BRIEFING

European Parliament

Unpacking 'commercial surveillance': The state of tracking

SUMMARY

Data is central to the EU's digital transformation. In combination with next generation connectivity and emerging technologies, the free flow of data should increase productivity and competitiveness, improve health and wellbeing, and make services more convenient. However, stakeholders are increasingly concerned that EU data protection rules limit innovation and are therefore jeopardising potential data-driven prosperity. They are calling for targeted weakening of the rules.

As the debate between data protection advocates and critics becomes more polarised, it is more important than ever to establish a baseline for discussions by taking stock of current tracking practices. This briefing illustrates how leading internet companies track users across devices and services. It reviews the latest research and answers questions such as: Do companies track my browsing behaviour? Do mobile phones and apps share data in idle mode? Is my phone listening? Are recent anti-tracking features effective? The briefing is thematically related to the e-privacy regulation, artificial intelligence act and European digital identity (e-ID) framework proposals.

After describing how companies use collected data, the briefing reviews the data processing and tracking practices of the incumbent advertising duopoly Google and Facebook, and also their strongest rival, Amazon. It examines each of these companies' general data processing practices as described in their privacy policies, and at least two specific tracking mechanisms they employ. Pointing to the fact that these three tech giants are nearly ubiquitous on devices, apps and websites, the briefing also shows that companies have started deploying ambient tracking technologies in the physical world as part of their vision of a digitalised future. A separate section looks at tracking on mobile apps, which is closely linked to mobile advertising – the biggest generator of digital advertising revenues. The two final sections underscore possible competition harms associated with Google's and Apple's recent privacy-minded initiatives and draw attention to the ongoing rule-making initiative of the US Federal Trade Commission on 'commercial surveillance and data security'.

IN THIS BRIEFING

EPRS | European Parliamentary Research Service

Author: Hendrik Mildebrath
Members' Research Service

PE 739.266 – December 2022

EN

1. Introduction

To achieve economic objectives and enhance user experience, many companies personalise products, processes and advertisement by approximating and predicting customer attributes and behaviours through profiling and data analysis. Companies collect the necessary input data for these operations through various methods. In addition to requesting information or observing offline behaviour, they collect data by logging and tracking user activities; by scraping, sharing, and trading data; as well as by merging or acquiring companies. Researchers have identified multiple fields of application that rely on tracking and profiling; these include marketing, customer analytics, credit scoring and personal finance, fraud prevention and risk management, employee monitoring, hiring and workforce analytics, insurance, healthcare, and e-learning. Many researchers agree that the key driver, if not the raison d'être behind tracking and profiling, is advertising, which funds the 'free' web. Companies and industry associations often herald big data analytics, frequently implying profiling, as the key to a new era of business intelligence and personalised services. It would help businesses improve operational efficiency and gain/maintain competitiveness. Users too could benefit from tracking and profiling, notably from enhanced functionality and convenience, such as better web searches, relevant online advertising, pertinent product recommendations, effective health services, and expedient fraud prevention. However, 'a single moment of engagement' oftentimes bundles 'interlocking and contradictory [processing] purposes' into the same package.

2. Data processing and tracking practices

Illustrating the extent of tracking across services, this section presents the data practices of global digital advertising leaders Google, Meta (formerly Facebook), and Amazon. In 2021, Google's ad revenues totalled US$209.49 billion worldwide, followed by Meta at US$114.93 billion and Amazon at US$31.16 billion. According to IAB Europe, already in 2016 behavioural targeting accounted for 66 % of all digital advertising and contributed to 90 % of growth in digital advertising. Studies claim that the click-through rate of targeted ads is 5.3 times higher on average than standard run-of-network advertising, and that in 2009 advertisers were willing to pay on average 2.68 times more for targeted ads compared with standard ones. A literature review by G. Johnson found that use of cookie technology increased advertising prices. Recently, researchers have begun questioning the effectiveness of personalised advertising and its advantages over contextual advertising.

a. Google

According to Statista, Google is forecast to earn US$244 billion from digital advertising in 2022, almost double the estimated earnings of its closest competitor, Facebook (US$136 billion). In its privacy policy, Google explains that it collects and combines information as individuals use its services, apps, websites, devices and other products embedded in third-party apps and sites. Google uses various technologies to collect information, including cookies, pixel tags, local storage, databases and server logs. When users are signed into their Google account, Google stores the collected information (including from third-party sites and apps) with the user's Google account. Google also 'collects' information on users from publicly accessible sources and from sources such as directory services, marketing partners and security partners. It also receives information from advertising partners to provide advertising and research services on their behalf.

Google stores data provided by users – such as phone numbers or payment information (**account and contact data**) – and 'collects' content users create, upload or receive from others (**content data**). Additionally, it collects certain platform data and platform-usage data, activity data, and location data. **Platform data and platform-usage data** include unique identifiers, device information and configuration settings, mobile network information (including carrier name and phone numbers), and IP addresses. Certain devices running on Google's Android operating system periodically contact Google servers to provide information about the device and connection to services. Additionally, Google collects **activity data** including search terms, videos watched,

purchase activities, voice and audio information, people with whom users communicate or share content, and activity on third-party sites and apps that use Google services. Google 'may' also measure interactions, such as how a user moves his/her mouse over an ad or if he/she interacts with the page on which the ad appears. Where users place and receive calls or send and receive messages via Google services, Google 'may' collect the user's phone number, the calling-party number, the receiving-party number and many other details. Google derives **location data** from GPS and sensor data (e.g. an accelerometer), IP addresses, user searches and place labels, and access point data.

Google uses the information it 'collect[s] from all [its] services' to provide, maintain, improve, and develop its services; to provide personalised services, including content and ads; to measures performance, not least to 'help' advertisers understand the performance of their ad campaigns; and to communicate with the user. Google also uses data collected through third-party sites and apps with Google services embedded in them for these purposes. It uses automated systems to analyse content provided by the user and algorithms to recognise patterns in data. It shares non-personally identifiable information publicly and with partners such as publishers and advertisers,. Google also allows certain partners to collect information from user browsers or devices for advertising and measurement purposes.

Scrutiny of data processing practices

In a 2018 study on Google Data Collection, D. C. Schmidt catalogues how much data Google collects about users and their most personal habits. A Google spokesperson said 'This report is commissioned by a professional DC lobbyist group, and written by a witness for Oracle in their ongoing copyright litigation with Google. So, it's no surprise that it contains wildly misleading information'.

Following consumer groups' complaints based on the findings of a report by the Norwegian Consumer Council, the Irish Data Protection Commission (DPC) announced that it would launch an investigation on how Google tracks its users. According to the DPC's 2021 annual report, 'The Inquiry set out to establish whether Google has a valid legal basis for processing the location data of its users and whether it meets its obligations as a data controller with regard to transparency. The DPC's preliminary draft decision was provided to Google in December 2021 for its submissions. In a similar vein, 10 consumer groups coordinated by the European Consumer Organisation (BEUC), filed data protection complaints against Google for 'unfairly steering consumers towards its surveillance system when they create a Google account'; see also the report entitled 'Fast track to surveillance'. The complaint references a US class action by Calhoun et al., alleging that Google tracked Chrome users, contrary to its policy statements. Court documents drew attention to little-known identifiers and revealed that Amazon was internally aware that their processing descriptors were worded in a way that confused users.

Google social login

Social integrations (for instance, the **social login**) are examples of embedded website components that facilitate tracking. The term social login refers to offering users the option to log into websites using their social network accounts such as Facebook, Google, Amazon, Twitter and LinkedIn. Independent websites ('relying parties') often outsource user identification to social identity providers (Facebook, Google, Twitter, etc.). As part of the sign-in process, the relying parties can request access to additional personal user data stored with the identity provider. A study from late 2021 analysed the privacy practices of websites that were listed in (the now 'retired') Alexa's Top 500 ranking and used the OAuth authorisation protocol. The researchers analysed login procedures used in Australia, Canada, Germany, India and the United States. They found that many websites allowed users to choose from multiple social login options, of which Google and Facebook were the most popular. The majority of relying parties requested one or more 'basic attributes' such as name and profile picture (96-99 %, depending on the country); in Germany in particular, relying parties requested identity data such as birthday and gender (16 %). Relatively similar patterns emerged across all five countries. Without further analysing functional necessity or usability benefits, the researchers found that relying parties often requested access to different categories of data, depending on the identity providers. In most cases, the first login option requested more data than other login options did, and design choices made a comparison of the options cumbersome. According to the researchers, their results suggest the presence of interface designs akin to dark patterns, that is, interfaces and user experiences that lead users into making unintended, unwilling and potentially harmful decisions.

While these practices allow independent websites to offload work to social sites and afford users convenience, the question about how they benefit identity providers is worth exploring. Already in 2011 F. Corella and K. P. Lewison determined that 'The social site is informed of every social login performed by the user, and may even be informed of the activities of the user in the Web site or application to which the user has logged in'. Facebook has indicated that such data transmission occurs towards the European Parliament. Conversely, Google emphasises in relation to its new identity services for the web that 'Data from Sign In With Google is not used for ads or other non-security purposes'– without specifying the scope of 'data from Sign In With Google' or explaining its prior practices. Essentially, login services are more reliable than other tracking mechanisms, since they request users to confirm their online identity and their providers enforce real-name policies. Reportedly, Janrain, the company that pioneered customer identity and access management (CIAM) and coined the term social login, recommended allaying customers privacy concerns by only requesting access to basic information and by 'build[ing] a supplemental strategy to collect everything else'. Additionally, social logins raise competition and various cybersecurity concerns.

Google Analytics

Many developers employ third-party integrations to understand how users engage with their websites and apps and to improve their services. Google Analytics (GA) is the most popular website analytics platform and, according to BuiltWith, over 35 million websites, including almost 50 % of their top 1 million domains, currently use GA. Not only can GA customers analyse visitor behaviour on a single site but also link information about on-site or in-app user activity with activity from other sites or apps. Google Analytics uses a combination of JavaScript, first-party cookies and pixel GIF images to send data about user activity to the Analytics server. The data is then processed further and reported to the website operator's instance of the Google Analytics tool. Google indicates that where a website uses its 'advertising services like AdSense, including analytics tools like Google Analytics, ... your web browser automatically sends certain information to Google'. The competing analytics providers PIWIK pro and Matomo indicate that Google Analytics, by default, uses data for own purposes. Apparently, the Google Analytics customer can disable data sharing with Google via the data sharing settings, but will then lose functionalities. Google clarifies that if customers enable the data sharing option for 'Google products & services' (and accept Google's Controller-Controller terms), Google 'can', for example, use the shared data to improve its advertising system.

Recently, the Austrian Data Protection Authority (Austrian SA) has looked into two cases relating to Google Analytics. In both instances, a website visitor (complainant) lodged a complaint against the website operators who had embedded Google Analytics into their website (first respondent) as well as Google LLC itself (second respondent), for transferring data to the US as part of Google Analytics' data processing operations. While the procedures mainly concerned the illegality of EU-US data transfers, the Austrian SA addressed possible further processing by Google. In neither case had the data subject (complainant) given Google permission to track and save activities across third-party websites using Google services (GA). Nevertheless, in the decision on the second case, the Austrian SA held that 'In exchange for allowing website operators to use the free version of Google Analytics, the second respondent [Google LLC] receives technical possibilities to collect data and further enrich the profiles of Google account holders. Therefore, it cannot be assumed that Google Analytics is a mere web analytics service for website operators'. Already in its decision on the first case, the Austrian SA (the decision-making body) emphasised that it had the right to perform an official review and carry out data protection audits as regards further processing by Google.

Pixel phone and Android operating system

While the preceding sections mainly dealt with app- and web tracking, this one focuses on the fact that Google also tracks users on Android operating systems and smart devices such as its Pixel phones. In a March 2021 study, D. J. Leith analysed the data Android sends from Google phones (Google Pixel) and compared it to data iOS sends from Apple iPhones. The study revealed that when in idle mode, a Pixel handset sends roughly 1 MB of data to Google every 12 hours while an iPhone

sends 52 KB to Apple, i.e. Google collects around 20 times more handset data than Apple does. The shared data includes persistent identifiers such as hardware serial number, SIM serial number, phone number and device ID. Even if the user explicitly opts out, iOS and Android transmit telemetry data. While Google warns in its settings that this may occur in support of essential services and system updates, Leith argues that 'the "essential" data collection is extensive, and likely at odds with reasonable user expectations'. Leith also highlights the presence of a number of pre-installed apps and services that make network connections, despite never having been opened or used. The collection of so much data (through different channels and at times containing common identifiers) – Leith observes – raises concerns about the linkage of collected data with other data sources. Additionally, the disclosure of IP addresses on average every 4.5 minutes enables Apple and Google to track the device's location over time. Google initially criticised the methods Leith used to obtain his findings, but later conceded that the majority of the research is accurate. Leith perceived Google's criticism as 'positively misleading'.

In a report from October 2021, D. J. Leith et al. analysed the Android-based operating systems by Samsung, Xiaomi, Huawei, Realme, Lineage OS and /e/OS. They found that except for the privacy-focused operating system /e/OS, 'even when minimally configured and the handset is idle these vendor-customized Android variants transmit substantial amounts of information to the OS developer and also to third-parties ... that have pre-installed apps'. They claimed that there is no way to opt out of this data collection, and that even if users reset resettable identifiers, the new identifiers can potentially be relinked to the same physical device or even user, notably when long-lived identifiers are sent in parallel through the same connection as resettable identifiers. Similarly, the researchers noted that there is potential for companies to link their separate data.

Google is increasingly offering smart devices and operating systems, perceived as providing further opportunities for ubiquitous tracking. Google has become a major provider of smart home solutions, wearables and services related to the automotive sector. Smart home solutions and wearables include Google Nest devices, the smart TV operating system Android TV (which has significant success on the market), the smart watch Pixel Watch, and the smartwatch operating system Wear OS. As regards the automotive sector, Google offers, for instance, the platform integration for automotive displays, Android Auto, as well as the autonomous driving technologies and services Waymo Driver (driving technology), Waymo One (ride-hailing service) and Waymo Via (trucking solution). For illustrations of IoT tracking, see the subsections on Amazon's Alexa and Ring.

b. Facebook

Facebook is forecast to generate the second-largest digital advertising revenue – €123.6 billion (US$136 billion) – in 2022. This subsection will give an overview of Meta's data processing practices as described by its privacy policy, and present two tracking mechanisms.

Meta's privacy policy is applicable to Meta products, including Facebook, Messenger, Instagram, Marketplace, Meta Platforms Technologies Products, such as virtual reality products and worlds, and Meta Business Tools such as analytics tools and social plugins embedded in third-party websites. Meta collects information provided by users and logs certain activity, platform, and network data it observes. It uses cookies, device identifiers, Meta-specific IDs (e.g. Family Device IDs), and software such as social plugins and the Meta Pixel to collect data. Additionally, it receives information from various partners (e.g. advertisers, independent websites and app developers), measurement vendors and third parties (e.g. publicly available sources, industry peers and government authorities) about user activities within its products and across third-party websites and apps. It collects information both directly and indirectly, irrespective of whether a user is logged in or even has a Meta account. By design, social media platforms facilitate the (re)sharing of information. Users can report abusive content. However, information can also be shared off-platform. Facebook maps users' social ties in a 'social graph', a term that CEO Mark Zuckerberg introduced in 2007 to describe a graph that maps entities (user, page, place, photo, post, etc.) and their relationships (friendships, check-ins, tags, relationships, ownership, attributes, etc.). This facilitates searches on Facebook and

enables independent developers to build applications based on Facebook's data. Meta continuously updates the technology. Meta also collects data relating to non-Facebook users and their devices and apps each time they visit publicly accessible Meta services, from business partners (information sharing) and from Facebook users (contact uploading).

Meta collects **information provided** by the users, including their account, contact and content data. It logs user activities, including content views, interactions, purchases, hashtags used, use of social plugins and Facebook login, as well as the time, frequency and duration of activities on Meta products (**activity data**). As part of activity data, it also logs what users see through Instagram's camera feature, background sound from voice-enabled features, and communication-related data (subject to the e-privacy rules). Additionally, it records and infers information from and about friends, followers and other connections. It also collects device, network, app, and browser information, including whether Meta's app is in the foreground or if the users' mouse is moving (**platform, platform usage and network data**). When users activate Meta's location services function, Meta collects **location data** from the GPS and, depending on the operating system, other device signals such as nearby Bluetooth or Wi-Fi connections. Even if locations services is turned off, Meta receives and uses location-related information from IP addresses, check-ins, events, metadata from photos and information about internet connections. From partners, measurement vendors and third parties, Meta obtains information about user activities within its products and across third-party websites and apps (**third party data**). This includes information from the websites users visit, the way they use partner products, their purchases and transactions data, as well as information on their interactions with ads. Other information partners share with Meta includes users' email addresses and device IDs (for the purposes of targeted advertising).

Scrutiny of data processing practices

At the request of the Belgian Data Protection Authority's predecessor, B. Van Alsenoy et al. drafted a report analysing Facebook's policies and terms and conditions. Referring to the report and with the backing of several EU SAs, the Belgian SA published two recommendations and in 2015 also filed a lawsuit against Facebook. The lawsuit is pending with the Brussels Court of Appeals after the CJEU confirmed in a preliminary ruling that the Belgian supervisory authority was competent to engage in legal proceedings, despite failing to qualify as lead supervisory authority.

Following the Cambridge Analytica scandal, US Senate and House Committees held hearings with Mark Zuckerberg about Facebook's data use. CEO Mark Zuckerberg acknowledged the collection of data on internet users without a Facebook account (ambiguously termed 'shadow profiles'). Following up, Facebook did not clarify how many data points it held for the average non-Facebook user. While Facebook stated it 'does not create profiles or track website visits for people without a Facebook account', it acknowledged that it 'receive[s] some information from devices and browsers that may be used by non-users', as well as 'browser and app logs'. Considering how the above statement is phrased, it is likely that this information is recorded in logs. Similarly, the EP LIBE Committee held hearings and received two written post-hearing clarifications. Facebook specified that it uses information associated with non-Facebook users to provide services, show ads about Facebook, protect security, and improve products. After the scandal, 'Facebook made changes to certain terms and policies and implemented restrictions on ... data access and sharing'. A recently unsealed transcript of a 2022 discovery hearing, which was part of a lawsuit relating to the Cambridge Analytica scandal, revealed opaque and scattered storage of user data as well as liberal re-use prompting the judicially appointed independent arbiter to wonder about the methods Facebook uses to comply with the GDPR.

In early 2022, Mozilla announced that it is collaborating with the non-profit newsroom The Markup to analyse Facebook's Pixel tracking network and understand the kinds of information it collects on sites across the web. The first results revealed that Facebook is receiving sensitive medical information from hospital websites.

Amongst other things, Meta uses the information it collects to personalise products and ads; provide and improve products; derive business insights (business intelligence); provide measurement, analytics and other business services; as well as to communicate with users. **Product personalisation** includes personalising features, content and recommendations, such as one's Facebook feed, Instagram feed, Stories and ads. To show ads and other sponsored or commercial content, Meta processes information, including users' profile information, their activity while they are on and off products, inferred user interests, and information about other connections. Its systems automatically process the information collected to assess and understand users' interests

and preferences (personalise products), and the policy explicitly mentions profiling with the aim of improving products. In its privacy and data-use business hub, Meta clarifies that certain solutions 'pool events at a person level to build richer understandings of what ads are relevant for people'. Meta's **measurement and analytics services** help partners understand things such as: what types of people interact with their content or use their services and how people interact with their content, websites, apps and services. Meta uses data obtained from partners, measurement vendors and third parties for most of the purposes mentioned in its privacy policy.

Meta shares information with partners, measurement and marketing vendors, service providers and third parties such as external researchers. To facilitate transactions on its products, Meta shares information with **businesses offering goods or services**. When users interact with third-party products or services that are integrated with Meta (e.g. products accessible through Facebook login or products embedded in Meta products), these third parties (**integrated partners)** receive information about users' activity when they interact with the service, as well as information from and about the device they are using. Sometimes these integrated partners ask users for permission to access certain additional information from their accounts. Unless users give permission, Meta does not share data with **publishers and advertisers** that 'by itself' can be used to contact and identify the user. Instead, Meta shares aggregate data, such as reports and data, to facilitate the sale and purchase of advertisement space (e.g. the users' advertising device IDs). Similarly, Meta provides **analytics customers**, who use Meta's analytics services to understand more about how people are using their content and features, with aggregate reports. The reports include the general demographics and interests of the people who interacted with the content. Meta outsources the aggregation of data to **measurement vendors** who then use this aggregated data to produce measurement and analytics reports. Meta shares information about users with companies that help market its company and products, measure the effectiveness of its marketing campaigns, and perform advertising research (**advertising vendors**). Meta explicitly states that it also shares information with **other Meta companies**, such as those providing WhatsApp, Novi and others. Meta also processes information that it receives about users from other Meta companies. Meta apparently also supports 'privacy-safe research' with user data.

The 'Like' button

The **Facebook Like button** has received considerable attention from researchers, including as regards its legal aspects. In 2019, it became the subject of legal proceedings before the CJEU (Case C-40/17, *Fashion ID*, 20 July 2019). In this context Facebook explained that when an internet user visits a website with the Like button or another embedded social plugin, the browser sends Facebook more or less comprehensive information, depending on whether the user is logged into Facebook or not. Even if the user is logged out of Facebook or if a visitor does not have a Facebook account, the browser sends Facebook information about the visited web page, the date and time, as well as other browser-related information. Ultimately, Facebook not only collects information about the liked content but also information from across web pages, which may feed into comprehensive user profiles and 'shadow profiles' (see box above). Facebook's social plugins also include the **share buttons**, as well as **embedded posts** and **comment boxes**. Facebook has recently announced changes to its social plugins in the European region, meaning that visitors would only see Like or Comment buttons, if they are both 1) logged into their Facebook account; and 2) have provided their consent to the 'App and Website Cookies' control.

Meta Pixel

Like Google, Meta offers an analytics solution enabling website operators to track visitors. Only preceded by Google's analytics technologies, Meta Pixel is the most widely distributed (5 %) analytics technology among BuiltWith's top 1 million sites and is live on nearly 6.5 million websites. Meta Pixel is a small piece of code (JavaScript) that website operators can integrate into their website. It enables customers to use Meta's advertising products to better reach people similar to their existing audience or website visitors (targeting), to optimise customers' ads to reach the

people most likely to take action (delivery) and to learn more about website visitors and the conversions that result from ads (measurement). When a user visits a website with Meta Pixel embedded in it, pre-defined actions trigger the reporting of events and cookie IDs to Meta, enabling Meta to track activities, which it may share with the Pixel business customer. Meta Pixel collects five types of data, including HTTP headers (including e.g. IP addresses), Pixel-specific data, such as pixel ID and the cookie, button click data, field names, and other optional data about events determined by developers and marketers. As noted in Meta's terms, Facebook also uses information obtained from websites that install Pixels for its own purposes, i.e. to personalise the features and content (including internet-based ads and recommendations) that it shows on and off its products. P. Bekos et al. demonstrate that seemingly ephemeral Pixel cookie tracking is often persistent (refresh of expiration date upon revisit) and that Meta can theoretically match the activity data collected with Facebook user profiles and account owners (real-name policy) if users access the third-party website containing Pixel by clicking on a link displayed on the Facebook platform (new tagging mechanism).

c. Amazon

Considering the possibilities offered by Amazon's 'data machine', A. Lipsman from eMarketer believes that Amazon will overtake Meta in total advertising revenue, possibly within five years. The privacy policy of Amazon's German website (the only European country website currently offering a non-binding English translation), describes how Amazon collects and processes personal information through its websites, devices, products, services, online and physical stores, and other applications (collectively referred to as Amazon Services). Amazon stores user-provided information in relation to Amazon Services. It automatically logs certain types of information about usage of Amazon Services and about the web browser or device used to access these services or content served by or on behalf of Amazon on other websites. It uses device identifiers, operational and advertising cookies, and other technologies to collect browsing, usage or other technical information to enhance the shopping experience, provide services, understand customer activity, display ads, and prevent fraud. Approved third parties also use these tools in connection with Amazon's display of ads to collect information about users. Third parties include search engines, providers of measurement and analytics services, social media networks and advertising companies. Amazon may also receive information from subsidiaries and third-party sources.

Amazon collects information **provided** by the user such as: the name and phone number submitted as part of the account information, downloaded or otherwise used content, voice recordings when users speak to Alexa, images and videos collected or stored in connection with Amazon Services, credit history information, and device log files and configurations. Information Amazon **automatically collects** and analyses includes IP addresses; computer, device and connection information; information about user interactions with content; **location** of users' devices or computers; device metrics including application usage and connectivity data; Amazon Services metrics; purchase and content use history; as well as the URL clickstream to, through, and from Amazon's website (**certain activity, platform, platform usage, and connection data**). As part of **information from third-party sources**, Amazon receives delivery and address information from e.g. carriers; page view information from certain merchants; and credit history information from credit bureaus. Amazon works with third parties, such as publishers, social media networks, search engines and advertising companies, to improve the relevance of the ads it serves. Some third parties provide Amazon with (allegedly) non-personally identifiable information about users from offline and online sources, which it 'may' use to provide users with more relevant and useful advertising.

Amazon processes personal information to operate, provide, and improve Amazon Services. This includes identifying preferences to personalise features, products, services and recommendations; processing personal information to provide and improve Amazon Services; and using personal information such as activity data to display interest-based advertisements. According to Amazon Germany's help page, Amazon does not use information that personally identifies users nor does it use personally identifiable information to serve interest-based ads.

Amazon Europe shares customers' personal information with Amazon.com, Inc. and certain subsidiaries. Additionally, Amazon Europe shares data with third parties – such as sellers and third-party applications providers (**transaction parties and supporting services**) – that offer services, products, applications or skills for use on or through Amazon Services. Additionally, it shares personal information with companies and individuals that perform functions on Amazon's behalf such as analysing data, providing marketing assistance, providing search results and links, and assessing and managing credit risk (**third-party service providers**). These service providers may use the shared information for no other purposes than to perform their functions. Amazon provides **ad companies** with information that allows them to offer their users more useful and relevant Amazon ads and to measure their effectiveness. In this context, Amazon does not share users' names or other information that 'directly' identifies them. Instead, Amazon uses an advertising identifier such as a cookie, a device identifier or a code derived from applying irreversible cryptography. Like many other internet companies, Amazon's privacy policy is worded in a way that allows them to collect, combine, analyse and share data extensively. It is worth noting that this does not guarantee the GDPR-compliance of their policies and practices.

Scrutiny of data processing practices

Amazon confirmed in its US Securities and Exchange Commission filing that the Luxembourg National Data Protection Commission (CNPD) issued a decision against Amazon Europe Core S.à.r.l. claiming that Amazon's processing of personal data violated the EU GDPR. On these grounds, the CNPD imposed a record fine of €746 million on Amazon regarding its targeted advertising system. Amazon appealed the decision and the Administrative Court responsible suspended the part of the decision threatening to add daily fines of 0.1 % (€746 000) if Amazon did not make changes to its data processes by 15 January 2022. The Administrative Court did not address the €746 million fine, either from a legal standpoint or in terms of the amount charged.

Instead of focusing on Amazon's web and app tracking capabilities, journalists and researchers draw particular attention to the ambient tracking capabilities of Amazon's connected devices and its tech-driven company acquisitions. Specifically, they have repeatedly criticised voice assistant Alexa and Ring smart home systems. These devices tie in with Amazon's vision of augmenting the future with **ambient intelligence**. Alexa's chief scientist Rohit Prasad explains that in 'this paradigm, our environment responds to our requests and anticipates our needs, provides information or suggests actions, and then recedes into the background'. Ambient intelligence focuses on integrating sensors and microprocessors with everyday objects and making them responsive. Already in 2006 K. de Vries predicted that ambient intelligent technologies are 'likely to make the development of successful profiling and personalization algorithms, like the ones currently used by internet companies such as Amazon, even more important than it is today'. Additionally, P. Ahonen et al. see an enhanced risk of companies and governments linking digital and social identities (neighbour, employee, student, etc.) for their own purposes. It is worth noting that like Amazon, Google and Meta are diversifying their core businesses and are respectively working towards ambient computing and the metaverse.

Alexa

Through innovation and technology-driven acquisitions (Yap, Evi and IVONA), Amazon was able to 'leapfrog' Apple and Google in the race for building a speech-enabled virtual assistant. Reportedly, by 2019 Amazon had already sold more than 100 million Alexa-enabled devices. As a key technology for 'making ... life easier', which is often deeply enmeshed with users' private lives, voice assistants are exceptionally polarising. Anecdotal evidence from US data access requests demonstrate Amazon's ability to compile intimate portraits of individual users. One reporter's dossier revealed that Amazon had collected more than 90 000 Alexa recordings of family members over 3.5 years– averaging about 70 daily. The recordings revealed that Alexa sometimes recorded longer conversations rather than stopping when the user command ended.

In a 2022 study, U. Iqbal et al. found evidence that Amazon personalises ads based on user interactions with Amazon Echo. The results 'indicate that (i) Amazon Echo user interactions are tracked by both Amazon and third-parties, (ii) Amazon used Amazon Echo interactions for ad targeting on-platform (e.g., audio ads) and off-platform (e.g., web ads), and (iii) Amazon computed

user interests from voice data in a way that was inconsistent with their public statements'. The researchers also noted inconsistencies in Amazon's responses to their data access requests. While Amazon– in line with what it states– may not literally be using the 'recordings', but rather transcripts thereof, results suggest that Amazon is 'processing voice recordings, inferring interests, and using those to target ads'. The distinction between voice recordings and transcribed recordings may not be meaningful to many users. It remains uncertain to what extent Amazon applies voice and speech analysis. A spokesperson from Amazon challenged the researchers' methodology but confirmed that Amazon uses voice data from Alexa interactions to inform ads. In response, the researchers pointed out that Amazon did not directly address their findings. In a 2021 cross-country large-scale study analysing 90 192 unique Alexa apps ('Skills'), C. Lentzsch et al. found that only 24.2 % of all Skills provided privacy policies and that 23.3 % of Skills' privacy policies did not fully disclose the data types associated with the permissions requested.

Scrutiny of voice assistants

Various voice assistants were found to record and transcript snippets outside of any activation. Transcripts were shared with external subcontractors for corrections. Ireland's Data Protection Commission (DPC) is the lead supervisory authority for Apple and Google, while the CNPD is the one for Amazon. As per their remit, these supervisory authorities engaged with the three companies concerned in 2019 and 2020. Reportedly, they did not open official investigations or impose meaningful fines. In 2020, a whistle blower encouraged 'proper investigation and action', as well as 'further sanctions – including fines and bans of processing – to prevent these big companies from initiating any similar programmes in the future'. Meanwhile, the EDPB published guidelines on virtual voice assistants. Amazon, Apple, Google, and Microsoft committed to making changes to their grading programmes.

Ring

Strategy Analytics estimated that Ring Doorbell was the most sold doorbell in 2021 globally, with 1.7 million units sold. A 2020 BBC data subject access request (DSAR) revealed that Ring stores detailed records of user activity, including camera events and app interactions. According to tests by Consumer Reports, some of Ring's doorbell cameras can record audio from about 20 feet away. The doorbells' cameras can perform a combination of face and body-shape analysis, to spot the differences between humans and other living things (camera-based motion detection).

Ring lists 14 ways Amazon can use collected data. This includes improving the service, protecting against fraud, personalising the user experience, identifying and authenticating the user, conducting consumer research and complying with 'relevant industry standards and policies'. Additionally, Ring uses information about users' online activities to provide them with personalised advertising. It explicitly states that Ring 'may use these automated technologies to collect information about your equipment, browsing actions, and usage patterns'. Ring anonymises some of the collected information and uses it to perform analytics (if not for other purposes). Ring also uses third-party services, such as Google Analytics, to analyse and understand individuals' web and app usage. It is not clear from Ring's policy whether Ring configured these services in a way that allows their providers to use the data for their own purposes. In early 2020, the Electronic Frontier Foundation discovered that the Ring doorbell app under investigation sent personally identifiable information to four analytics and marketing companies, including Facebook. It is worth noting that Ring reserves the right to store, alter and use audio, video, images and other content that a user posts to Ring's Neighbour app, but stipulates that the user must warrant that the posting and use of the content does not violate the privacy and other rights of any person. Reportedly, Ring used doorbell videos posted to its app for ads in 2019.

In its privacy policy, Ring clarifies that the user is 'solely' responsible for ensuring their compliance with the applicable law when using Ring's products or services (e.g. display a notice). These may create up to two rectangular privacy zones, with a minimum size of approximately 15 % of the screen. Ring will not display or record what is happening in these zones. This feature is not available in Ring's first generation doorbell. Additionally, users can switch off audio streaming and recording. Nevertheless, the graphics in the positioning guides depict the field views of Ring's doorbells capturing footage of events occurring on public property. Researchers take issue with the fact that

Ring shifts the compliance onus to users. In a recent case, the Oxford country court held that two Ring devices invaded the neighbour's privacy.

While Amazon's tech acquisitions continue to raise privacy concerns, it is worth noting that Amazon too offers a range of technologies that may enable it to track users across third-party sites and apps. These include 'Login with Amazon', 'Alexa for Video Publishers', 'Amazon Simple Email Service (SES)', and 'Amazon Attribution'. The terms of use of Amazon's Sizmek Ad Suite explicitly state that it may use digital advertising technologies in relation to websites or applications of customers, their respective (downstream) customers, and other online presence.

The US debate on Ring doorbells

In the US, digital rights groups, journalists and researchers have raised concerns reaching beyond commercial tracking, such as:

- controversial cooperation with law enforcement that Ring itself had submitted to an external audit;
- the audio and video streaming and recording of protected third parties, e.g. moving on public property;
- the dissemination and sharing of recorded content on platforms;
- a possible reality show based on footage from Ring devices;
- improper access to captured content by Ring employees;
- cybersecurity vulnerabilities that allow third parties to access stored and streamed data;
- possible augmentation of Ring products with biometric surveillance capabilities.

d. Tracking on mobile apps

This final sub-section reviews tracking and profiling practices of native mobile apps available in the iOS and Android app stores. Many companies have developed dedicated apps for services they previously exclusively offered over the internet, not least because it allows them to leverage functionalities and resources of mobile platforms and frees them from the limitations posed by web and browser-related standards. As reported by IAB, mobile ad revenues now account for the majority of internet ad revenues plateauing at around 70 % of the internet-based ad market. Between December 2019 and February 2020, K. Kollnig et al. analysed 12 000 free apps from each of the iOS and Android platforms. They discovered that Android apps contained on average 3.8 tracking libraries, whereas iOS apps contained on average 3.1 tracking libraries. They discovered that 81.44 % of the Android apps and 68.46 % of iOS apps share data with tracking companies before any user interaction, which 'suggests potentially widespread violations of applicable data protection law'. The overwhelming majority of apps share data with tracking companies owned by Alphabet, Google's parent company. Conversely, Apple can collect tracking data from more than two thirds of iOS apps. The next in line in this regard is Facebook, with a similar presence on both iOS and Android platforms. IOS apps – such as Bluetooth, Calendar and Camera – consistently requested more opt-in permissions when comparing the permissions that exist on both platforms. 'Once a permission is granted, an app can usually access sensitive data anytime without the user's knowledge'. It is worth noting that since the 2019-2020 analysis, Apple has taken privacy measures as described in the next section and K. Kollnig et al. published a follow-up study that took these developments into account. The researchers concluded that Apple's changes 'make tracking individual users more difficult, they motivate a countermovement [by ostracized actors], and reinforce exiting market power of gatekeeper companies with access to large troves of first-party data'. '[T]racking companies, especially larger ones with access to large troves of first-party data, can still track users behind the scenes'.

4. A privacy-minded paradigm shift?

Recently, Google and Apple have announced and implemented restrictions on web- and app-tracking, sparking widespread criticism. Apple began blocking third-party cookies in its Safari **browser** through its Intelligent Tracking Prevention (ITP) initiative in September 2017. Google announced a similar initiative, known as Privacy Sandbox for its Chrome browser (CPS), but postponed its implementation to the end of 2023 after considerable pushback from the industry. Both Apple and Google have been criticised for self-preferencing, introducing double standards, and causing harm to competition. On 11 February 2022, the UK's Competition and Markets Authority (CMA) accepted legally binding commitments from Google to address the CMA's competition concerns and support a competitive ad-funded web. The CMA suggested that Apple too should

consider using privacy-preserving technologies that mitigate potential competition harms. However, '[u]nlike Google, Apple does not have a significant position in online search and display advertising which could be advantaged by these privacy changes'. Similarly, Apple introduced App Tracking Transparency (ATT), restricting third-party tracking options on its mobile **devices**, while Google launched its Privacy Sandbox for Android (APS). Again, industry players criticised them for self-preferencing and for applying double standards. Based on preliminary information, the UK CMA considers that both APS and ATT could similarly harm competition, but that self-preferencing could be more acute for APS than for ATT. Google has indicated that it intends to apply to its proposed APS the principles of the commitments made to the CMA regarding the CPS. Google's initial solutions were criticised by privacy advocates. It is hard to predict what the reactions of the advertising industry and the long-term implications will be. In terms of privacy, these initiatives may present a turning point, a bargaining phase, or window dressing for self-preferencing leading to more market shares and data.

Is my phone listening? Maybe.

Facebook (under oath) and Google denied using microphones to inform ads or (certain) other functions without consent. Computer scientists such as J.L. Kröger and P. Raschke, S. Abhishek Anand et al., and S. Bayerl demonstrated that it is possible to program apps that collect motion sensor data and reconstruct spoken words or phrases, as well as disseminate the apps via app stores. Back in 2016, cybersecurity expert D. Lodge built an app that surreptitiously collected audio recordings from the microphone of an Android smartphone. Co-author J.L. Kröger tweeted 'At the bottom line, whether sensitive information is extracted from private conversations or collected from other sources does not make much difference to the possibilities of data exploitation and the entailing consequences for the data subject. ... Therefore, whether justified or not, the suspicions examined ... lead to a very fundamental question: What degree of surveillance should be considered acceptable for commercial purposes like targeted advertising'.

5. FTC deep dive into 'commercial surveillance'

In a joint statement accompanying an order for information on privacy practices addressed to social media and video streaming providers three US Federal Trade Commissioners (FTC) stated that:

> *Despite their central role in our daily lives, the decisions that prominent online platforms make regarding consumers and consumer data remain shrouded in secrecy. Critical questions about business models, algorithms, and data collection and use have gone unanswered. ... It is alarming that we still know so little about companies that know so much about us. ... Never before has there been an industry capable of surveilling and monetizing so much of our personal lives. ... This constant access allows these firms to monitor where users go, the people with whom they interact, and what they are doing. But to what end? Is this surveillance used to build psychological profiles of users? Predict their behavior? Manipulate experiences to generate ad sales? Promote content to capture attention or shape discourse? ...*

Two years on, the FTC is considering regulatory action on commercial surveillance and data security. It received over 11 000 comments and over 1 000 posts in response to its public consultation.

eprs@ep.europa.eu (contact)

www.eprs.ep.parl.union.eu (intranet)

www.europarl.europa.eu/thinktank (internet)

http://epthinktank.eu (blog)