Publication date:
12 Jun 2019
Author:
Eden Zoller
Omdia Distinguished Analyst & Principal Analyst, Smart Living

Making Data Privacy an Asset in the AI Era: Mastering Market Dynamics

Brought to you by Informa Tech

Table of Contents:

Table of Figures:

Summary

In brief

The breadth and depth of consumer data insights are proliferating and will become ever more granular going forward. This is being driven by increasing connectivity across multiple consumer touchpoints and domains, along with advances in artificial intelligence (AI). This has the potential to benefit service providers and consumers in terms of service innovation and advanced personalization. However, the benefits can only be realized if service providers and other stakeholders treat data privacy as a long game and plan now for the challenges ahead. In the first of a two-part report series, we examine the complex forces impacting data privacy. It is only by understanding these forces that we can formulate the optimum privacy strategies and best practices, which will be the subject of our follow-on report.

Ovum view

- **The rise of "data fracking" has compromised privacy**. The prevailing approach to leveraging data insights is the equivalent of fracking. "Data fracking" is a high-pressure, intrusive process that operates in the margins of regulation and consumer acceptance to maximize the volume and variety of data sets and the speed of extraction.
- **The data super platform/AI hyper-scaler mix puts intense pressure on data privacy.** Powerful consumer tech firms such as Facebook and Google are so adept at data collection at scale that they have become data super platforms. These players are also AI hyper-scalers, and the data imperatives that drive them make privacy vulnerable. This is exactly why they should do more to protect and respect data privacy.
- **Moves are afoot to link data privacy to competition**. Certain regulatory authorities, government ministers, and industry bodies are looking at data privacy in the context of anticompetitive practices, particularly in Europe. It is an interesting, albeit problematic, line of enquiry, but one that could have far-reaching implications if regulators and governments get behind it.
- **AI is changing the data privacy stakes.** AI-powered analytics, devices, and services will be able to generate increasingly broad and deep data sets, at scale and speed. AI machine learning (ML) models can detect hitherto unforeseen patterns and connections, and make increasingly accurate predictions that would have been unimaginable even a few years ago**.** This introduces new complexity and challenges for data privacy.
- **Data privacy laws are not keeping pace with AI.** AI developments are moving quickly, and existing laws are struggling to keep pace and to factor in AI impacts. For example, AI systems could use data for new and often unforeseen purposes beyond the original scope. This scenario is at odds with the European Union's (EU) General Data Protection Regulation (GDPR) law that data should be used for specified and explicit purposes and not used in a way that someone would not expect.

Recommendations for consumer service providers

- **Data privacy in the digital economy must be rebalanced: from "data fracking" to "data friending."**Consumer service providers need to adopt a proactive, relationship-first approach to data privacy that is based on genuine transparency and consumer control. This is a "data friending" strategy, where the focus is on engaging with consumers to build a positive relationship around data privacy. In this scenario, data privacy becomes an asset.
- **Embrace the principles of GDPR, even if you don't have to.** The EU's GDPR is by no means the only regulatory framework for data privacy but it is a good reference point. Its core principles bring accountability, transparency, and consent center stage in ways that previous privacy regulation has not, and we recommend that all organizations embrace its tenets, even if they are not required to by law.
- **Be mindful of AI impacts on data privacy – even if a collision between the two is not immediate.** AI presents challenges to privacy, and service providers must do their best to be aware of what the flash points could be. This will not always be easy as AI impacts on privacy will keep emerging – and evolving. However, it is clear that "explainability" and transparency will be even more critical for data privacy in the AI era and those that adopt this mantra will be on solid ground. Service providers should also look to collaborate with industry bodies and other organizations to track the impact and how best to respond.
- **AI can bring benefits to data privacy.** There is a very real risk that AI will hit data privacy in negative ways, but it can also be a force for good, and service providers can leverage it in ways that bring benefits both internally and externally – for example, the use of AI to monitor data patterns and detect anomalies and potentially fraudulent activity in real time, and the use of AI to continuously monitor a company's collection and use of consumer data.

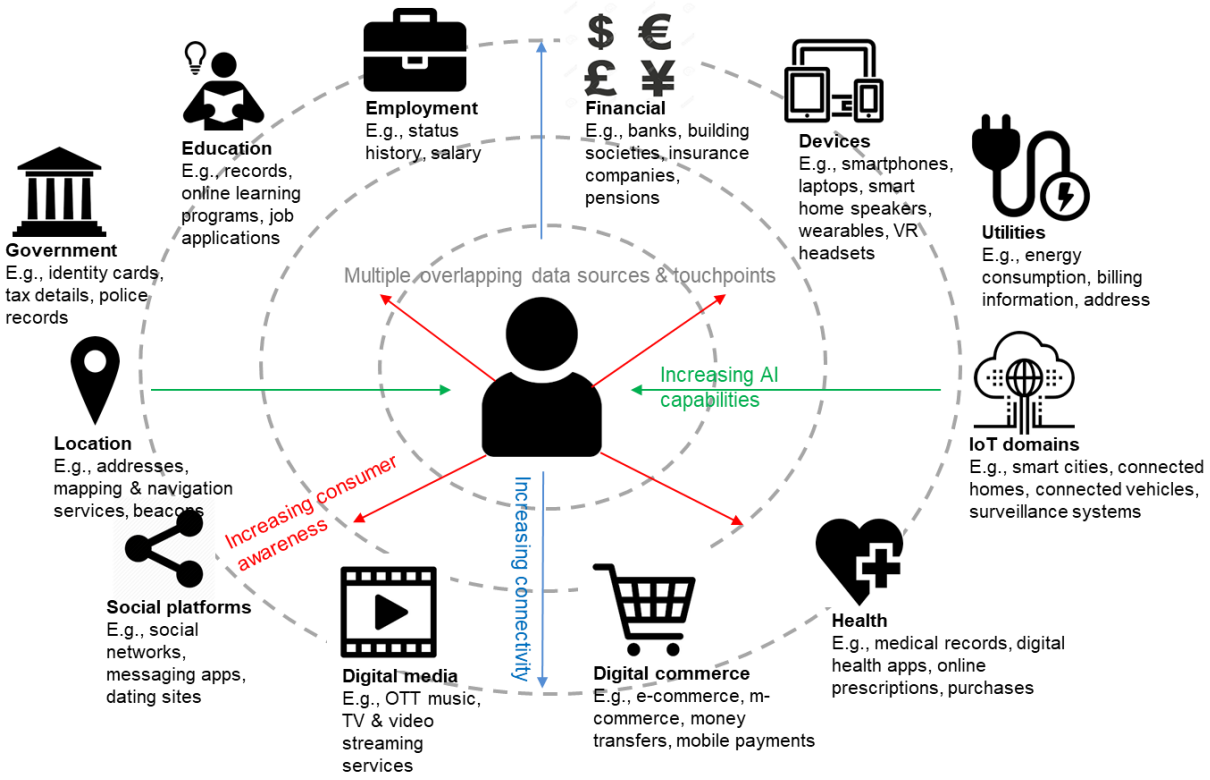
The consumer data explosion

The "datafication" of everything

Consumer data is proliferating in unprecedented ways and at increasing velocity, driven by expanding connectivity, consumer touchpoints and devices, data sources, and AI capabilities, as shown in Figure 1. We live in a data-driven world where our digital footprint is ever widening and covering new ground, leaving a trail that is hard to eradicate and that others are keen to track. Taken together, these trends are raising the stakes in consumer data privacy.

Figure 1: Increasing connectivity, touchpoints, and AI capabilities driving the data storm

Figure 1: Increasing connectivity, touchpoints, and AI capabilities driving the data storm1



Data is growing exponentially

In 2012, Facebook revealed that it processed 2.5 billion pieces of content and in excess of 500 terabytes of data each day. At this time, it generated 2.7 billion Likes and 300 million photos per day, and it scanned around 105 terabytes of data each half hour (it has not subsequently updated these figures). In 2013, IBM estimated that the world produced 2.5 quintillion bytes of data per day, and that 90% of all data had been produced in two years from 2013. These estimates go back several years and are staggering, and there is no doubt that global data volumes will have increased exponentially since 2013 and will continue to do so going forward.

Online services of various kinds generate massive amounts of data as consumers interact with digital services and commerce, and also generate their own content. Ovum has illustrated the scope of this activity in Table 1 with highlights from just a few of the most conspicuous online players.

Table 1: Selected AI technologies and impacts on consumer services

Company	Selected stats
Google Search	Around 3.05 billion searches globally per day
Google Maps	Google Maps has over 1 billion monthly active users worldwide, is subject to 25 million updates per day, and covers 99% of the world
YouTube (Google)	Over 1.9 billion logged in users visit YouTube each month Users watch over 1 billion hours of video per day In 2015, 400 hours of content were uploaded to YouTube every minute (globally), up from 300 hours of content uploaded every minute in 2014 Based on the above trajectory, approximately 800 hours of content are uploaded every minute today (2019)
Facebook	1.52 billion daily active users as of 4Q18, and 2.32 billion monthly active users 2.7 billion people use Facebook, Instagram, WhatsApp, or Messenger each month; more than 2 billion people use at least one of the services every day on average
Instagram (Facebook)	1.06 billion monthly active users as of 4Q18 4.37 billion messages sent per month Over 100 million-plus photos and videos uploaded to Instagram per day (globally)
Twitter	321 million average monthly active users as of 4Q18 Over 500 million Tweets are sent each day
WordPress	Popular online publishing platform currently powering more than 32% of the web WordPress users produce about 70 million new posts and 77 million new comments each month Over 409 million people view more than 20 billion WordPress pages each month

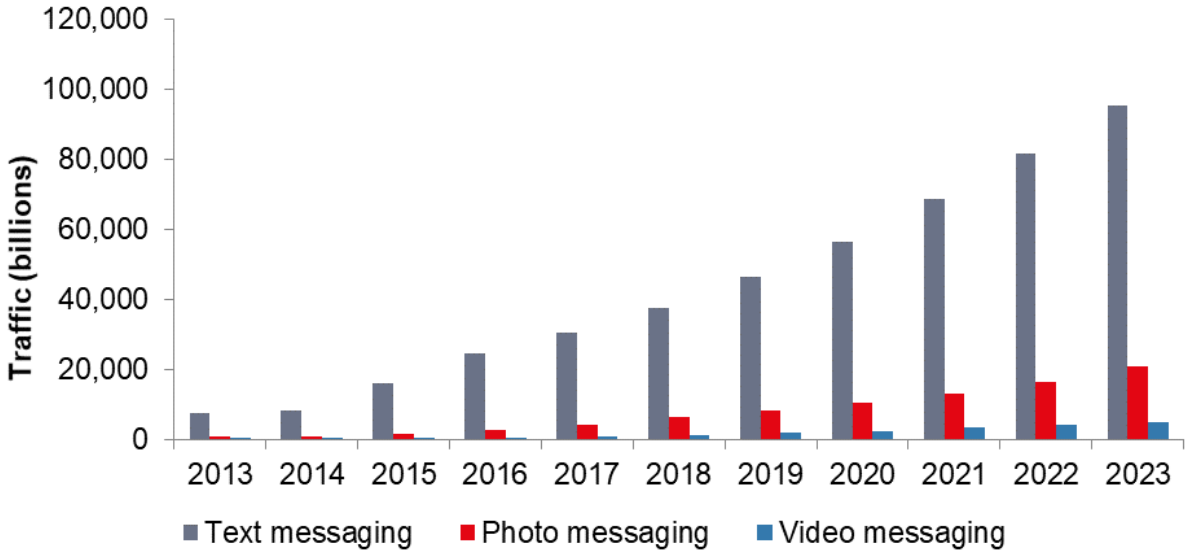
Source: Ovum, company data

The pivotal role of mobile

Mobile devices are becoming the premier platforms through which consumers interact with digital services and through which they generate data. Ovum predicts that the installed base of mobile phones (smartphones and feature phones) will increase from 6.23 billion at the end of 2018 to 6.54 billion by the close of 2022. Mobile messaging services alone are huge generators of personal data: Ovum estimates that total annual mobile OTT messaging traffic will increase from 35.89 trillion messages in 2017 to 121.56 trillion messages in 2023, as seen in Figure 2.

Figure 2: OTT global mobile messaging traffic by content type

Figure 2: OTT global mobile messaging traffic by content type2

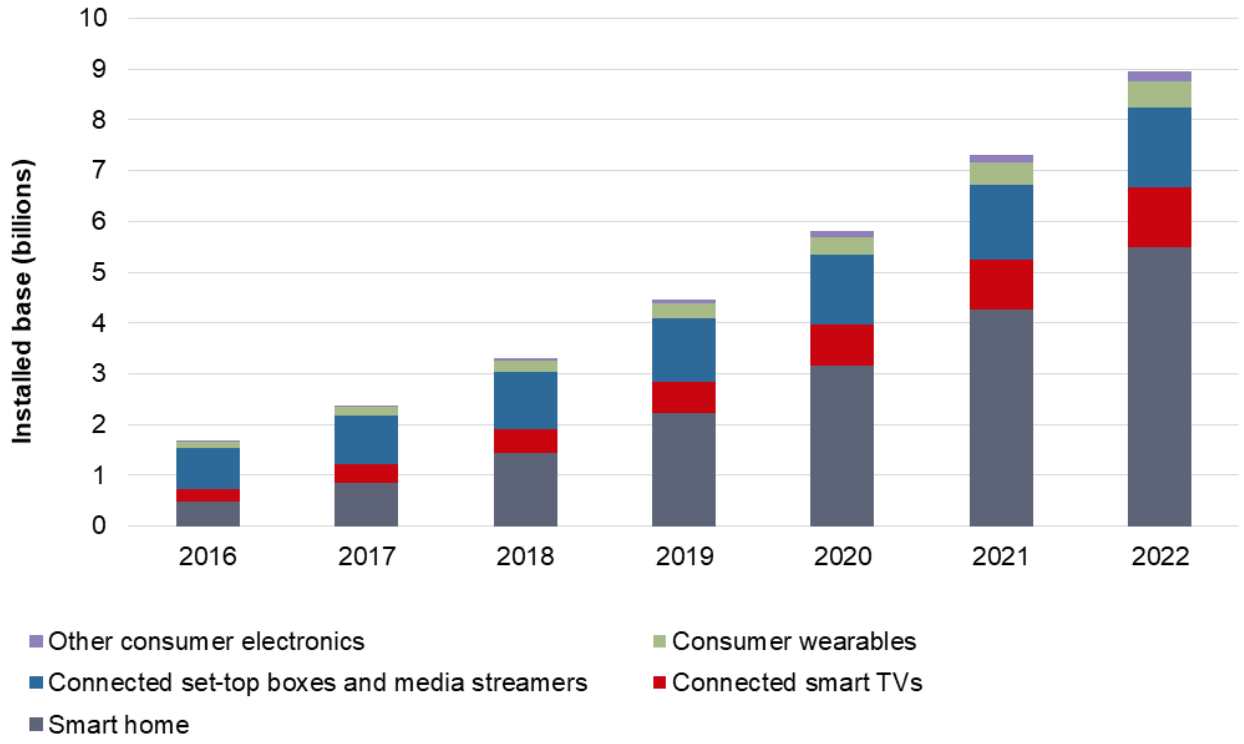


The IoT data floodgates are open

The wider IoT domain is enabling and producing an unstoppable deluge of new data. Ovum expected the installed base of connected IoT consumer electronics devices to rise from 3.31 billion at the end of 2018 to 8.96 billion by the close of 2018 (see Figure 3), with the device spectrum spanning wearables, connected smart TVs, set-top boxes, media streamers, and smart home devices. The latter covers a wide spectrum: interactive audio speakers, connected household appliances, security devices, and utility devices that allow control of functions such as lighting, heating, or energy monitoring. Then there is the wider IoT ecosystem to consider, particularly the contribution from the explosion of sensor networks on roads, in smart cities, and in connected vehicles. Ovum predicts the global installed base of connected vehicles will rise from a total of 115 million at the end of 2018 to 308 million by 2022. All these devices and touchpoints are generating new and deeper data insights.

Figure 3: Global installed base for selected consumer electronics devices

Figure 3: Global installed base for selected consumer electronics devices3



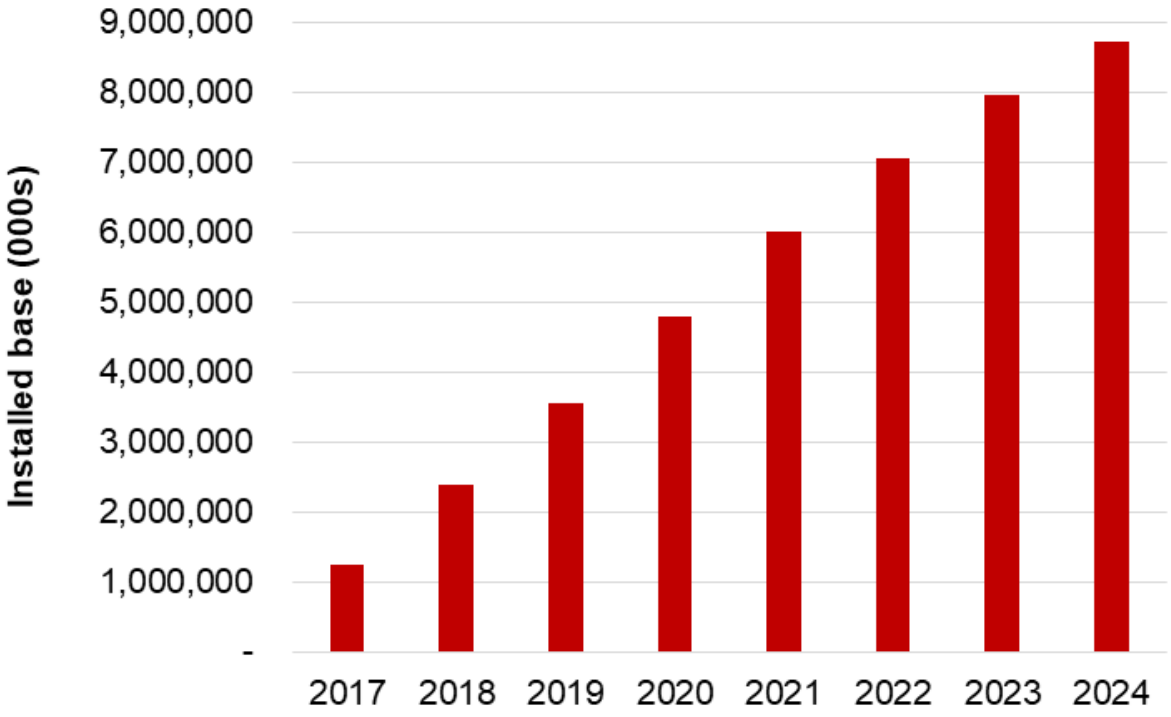
AI's impact on consumer services and devices is proliferating

Bringing ever more granular data insights

AI technologies are being deployed across the consumer domain, a trend that will accelerate and often without consumers even being aware of it. AI assistants are becoming a mainstream touchpoint for consumer interactions with AI, mainly from smartphones but increasingly from other connected devices and platforms. Ovum predicts that the global installed base for AI assistants across all device types will grow from 2.39 billion in 2018 to 8.73 billion by the end of 2024, as shown in Figure 4. AI assistants today leverage ML (e.g., for recommendations and predictive actions/next steps), AI voice technologies (e.g., speech recognition and natural language processing) and, increasingly, visual AI (e.g., image recognition and object detection). Data insights of some kind are generated each time a user interacts with an AI assistant. Exposure to user data also provides inputs and training opportunities for underlying ML models on which AI assistants are built, enabling them to become smarter. AI assistants have the potential to become the ultimate profiling tools.

Figure 4: Rapid growth ahead for global installed base of AI assistants

Figure 4: Rapid growth ahead for global installed base of AI assistants4



AI brings new complexity to data privacy

New AI-powered scenarios and data insights beckon

AI assistants are a very conspicuous way for AI to impact consumers, but there are a multitude of other scenarios where AI technologies are already at play or will be soon, as summarized in Table 2. The table highlights only a few examples, but what jumps out is how advances in AI technology are leveraging completely new types of consumer data insights. For example, the way that computer vision is driving developments in facial recognition, along with image, object, and gesture detection. Cognitive computing is driving advances in emotion detection in increasingly nuanced ways, enabling AI systems to interpret happiness, sadness, anger, and so on. At the same time, AI-powered analytics can also interrogate and extract meaning from unstructured data (e.g., text messages, images, videos, audio, blogs, and IoT/sensor data), and increasingly in real time or close to it. The bottom line is that advanced AI analytics can detect hitherto unforeseen patterns and connections, and make increasingly accurate predictions that would have been unimaginable even a few years ago.

Table 2: Selected AI technologies and impacts on consumer services

	Media and entertainment	**Smart home**	**Commerce**	**Automotive**
Voice technologies (e.g., voice recognition and natural language processing)	Voice-controlled TV and games	Voice interactions with/control of home devices and appliances	Voice-driven shopping, identity verification, and payments authentication	Voice-activated vehicle locking/unlocking; in-vehicle service activation
Immersive AI (e.g., AR, VR, and holograms)	AR-enhanced TV and video, and AR/VR games	Hologram-based digital assistants	Augmented online shopping and mobile apps, AR/VR enhanced in-store experiences; identity verification, and payments authentication	AR-enhanced data/information on head-up displays (HUDs)
Visual recognition (e.g., facial, image and object, and gesture)	Facial recognition for identity verification for restricted content, and to gauge audience reaction to visual media	Visual interactions with device screens; control of devices and appliances via gesture	Visual product search and recommendations based on appearance; interactions with products shown on displays	Object/obstruction detection
Sensor technologies	Motion effects in games	Security system activation	Indoor mapping at retail venues and footfall analysis	Road obstruction detection, high definition, and 3D mapping; autonomous driving
Predictive computing/analytics	Content recommendations; content development based on consumer preferences or reactions	Optimum times for automated lights to go on/off and heating program	Offers on the fly, real-time pricing optimization; contextual recommendations, and automatic reordering	Predictive traffic conditions, weather conditions for journey, suggested points of interest, and vehicle maintenance alerts

Emotion detection	Content recommendations according to mood; assessment of consumer reaction to content based on emotional state	Inhabitant emotion detection to set optimum lighting, audio volumes, and temperature	Gauge degree of consumer engagement with, and attitudes toward, a product/service based on emotional state	Driver sentiment/emotion detection (e.g., fatigue and agitation) to prescribe actions/remedies to alleviate state

Source: Ovum

The dangers of identifying the unidentifiable

One of the core tenets of data privacy is that consumer data insights shared with third parties are anonymized so that individuals cannot be detected. However, AI systems have the potential to detect individual data subjects from aggregated, anonymized data, which has serious data privacy implications. To function at their best, ML systems need big representative data sets that contain variety and depth of data, some of which can be highly sensitive, such as personal photos, emails, and even medical information. ML systems interrogate the data and learn how to link data elements and make inferences to surface patterns and relationships. This process can produce insights that are highly beneficial to society; for example, building ML-based models for cancer detection that learns from sensitive patient data. However, ML systems can work to identify individuals or attribute personal data to them from what starts as unidentifiable data.

Differential privacy can help reduce the risks

There is clearly a need for ML models to access data in order to train and carry out tasks, but they must do so in a way that simultaneously preserves data privacy. Implementing solutions that balance these needs is clearly no mean feat but, on a positive note, developments are in motion to help achieve this. One approach is based on the concept of differential privacy. This is not entirely new and is a complex algorithmic technique that, put very simply, works by inserting random data into a data set but in a way that does not change the aggregate meaning of that data. The introduction "noise" does make the data slightly less precise, but the trade-off is better data privacy. Differential privacy has been adopted in various ways by Microsoft, Google, and Apple. Google uses the technique in the AI-based Smart Reply feature in Gmail to prevent sensitive, identifiable personal data being surfaced. In March 2019, Google introduced a differential privacy module for TensorFlow, its popular ML framework that is used by developers to build ML applications.

Data privacy laws will need to adapt

AI technologies are developing quickly, and so too is their impact on consumer services. Existing consumer data privacy regulations are struggling to keep pace, which is inevitable and will continue as AI further evolves. This means that existing data privacy frameworks will need to evolve and in some cases new principles may well be required. However, it must be approached in a way that strikes a fair balance between protecting individual data privacy rights and ensuring that AI can bring benefits to society.

When examining AI in the privacy context, it is also critical regulators do so in consultation with the wider AI community, as it is highly unlikely that regulators working in isolation will understand all the issues at stake. Legal authorities should draw on the expertise and experience of technical and industry bodies, universities and research institutes, enterprises, telcos, and consumer tech and commerce players. (Ovum examines AI

in the context of data privacy regulations later in the report section, "Data privacy legal regimes are a work in progress").

But AI can be a force for good in the data privacy domain

Although there is no doubt that AI poses fresh challenges for consumer data privacy, it can also bring benefits and this should not be overlooked. The most immediate scenario is data security, which of course has a privacy element. AI-powered analytics can be a highly effective way to monitor data patterns and detect anomalies, and potentially fraudulent activity in real time, while AI-powered security systems can provide automated procedures to prevent fraud or inappropriate use of data. Service providers could give users AI-powered tools that alert them in real time of scams, suspicious websites, or other malicious activity that could harm or infringe their data. There is scope for AI-powered privacy tools that remember an individual's privacy preferences and make them consistent across a single organization's services, or even across multiple sites, although the latter is a much more challenging proposition.

The consumer view: concern and uncertainty

Attitudes and actions are not as contrary as they may seem

No understanding of data privacy is complete without considering the consumer perspective, which is not always straightforward as attitudes and behavior can appear confused and at times inconsistent. Most consumers are concerned about data privacy at some level but don't seem to alter their behavior as a result, although this is starting to change. However, the discrepancy is in fact understandable. People may be aware that data privacy is an issue in an abstract way, but do not really understand what the problems are or what is at stake. They may feel uneasy about data privacy but are confused and don't know what to do about it. In this situation it can be easier to do nothing and carry on as normal. There are of course other consumer responses to data privacy, which we will study more closely in relation to consumer data privacy segmentation in the forthcoming, complementary report in this series (*Making Data Privacy an Asset in the AI Era: Best Practice*).

However, even for those consumers that want to be more proactive about data privacy, the reality is that it is made difficult for them to do so. Regulations such as GDPR are meant to give consumers more control over data privacy, but the way this has fed through online is with a bewildering deluge of forms, settings, and small print, and the prospect of being blocked if consent is not granted. Faced with this scenario, people give in, accept the terms, and carry on as before. GDPR may have made digital service providers and enterprises more accountable, but it has done little to explain data privacy dynamics to consumers or to make taking control of their data any easier.

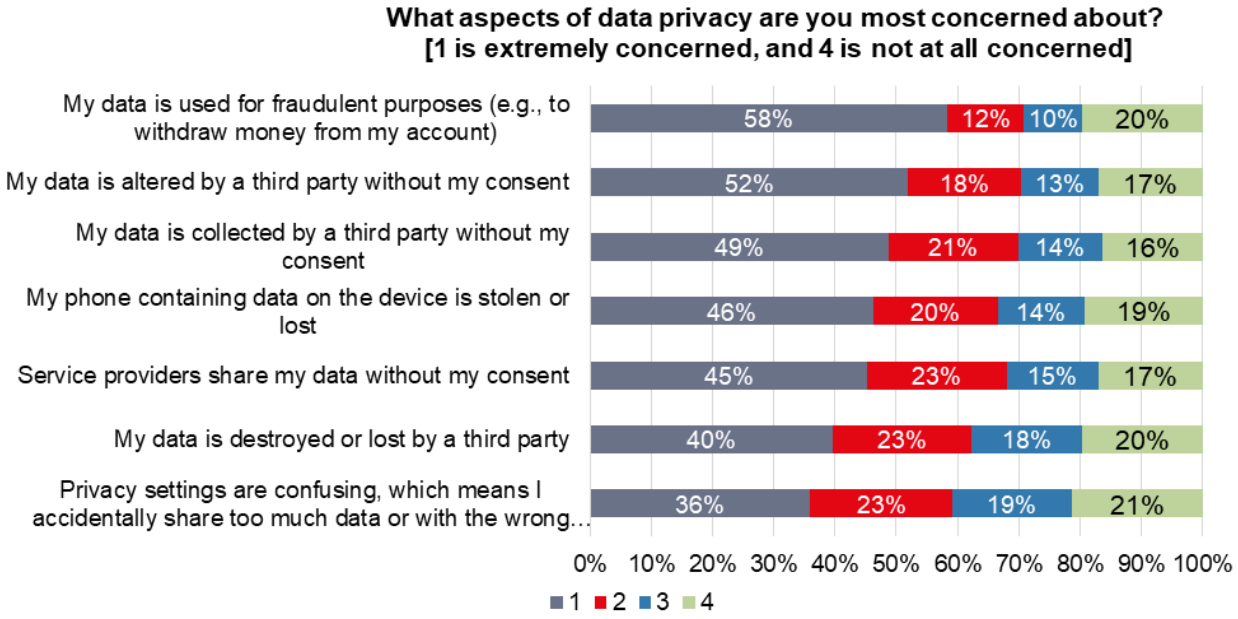
Concerns are becoming entrenched

When it comes to what worries people about data privacy, the most acute pain points are becoming entrenched, as shown in Ovum's Consumer Digital Insights survey, and indeed other industry surveys. The

two areas of data privacy that worry consumers the most relate to security and consent – Ovum's Consumer Digital Insights survey found that the fraudulent use of personal data is the top concern for almost 60% of respondents, as shown in Figure 5. Concerns relating to consent are likewise front of mind and include data collection without permission and altering or sharing personal data without consent.

Figure 5: Data privacy pain points for consumers

Figure 5: Data privacy pain points for consumers5

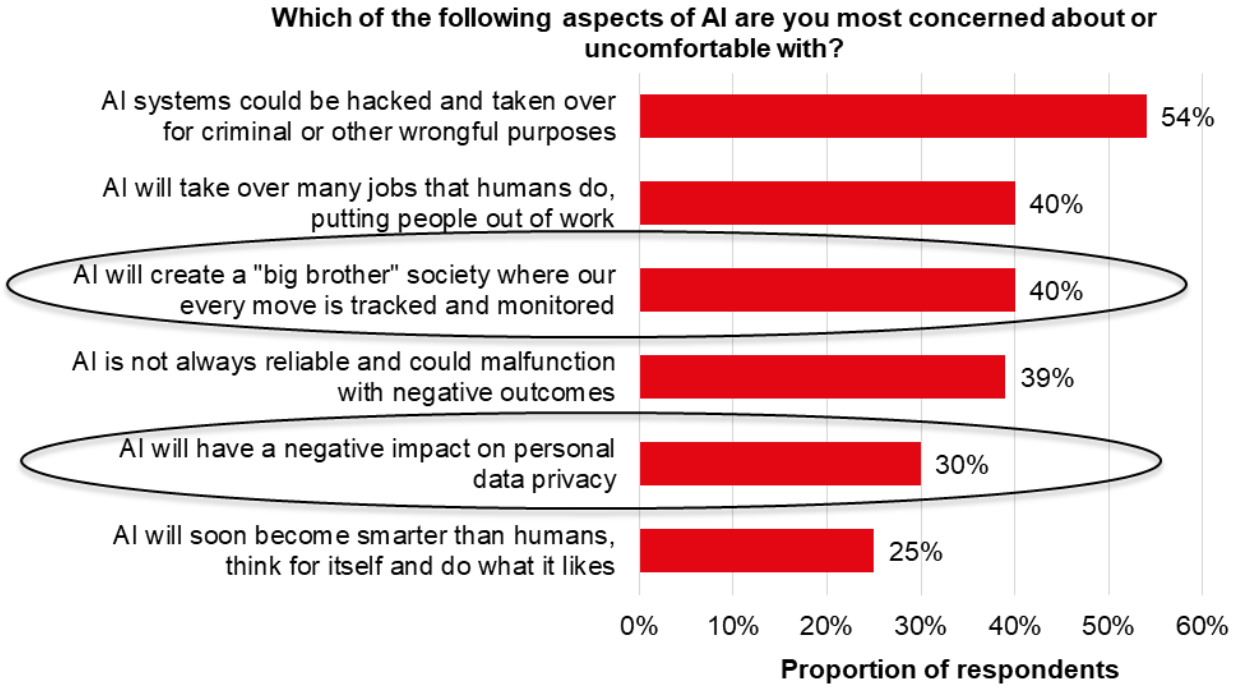


AI is adding a new dimension to consumer disquiet

AI is raising the stakes for data privacy and this is starting to filter through to consumers. The average consumer may not have a deep knowledge of AI but they know enough to be aware that it can impact privacy. When Ovum asked people about the aspects of AI that made them uncomfortable or anxious, 30% of survey respondents believe it will have a negative impact on data privacy in general, while 40% felt AI would lead to constant tracking and monitoring of activities (see Figure 6). Their concerns are justified and, if not addressed, consumers' fears about privacy will magnify and further intensify their unease with data privacy.

Figure 6: Data privacy-related issues firmly in the frame of consumer concerns with AI

Figure 6: Data privacy-related issues firmly in the frame of consumer concerns with AI6

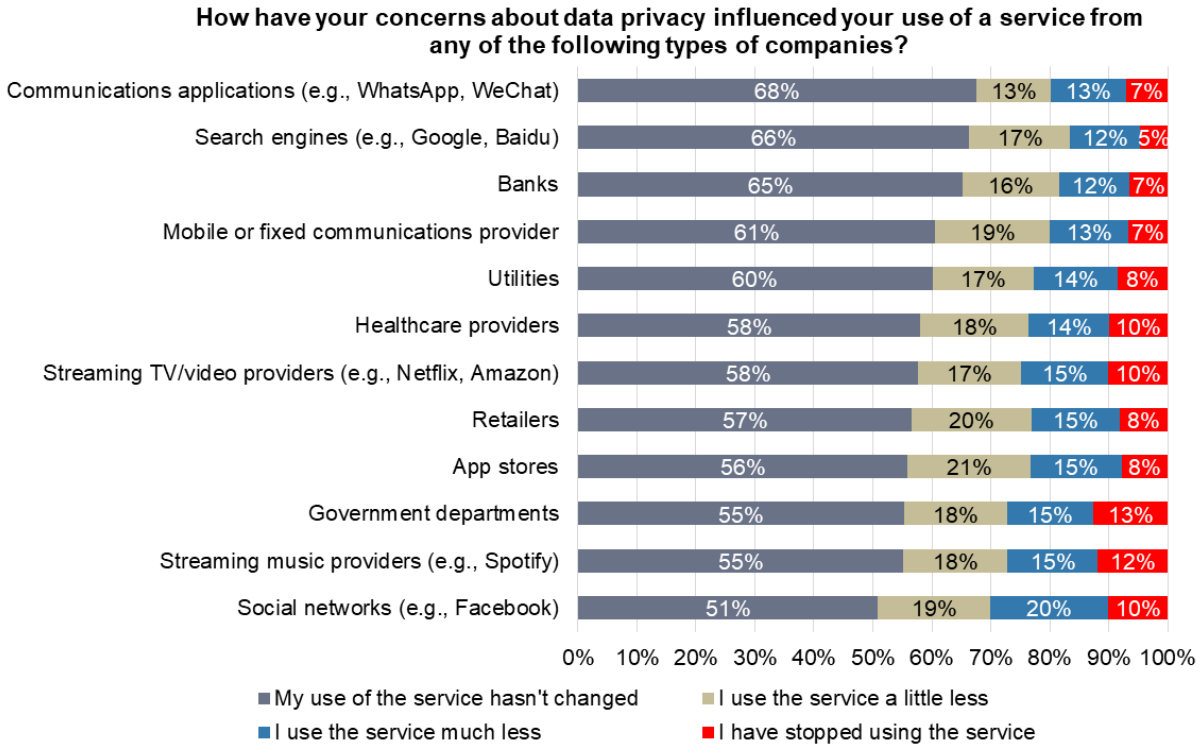


Change is in the air: service providers beware

Although concerns over data privacy have not yet caused a radical shift in consumer behavior, it is certainly eroding trust in service providers and has prompted some people to change how they interact with digital services. Ovum's Consumer Digital Insights survey shows that a small but significant number of consumers are less tolerant of companies that do not respect data privacy or have weak safeguards and, as a result, are reducing their usage of these companies' services or stopping usage altogether (see Figure 7). Every type of service provider and organization listed in the survey has been affected – no one is immune. In the case of social platforms, concerns over data privacy have prompted 10% of respondents to stop using a social platform's service, while 20% use the service a lot less. Even a small portion of users defecting or using services less can have a major impact on the bottom line. Moreover, if the trend continues and is not addressed, service providers are looking at a serious revenue drain due to their poor data privacy standing.

Figure 7: Impact of data privacy concerns on service usage

Figure 7: Impact of data privacy concerns on service usage7



Privacy is caught in the digital economy crosshairs

One of the key reasons that data privacy has taken center stage is because data is an economic issue. The idea that data is a currency is not new – the World Economic Forum tracked its emergence back in 2011 with its paper, *Personal Data: The Emergence of a New Asset Class*. However, what has changed is the value and volume of data, and its corresponding primacy in driving digital services and commerce.

Data drives multiple business models

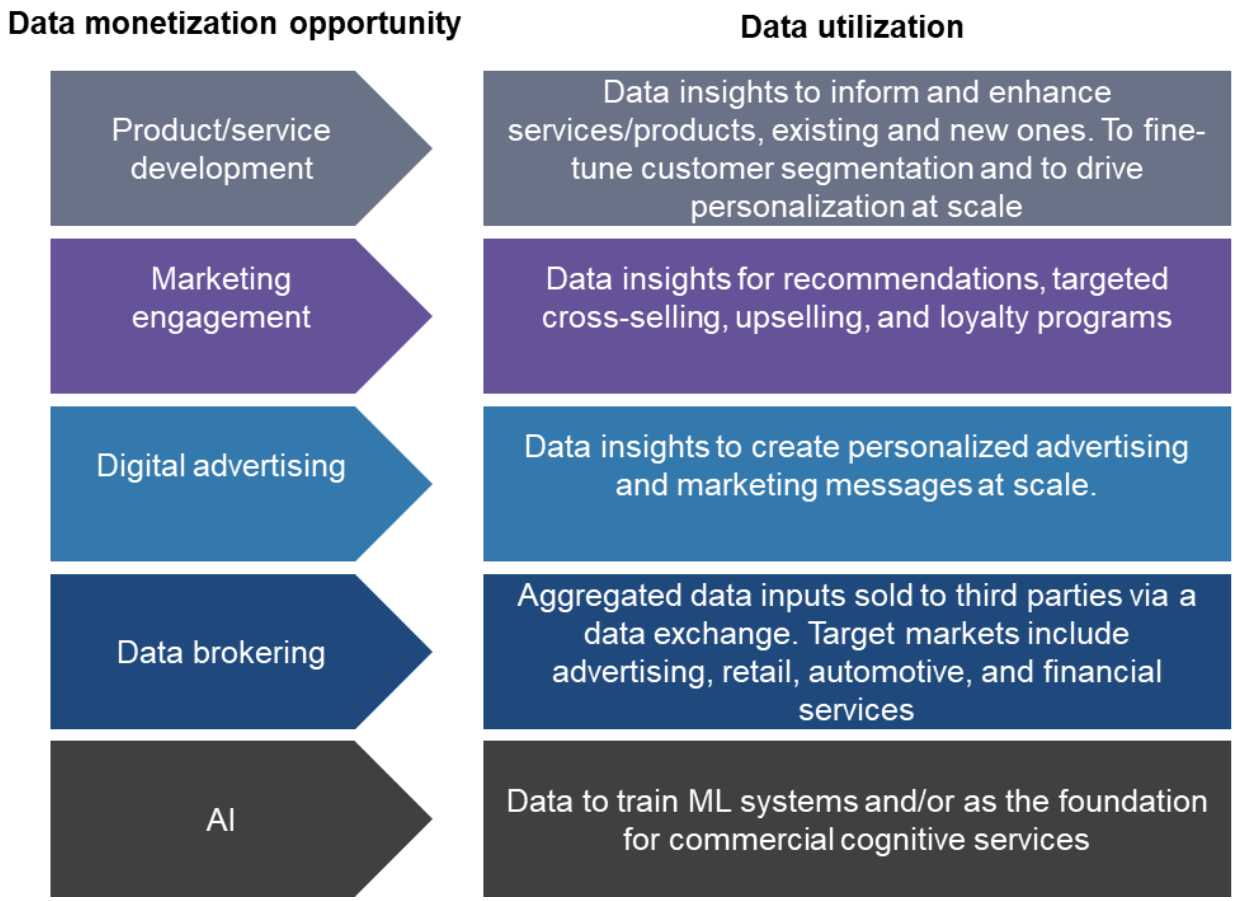
Data is the foundation of an increasing range of business models and monetization opportunities, which creates an unstoppable need to keep the data flowing. A few of the more significant data monetization opportunities are highlighted in Figure 8. Data insights can be used to drive product or service innovation and differentiation, which open up monetization opportunities. Digital advertising business models are based on increasingly deep and accurate data sets. Data insights feed the predictive analytics that fuel recommendations, cross selling, and upselling.

Data in the context of AI is a huge potential revenue opportunity. Data powers the machine learning that is being used to develop a range of new AI services and, in turn, new revenue. Associated with this is the rise of AI-powered cognitive services from the likes of Microsoft and IBM, among others, which are being made available to third parties on a commercial basis.

Some players are seeking ways to expose their own data assets to vertical markets, including mobile operators. Others are adopting a brokering model, aggregating data from a variety of data inputs, and selling them to third parties. For example, HERE Technologies has a data marketplace for the automotive and other industries in need of location data, and Acxiom has a data brokering proposition for the advertising market. A smaller number of players have also developed B2C data brokering propositions (for example, Digi.me).

Figure 8: Selected examples of core data monetization opportunities

Figure 8: Selected examples of core data monetization opportunities8



Data privacy in the digital economy needs to be rebalanced

The rise of "data fracking"

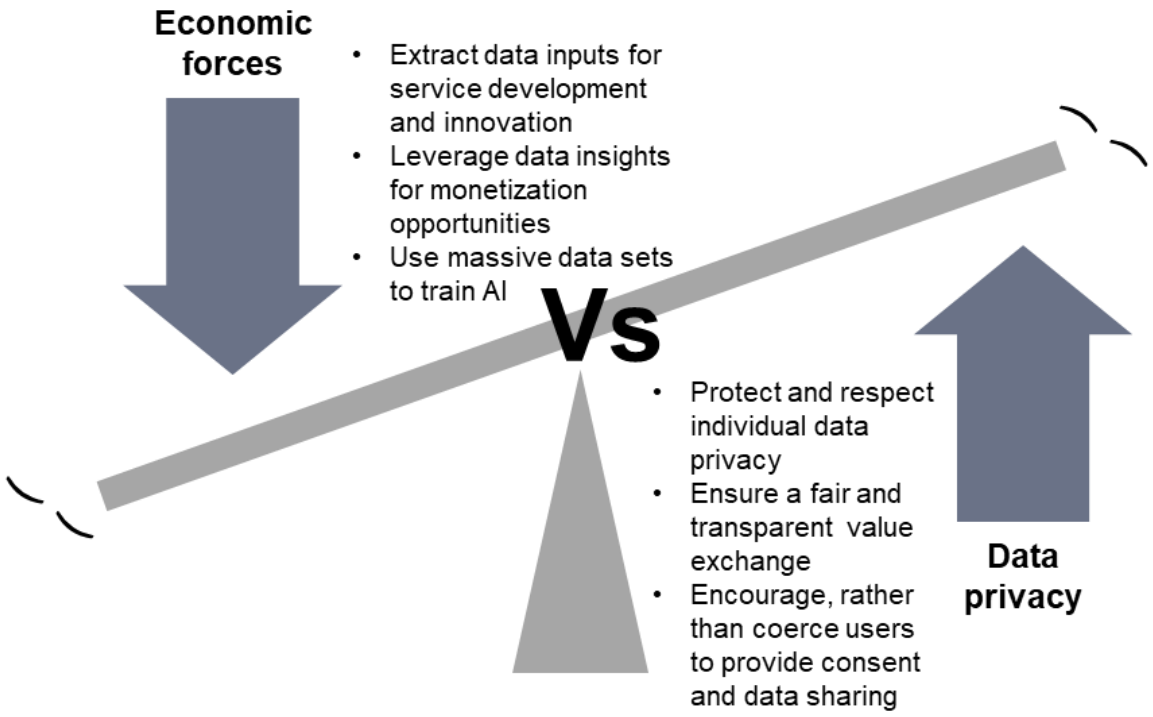
In principle, there is nothing wrong with leveraging data insights to improve digital services and to drive monetization opportunities – when handled correctly and responsibly this can benefit all stakeholders, including consumers. However, for several years the prevailing approach has been one of "data fracking." Ovum first coined this term in 2013 and, just like hydraulic fracking, it describes a high-pressure, intrusive approach to data extraction that can have negative outcomes and long-term, knock-on effects. "Data fracking" involves operating in the margins of regulation and consumer acceptance to maximize the volume and variety of data sets and the speed of extraction. "Data fracking" has gone hand in hand with a personal data land grab as players vie for a winning position in the digital economy.

Redressing the balance: from "data fracking" to "data friending"

The prevailing approach to data extraction needs to be reevaluated and a balance struck whereby consumer data is respected and protected, but in a way that does not prevent the opportunity to use it for service innovation and monetization purposes. This is a difficult balancing act to achieve, as captured in Figure 9, but it is in no one's interest to alienate consumers to the point where data consent and permissions are withdrawn.

Figure 9: The precarious data privacy balancing act

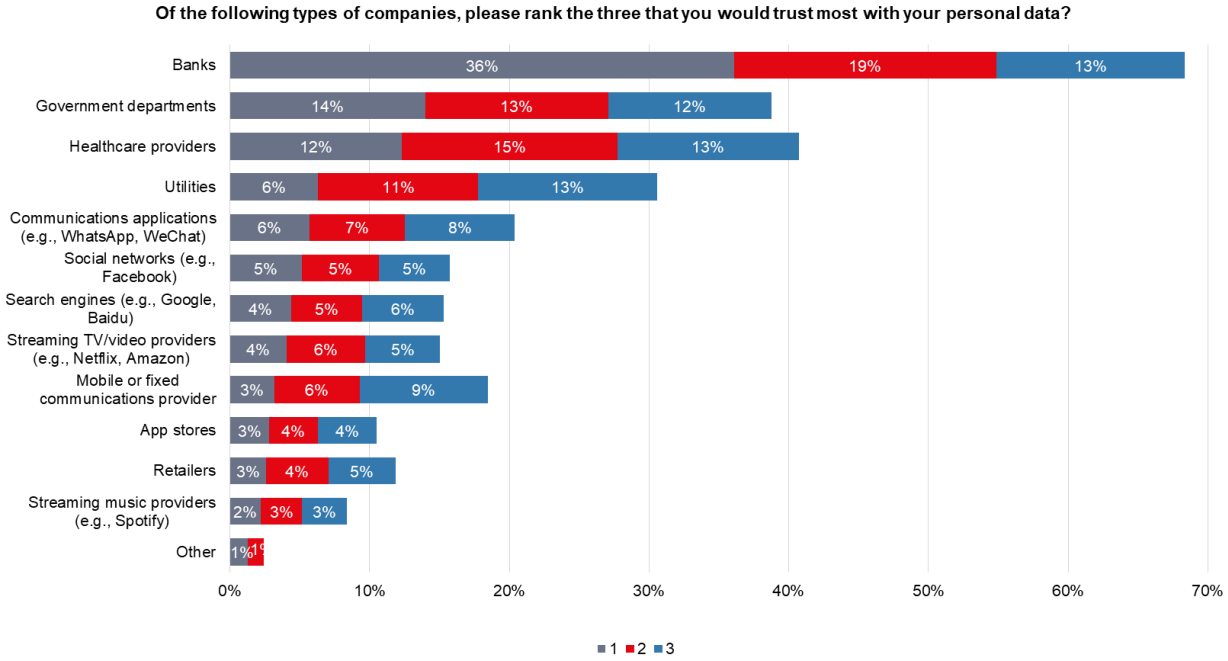
Figure 9: The precarious data privacy balancing act9



Protective measures such as those laid down by the EU'sGDPR are trying to redress the balance and this is a positive development. But consumer service providers need to go beyond regulatory box ticking and adopt a proactive, relationship-led approach to data privacy that is based on genuine transparency and consumer control. This is the "data friending" strategy, where the focus is on proactively engaging with consumers to build a positive relationship around data privacy. In this scenario, data privacy becomes an asset. But the shift to a consumer-centric, data friending approach will not happen overnight because consumer trust in the data privacy credentials of most service providers is extremely low, as shown in Figure 10. Service providers have their work cut out rebuilding trust. This is the subject of the forthcoming, complementary report in this two-part series entitled *Making Data Privacy an Asset in the AI Era: Best Practice*.

Figure 10: Very low trust status for company use of personal data outside of banks

Figure 10: Very low trust status for company use of personal data outside of banks10



Data super platforms and AI hyper-scalers pile on the pressure

A powerful mix that is often one and the same

A small number of companies have become so adept at data collection at scale that they have become what amounts to a data super platform. This is writ large with the big consumer tech and commerce firms such as Facebook and Google, among others, whose business models all rely on data and that over the years have amassed data assets that others cannot match. No single player can match Facebook's social graph, Google's search data, or Amazon's insights into shopping behavior. Facebook has traditionally been an overt data fracker, constantly pushing the boundaries of what is acceptable for privacy in the quest to leverage data insights for commercial purposes. Facebook's track record in data privacy is littered with breaches and infringements, to the point where it seems like Facebook views this as a cost of doing business. However,

the net result is that Facebook's poor data privacy credentials have mired the company in scandal, invited regulatory and government scrutiny, and caused reputational damage that is eroding trust among Facebook stakeholders, particularly consumers.

The aforementioned consumer tech and commerce companies also have growing AI prowess and service ecosystems that are powered by AI, which will deepen going forward, as will their need to tap into increasingly large data sets to train ML models that keep their services ahead of the curve. Consumer tech players such as Google and Facebook are what Ovum's sister company, Tractica, calls AI hyper-scalers. This, combined with their status as data super platforms, makes them a force to be reckoned with, and the data imperatives that drive these companies puts pressure on privacy, making it even more vulnerable. In this context, companies like Google and Facebook should be at the forefront of protecting and respecting data privacy. Their data privacy practices should not just be driven by economic imperatives but also tempered by social responsibility and a desire to make the digital economy more robust and better able to flourish for the benefit of all stakeholders.

Data privacy legal regimes are a work in progress

GDPR is a solid reference point and framework for debate

The EU's GDPR is by no means the only regulatory framework for consumer data privacy but it is one of the more rigorous laws in force today with a wide, harmonized regional remit: the EU and those companies that do business in it. GDPR is a regulatory regime to watch, with other regulatory authorities assessing it as a potential blueprint for how they might move forward, in part if not in its entirety. In the US this can be seen in the California Consumer Privacy Act. For this reason, we will use GDPR as a reference point to think about consumer data privacy regulation and whether it has gone far enough, although our analysis in this report will be at a high level rather than a technical or legal deep dive.

There are two overarching cornerstones of GDPR: the rights of the individual (the data subject) that are captured in eight core principles, alongside six core principles for those organizations that process, control, and store data. We have summarized both sets of principles in Figures 11 and 12.

There are of course other important elements of GDPR including:

- privacy by design: companies should take data privacy into account from the ground up when developing and designing products and services, thinking about how data privacy fits into, and is supported across, all stages of a product/service lifecycle
- accountability: organizations must demonstrate compliance with the GDPR principles. Companies that fail to comply will be in violation of the law and could face fines of as much as 4% of the company's global revenue or €20m, whichever is higher.

Figure 11: GDPR's six core principles for organizations

Figure 11: GDPR's six core principles for organizations11

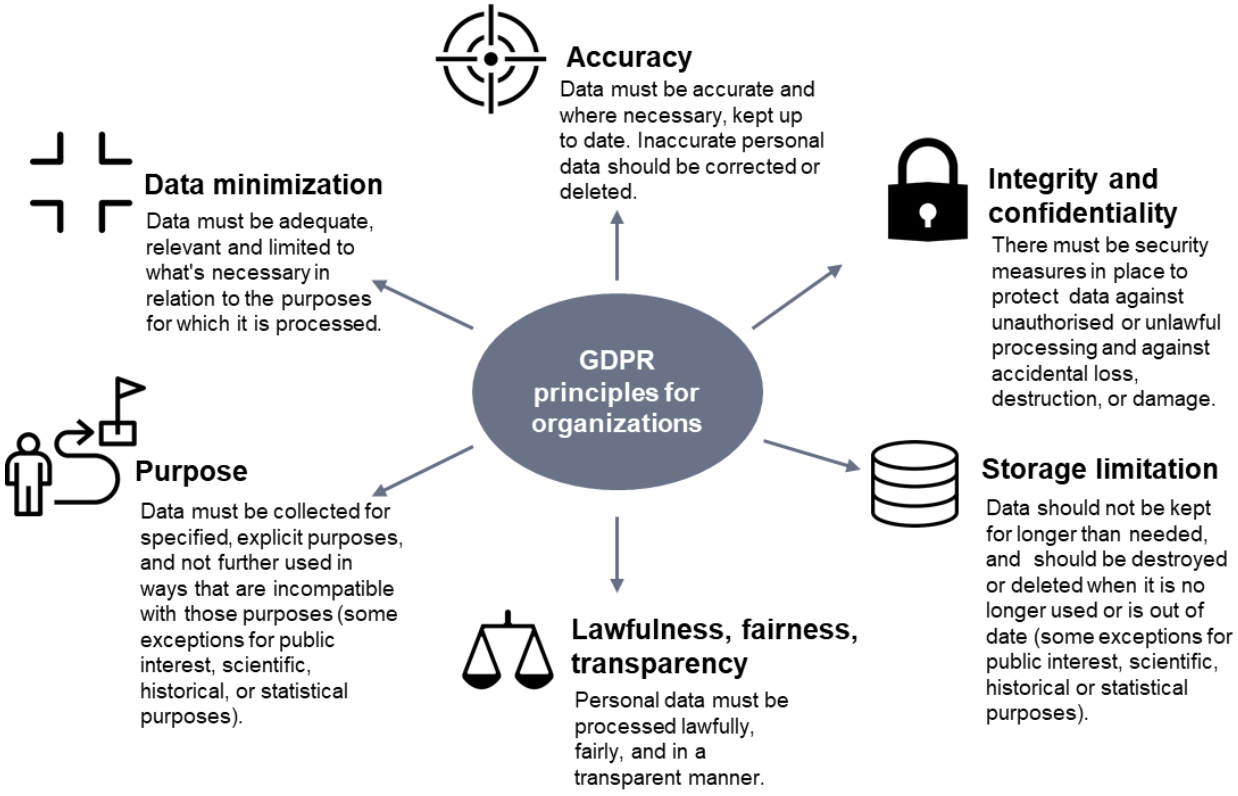
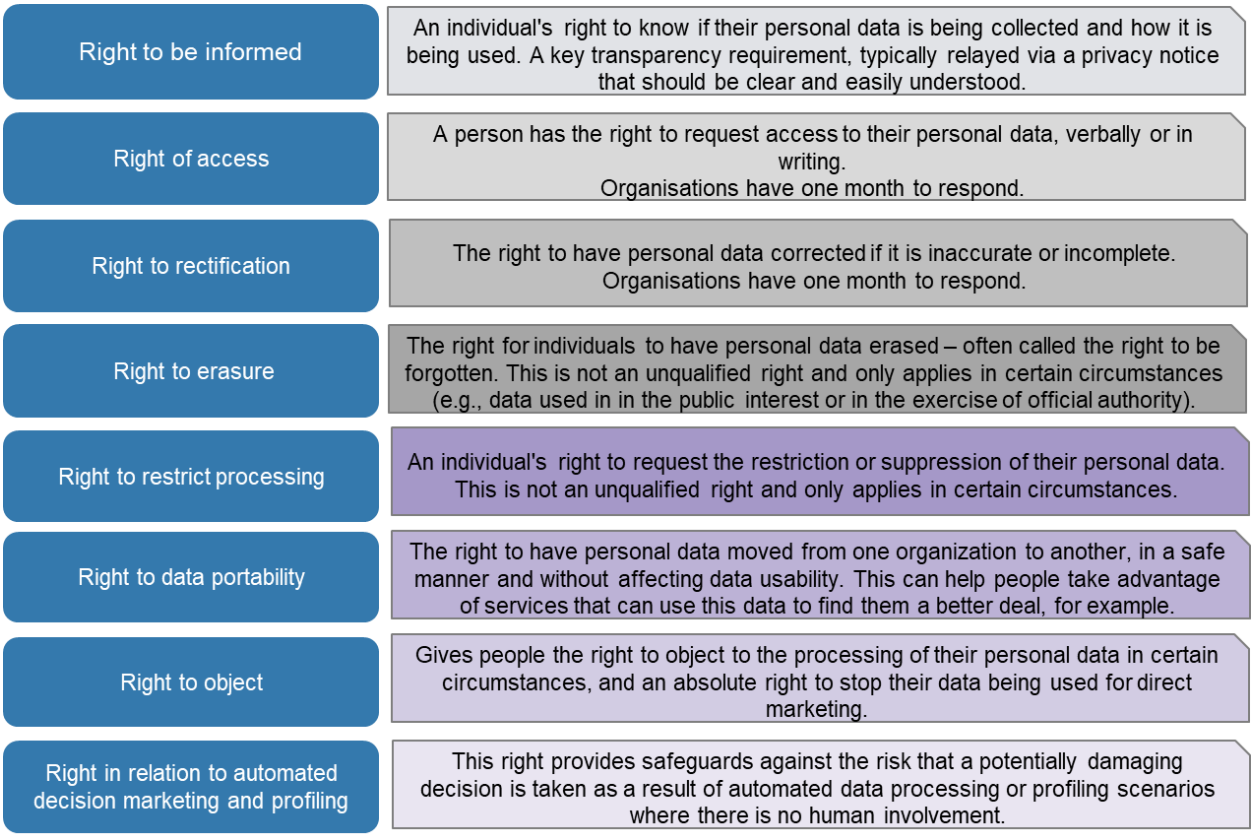


Figure 12: GDPR's eight core rights for individuals

Figure 12: GDPR's eight core rights for individuals12



GDPR and similar should be viewed as the start, not the finish

Ovum's overall view is that GDPR is a positive development and a big step in the right direction. It has heightened the debate on data privacy and tightened existing regulations like never before. Its core principles bring accountability, transparency, and consent to center stage in a way that previous data privacy regulation has not, and we recommend that all organizations embrace its tenets, even if they are not required to by law.

However, GDPR is not perfect and service providers and other organizations should treat it as the starting point not the endgame for consumer data privacy. One area where this stands out is in how GDPR relates to ongoing developments in AI, although, to be fair, this is an issue for pretty much all existing consumer data privacy and protection law. The impact of AI technologies and analytics on data is far reaching and fast moving, to the point where regulations can barely keep up.

AI brings fresh challenges for data privacy regulations

Some outcomes will be unforeseen or hard to explain – the AI black box

Certain AI impacts in the data privacy context are not yet fully understood; for example, in scenarios where AI will use data for new and often unforeseen purposes beyond the original scope. This could occur where advanced AI algorithms act in ways not initially anticipated or directed by programmers, which could surface unexpected data sets and insights. In some cases, programmers and data scientists and other experts may not fully understand how these outcomes were reached.

AI rubs up against data purpose and minimization rules in multiple scenarios

This kind of AI black box scenario goes against the principles of data transparency, and is certainly at odds with GDPR stipulations on data minimization and data purpose. A simpler example in this context is the use of AI to analyze and determine policy outcomes – in other words, the use of AI for automated decision-making. For example, AI could be used to help process insurance claims and both insurance firms and consumers need to know why and how decisions are reached – saying a claim is refused because the AI says no is not the way forward.

Data minimization principles and the need for transparency could pose a challenge to consent in the AI context. GDPR states that consent "should be given by a clear affirmative act establishing a freely given, specific, informed, and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her." But strict adherence to this could be difficult if an organization is not fully aware of how an AI system may eventually use the data.

The ML conundrum

The GDPR principle of data minimization and storage limitation could likewise be an issue for AI ML systems, which need access to large pools of data for training and to allow models to perform in an optimum way. In the context of machine learning, it can be hard to know precisely how long systems will need access to data to be fully trained, which is challenging in the context of storage limitation, whereby data can only be held for as long as is strictly necessary after which it should be destroyed. AI systems may also need to retain data for governments and for monitoring of models, and to ensure they are working at an optimal level. However, the need to retain data also raises a red flag for GDPR protection of an individual's right to be forgotten and the data retention limitations. For example, in the case of facial recognition, ML systems could require prolonged access to a very large database of people's photos in order to become as accurate as possible.

Conversely, denying ML algorithms access to complete data sets could introduce bias. If AI systems are restricted to limited data sets, the information they can work with is by nature partial, which could in turn result in narrow or biased outcomes.

There are caveats in storage limitation for scientific purposes and AI training could fall into this remit, although it has yet to be determined. Another issue is that it can be challenging to determine from the outset exactly what a ML algorithm may learn during the training purpose, which could sit uneasily with the use of data for defined and specific purposes (data minimization).

Linking data dominance and data privacy to competition

Data privacy regulations such as GDPR have certainly shaken up attitudes to data privacy and forced companies to fall in line with regulations to better protect and respect consumers. However, many feel that the GDPR regulations do not go far enough, particularly in the context of data super platforms such as Facebook. EU Competition Commissioner Margrethe Vestager is one of a vocal but growing minority that are starting to look at data privacy in the context of competition. Germany's antitrust regulator is doing the same, with its lens on Facebook in particular.

The linking of data, privacy, and competition may at first seem surprising as anticompetitive practices are typically linked to pricing issues, and many data super platforms such as Facebook and Google offer digital services to users (if not enterprises) for free. The thinking behind linking data privacy to competition is that data is an asset, and firms like Facebook have control over such massive troves of data that they are able to dominate the digital economy to the point that others cannot compete. This is one reason why Facebook's plans to merge the messaging platforms of Facebook Messenger, WhatsApp, and Instagram caused concern, along with data privacy red flags in the context of how data might be shared and utilized across the blended properties. Germany's national competition regulator (Bundeskartellamt) ruled that Facebook could not combine user data from its WhatsApp, Instagram, and Facebook Messenger apps without user consent. The authority applied the same ruling to user data that Facebook collects from third-party sites. Facebook has contested the ruling as unfair.

Data super platforms can also use data to lock users into their service, and their data dominance can encroach on consumer privacy in the interests of gaining yet more data or other business benefits in a way that is not based on a fair or transparent value exchange.

The linking of data privacy and competition is an interesting line of enquiry, but it does have holes (e.g., GDPR stipulations on data portability) and needs further exploration and validation if it is to get widespread support. Also, many believe it is inappropriate to use antitrust enforcement to address data privacy. However, it could have far-reaching implications if regulators and government get behind it and deem that firms like Facebook do indeed have undue data dominance to the point where it is anticompetitive. What could happen as a result is not clear, although Andreas Mundt, the president of Germany's antitrust authority, likened his office's ruling to an internal breaking up of Facebook's data trove.

Citation policy

Request external citation and usage of Omdia research and data via citations@omdia.com.

Omdia consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help you. For more information about Omdia's consulting capabilities, please contact us directly at consulting@omdia.com.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the "Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together "Informa Tech") and represent data, research, opinions or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an "as-is" and "as-available" basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness or correctness of the information, opinions and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees and agents, disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial or other decisions based on or made in reliance of the Omdia Materials.

CONTACT US

omdia.com

askananalyst@omdia.com