The use of personal data to train artificial intelligence systems

The Regional High Court (*Oberlandesgericht)* of Cologne, in its Judgment of 23 May 2025 (15 UKl 2/25), considers that the company that owns two social media platforms that combines and uses (for the training of an artificial intelligence system) personal data made public by platform users does not infringe either the Digital Markets Act or the General Data Protection Regulation.

ÁNGEL GARCÍA VIDAL
Professor of Corporate & Commercial Law, University of Santiago de Compostela
Academic counsel, Gómez-Acebo & Pombo

1. Introduction

Can a company that owns two social media platforms combine personal data made public by users of the platforms when the result of the combination is used to train an artificial intelligence system? Is this an unfair practice by an access *gatekeeper* contrary to the Digital Markets Act? Is personal data protection legislation being complied with? These are all questions that have been raised in an interesting case brought before the Regional High Court *(Oberlandesgericht)* of Cologne and that has led to a judgment (of 23 May 2025, 15 UKl 2/25[1]) that is very relevant insofar as it interprets EU law directly applicable in Spain.

[1] The full text can be found at this *link*.

In the case before the German court, the claimant sought an interim injunction prohibiting the defendant from processing personal data published by consumers on Facebook and Instagram for the development and improvement of artificial intelligence systems, on the grounds that doing so would infringe the Digital Markets Act and the General Data Protection Regulation.

2. The Digital Markets Act and the combination of personal data

As is well known, the Digital Markets Act (Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September on contestable and fair markets in the digital sector) applies to basic platform services provided or offered by *gatekeepers* to business users established in the Union or to end users established or located in the Union, where *gatekeeper* means an undertaking designated as such by the European Commission on the basis that it has a significant impact on the internal market, provides a core platform service that is an important gateway for business users to reach end users, and enjoys an entrenched and durable position, in its operations, or it is foreseeable that it will enjoy such a position in the near future. For these purposes, *core platform services* are online intermediation services, online search engines, online social networking services, video-sharing platform services, number-independent interpersonal communications services, operating systems, web browsers, virtual assistants, cloud computing services and online advertising services.

In the case referred to in the German court judgment here analysed, it is undisputed that the defendant has the status of *gatekeeper* and that the social media platforms involved are core services, and what is at issue is whether the defendant has complied with the prohibition imposed on it by Article 5(2)(b) of the Digital Markets Act, according to which *gatekeepers* shall refrain from combining "personal data from the relevant core platform service with personal data from any further core platform services or from any other services provided by the *gatekeeper* or with personal data from third-party services".

In the opinion of the Regional High Court of Cologne, although in this case there is a combination of data from Facebook and Instagram into a single data set for training an artificial intelligence system, there is no infringement of Article 5(2)(b) of the Digital Markets Act because there is no combination within the meaning of that provision.

Although the wording of Article 5(2)(b) of the Digital Markets Act expressly refers to the combination of 'personal data', there is no legal definition of the concept of *data combination*, which raises the question of whether the provision applies whenever personal data are merged or whether, on the contrary, it is necessary for the merged data to relate to the same person.

The Regional High Court of Cologne takes the view that such a combination can only be deemed to exist when the merged data refer to the same user. In the Court's view, if any combination of data were prohibited, even if such data are not linked to a specific person, the wording in Article 5(2) of the Digital Markets Act would be meaningless when it provides that the prohibitions laid down therein do not apply when "the end user has been presented with the specific choice and has given consent within the

meaning of Article 4, point (11), and Article 7 of Regulation (EU) 2016/679".

Furthermore, it is also noted that the antecedents of the Digital Markets Act confirm this interpretation. It should be recalled that the Digital Markets Act establishes a series of obligations and prohibitions that had already given rise to the imposition of penalties by competition authorities, as well as to various judgments of the Court of Justice of the European Union (CJEU). In this regard, the prohibitions in Article 5(2) of the Digital Markets Act find their closest precedent in the case decided by the CJEU in its Judgment of 4 July 2023, C-252/21, Facebook (ECLI:EU:C:2023:537), concerning the combination of personal data extracted from different platforms. Indeed, as the Regional High Court of Cologne points out, although the aforementioned CJEU judgment is subsequent to the adoption of the Digital Markets Act, the issue had already been raised before the German competition authorities and was taken into account by the EU legislature.

It can therefore be concluded that the Digital Markets Act seeks to prohibit the combination of personal data from different platforms or social media in order to prevent the creation of customer profiles for personalised advertising, but not the combination of personal data from different sources with the intention of creating a corpus of data with which to "feed" an artificial intelligence system.

3. The use of personal data to train artificial intelligence from the perspective of the General Data Protection Regulation

3.1. When analysing the combination of data from the two social media platforms it becomes necessary to examine, from the point of view of personal data protection law, whether the processing of such data has any legal basis.

In this regard, the German court ruling now under discussion highlights that the AI Act —Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence— has not affected the General Data Protection Regulation's rules in this regard, and no specific legal basis has been approved to justify the processing of personal data for the training of artificial intelligence. Therefore, in order to process the data, it is necessary to rely on one of the legitimising bases of the General Data Protection Regulation.

A legal basis is required to legitimise the processing of personal data to train artificial intelligence

The German court, contrary to the claimant's arguments, is of the opinion that the use of personal data to train artificial intelligence is covered by Article 6(1)(f) of the General Data Protection Regulation, according to which the processing of personal data is lawful if it "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". This is in line with the European Data Protection Board in its recent Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, where it also concludes that Article 6(1)(f) constitutes an appropriate basis for training artificial intelligence models with data sets containing personal data.

3.2. In applying this provision, the Court of Justice, in its judgments of 17 June 2021, C-597/19, and 4 July 2023, C-252/21, has held that "that provision lays down three cumulative conditions so that the processing of personal data covered by that provision is lawful, namely, first, the pursuit of a legitimate interest by the data controller or by a third party; second, the need to process personal data for the purposes of the legitimate interests pursued; and third, that the interests or fundamental freedoms and rights of the person concerned by the data protection do not take precedence over the legitimate interest of the controller or of a third party".

In the present case, the Regional High Court of Cologne considers that all these requirements allowing the processing of personal data for the purpose of training artificial intelligence are met:

a) With regard to legitimate interest, the German court recalls that the Court of Justice has held that economic interests, among others, are legitimate (Judgment of the Court of Justice of 4 October 2024, C-621/22). In this case, the interest lies in the intention to use the capabilities of generative artificial intelligence to provide a conversation assistant that, by adapting to specific regional customs, offers, for example, real-time responses for chats, assistance in organising and planning holidays, or even assistance with text writing.

The German court takes the view that data processing is necessary to satisfy the legitimate interests pursued by the data controller

b) With regard to the requirement that the processing of personal data be necessary for the satisfaction of the legitimate interest pursued, the Court of Justice — Judgment of 9 January 2025, C-394/23— stated that it is for the national court to ascertain "whether the legitimate interest pursued by the processing of the data can reasonably be achieved just as effectively by other means less restrictive of the fundamental freedoms and rights of data subjects, since such processing must be carried out only in so far as is strictly necessary for the

attainment of that legitimate interest". Similarly, the requirement relating to the necessity of the processing must be examined in relation to the principle of data minimisation, enshrined in Article 5(1)(c) of the General Data Protection Regulation, according to which personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (Judgment of the Court of Justice of the European Union of 12 September 2024, C-17/22 and C-18/22).

On this basis, the German court considers, in summary proceedings appropriate to interim relief applications, that the processing of personal data included in social media platforms is necessary for the training of artificial intelligence and complies with the aforementioned principle of data minimisation. In its view, there are no less intrusive options that are suitable for adequately satisfying the legitimate interest pursued, considering that anonymisation of the data would be impracticable *(unpraktikabel)* and that the use of so-called *Flywheel-Daten* (referring to user behaviour, visits, clicks, etc.) would drastically reduce the volume of usable data and would not meet the need to train generative artificial intelligence models with "vast amounts of text, images, videos and other data" (a need explicitly recognised in recital 105 of the AI Act).

c) Lastly, the Regional High Court of Cologne also concludes that the requirement that the interests or fundamental rights and freedoms of the data subject in data protection do not override the legitimate interests of the data controller or a third party is fulfilled. It bases this conclusion, among other arguments, on the fact that the data subjects could reasonably expect such processing when the data was collected; that only data from platforms that are open or public (and therefore also accessible via search engines) are used for training artificial intelligence; and on the fact that the defendant has implemented many of the measures recommended by the European Data Protection Board in its aforementioned Opinion 28/2024, paragraph 96 et seq., to mitigate the impact on the rights of data subjects.

Among the measures adopted in the specific case is the deletion of certain data (removal of full names, email addresses, telephone numbers, national identification numbers, user IDs, credit/debit card numbers, bank account numbers, vehicle registration numbers, IP addresses and postal addresses) and its collection in an unstructured manner. The court is aware that, despite everything, this does not mean anonymisation of the data, as the faces of identifiable persons remain in the photographs,

but the risk of identification is considerably reduced by the deletion of the aforementioned data. Also relevant is the measure allowing users to prevent their data from being included in the training data set by deleting the public nature of their account or their posts on social media, even retroactively.

3.3. Despite all of the above, users sometimes enter personal data of third parties on the social media accounts involved in the case, such as photographs in which those third parties appear. Those other people cannot modify the public nature of the accounts or make use of the possibility offered by the social media platform owner to account holders to object to the processing of data for the purposes of training artificial intelligence.

However, the Regional High Court of Cologne takes the view that this situation does not affect the lawfulness of the processing of the data on the basis of Article 6(1)(f) of the General Data Protection Regulation. In its view, the interests of those third parties do not prevail because the possible harm to them would arise if the artificial intelligence system were to provide, in response to a specific question, personal data relating to those third parties. And, according to the German court, that risk is very low, since the artificial intelligence system is not repeatedly trained on the data set in question, which means that there is a high probability that information about a specific individual will be lost in the vast amount of data and, consequently, that such information will not appear unchanged in an provided by the artificial intelligence system.

4. Conclusion

As can be seen, this is an important court ruling that addresses a highly significant issue and facilitates the use of personal data for training artificial intelligence systems. All the same, further developments in this matter will have to be monitored given that the ruling was adopted in the context of an application for interim relief.