Douwe Korff
Emeritus Professor of International Law, London Metropolitan University
Associate, Oxford Martin School, University of Oxford

The legal bases for processing of non-sensitive and sensitive data including the "legitimate interest" legal basis & the use of personal data to "train" an AI model*

1. Introduction

On 8 October, the European Data Protection Board (EDPB) issued guidelines on the processing of personal data on the basis of Article 6(1)(f) of the EU General Data Protection Regulation (GDPR).[1] This Note is a quick immediate response to the EDPB comments in that document relating to the processing of certain special categories of personal data that enjoy special protection under the GDPR, commonly referred to as "**sensitive data**". These are:

> personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data [when used] for the purpose of uniquely identifying a natural person, data concerning health [and] data concerning a natural person's sex life or sexual orientation …
>
> (See Article 9(1), discussed below, at 2 and 3)

Specifically, the EDPB appears to suggest that such data can be processed on the basis of the "legitimate interest" legal basis set out in Article 6(1)(f) of the GDPR, provided certain "**additional conditions**" for processing of sensitive data contained in Article 9(2) GDPR are met:

> In qualifying the nature of the data to be processed, the controller should pay special attention to, among other things … [t]he fact that special categories of personal data enjoy additional protection under Article 9 GDPR and, that the processing of special categories of personal data ("sensitive data") is only allowed under **specific additional conditions** set out in Article 9(2) GDPR.[2] In this regard, it should be kept in mind that a set of data that contains at least one sensitive data item is deemed sensitive data in its entirety, in particular if it is collected en bloc without it being possible to separate the data items from each other at the time of collection.[3] Further, it should be recalled that data are deemed sensitive if such data allow information falling within one of the categories referred to in Article 9(1) GDPR to be revealed.[4] It is irrelevant whether or not the information revealed by the processing operation in question is correct and whether the controller is acting with the aim of obtaining information that falls within one of the special categories referred to in that provision.[5] Hence, according to the jurisprudence of the CJEU, the relevant question is

---

* This note expands on and in some respects amends an earlier one that covered only the issue of legal bases for the processing of sensitive data, in particular "legitimate interest", available at: https://www.ianbrown.tech/2024/10/11/edpb-guidance-on-sensitive-data-and-legitimate-interest-processing-needs-more-clarity/

[1] EDPB, Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR, Version 1.0 (i.e., the version "adopted for public consultation"), adopted on 8 October 2024 (hereafter: "**the Guidelines**"), available at: https://www.edpb.europa.eu/system/files/2024-10/edpb_guidelines_202401_legitimateinterest_en.pdf#page26

[2] It should be reiterated that meeting the conditions laid down in Article 9(2) GDPR does not automatically fulfil the conditions of Article 6(1)(f) GDPR. **If this legal basis for processing is to be used, the controller must satisfy the requirements of both GDPR provisions when it processes special categories of personal data**. [original footnote, emphasis added]

[3] CJEU, judgment of 4 July 2023, Case C-252/21, *Meta v. Bundeskartellamt* (ECLI:EU:C:2023:537), para. 89. [original footnote]

[4] Ibid., para. 68. [original footnote]

[5] Ibid., para. 69. [original footnote]

> whether it is objectively possible to infer sensitive information from the data processed, irrespective of any intention of actually doing so.
>
> (p.14, emphasis added)

Below, I explain why, in my opinion, this is not clear enough, both in general (sections 2 and 3) and in particular also in relation to the use of personal data "scraped" from the Internet, to "train" AI models (section 4). I hope my comments can feed into the public consultation that will result in a second, final version of the guidelines.

2. The legal bases for processing non-sensitive and sensitive data under the GDPR

2.1 The rules

Article 6(1) of the GDPR stipulates, as a general rule, that personal data may only be processed if one of the following legal bases applies:

> (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
>
> (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
>
> (c) processing is necessary for compliance with a legal obligation to which the controller is subject;
>
> (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
>
> (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
>
> (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

The consent referred to in Article 6(1)(a) must meet the following conditions in order to be valid:

> [It must be consist of a] freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.
>
> (Article 4(11). See also Article 7 for further conditions for consent.)

Article 6(1) adds that:

> Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.

And the third paragraph of Article 6 stipulates that:

> The basis for the processing referred to in point (c) and (e) of paragraph 1 shall be laid down by:
>
> (a) Union law; or
>
> (b) Member State law to which the controller is subject.

Furthermore:

> The purpose of the processing shall be determined in that legal basis [i.e., law] or, as regards the processing referred to in point (e) of paragraph 1, shall be necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. … The Union or the Member State law [in question] shall meet an objective of public interest and be proportionate to the legitimate aim pursued.

The GDPR also contains a number of special rules – i.e., a prohibition subject to a number of specific exceptions – on the processing of sensitive data, set out in Article 9 as follows:

Article 9

Processing of special categories of personal data

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

2. Paragraph 1 shall not apply if one of the following applies:

(a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;

(b) processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorised by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject;

(c) processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent;

(d) processing is carried out in the course of its legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or to former members of the body or to persons who have regular contact with it in connection with its purposes and that the personal data are not disclosed outside that body without the consent of the data subjects;

(e) processing relates to personal data which are manifestly made public by the data subject;

(f) processing is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity;

(g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;

(h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;

(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy;

(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

2.2 The relationship between Articles 6 and 9

The basic legal rule is that a special rule applies over and above a more general rule (*lex specialis derogat lex generalis*). Therefore, the special rules in Article 9 (the prohibition with specific exceptions) apply over and above the rules on legal bases in Article 6 when it comes to processing of sensitive data. However, it is important to differentiate in this regard between special rules that merely add further conditions to a certain matter (here: to processing of sensitive data) and special rules that impose special restrictions on that matter (i.e., here again: on processing of sensitive data).

If a special rule merely adds additional conditions to a general rule in relation to a certain matter (here: processing of sensitive data), the processing to which the special rule relates must meet the requirements of the general rule and the additional requirements of the special rule.

But if a special role limits the application of a general rule in relation to a special matter (here: processing of sensitive data) to certain defined situations, then processing of personal data (here: sensitive data) in those special situations may only take place in those situations (and subject to the conditions set out in the special rule in relation to those situations), even if the general rule allowed processing of personal data (*in casu*: of non-sensitive data) more broadly.

Below, I set out the implications.

In most respects, the special rules in Article 9 can indeed be seen as imposing **additional conditions** on processing of personal data on the legal bases set out in Article 6 when sensitive data are processed, either in general or in specific contexts. Thus:

2.3 Article 9(2)(a) adds to the Article 6(1)(a) conditions for valid ***consent*** ("*freely given, specific, informed and unambiguous*"), the **additional condition** that the consent for processing of sensitive data must also be "*explicit*" (and allows the Member States to forbid processing of sensitive data even with such consent in certain context such as, e.g., employment relationships);*

*Note: Processing that is "*necessary for the performance of a **contract** to which the data subject is party*" or "*in order to take steps at the request of the data subject prior to entering into a contract*" (i.e., credit checks) (Article 6(1)(b)) is of course also based on consent (a contract being defined, at least in Continental European legal systems, as a matching of [expressed] wills) and will effectively always be "explicit". In any case, the requirement of necessity will of course have to apply particularly strictly in relation to the use of sensitive data in such contexts.

- Article 9(2)(g) relates to contexts covered by Article 6(1)(c) and (e) (i.e., "*processing [that] is necessary for compliance with a **legal obligation** to which the controller is subject*" and "*processing [that] is necessary for the performance of a **task carried out in the public interest** or in the **exercise of official authority** vested in the controller*") and adds to the requirement set out in Article 6(3) that such processing must be based on a Union or Member State law that "*meet[s] an objective of public interest and [is] proportionate to the legitimate aim pursued*", the **additional condition** that processing of sensitive data in such cases must be "*necessary for reasons of substantial public interest*" and that the law in question must "*provid[e] for appropriate safeguards for the fundamental rights and the interests of the data subject*";

- Article 9(2)(b) also relates to contexts covered by Article 6(1)(c) and (e) (here more specifically: ***employment and social security and social protection law***), but again also adds to the requirement that the processing must be covered by a Union or Member State *law that "meet[s] an objective of public interest and [is] proportionate to the legitimate aim pursued*", the **additional condition** that if sensitive data are processed the law in question must *"provid[e] for appropriate safeguards for the fundamental rights and the interests of the data subject*";

- Article 9(2)(h) also relates to contexts covered by Article 6(1)(c) and (e) (here more specifically: ***health and social care***), but Article 9(3) again also adds to the requirement that the processing must be covered by a Union or Member State *law that "meet[s] an objective of public interest and [is] proportionate to the legitimate aim pursued*", the **additional condition** that if sensitive data are processed this must be done "*by or under the responsibility of a professional subject to the obligation of professional secrecy under Union or Member State law or rules established by national competent bodies or by another person also subject to an obligation of secrecy under Union or Member State law or rules established by national competent bodies*";

- Article 9(2)(i) also relates to contexts covered by Article 6(1)(c) and (e) (here more specifically: ***threats to health and medical safety***) and adds the **additional condition** that

the relevant law must "*provide[ ] for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy*"; and

- Article 9(2)(c) adds to the legal basis in Article 6(1)(d) covering "*processing [that] is necessary in order to protect the vital interests of the data subject or of another natural person*", the **additional condition** "*where the data subject is physically or legally incapable of giving consent*".

Furthermore:

- Article 9(2)(j) sets out **additional conditions** in relation to processing sensitive data for archiving and research purposes in that it makes clear that such processing must comply with the "compatability" principle set out in Article 5(1)(e) and the special rules on archiving and research in Article 89, and **in addition** must not only be based on a law that is "*proportionate to the aim pursued*", but that must also "*respect the essence of the right to data protection*" and again "*provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject*".

This leaves three further situations in relation to which Article 9 allows for the processing of sensitive data:

- Article 9(2)(d) allows for processing of sensitive data on their members, former members and regular contacts by political, philosophical, religious or trade union bodies;
- Article 9(2)(e) allows for processing of sensitive data which were "manifestly made public by the data subject"; and
- Article 9(2)(f) allows for processing of sensitive data when that is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity.

As noted below, these three situations can be seen as situations covered by the "legitimate interest" legal basis set out in Article 6(1)9f) GDPR – but rather than imposing "additional conditions" on the application of that legal basis (in a way that would be similar to the above cases), they limit its scope of application.

3. Issues

I have two issues with the Guidelines.

3.1 Processing of sensitive data by public authorities

First of all, as noted at 2.1, above, the final sentence of Article 6(1) stipulates clearly and unequivocally that the "legitimate interest" legal basis "*shall not apply to processing carried out by public authorities in the performance of their tasks*". In that regard, the Guidelines say the following:[6]

> **Processing by public authorities**
>
> Article 6(1), second indent, of the GDPR states that the legal basis under Article 6(1)(f) shall not apply to processing carried out by public authorities in the performance of their tasks.

[6] Guidelines, section IV.2, paras. 98 – 99.

> Recital 47 of the GDPR clarifies the reason: "it is for the legislator to provide by law for the legal basis for public authorities to process personal data". Such provision indeed relates to the fact that, as a general rule, processing undertaken by public authorities falls under the scope of their tasks and missions provided for by EU or Member State law.
>
> Nevertheless, these provisions do not prevent from relying, in exceptional and limited cases, on Article 6(1)(f) GDPR when the processing is not linked to or does not relate to the performance of their specific tasks or the exercise of their prerogatives as public authorities, but concerns, where permitted by the national legal system, other activities that are lawfully carried out. Relying on Article 6(1)(f) GDPR in such exceptional cases should be documented internally. In no circumstances, public authorities may rely on Article 6(1)(f) for processing activities falling within the scope of the performance of their tasks.

The second paragraph appears to be, or at least gets close to being, *contra legem*. What are the "*exceptional and limited cases … when the processing is not linked to or does not relate to the performance of their specific tasks or the exercise of their prerogatives as public authorities, but concerns, where permitted by the national legal system, other activities that are lawfully carried out [by such authorities]*"? If they are "permitted by the national legal system", does that not mean that they are "in the performance of [the authorities'] tasks"?

At the very very least, the EDPB should provide clear examples of such "exceptional cases" in the revised guidelines.

3.2 Processing of sensitive data on the basis of the "legitimate interest" legal basis

Secondly, as concerns the relationship between Article 6 and 9, my analysis at 2, above, shows that:

- Article 9(2), paras. (a), (b), (c), (g) and (h), between them, do indeed impose "additional conditions" on the legal bases of consent (Article 6(1)(a)), compliance with a legal obligation (Article 6(1)(c)), vital interest (Article 6(1)(d)) and public interest tasks/official authority (Article 6(1)(e)), and on the rules relating to archiving and research; and
- Article 9 does not as such affect the legal basis of processing that is necessary for a contract or pre-contractual measure (Article 6(1)(b)), because the consent for such processing is always already "explicit" (although the necessity of the use of sensitive data in such context should be strictly assessed).

This leaves the three exemptions noted above, contained in Article 9(2)(d), (e) and (f):

- processing of sensitive data on their members, former members and regular contacts by political, philosophical, religious or trade union bodies;
- processing that relates to sensitive personal data which are manifestly made public by the data subject; and
- processing of sensitive data that is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity.

It can be argued that these are three contexts that in relation to non-sensitive data are covered by Article 6(1)(f), "legitimate interest". But if that is so, they constitute **the only three contexts** in which sensitive data can be based on that legal basis. In other words:

The "legitimate interest" legal basis for processing of personal data can only be relied on when the data include sensitive data in these three contexts: that the sensitive data relate to members and "close contacts" of political, philosophical, religious or trade union bodies and are only processed by those bodies; that they have been "manifestly" made public by the data subject; or that the sensitive datum or data are (strictly) necessary in a legal claims context.

In my opinion, this too should be spelled out in the second version of the Guidelines.

4. Implications in relation to the use of personal data "scraped" off the Internet to "train" AI models

AI models are often "trained" on data including personal data that are "scraped" off the Internet,[7] at times by deliberately bypassing web scraping protection mechanisms.[8] This comes within the scope of Article 9 GDPR, either if the data that are collected directly "reveal" a sensitive characteristic such as the ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or sexual orientation of any of the data subjects, or if inferences can be drawn from the (massive) data collections as to such matters.

As to the first: it is in practice almost impossible to not "catch" such directly revealing sensitive data in the "dragnet" of Internet scraping. In that regard, the EDPB rightly observed, with reference to the CJEU case-law, that:[9]

> it should be kept in mind that a set of data that contains at least one sensitive data item is deemed sensitive data in its entirety, in particular if it is collected en bloc without it being possible to separate the data items from each other at the time of collection.

More specifically: a controller cannot claim to not be processing sensitive data if they filter out those data immediately after they are collected (scraped up) and before using them in the training of the model, because they are still collected, which is a form of processing – to which therefore the whole of the GDPR applies.[10]

---

[7] See, e.g., Webautomation, *How to Train Your AI Model With Web Data Using Web Scraping*, 23 June 2023, available at: https://webautomation.io/blog/how-to-train-your-ai-model-with-web-data-using-web-scraping/

[8] See, e.g., this sponsored ad: "*Bright Data's Data Collector is a no code web scraping solution that extracts real-time public data from online platforms and delivers it to businesses on autopilot in different formats. It is especially useful when collecting data from websites that protect themselves against scraping. Using proxies and other techniques, Bright Data can bypass web scraping protection mechanisms.*"
https://research.aimultiple.com/machine-learning-web-scraping/
The main link in the above is to: https://brightdata.com/products/web-scraper

[9] Guidelines, p. 14. See the full quote with references on pp. 1 – 2, above.

[10] This was – and is – an issue in relation to the: USA interception of Internet communications in bulk, although that was glossed over in the drafting and adoption of the adequacy decision on that third country. It should also have been taken into account in relation to the UK now that it is also a third country in EU terms, because the intercepting of the Internet communications from the data carriers is in fact done by the USA and the UK working together. See: Ian Brown & Douwe Korff, Exchanges of personal data after the Schrems II judgment, study carried out at the request

And as to the second: it has been repeatedly shown that sensitive characteristics can often be deduced, with a high degree of certainty, from large datasets containing seemingly only non-sensitive data.[11] In that regard, the EDPB (again rightly) observed, again with reference to the case-law, that:[12]

> [data are deemed sensitive if] it is objectively possible to infer sensitive information from the data processed, irrespective of any intention of actually doing so.

In my opinion, the scraping of large amounts of personal data from (semi-)public sources and in particular the Internet should therefore (almost?) always be assessed with reference to Article 9 GDPR, rather than only to Article 6.[13]

And the only Article 9 GDPR exceptions to the in-principle prohibition on the processing of sensitive data in the first paragraph of that article that are relevant to the scraping of personal data from the Internet in order to train commercial AI models[14] **are the ones set out in clauses (a), (b) and (e) of the second paragraph: consent, necessity for a contract or a pre-contract (credit) check, and the data having been "manifestly made public by the data subject".**

The Court of Justice of the European Union (CJEU or "the Court") has addressed the issue of large data collection, aggragating and analysing by Meta (previously Facebook) in general in its 2023 judgment in the case of *Meta Platforms and Others (General terms of use of a social network)* (hereafter: "*Meta GC*"),[15] and in its very recent judgment in the case of *Schrems v. Meta*[16] in relation to the questions of when sensitive data can be said to have "manifestly" been made public by the data subject, and what the implications are. I will look at these judgments next.

---

of the European Parliament's Civil Liberties (LIBE) Committee into the future of EU – US flows of personal data, July 2021, available at:
https://www.europarl.europa.eu/RegData/etudes/STUD/2021/694678/IPOL_STU(2021)694678_EN.pdf

[11] See, e.g., Montreal AI Ethics Institute, *When Algorithms Infer Pregnancy or Other Sensitive Information About People*, 2 November 2020, available at:
https://montrealethics.ai/when-algorithms-infer-pregnancy-or-other-sensitive-information-about-people/
"*Machine learning can ascertain a lot about you — including some of your most sensitive information. For instance, it can predict your sexual orientation, whether you're pregnant, whether you'll quit your job, and whether you're likely to die soon*."

[12] Guidelines, p. 14. See again the full quote with references on pp. 1 – 2, above.

[13] Cf. my blog article on what I called "*A Bad Belgian Decision*" that split the training and use of AI models into two phases and allowed the first "training" phase to take place on the basis of "legitimate interest" even though the actual use of the model in the "second phase" (deployment) could not be allowed on that legal basis; and the follow-up article in which I noted that in fact the French data protection authority took the same (in my opinion, erroneous) view:
https://www.ianbrown.tech/wp-content/uploads/2024/04/KORFF-D-A-bad-Belgian-decision-240411-EDITED-1.pdf
https://www.ianbrown.tech/wp-content/uploads/2024/04/KORFF-D-AI-the-GDPR-a-follow-up.pdf

[14] In this short note, I am leaving aside AI models developed for public sector use.

[15] CJEU Grand Chamber judgment in the case of *Meta Platforms Inc., formerly Facebook Inc., Meta Platforms Ireland Ltd, formerly Facebook Ireland Ltd and Facebook Deutschland GmbH v. Bundeskartellamt* (Case C-252/21) of 4 July 2023, ECLI:EU:C:2023:537.

[16] CJEU judgment in the case of *Maximilian Schrems v. Meta Platforms Ireland Ltd, formerly Facebook Ireland Ltd* (Case C-446/21) of 4 October 2024, ECLI:EU:C:2024:834

(i) What are the general conditions for the bulk collection of personal data to train an algorithm or AI model?

In *Schrems v. Meta*, the Court confirmed that:[17]

> Meta Platforms Ireland collects the personal data of Facebook users, including Mr Schrems, concerning those users' activities both on and outside that social network, including in particular data relating to online platform visits and third-party websites and apps, and also follows users' navigation patterns on those sites through the use of social plug-ins and pixels embedded in the relevant websites –

and that Meta aggregates and analyses those data, without distinction as to type of data, in order to offer the data subject/Facebook-Meta user personalised advertising.[18]

The Court also reiterated the view already expressed in *Meta GC*, i.e., that:[19]

> such processing is particularly extensive since it relates to potentially unlimited data and has a significant impact on the user, a large part – if not almost all – of whose online activities are monitored by Meta Platforms Ireland, which may give rise to the feeling that his or her private life is being continuously monitored.

On this general issue, the Court, in the recent case, *Schrems v. Meta*, linked this to the data minimisation principle set out in Article 5(1)(c) GDPR as follows:

> [Personal data shall be] kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; …[20]

It noted that it had already previously held that:[21]

> [I]n the light of the principle of data minimisation provided for in Article 5(1)(c) of the GDPR, the controller may not engage in the collection of personal data in a generalised and indiscriminate manner and must refrain from collecting data which are not strictly necessary having regard to the purpose of the processing (judgment of 24 February 2022, *Valsts ieņēmumu dienests (Processing of personal data for tax purposes)*, C 175/20, EU:C:2022:124, paragraph 74).

The Court then set out a major general principle relating to the scraping of large amounts of personal data for advertising purposes, as follows:[22]

> **[T]he principle of data minimisation provided for [in Article 5(1)(c) GDPR] precludes all of the personal data obtained by a controller, such as the operator of an online social network platform, from the data subject or third parties and collected either on or outside that platform, from being aggregated, analysed and processed for the purposes of targeted advertising without restriction as to time and without distinction as to type of data.**

---

17 *Schrems v. Meta* (previous footnote), para. 59.

18 *Idem*, para. 83.

19 *Idem*, para. 62, with reference to *Meta GC*, para. 118.

20 The remainder of the clause relates to processing for archiving, scientific or historical research or statistical purposes and need not be discussed here.

21 *Schrems v. Meta* (footnote 16, above), para. 59.

22 *Idem*, para. 65, repeated *verbatim* in the conclusions (para. 84, at 1), emphasis added.

Suffice it to note that this is a major, newly-clarified principle that will have to lead to significant changes to the collecting of personal data in bulk and essentially *sine die*, for the purpose of offering "targeted" "personalised" advertising.[23]

In my opinion, the same applies to other collections in bulk over extended periods, including the scraping of personal data from the Internet, to train commercial AI models,[24] for similar (i.e., commercial) purposes such as AI models underpinning gaming or gambling (that grab personal data from players), educational software (that grabs sensitive data from pupils) or dating algorithms (that grab sensitive data from people seeking relationships): unless clearly limited as to (relevant and necessary) [sensitive] data and in time, such bulk data processing is incompatible with the GDPR.

(ii) When can sensitive data be said to have "manifestly" been made "public" by the data subject, and what are the implications?

In *Meta GC*, the Court had already held that:[25]

> Article 9(2)(e) of the GDPR must be interpreted as meaning that, where the user of an online social network visits websites or apps to which one or more of the categories set out in Article 9(1) of the GDPR relate, the user does **not** manifestly make public, within the meaning of the first of those provisions, the data relating to those visits collected by the operator of that online social network via cookies or similar storage technologies.

This is only different if the data subject has explicitly (and freely, specifically and unambiguously) indicated that they wanted the information to be (made) public, i.e., if they explicitly consented in accordance with Article 9(2)(a) GDPR. In the rather convoluted phrasing of the Court:[26]

> Where he or she enters information into such websites or apps or where he or she clicks or taps on buttons integrated into those sites and apps, such as the 'Like' or 'Share' buttons or buttons enabling the user to identify himself or herself on those sites or apps using login credentials linked to his or her social network user account, his or her telephone number or email address, **that user manifestly makes public, within the meaning of Article 9(2)(e), the data thus entered or resulting from the clicking or tapping on those buttons *only in the circumstance where he or she has explicitly made the choice beforehand***, as the case may be on the basis of individual settings selected with full knowledge of the facts, to make the data relating to him or her publicly accessible to an unlimited number of persons.

It does not suffice that the platform settings pre-select ("pre-tick") such choices; rather, the user has to select them "*with full knowledge of the facts*". Similarly, the user must have been clearly made aware that clicking on a 'Like' or 'Share' button (etc.) means she or he is consciously and "manifestly" making this information public.

---

[23] The Belgian decisions and the French views that I criticised in my earlier blog posts (footnote 13, above) will certainly need to be reviewed.

[24] See footnote 14, above.

[25] *Meta GC* (footnote 16, above), para. 84, repeated *verbatim* in the conclusions (para. 155, at 3), emphasis added.

[26] *Idem*, para. 85, also repeated *verbatim* in the conclusions (para. 155, at 3), emphases added.

But what if a person makes a sensitive matter public outside of a platform? Does this mean that the platform can then (more) freely use this datum?

This question arose in *Schrems v. Meta*, in which it was held by the Austrian Oberster Gerichtshof (Supreme Court) that:[27]

> Mr Schrems discloses his sexual orientation in public. In particular, on the occasion of a panel discussion in which he participated in Vienna on 12 February 2019, at the invitation of the Representation of the European Commission in Austria, Mr Schrems referred to his sexual orientation for the purpose of criticising Facebook's processing of personal data, including the processing of his own data. However, and as also stated by Mr Schrems on that occasion, he has never mentioned that aspect of his personal life on his Facebook profile.

The Court held that:[28]

> [I]t is apparent from the order for reference that the panel discussion organised in Vienna on 12 February 2019, in the context of which Mr Schrems made a statement about his sexual orientation, was accessible to the public, who could obtain a ticket to attend the event, subject to seating availability, and that it was streamed. Moreover, a recording of the round table was subsequently published as a podcast, as well as on the Commission's YouTube channel.
>
> In those circumstances, and subject to verifications which it is for the referring court to carry out, the possibility cannot be ruled out that that statement, although forming part of a broader discussion and made solely for the purpose of criticising the processing of personal data by Facebook, constitutes an act by which the person concerned in any event manifestly made his sexual orientation public within the meaning of Article 9(2)(e) of the GDPR.

Note the cautious language: "*the possibility cannot be ruled out*" that Schrems, by being open about his sexual orientation, "manifestly" made the fact public; that is still a matter for a domestic court to determine in the light of all the specific circumstances.

Moreover, in any case, and "*contrary to the contentions of Meta Platforms Ireland*":[29]

> **[T]he fact that the data subject has manifestly made public data concerning his or her sexual orientation … alone does not … authorise the processing of other personal data relating to that data subject's sexual orientation.**
>
> **Thus, it would be contrary to the restrictive interpretation that should be made of Article 9(2)(e) of the GDPR to find that all data relating to the sexual orientation of a person fall outside the scope of protection under Article 9(1) thereof solely because the data subject has manifestly made public personal data relating to his or her sexual orientation.**

This is a crucial limitation on the "manifestly made public" exception to the in-principle prohibition on the processing of sensitive data in the GDPR, and applicable to all sensitive data: if someone is open about a sensitive aspect of themselves (which can be any of the ones listed

---

27 *Schrems v. Meta* (footnote 16, above), para. 32.

28 *Idem*, paras. 78 – 79.

29 *Idem*, para. 83, emphasis added. See paras. 80 – 82 for the wider reasoning and earlier case-law.

in Article 9(1) including their ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or sexual orientation),[30] that does not create a *carte blanche*: controllers may not simply assume that because that aspect of the person is in the open, they have a free for all in relation to the collecting of anything else relating to that aspect of that person. In particular, it does not provide a license to use the sensitive data to train an AI model.

(iii) What about the other exceptions to the in-principle prohibition on the processing of sensitive data that can be seen as special applications of the "legitimate interest" legal basis?

In fact, the Court goes further:[31]

> **Moreover, the fact that a person has manifestly made public information concerning his or her sexual orientation does not mean that that person has given his or her consent within the meaning of Article 9(2)(a) of the GDPR to processing of other data relating to his or her sexual orientation by the operator of an online social network platform.**

In other words, the fact that a person is open about a sensitive aspect of her- or himself does not mean they give consent to every- and anyman (or woman) to start collecting other information about that aspect of the person's life. – let alone use that datum to train an AI model. **That openness does not constitute (explicit) consent within the meaning of Article 9(2)(a) GDPR.**

This is also relevant in relation to contracts or pre-contract (credit) checks. **The mere fact that a person seeking to enter into a contact (or agreeing to a credit check if they ask for credit) is open about a sensitive aspect of her- or himself does not mean that the other party, or the credit agency, can take this sensitive matter into account in entering (or refusing to enter) into the contract or in such checks.** Rather, it remains the case that such sensitive information may only be processed in these contexts if those data are (*strictly*) necessary for the contract or the check – which would only apply in special circumstances.

Contracts relating to the use of sophisticated software – e.g., gaming, gambling, education or dating apps, to use my previous examples – often stipulate that the provider of the app may use the user's data to (further) train the model (these days almost always an AI model); and the user data will often include sensitive data, or allow sensitive aspects of the user to be inferred or deduced.

In my opinion, the collecting and further processing of app user data is <u>not</u> necessary for the fulfilment of a contract that is simply aimed at providing the relevant gaming, gambling, educational or relationship-finding opportunity. The fact that a user, when using the app, may reveal one or more sensitive aspects of himself does not alter this: especially in relation to sensitive data, clauses allowing the use of uer data for the training of the AI model violate the GDPR: the app provider asks for more (and sensitive) data than is necessary for the contract.[32]

---

30 See the full list on p. 1, above.

31 *Schrems v. Meta* (footnote 16, above), para. 83, emphasis added.

32 By contrast, the Belgian and French authorities felt that training of an AI model was a purpose in itself, that could be split from the overall purpose of the app, and based on the "legitimate interest" legal basis. In my criticism (footnote 13, above), I called this disingenuous and wrong.

5. Conclusions

1. In the final, post-consultation version of the guidelines, the EDPB should clarify what is meant by "*exceptional and limited cases … when the processing [of personal data] is not linked to or does not relate to the performance of their specific tasks or the exercise of their prerogatives as public authorities, but concerns, where permitted by the national legal system, other activities that are lawfully carried out [by such authorities]*" and that can therefore, in its view, be done by such authorities under the "legitimate interest" legal basis of Article 6(1)(f) GDPR. At first glance, this appears to be, or to get close to being, *contra legem*.

2. The "legitimate interest" legal basis for processing of personal data can only be relied on when the data include sensitive data in these three contexts:

   - that the sensitive data relate to members and "close contacts" of political, philosophical, religious or trade union bodies and are only processed by those bodies;
   - that they have been "manifestly" made public by the data subject; or
   - that the sensitive datum or data are (strictly) necessary in a legal claims context.

3. The scraping of large amounts of personal data from (semi-)public sources and in particular the Internet should (almost?) always be assessed with reference to Article 9 GDPR, rather than only to Article 6. And the only Article 9 GDPR exceptions to the in-principle prohibition on the processing of sensitive data in the first paragraph of that article that are relevant to the scraping of personal data from the Internet in order to train commercial AI models are the ones set out in clauses (a), (b) and (e) of the second paragraph:

   - consent;
   - necessity for a contract or a pre-contract (credit) check; and
   - the data having been "manifestly made public by the data subject".

4. "*[T]he principle of data minimisation provided for [in Article 5(1)(c) GDPR] precludes all of the personal data obtained by a controller, such as the operator of an online social network platform, from the data subject or third parties and collected either on or outside that platform, from being aggregated, analysed and processed for the purposes of targeted advertising without restriction as to time and without distinction as to type of data.*" (*Schrems v. Meta*, para. 65)

   And the same applies to other collections in bulk over extended periods, including the scraping of personal data from the Internet, to train commercial AI models, for similar (i.e., commercial) purposes such as AI models underpinning gaming or gambling (that grabs personal data from players), educational software (that grabs sensitive data from pupils) or dating algorithms (that grab sensitive data from people seeking relationships: unless clearly limited as to (relevant and necessary) [sensitive] data and in time, such bulk data processing is incompatible with the GDPR.

5. If someone is open about a sensitive aspect of themselves (which can be any of the ones listed in Article 9(1) including their ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or sexual orientation), that does not create a *carte blanche*: controllers may not simply assume that because that aspect of the person is in the open, they have a free for all in relation to the collecting of anything else relating to that aspect of that person. (*Schrems v. Meta*, para. 83). In particular, it does not provide a license to use the sensitive data to train an AI model.

6. More in particular: the collecting and further processing of app user data is not necessary for the fulfilment of a contract that is simply aimed at providing the relevant gaming, gambling, educational or relationship-finding opportunity. The fact that a user, when using the app, may reveal one or more sensitive aspects of himself does not alter this: especially in relation to sensitive data, clauses allowing the use of user data for the training of the AI model violate the GDPR: the app provider asks for more (and sensitive) data than is necessary for the contract.

I hope that the European Data Protection Board will consider my analyses and conclusions and, to the extent they agree with them, will reflect them in it second (post-consultation) version of the guidelines.

- o – O – o -

Douwe Korff (Prof.)
Cambridge (UK), 16 October 2024