Centre for Information Policy Leadership
HUNTON

Rethinking Sensitive Data in the Age of AI

September 2025

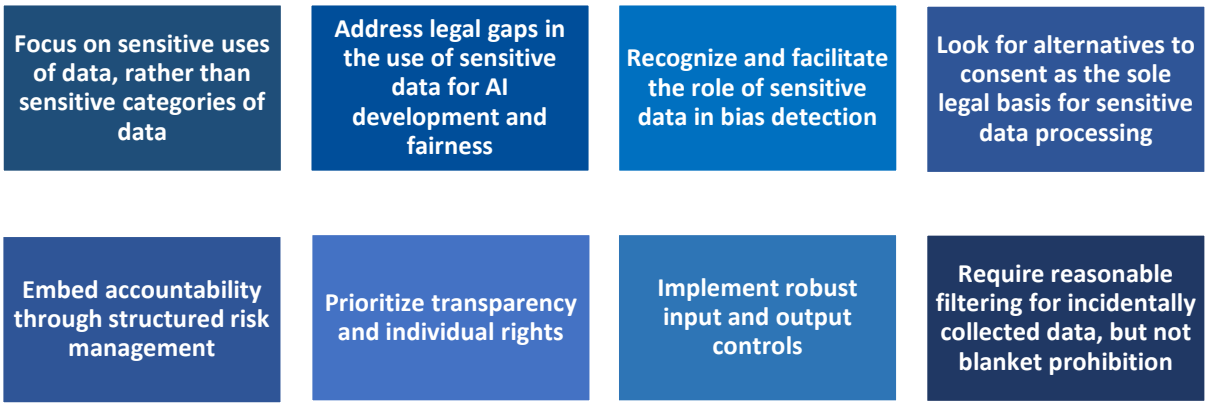
Centre for Information Policy Leadership
HUNTON

Rethinking Sensitive Data in the Age of AI

Contents

Rethinking Sensitive Data in the Age of AI

To facilitate and enable the responsible use of sensitive data in AI systems, CIPL recommends:



I. Introduction and Background

The concept of sensitive or special category data originated in early European privacy laws dating back to the 1970s.[1] In the context of twentieth-century European history, lawmakers understood that certain types of personal information are particularly vulnerable to misuse, requiring stricter legal protections.[2] Since then, many jurisdictions have incorporated the concept into their privacy laws, although the definitions and regulatory obligations vary.

The General Data Protection Regulation (GDPR) offers one of the most conservative frameworks, setting out specific categories such as racial or ethnic origin, political opinions, or health status to qualify as sensitive, and prohibits the processing of such data, unless specific exemptions under Article 9(2) GDPR apply. In the United States, while federal law does not explicitly define sensitive data for general data processing, several state privacy laws establish their own definitions and regulations, generally giving individuals the ability to limit its use and disclosure,[3] or requiring specific consent.[4] Other jurisdictions have adopted different approaches. For example, Singapore and Hong Kong do not formally define categories of sensitive data; instead, regulators have issued guidance indicating that certain types of information require greater protection.[5] Canada's PIPEDA acknowledges that the sensitivity of data may be context-specific and mandates safeguards proportionate to the data's sensitivity.[6]

There is consensus that sensitive data generally receives greater protection because its misuse can cause severe harm, such as discrimination, in the context of employment or public services. Recital 51 of the GDPR explains that such data are "by their nature, particularly sensitive in relation to fundamental rights and freedoms" and, thus, "merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms." The Court of Justice of the European Union (CJEU) used similar language when contemplating more stringent protection for certain data types.[7] Likewise, the Council of Europe emphasizes that sensitive data processing poses risks not only to discrimination but also of harm to an individual's dignity, physical

integrity, and most intimate personal sphere.[8] Consequently, the use of such data must come with additional obligations or safeguards to protect individuals.

Tension can arise in a data-driven world that relies on often large and certainly diverse datasets to provide services and develop new products.[9] As Advocate General Jääskinen put it in his opinion in *Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González*, case: "*The internet has revolutionized our lives by removing technical and institutional barriers to dissemination and reception of information, and has created a platform for various information society services. These benefit consumers, undertakings, and society at large. This has given rise to unprecedented circumstances in which a balance has to be struck between various fundamental rights, such as freedom of expression, freedom of information, and freedom to conduct a business, on one hand, and protection of personal data and the privacy of individuals, on the other.*"[10]

The *Costeja* case already demonstrated the limitations of a category-driven approach to defining personal data, in the context of search engines relying on the collection and indexing of large amounts of data from the internet, which may include data fitting a "sensitive" category.[11] Search engines function primarily as tools that organize information from third-party websites without actively distinguishing or controlling the personal data they process; it therefore becomes challenging to identify and protect such data based solely on predefined categories. This has come further to the fore in the context of algorithmic fairness and AI. Apart from the incidental collection of potentially sensitive data categories where data is scraped from the web for AI training purposes, sensitive data plays an intentional and important role in effectively detecting and mitigating bias or discrimination in AI systems.

Without these data categories, organizations may be unable to uncover disparities in how AI models perform across different demographic groups, making it impossible to ensure fairness and equal benefits of AI across all communities. For instance, in order to ensure a bank's AI system is not used to assess whether a customer is creditworthy enough to apply for a mortgage in a way that disproportionally denies mortgages to people with a certain ethnicity, the developer of the AI system needs to be able to distinguish the ethnicity of the people about whom its AI system makes decisions.[12] Regulators such as the UK's ICO acknowledge that sensitive data may be necessary to assess discrimination risks, evaluate model performance, and retrain models accordingly.[13] The categorical restrictions many data protection laws place on sensitive data processing, such as requiring specific consent, coupled with an increasingly broad interpretation of the concept of sensitive data, can place organizations in a position of being unable to include sensitive data in AI training datasets to the detriment of the performance of the model, where such consent is not obtainable, for example.

II. DISCUSSION

A. The role of sensitive data in AI training

Certain AI applications require the processing of sensitive data to function effectively, fairly, and safely. However, current legal frameworks can limit the ability to collect and use this data, even where it is necessary and proportionate. For example:

1. Healthcare chatbots in hospital settings

Deploying AI chatbots in healthcare environments (e.g., hospitals or clinics) for tasks such as appointment scheduling, symptom screening, or triage presents legal and operational difficulties due to the potentially sensitive nature of health data. Example use cases:

- Developing an AI-driven chatbot that can accurately triage patients based on their reported symptoms would ideally leverage detailed patient interactions.
- To monitor patient conditions over time, chatbots would need to link multiple interactions across visits.

However, regulations prohibit the collection and use of health data without explicit consent, potentially limiting access to the data necessary for the development of effective, reliable chatbot systems for healthcare settings.

2. Connected and autonomous vehicles

Connected cars and autonomous vehicles must continuously gather and process large volumes of sensitive data in real-world environments. This includes both data from within the vehicle and from the surrounding environment:

- Capturing data from drivers and passengers (e.g., heart rate, voice commands, driving patterns) to detect emergencies and improve safety.
- Processing external data – such as images of pedestrians, children,[14] and individuals near locations that might reveal information falling within the category of sensitive data (e.g., hospitals, religious sites, political events) – to inform real-time decision-making and model training.[15]
- Evaluating pedestrian characteristics (e.g., children who may behave unpredictably, recognizing an individual in a wheelchair versus a person sitting down) to improve safety and situational awareness for autonomous driving systems.

While such data is essential for improving vehicle safety, navigation, and user experience, when such data is stored, shared, or used to refine AI systems for improving situational awareness or navigation, it raises serious concerns. Obtaining explicit consent from all involved parties, including passengers, pedestrians, and other road users will often not be possible, since these individuals have no direct interaction with the vehicle and are unlikely to even be aware of the collection of their data.

To illustrate the importance of sensitive data in AI applications, the table below provides additional use cases.

Examples of Sensitive Data in AI			
Use Case	**Purpose of the AI System**	**Types of Sensitive Data**	**Data Source**
Personalized therapy chatbot	To provide mental health support, personalized therapeutic exercises, and cognitive behavioral therapy techniques to users.	Health data: Mental health conditions, emotional state, user's self-reported feelings, and chat transcripts.	User inputs and conversations within the application.

Early health issue detection	AI-powered full body scanning platform for preventative healthcare.	Health and biometric data.	Full-body scans.
Loan application assessment to find qualified borrowers with limited credit history	To analyze alternative financial data to determine a person's creditworthiness and eligibility for a loan, especially for those with limited credit history.	Financial data: Mobile phone payment history, apartment rental payment history, utility bill payments, and transaction details from non-traditional data sources.	Third-party data providers, mobile phone companies, property management companies, and utility companies.
Hiring and recruitment to seek out applicants from underrepresented groups	To proactively identify and find qualified minority applicants in the job market and to analyze existing applicant pools for diversity metrics.	Personal data: Racial or ethnic origin (can be inferred from names or demographic information), as well as details on any affiliations with minority groups.	Job application platforms, professional networking sites, and candidate resumes.
Social media content moderation	To automatically identify and remove content that violates platform policies, such as hate speech or discriminatory content.	Racial or ethnic origin, political opinions, and religious beliefs (as these are often the subject of hate speech).	User-generated content, including text, images, and videos posted on the platform.
Fitness and health tracking	To analyze user activity, heart rate, and sleep patterns to provide personalized health insights and recommendations.	Health data: Heart rate, sleep data, and physical activity levels.	Smartwatches, fitness trackers, and mobile health apps.
Accessibility feature	To provide real-time translation of spoken language to sign language or to describe images for users with visual impairments.	Health data: Disabilities (visual or hearing impairment), as well as biometric data (e.g., facial movements or sign gestures).	User-provided data, video, or image inputs, and user settings.
Healthy lifestyle habits	To create personalized nutrition and diet plans, and to track dietary intake to achieve health goals.	Health data: Dietary restrictions, weight, height, body mass index, food preferences, and a person's health goals.	User inputs into a mobile application or fitness tracker.
Minority and student support	To analyze and predict academic performance, and to offer tailored support for minority or at-risk students.	Personal data: Racial or ethnic origin, as well as education data (which can be considered sensitive).	School records, student demographic information, and academic performance data.

Developing autonomous vehicles sensor systems	Anonymization of faces or license plates captured during data collection for autonomous vehicle sensory development.	Biometric (facial data).	Autonomous vehicle/connected car camera external feed.

To better balance innovation and privacy, and to ensure the development of accurate, fair, and reliable AI models, regulators and policymakers should adopt more targeted approaches to the regulation of sensitive data.

B. The need for a data use approach

Current data protection laws present significant limitations in establishing lawful bases for the processing of sensitive data, which in turn restricts AI developers from using these data types to create accurate and fair models. For example:

- Legal grounds for processing sensitive data tend to be narrowly defined, and in many jurisdictions, such as Australia[16] and Japan[17], consent is the default requirement for collecting and processing sensitive data unless an exemption applies.
- Guidance from the Office of the Australian Information Commissioner on developing generative AI models makes it clear that any sensitive data collected inadvertently without consent should typically be deleted.[18]
- Similarly, Japan's Personal Information Protection Commission has, for example, warned OpenAI against collecting sensitive personal data without user consent, emphasizing the need to remove or anonymize such data if acquired.[19]
- The Italian Data Protection Authority, the Garante, sent a formal warning to GEDI after the media group planned to rely on legitimate interests to share editorial content, including sensitive data, with OpenAI for AI training.[20]

These examples reflect a trend toward a restrictive regulatory approach to sensitive data, especially in AI, and a preference for consent as a legal basis despite its limitations.[21]

Some data protection frameworks and laws, on the other hand, recognize the need to process sensitive data for non-discrimination purposes and therefore provide broader permissions. Several jurisdictions outside the EU, including Bahrain,[22] Ghana,[23] Jersey,[24] South Africa,[25] and the Dubai International Financial Centre (DIFC)[26] have incorporated legal exceptions allowing sensitive data to be used when necessary to protect individuals from discriminatory decisions.

In the EU, Article 10 of the EU AI Act requires that the datasets used for high-risk systems are relevant and sufficiently representative, and have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended. Article 10(5) specifically recognizes the necessity for the inclusion of sensitive data in the development process of AI and provides permission for limited purposes and with requirements for suitable safeguards, but only in the context of high-risk applications.[27]

While encouraging, Article 10(5) of the EU AI Act presents the following challenges:

1. The exception applies exclusively to AI system providers, excluding deployers who also play critical roles in addressing bias.

2. It restricts processing of sensitive categories of data to one-off training, validation, and testing datasets, overlooking biases that may arise from algorithm design choices or deployment contexts over time.[28]
3. Additionally, this exception is limited to high-risk AI systems, leaving organizations working with lower-risk AI without similar legal grounds to use sensitive data to test for bias under the GDPR.[29]

Policymakers should clarify that bias detection aimed at preventing discrimination constitutes a substantial public interest beyond just high-risk AI, and that there is a range of important public interests that support the use of arguably sensitive data for AI-related purposes.

Separately, the EDPB's ChatGPT Taskforce recommends filtering out sensitive data at collection or deleting it immediately afterward, before model training.[30] Similarly, the CNIL requires controllers to implement automated measures to exclude sensitive data, including avoiding inherently sensitive environments (e.g., health forums),[31] applying anonymization or pseudonymization processes immediately after collection,[32] and promptly deleting it if incidentally collected.[33]

These approaches strongly restrict the collection of sensitive data for training generative AI systems. In many cases, sensitive information may only be inferred when multiple data points are (purposefully) combined, for example, and collecting consent may be practically impossible. Equally, the requirement to delete sensitive data once the bias has been corrected (including in Article 10(5) EU AI Act), prevents bias monitoring throughout the AI system's lifecycle.

In Europe, the recent debate around the European Health Data Space illustrates the challenges of relying on consent to process sensitive data. The proposal sparked debate over whether an opt-in or opt-out consent model would be most appropriate. Advocates for an opt-in approach emphasized the importance of having individuals actively agree to having their data shared[34]; in contrast, supporters of the opt-out approach warned that actively obtaining consent risks introducing selection bias, reducing the representativeness of datasets, and ultimately compromising research outcomes, one of the regulation's primary goals.[35] The final agreement favored an opt-out mechanism, with safeguards to ensure that data is used only for legitimate purposes and processed in a manner that prevents re-identification, as well as exceptions for the purposes of public interest, where the broader societal benefits of data use may outweigh individual risks.[36]

The societal benefits of technology, and the corresponding need to interpret law in light of those benefits, have been contemplated in other contexts before. In *Authors Guild v. Google*, Inc., the CJEU had to consider the digitization of millions of books without previous permission in light of the transformative public value of the project, including improved access for researchers, educators, and underserved communities.[37] Similarly, in the context of search engines, Advocate General Jääskinen recognized that legitimate interests could justify data processing that makes information more accessible and supports innovation, as long as fundamental rights are respected.[38]

In *GC and Others v. CNIL* Advocate General Szpunar considered that the prohibitions of Articles 9 and 10 would require search engine operators to vet all published articles before displaying a link, which is "neither possible nor desirable",[39] and that because the search engine did not cause the sensitive data to appear on the internet pages, but intervenes afterward, it is instead required to verify whether it had an exemption to process the sensitive data, if it receives a dereferencing request by the data subject.[40] Analogous to search engines, generative AI developers process and reproduce such data only after it is published online. General-purpose large language models are predominantly trained on vast amounts of publicly available data gathered from the internet, which

inevitably but unintentionally includes sensitive information due to its presence online. Imposing limitations on the collection and use of such publicly accessible sensitive data can ultimately also limit the development of well-functioning general-purpose AI models by restricting their ability to learn from the broad and diverse sources necessary for their functionality.

C. The importance of sensitive data in AI models

As illustrated above, beyond the incidental presence of sensitive data in AI models, there are many examples, where sensitive data is also critical for responsible AI development. It is necessary to include sensitive data in training datasets to ensure that systems are accurate, representative, and aligned with broader legal obligations. Data quality, integrity, and fairness are not merely technical ideals but legal requirements under GDPR and other data protection laws. The exclusion of sensitive data could compromise the accuracy and utility of AI models, leading to discriminatory outcomes or failure to meet duties under non-discrimination law, duty of care, and contractual obligations. In many contexts, the inclusion of such data is essential to comply with legal obligations beyond the choice of legal basis for processing.

With data protection frameworks generally defining sensitive data through an objective lens, focusing on fixed categories with narrow legal bases instead of the context and purpose of data processing, they are ill-equipped to manage the nuanced risks and opportunities presented today. Certain uses of traditionally defined "sensitive data" may not in fact pose any meaningful risk in a given context, yet current frameworks may still require organizations to treat them as high-risk, resulting in disproportionate restrictions. As a result, this approach carries the risk of both under- and over-regulation, either by burdening harmless processing with excessive compliance obligations or failing to safeguard individuals from genuinely harmful practices.

Policymakers should therefore consider the importance of using sensitive data beyond a high-risk context as in the example of the EU AI Act. At a minimum, bias mitigation aimed at ensuring fair and well-functioning AI should be recognized as a legitimate public interest across all AI applications, enabling developers and deployers to use sensitive data with appropriate safeguards.

A purpose-based approach is already reflected in parts of certain privacy laws. For instance:

- Article 9 GDPR only restricts biometric data for the purpose of identifying individuals, and Section 6 of the Australian Privacy Act 1988 limits biometric information used for the purpose of automated biometric verification or biometric identification. These provisions aim to prevent ordinary photographs from being automatically classified as sensitive biometric data unless used to uniquely identify a specific individual.[41]
- The GDPR's alternative legal bases for processing sensitive data, including scientific research, public health, and substantial public interest, are also inherently purpose-driven. Article 11 of Brazil's LGPD includes protecting life, health, and ensuring fraud prevention.
- Finally, in the United States, many states' comprehensive privacy laws contain exceptions to protect against a variety of illegal and malicious activities; conduct basic internal research to develop, improve, or repair technologies; or provide products or services specifically requested by consumers, among other purposes.[42]

These examples demonstrate that incorporating the data use intent as a key factor, and a model which considers both the nature of the data and the context and purpose of its use to determine its sensitivity, would not be inconsistent.[43]

A data use approach offers a more proportionate and future-proof alternative:

- It enables protections to be tailored to the actual risk presented by a particular use of data, allowing organizations to adopt contextually appropriate mitigation strategies.
- The limitations of a category-driven approach and the continual need for the creation of new exemptions as novel use cases emerge are exemplified in the use of sensitive data to reduce bias in AI systems. A use-based framework, particularly one that recognizes justifications for data use outside of consent, provides a more agile and effective way to adapt to evolving technologies and risks.
- It helps resolve inconsistencies in which categories of data are sensitive across jurisdictions.[44] For instance, geolocation data is considered sensitive under many US laws but not under the GDPR, while the GDPR includes categories like philosophical beliefs that are largely absent from US law.

Finally, requiring organizations to assess the full context of their data processing, considering the nature of the data, the purpose of processing, who is carrying it out, and what mitigating measures are in place, offers a more accountable and effective approach to regulating sensitive data. Rather than relying on a static list of predefined data types, this model obliges organizations to carry out risk assessments, justify their decisions, and implement robust safeguards proportionate to the actual risk posed. It ensures that additional regulatory burdens are imposed only when warranted by the context. While risk assessments may vary in form, focusing on two key factors may help: data sensitivity and the potential impact on individuals. This should be part of a broader, holistic, context-driven evaluation that recognizes that risk depends not just on the data itself, but on how it is used. For instance, whether data is used to train or deploy AI, whether the system makes decisions or recommendations, and the potential consequences of those outputs should all shape the controls put in place to minimize harm. As both sensitivity and impact increase, organizations should be expected to implement stronger controls. This model not only improves individual protection but also incentivizes more responsible and transparent data governance. AI governance should protect individuals, not data.

To support organizations in operationalizing a purpose-based framework, regulators should provide comprehensive guidance, including illustrative use cases, decision trees, and risk assessment templates. This will enable organizations to confidently determine when sensitive data is justified based on the context, purpose, and societal benefit.

III. RECOMMENDATIONS

The following CIPL recommendations are intended to facilitate and enable the responsible use of sensitive data in AI systems:

1. **Focus on sensitive uses of data, rather than sensitive categories of data**: Legislators/policymakers should move away from rigid definitions of "sensitive data" and instead focus on high-risk or sensitive *uses* of data. All processing should be subject to a risk assessment that considers the data type, context, purpose, actors involved, and available mitigation measures. While some types of data might be deemed "likely sensitive", an initial assessment of "sensitivity" should be able to be disproved through proper risk assessments and the appropriate implementation of risk mitigation strategies.
2. **Address legal gaps in the use of sensitive data for AI development and fairness**: Legislators and policymakers should ultimately consider revising existing privacy or AI laws to address

legal gaps that limit the responsible use of sensitive data in AI development essential to ensure non-discrimination and fairness in AI outcomes, complying with other legal obligations (e.g., equality laws, duties of care), or developing AI systems that inherently rely on sensitive data (e.g., health monitoring, assistive technologies). Where such processing is necessary and proportionate, the law should permit it under clear conditions, supported by strong safeguards, accountability measures, and transparency requirements.

3. **Recognize and facilitate the role of sensitive data in bias detection**: Legislators and policymakers should recognize the critical role of sensitive data in bias detection and mitigation, and explicitly permit the use of sensitive data for fairness auditing and algorithmic impact assessments, even in non-high-risk AI systems. This use should be considered a substantial public interest, provided appropriate safeguards are in place.
4. **Look for alternatives to consent as the sole legal basis for sensitive data processing:** Legislators/policymakers should enable the processing of sensitive data under lawful bases beyond consent, recognizing that consent is often impractical or insufficient for complex or large-scale uses such as AI development. Such bases should be available when organizations meet higher thresholds of need for the data and standards of demonstrable organizational accountability. These include conducting a thorough balancing test, considering the rights and interests of individuals, applying appropriate safeguards, and documenting the rationale and mitigation measures adopted.
5. **Embed accountability through structured risk management:** Organizations should implement internal processes to assess and manage the risks of sensitive or high-risk data use. This includes evaluating the context and data involved, identifying appropriate safeguards for individuals, documenting legal bases for processing, and considering the broader societal or public interest. Stricter protections should be applied where sensitivity or risk is higher.
6. **Prioritize transparency and individual rights**: During model training and fine-tuning, organizations should prioritize transparency and apply individuals' rights as appropriate under applicable data protection laws. Blanket exclusion or consent-based mechanisms, particularly in domains such as health data or public research, could undermine valuable societal outcomes. Therefore, organizations should ensure clear communication about data use and apply appropriate privacy-preserving defaults and safeguards.
7. **Implement robust input and output controls**: Organizations developing or deploying AI models should first implement controls at the input stage to filter or restrict potentially harmful or unauthorized requests. Following this, controls at the output stage should be applied to prevent the generation of harmful content, such as unauthorized profiles, retrieval of sensitive personal details, or producing likenesses of individuals without their consent.
8. **Require reasonable filtering for incidentally collected data, but not blanket prohibition**: Exhaustively identifying and excluding all personal or sensitive data from initial training datasets is often infeasible, may increase risks through unnecessary additional processing, and limit access to sufficiently diverse datasets. Regulators should clarify what constitutes reasonable technical measures to limit unnecessary data use and provide examples of acceptable technical solutions.

ABOUT CIPL

The Centre for Information Policy Leadership (CIPL) is a global privacy and data policy think tank within the Hunton law firm that is financially supported by the firm, 85+ member companies that are leaders in key sectors of the global economy, and other private and public sector stakeholders through consulting and advisory projects. CIPL's mission is to engage in thought leadership and develop best practices for the responsible and beneficial use of data in the modern information age. CIPL's work facilitates constructive engagement between business leaders, data governance and security professionals, regulators, and policymakers around the world. For more information, please see CIPL's website at www.informationpolicycentre.com. Nothing in this document should be construed as representing the views of any individual CIPL member company or Hunton. This document is not designed to be and should not be taken as legal advice.

---

[1] Karen McCullagh, "Data Sensitivity: Proposals for Resolving the Conundrum", 13 April, 2007, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1378121.
[2] Georgieva/Kunder, The General Data Protection Regulation, Art. 9, p. 369
[3] See for example, California Privacy Rights Act, Section 1798.121(a).
[4] See for example, Virginia Consumer Data Protection Act, Section 59.1-578(5) and Colorado Privacy Act Section 6-1-1308(7).
[5] Office of the Privacy Commissioner for Personal Data, Hong Kong, Guidance on Preparing Personal Information Collection Statement and Privacy Policy Statement, available at https://www.pcpd.org.hk/english/publications/files/GN_picspps_e.pdf; Personal Data Protection Commission Singapore, Advisory Guidelines on The Personal Data Protection Act for NRIC and Other National Identification Numbers, available at https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/advisory-guidelines-for-nric-numbers---310818.pdf.
[6] Personal Information Protection and Electronic Documents Act, Schedule 1, Principle 4.3.4Principle 4.7.
[7] Case C-136/17, GC and Others v CNIL, paragraph 44.
[8] Council of Europe, Explanatory Report to the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, available at https://rm.coe.int/cets-223-explanatory-report-to-the-protocol-amending-the-convention-fo/16808ac91a.
[9] It is encouraging that Section 74 of the UK's Data (Use and Access) Act 2025 grants the Secretary of State the power to adjust the list of sensitive data categories and the processing activities that would be subject to the prohibition of processing.
[10] Opinion of Advocate General Jääskinen, Case C-131/12, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, https://curia.europa.eu/juris/document/document.jsf?docid=138782&doclang=EN, para 2.
[11]AG Jääskinen, in para 84, argued, that: "The internet search engine service provider merely supplying an information location tool does not exercise control over personal data included on third-party web pages. The service provider is not 'aware' of the existence of personal data in any other sense than as a statistical fact web pages are likely to include personal data. In the course of processing of the source web pages for the purposes of crawling, analysing and indexing, personal data does not manifest itself as such in any particular way".
[12] Indrė Žliobaitė and Bart Custers, "Using Sensitive Personal Data May Be Necessary for Avoiding Discrimination in Data-Driven Decision Models", 2016, available at https://link.springer.com/article/10.1007/s10506-016-9182-5.
[13] Information Commissioner's Office (ICO), What about fairness, bias and discrimination?, 16 March, 2023, available at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/?search=sensitive.
[14] Some data protection laws including in China's (Personal Information Protection Law, Article 28), Egypt (Egypt's Child Law No. 12 of 1996, Article 12) and Ghana (The Data Protection Act, 2012, Article 37(1)(a)) consider children's data as sensitive.

[15] European Data Protection Supervisor, "TechDispatch #3: Connected Cars", 2019, available at https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-3-connected-cars_en.
[16] Privacy Act 1988, Privacy Principle 3.
[17] Act on the Protection of Personal Information, Article 20.
[18] Office of the Australian Information Commissioner, "Guidance on privacy and developing and training generative AI models," 23 October, 2024, available at https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/guidance-on-privacy-and-developing-and-training-generative-ai-models.
[19] Michihiro Nishi et al., "Japanese law issues surrounding Generative AI: ChatGPT, BARD and beyond", 5 October, 2023, available at https://www.cliffordchance.com/content/dam/cliffordchance/briefings/2023/10/japanese-law-issues-surrounding-generative-ai-chatgpt-bard-and-beyond.pdf.
[20] Italian Data Protection Authority (Garante per la protezione dei dati personali), Order of 27 November 2024, available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10077129.
[21] For more on the limitations of consent, see CIPL paper "The Limitations of Consent as a Legal Basis for Data Processing in the Digital Society", available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_bkl_limitations_of_consent_legal_basis_data_processing_dec24.pdf.
[22] Bahrain Personal Data Protection Act of 2018, Article 5(8).
[23] Ghana Data Protection Act, 2012, Section 37(8).
[24] Jersey Data Protection Law 2018, Schedule 2, Part 2, Section 18.
[25] South Africa Protection of Personal Information Act 2013, Chapter 3, Part B, Section 29.
[26] Dubai Data Protection Law, DIFC Law No. 5 of 2020, Article 11(k).
[27] Regulation 2024/1689, Article 10 and particularly 10(5).
[28] Marvin van Bekkum, "Using sensitive data to de-bias AI systems: Article 10(5) of the EU AI act", April 2025, available at https://www.sciencedirect.com/science/article/pii/S026736492500010X.
[29] Mark R Leiser, "Bias Mitigation Under the EU's AI Act and the GDPR", 24 June, 2025, available at https://digidata.substack.com/p/bias-mitigation-under-the-eus-ai.
[30] European Data Protection Board, Report of the work undertaken by the ChatGPT Taskforce, 23 May, 2024, available at https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf.
[31] Commission Nationale de l'Informatique et des Libertés, The legal basis of legitimate interest: focus sheet on the measures to be taken in the event of data collection by web scraping, 19 June, 2025, available at https://www.cnil.fr/fr/focus-interet-legitime-collecte-par-moissonnage.
[32] ibid.
[33] Commission Nationale de l'Informatique et des Libertés, Ensuring the lawfulness of the data processing - Defining a legal basis, 7 June, 2024, available at https://www.cnil.fr/en/ensuring-lawfulness-data-processing-legal-basis.
[34] Assessing the pulse of the European Health Data Space proposal, December 2023, available at https://www.eurordis.org/jelena-malinina-discusses-ehds/?utm_source=chatgpt.com
[35] Luca Marelli et al., "The European health data space: Too big to succeed?", September 2023, available at https://pmc.ncbi.nlm.nih.gov/articles/PMC10448378/.
[36] European Health Data Space: Council adopts new regulation improving cross-border access to EU health data, January 2025, available at https://www.consilium.europa.eu/en/press/press-releases/2025/01/21/european-health-data-space-council-adopts-new-regulation-improving-cross-border-access-to-eu-health-data/
[37] Authors Guild v. Google, Inc., available at https://cases.justia.com/federal/appellate-courts/ca2/13-4829/13-4829-2015-10-16.pdf?ts=1445005805.
[38] Opinion of Advocate General Jääskinen, supra note 10, para 95.
[39] Opinion of Advocate General Szpunar, Case C-136/17, GC and Others v. CNIL, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62017CC0136, para 49.
[40] ibid para 56.

[41] Catherine Jasserand, "Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data", August 2018, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3230342.
[42] See for example, Virginia Consumer Data Protection Act, Section 59.1-582.
[43] Paul Quinn and Gianclaudio Malgieri, "The Difficulty of Defining Sensitive Data – The Concept of Sensitive Data in the EU Data Protection Framework", 16 October 2020, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3713134.
[44] Daniel J Solove, "Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data", 21 January, 2024, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4322198.