitions of it, the core of privacy has always been the private nature of personal information. The Universal Declaration of Human Rights of 1948 states that no one's privacy should be arbitrarily interfered with. Coll (2014) even describes privacy as a prerequisite for the rapid development of the economy in the digital age. However, setting aside biometric and physical information, due to the presence of algorithms, artificial intelligence can analyze and predict users' values, preferences, and even political leanings based on latent information (Y. Liu et al., 2022). Coupled with the development of smart homes, environmental assisted applications for living (AAL) are widely used, and cameras can capture most of the visual information in users' lives, thus people have no privacy in the electronic age (Carpentieri et al., 2022).

Scholars in the past have differed views on the impact of the Internet on privacy. According to Y. Liu et al. (2022), despite the risk of privacy breaches associated with the use of social media, in an era of highly electronic society, these information technologies can indeed help users stay connected with others and share their status, thus avoiding data silos. Carpentieri et al. (2022) also argue that the privacy threat posed by information technology is undeniable, but that smart environments such as smart homes created through information technology can indeed enhance people's well-being. However, there are also many voices that have a negative attitude towards this issue. Elvy (2017) argues that the so-called conveniences and benefits that users get in the process of using the Internet are exchanged for their own privacy. In reality, users know very little about the extent to which their privacy is exposed on the internet. These platforms not only collect and utilize data without the users' knowledge but may also disclose it to third parties, posing even greater risks. Similarly, according to Westin (2003), as Internet usage has gradually increased, there has been a growing concern about privacy breaches, especially in the case of children's information disclosure and economic information disclosure.

Although the impact of the Internet has been mixed from past research, it is undeniable that the problem of privacy leakage does exist and is becoming more and more serious with the development of information technology. The widespread use of artificial intelligence has improved the information utilization rate of enterprises. Media companies use algorithmic models to analyze data to predict user preferences and behaviors and provide users with a better user experience (Khoo, 2023). These similar models make firms an indispensable means of reaping commercial benefits (Rauhofer, 2008). Global digitization has enabled the rapid integration of textual, image, and even video information for all Internet-connected users, laying the foundation for enterprise data analysis (Carpentieri et al., 2022).

Previous scholars have studied the impact of privacy breaches and the reactions of stakeholders. The consequences of privacy leaks are not only for the users themselves, but also for public discourse, with implications for state surveillance and market development (Lovink, 2008). Users have also reacted differently to this phenomenon, with some expressing concerns about privacy breaches, but also some making conscious disclosures of privacy in order to reach intimacy due to cultural and social factors (Balleys and Coll, 2017). How privacy is eroded as a social good in a democratic society is a necessary issue for discussion (Westin, 2003). As the main technology in the data economy, the existence of algorithms is the main reason for the demise of privacy. This paper will take the algorithm to demonstrate how convergent media collects and uses information from three aspects: recommendation system, advertising target, and social media functions.

# 4. Recommender system

Among the various algorithms, recommendation algorithms are the core technical foundation used by enterprises to maintain user experience and rapid expansion. The recommendation algorithm is based on user portraits, and it is known for labeling and classifying users and improving services through personalized recommendations. Its modeling factors are usually related to the user's main attributes, and it is usually labeled for the user's preferences and behaviors to establish virtual expressions (Miao & Li, 2022).

A strong example of this is the application of recommendation algorithms in short-form video applications. TikTok and Instagram Reels have greatly promoted the personalized distribution of content through recommendation algorithms to enhance the user's experience. According to Miao and Li (2022), the recommendation algorithm used by short video apps is based on user portraits. Likes, comments, and shares on the app interface, as well as video completion rates, are all included in the short video app's database. This information can be used to provide a complete picture of the user's behavioral preferences and social interactions. This is a solid foundation for efficient content production. In addition, the recommendation algorithm used in short videos has strong real-time performance, which can be adjusted and recommended as soon as possible through the user's current behavior, which plays a core role in the growth of user stickiness.

Another strong example is Netflix's proprietary recommendation algorithm, the Netflix recommender system (NRS). According to Bartlett (2021), Netflix is known for its recommendation capabilities, using models derived from complex behavioral analysis based on user data. As a set of proprietary algorithms, NRS is mainly used to recommend content to users and filter users' experiences on various features of the platform. NRS is a set of hybrid algorithms, including collaborative filtering and content-based filtering, among others (Khoo, 2023). It is almost entirely driven by machine learning, collecting information about the user's interaction with the platform for filtering and extrapolation (Pajkovic, 2022). The data source of NRS comes from several sources, starting with the user's interaction such as viewing history, watching time while watching content, pause behavior, and skipping behavior. This is primarily used to recommend specific content to users that is in line with their preferences. The second is the user's device information, including the type of device such as mobile phone, computer, TV, device model and even the geographical location of the device. This is primarily used to accommodate the availability and compliance of content across regions. Once this data is collected, NRS extracts content from it that can be labeled and

fed into the model to generate personalized recommendations. These recommendations are not only used for the content itself, but also in the ordering and contextual awareness of the content, aiming to provide users with the most appropriate content for the current situation. Netflix is very confident in its deep personalization algorithm, and they believes that under the role of NRS, each user's profile will correspond to a highly adapted product.

The working mode of the recommendation system used in different categories is similar, which is based on the user's use behavior and the user's basic information for feature extraction. This method of data processing and analysis ensures maximum user satisfaction with content recommendations. However, these business model changes have come at the expense of user privacy. The data stored by users in the system is fragmented, but it can be aggregated and linked by algorithms to form a relatively complete personal information (Rauhofer, 2008). This means that simply by allowing users to use these software for entertainment, companies can portray users who have certain preferences under the influence of certain culture in a certain geographical location, which is a high degree of self-exposure for users.

# 5. Advertising Target

Similarly, the combination of the internet and artificial intelligence is also used in business advertising targeting systems for commercial purposes (Bartlett, 2021). Today's business targeting systems rely on algorithms and machine learning to accurately deliver ad content to the most interested customer groups, making them more interested in buying a specific product or service, thereby increasing the effectiveness and conversion rate of ads.

A prime example of this is Google's use of the behavior value reinvestment cycle (BVRC). As a conceptual framework, BVRC is often used in marketing and product development to increase customer engagement and loyalty. It is a description of a cycle. In this cycle, user behavior is recorded and stored, and then the algorithm analyzes this data to gain relative insights. Businesses then invest in these insights to reap higher value returns. Google's ad targeting system using BVRC ensures that ads are seen at the right time by people who are genuinely interested. According to Google, this has become an automated futures market. Decide whether to push ads to the user's interface by predicting what users will do at a specific time (Bartlett, 2021).

Google's search engine, email and other items are used to collect information left by users in these applications, including search history, browsing history, and click behavior. Using data analytics tools and machine learning algorithms, this data is distilled into valuable insights. For example, if a user chooses to click or ignore a specific piece of content, it will be possible to determine how receptive the user is to the ad that may be served, or the format and content that should be chosen. Based on these insights, Google makes changes and adjustments to its advertising algorithms and delivery strategies to improve the accuracy of its delivery. This is beneficial for the click-through rate and conversion rate of the ad. When the improved ad is put back into the market, user feedback and behavior are collected again. This constant cycle is key to optimizing ad serving. This feedback loop ensures that the content of the ad is always closely aligned with the user's needs, improving the user experience and increasing the Investment Return on Investment (ROI) of the ad server. This efficient use of BVRC has enabled Google's ad targeting system to remain a leading player in the market.

However, this is undoubtedly unreasonable for users. In the course of BVRC's operation, users do not even actively provide any information to specific collectors, but only carry out routine operations on the platform. This data is not only recorded, but also sold to advertisers to construct automated futures markets (Bartlett, 2021). According to Bucher (2017), users are not unaware of the platform's behavior of delivering customized ads, but they are also afraid of the accuracy of the ads. This means that companies exploit users' privacy to irrationally guide their consumer behavior, offering minimal protection for user privacy.

# 6. Social Media Fouctions

The functions of social media are constantly increasing and strengthening with the development of information technology. These functions extend from text to visual and geographically located, which means that the scope of information collected by the platform for users is gradually expanding. In recent years, location-based online services have grown rapidly, and more and more social software has introduced the ability to share geolocation (Ghosh & Singh, 2022). This means that users can post content while showing their geographic location. This is a friendly design for cross-socializing with users in the same geographic location, while also helping people find the location of this content through a map. However, this is undoubtedly another form of information leakage for users. Facebook allows users to share their location and by default, everyone in the user list, regardless of their relationship proximity, can view the user's location. However, this kind of sharing is highly real-time, which means that the spatial information of the user is public to a certain extent. Also, by combining location data such as home addresses or school locations with basic user information, people can analyze additional details, such as the places a user frequently visits. This is a great risk to user security. Additionally, as the platform hosting this sensitive data, Facebook's efforts in protecting user privacy have been criticized as insufficient. In 2019, Facebook was accused of having more than 50 million users' information acquired by a data analytics firm. At the same time, Facebook has also admitted to selling access to user data to tech companies such as Amazon (DW.com, 2019). The function of social media platforms as public spaces has blurred the boundaries of privacy (Y. Liu et al., 2022).

In fact, with the development of information technology and the continuous expansion of its application, new policies have emerged to regulate the leakage of user privacy by the media (Imana et al., 2022). Privacy disclosure has been debated since the mid-90s of the 20th century, when the Internet and wireless communication devices were emerging. Social values and environmental conditions are changing, but the protection of privacy should be seen as an important goal (Westin, 2003). Although major platforms have privacy statements and set them as essential information for users, in reality, these privacy statements are very complex and not easy for ordinary users to read and understand. According to Bartlett (2021), an empirical survey of platform privacy policies showed that users took at least 45 minutes to understand these terms, and they had to choose whether to agree or forgo using the platform. This is unfair to the users. In addition, many countries have policies on Internet privacy protection, such as the Digital Services Act in the United

States and the General Data Protection Regulation in Europe. However, the implementation of these policies is limited to a certain extent due to issues of jurisdiction. In the context of globalization, how to implement it has also become a difficult problem. However, data breaches represent a loss of privacy for users and the possibility of algorithmically shaped behavior and intentions. Therefore, how to rectify and prevent the spread of this phenomenon is still an important issue to be considered for the sustainable development of the Internet in the future.

# 7. Conclusion

Scott McNealy's statement that 'Privacy is dead – get over it' is not a lie. Information technology is double-edged. The rise of the data economy, which capitalizes on personal information, inevitably leads to the side effect of privacy erosion. Although the online environment provides a place for communication and entertainment that is easy to access information and free from the constraints of time and space, algorithms are indeed taking away people's privacy and using this privacy to influence people's perception and the way they approach the world (Bartlett, 2021). In the digital world, privacy breaches are multidimensional (Carpentieri et al., 2022). When these fragments are combined, a complete portrait of the individual can be constructed (Rauhofer, 2008). The application of algorithms permeates almost every part of people's leisure life, whether it is online entertainment, shopping or dating, it will leave huge data nodes. Therefore, how to prevent the intensification and spread of privacy leakage is still a problem worth thinking about in the future.

# References


[1] Ariadna Matamoros-Fernández, Joanne E. Gray, Louisa Bartolo, Jean Burgess, & Nicolas Suzor. (2021). What's "Up Next"? Investigating Algorithmic Recommendations on YouTube Across Issues and Over Time. Media and Communication, 9(4), 234–249.

https://doi.org/10.17645/mac.v9i4.4184.

[2] Balleys, C., & Coll, S. (2017). Being publicly intimate: teenagers managing online privacy. Media, Culture and Society, 39(6), 885–901.

https://doi.org/10.1177/0163443716679033.

[3] Bartlett, M. (2021). Beyond Privacy: Protecting Data Interests in the Age of Artificial Intelligence. Law, Technology and Humans, 3(Issue 1), 96–108.

[4] Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. Information Communication and Society, 20(1), 30–44.

https://doi.org/10.1080/1369118X.2016.1154086.

[5] Carpentieri, B., Castiglione, A., De Santis, A., Palmieri, F., & Pizzolante, R. (2022). Privacy-preserving Secure Media Streaming for Multi-user Smart Environments. ACM Transactions on Internet Technology, 22(2).

https://doi.org/10.1145/3423047.

[6] Coll, S. (2014). Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. Information Communication and Society, 17(10), 1250–1263.

https://doi.org/10.1080/1369118X.2014.918636.

[7] DW.com. (2019, July 13). US regulators approve $5 billion Facebook fine. Deutsche Welle.

https://www.dw.com/en/facebook-faces-5-billion-fine-over-privacy-violations/a-49575702.

[8] Elvy, S.-A. (2017). Paying for Privacy and the Personal Data Economy. Columbia Law Review, 117(Issue 6), 1369–1460.

[9] Ghosh, I., & Singh, V. (2022). "Not all my friends are friends": Audience-group-based nudges for managing location privacy. Journal of the Association for Information Science & Technology, 73(6), 797–810.

https://doi.org/10.1002/asi.24580.

[10] Imana, B., Korolova, A., & Heidemann, J. (2022). Having your Privacy Cake and Eating it Too: Platform-supported Auditing of Social Media Algorithms for Public Interest. Proceedings of the ACM on Human-Computer Interaction 7, CSCW1, Article 134 (April 2023), 33 Pages.

https://doi.org/10.1145/3579610.

[11] Klinger, U., & Svensson, J. (2018). The End of Media Logics? On Algorithms and Agency. New Media and Society, 20(12), 4653–4670.

https://doi.org/10.1177/1461444818779750.

[12] Khoo, O. (2023). Picturing Diversity: Netflix's Inclusion Strategy and the Netflix Recommender Algorithm (NRA). TELEVISION & NEW MEDIA, 24(3), 281–297.

https://doi.org/10.1177/15274764221102864.

[13] Lovink, G. (2008). The society of the query and the Googlization of our lives: a tribute to Joseph Weizenbaum. Karlsruhe institute of technology.

[14] Liu, Y., Tse, W. K., Kwok, P. Y., & Chiu, Y. H. (2022). Impact of Social Media Behavior on Privacy Information Security Based on Analytic Hierarchy Process. Information (2078-2489), 13(6), 280.

https://doi.org/10.3390/info13060280.

[15] Miao, R., & Li, B. (2022). A user-portraits-based recommendation algorithm for traditional short video industry and security management of user privacy in social networks. Technological Forecasting & Social Change, 185.

https://doi.org/10.1016/j.techfore.2022.122103.

[16] Pajkovic, N. (2022). Algorithms and taste-making: Exposing the Netflix Recommender System's operational logics. Convergence, 28(1), 214–235.

https://doi.org/10.1177/13548565211014464.

[17] Rauhofer, J. (2008). Privacy Is Dead, Get over It: Information Privacy and the Dream of a Risk-Free Society. Information & Communications Technology Law, 17(Issue 3), 185–198.

[18] Westin, A. F. (2003). Social and Political Dimensions of Privacy. Journal of Social Issues, 59(2), 431–453.

https://doi.org/10.1111/1540-4560.00072.