

El uso de datos personales para entrenar sistemas de inteligencia artificial

El Tribunal Superior Regional (*Oberlandesgericht*) de Colonia, en su Resolución de 23 de mayo del 2025 (15 UKI 2/25), considera que la compañía titular de dos redes sociales que combina y usa (para el entrenamiento de un sistema de inteligencia artificial) los datos de carácter personal que han hecho públicos los usuarios de las redes no infringe ni el Reglamento de mercados digitales ni el Reglamento General de protección de datos.

ÁNGEL GARCÍA VIDAL

Catedrático de Derecho Mercantil de la Universidad de Santiago de Compostela
Consejero académico (asesor externo) de Gómez-Acebo & Pombo

1. Preliminar

¿Puede la compañía titular de dos redes sociales combinar los datos —de carácter personal— que han hecho públicos los usuarios de las redes cuando el resultado de la combinación se usa en el entrenamiento de un sistema de inteligencia artificial? ¿Estamos ante una práctica des-

leal de un guardián de acceso contraria al Reglamento de mercados digitales? ¿Se respeta la normativa de protección de datos de carácter personal? Todas estas son cuestiones que se han planteado en un interesante asunto que se ha presentado ante los tribunales alemanes y que ha dado lugar a la resolución del Tribunal Superior Regional (*Oberlandesgericht*) de Colonia

de 23 de mayo del 2025 (15 UKI 2/25)¹. Se trata de un pronunciamiento muy relevante, en la medida en que aplica el Derecho de la Unión Europea, que es también normativa vigente en España.

En el caso afrontado por el tribunal alemán, el demandante solicita una medida cautelar que prohíba al demandado tratar los datos personales publicados por los consumidores en los servicios de Facebook e Instagram para el desarrollo y la mejora de sistemas de inteligencia artificial, por entender que, actuando de ese modo, se estaría infringiendo el Reglamento de mercados digitales (Digital Services Act), así como el Reglamento General de protección de datos.

2. El Reglamento de mercados digitales y la combinación de datos personales

Como es sabido, el Reglamento de mercados digitales —Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo, de 14 de septiembre, sobre mercados disputables y equitativos en el sector digital— se aplica a los servicios básicos de plataforma prestados u ofrecidos por guardianes de acceso a usuarios profesionales establecidos en la Unión o a usuarios finales establecidos o situados en la Unión, entendiéndose por *guardián de acceso* una empresa designada como tal por la Comisión Europea por tener una gran influencia en el mercado interior, por prestar un servicio básico de plataforma consistente en ser una puerta de acceso importante para que los usuarios profesionales lleguen a los usuarios finales y por tener —o ser previsible que la va a alcanzar próximamente— una posición afianzada y duradera

en lo que respecta a sus operaciones. Y, a estos efectos, son *servicios básicos de plataforma* servicios de intermediación en línea, motores de búsqueda en línea, servicios de redes sociales en línea, servicios de plataforma de intercambio de vídeos, servicios de comunicaciones interpersonales independientes de la numeración, sistemas operativos, navegadores web, asistentes virtuales, servicios de computación en la nube y servicios de publicidad en línea.

En el caso al que se refiere la resolución judicial alemana ahora analizada, es pacífico que la demandada tiene la condición de guardián de acceso y que las redes sociales implicadas son servicios básicos, y lo que se discute es si ha cumplido la prohibición que a tales empresas les impone el artículo 5.2b del Reglamento de mercados digitales, según la cual, los guardianes de acceso se abstendrán de «combinar datos personales procedentes de los servicios básicos de plataforma pertinentes con datos personales procedentes de cualesquiera servicios básicos de plataforma adicionales o de cualquier otro servicio que proporcione el guardián de acceso o con datos personales procedentes de servicios de terceros».

Pues bien, a juicio del Tribunal Superior Regional de Colonia, aunque en este caso hay una combinación de los datos de las redes sociales Facebook e Instagram en un único conjunto de datos para entrenar un sistema de inteligencia artificial, no se infringe el citado artículo 5.2b del Reglamento de mercados digitales porque no existe una combinación en el sentido de dicha disposición legal.

¹ Su texto íntegro puede verse en este [enlace](#).

Aunque el tenor literal del artículo 5.2b del reglamento se refiere expresamente a la combinación de «datos personales», falta una definición legal del concepto de com-

Se requiere una base jurídica que legitime el tratamiento de datos personales para entrenar la inteligencia artificial

binación de datos, lo que genera la duda si el precepto se aplica siempre que se fusionen datos personales o si, por el contrario, es necesario que los datos fusionados se refieran a una misma persona.

Pues bien, el Tribunal Superior Regional de Colonia considera que sólo cabe entender que existe tal combinación cuando los datos que se fusionan se refieren a un mismo usuario. A juicio del tribunal, si quedara prohibida toda combinación de datos, aunque no estén vinculados a una determinada persona, carecería de sentido la regulación del artículo 5.2 del Reglamento de mercados digitales cuando dispone que las prohibiciones en él establecidas no operan cuando «se le haya presentado al usuario final esa opción específica y éste haya dado su consentimiento en el sentido del artículo 4, punto 11, y del artículo 7 del Reglamento (UE) 2016/679».

Además, también se destaca que los antecedentes del reglamento confirman esta interpretación. Recuérdese, en efecto, que el Reglamento de mercados digitales tipifica una serie de obligaciones y prohibiciones que ya habían dado lugar previamente a actuaciones sancionadoras de las autoridades de defensa de la competencia, así

como a distintas sentencias del Tribunal de Justicia de la Unión Europea (TJUE). Y, en este sentido, las prohibiciones del artículo 5.2 del reglamento encuentran su antecedente próximo en el caso resuelto por el Tribunal de Justicia en su Sentencia de la Gran Sala de 4 de julio del 2023, C-252/21, *Facebook* (ECLI:EU:C:2023:537), referido a la combinación de datos de carácter personal extraídos de distintas plataformas. En efecto, como re-

cuerda el Tribunal Superior Regional de Colonia, aunque la citada sentencia del Tribunal de Justicia es posterior a la aprobación del Reglamento de mercados digitales, la problemática ya se había suscitado ante las autoridades alemanas de defensa de la competencia y fue tenida en cuenta por el legislador de la Unión.

Cabe concluir, por tanto, que el Reglamento de mercados digitales lo que busca es prohibir la combinación de datos personales de distintas plataformas o redes sociales para evitar que se creen perfiles de clientes para publicidad personalizada, pero no, en cambio, la combinación de datos personales procedentes de distintas fuentes, con la intención de crear un corpus de datos con el que «nutrir» un sistema de inteligencia artificial.

3. La utilización de datos personales para el entrenamiento de la inteligencia artificial desde la perspectiva del Reglamento General de protección de datos

3.1. Analizando la combinación de datos procedentes de las dos redes sociales, desde el punto de vista de la normativa de protección de los datos de carácter personal, debe examinarse si

existe algún fundamento jurídico en el que asentar la licitud del tratamiento de dichos datos.

A este respecto, la resolución judicial alemana ahora comentada destaca que el Reglamento de inteligencia artificial —Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio del 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial— no ha afectado en este punto a la regulación del Reglamento General de protección de datos, sin que se haya aprobado una base jurídica específica que justifique el tratamiento de datos personales para el entrenamiento de la inteligencia artificial. Por ello, para poder realizar el tratamiento de los datos es necesario basarse en alguna de las bases legitimadoras del Reglamento General de protección de datos.

El tribunal alemán, en contra de lo alegado por la parte demandante, considera que la utilización de datos personales para entrenar la inteligencia artificial queda amparada por el artículo 6.1f de dicho reglamento, según el cual el tratamiento de datos personales es lícito si «es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales, en particular cuando el interesado sea un niño». De este modo se sigue el mismo planteamiento que el Comité Europeo de Protección de Datos en su reciente Dictamen 28/2024 sobre determinados

aspectos de la protección de datos relacionados con el tratamiento de datos personales en el contexto de los modelos de inteligencia artificial, donde también considera que el citado artículo 6.1f constituye una base adecuada para entrenar modelos de inteligencia artificial con conjuntos de datos que contienen datos personales.

- 3.2. Al aplicar este precepto, el Tribunal de Justicia —en sus sentencias de 17 de junio del 2021, C-597/19, y de 4 de julio del 2023, C-252/21— ha declarado que «esta disposición establece tres requisitos acumulativos para que el tratamiento de datos personales resulte lícito, a saber, en primer lugar, que el responsable del tratamiento o el tercero persigan un interés legítimo; en segundo lugar, que el tratamiento de los datos personales sea necesario para la satisfacción de ese interés legítimo, y, en tercer lugar, que no prevalezcan los intereses o los derechos y libertades fundamentales del interesado en la protección de los datos sobre el interés legítimo del responsable del tratamiento o de un tercero».

Pues bien, en el caso concreto, el Tribunal Superior Regional de Colonia considera que concurren todos estos requisitos que permiten el tratamiento de los datos personales para entrenar la inteligencia artificial:

- a) Por lo que respecta al interés legítimo, el tribunal alemán recuerda que el Tribunal de Justicia ha declarado que tienen tal condición de legítimos, entre otros, los intereses económicos (Sentencia del Tribunal de Justicia de 4 de octubre del 2024, C-621/22). Y, en

este caso, el interés reside en la intención de utilizar las capacidades de la inteligencia artificial generativa para proporcionar un asistente de conversación que, adaptándose a las costumbres regionales concretas, ofrezca, por ejemplo, respuestas en tiempo real para chats, asistencia para organizar y planificar vacaciones o incluso asistencia con la redacción de textos.

- b) A propósito de la exigencia de que el tratamiento de los datos personales sea necesario para la satisfacción del interés legítimo que se persigue, el Tribunal de Justicia —Sentencia de 9 de enero del 2025, C-394/23— ha declarado que el tribunal nacional correspondiente debe comprobar «si tal interés legítimo del tratamiento de los datos puede alcanzarse razonablemente de manera tan eficaz por otros medios menos atentatorios respecto de las libertades y los derechos fundamentales de los interesados, puesto que tal tratamiento debe efectuarse sin sobrepasar los límites de lo estrictamente necesario para la satisfacción de ese interés legítimo». Y, de igual modo, el requisito relativo a la necesidad del tratamiento debe examinarse en relación con el principio de minimización de datos, consagrado en el artículo 5.1c del Reglamento General de protección de datos, conforme al cual los datos personales deben ser adecuados, pertinentes y limitados a lo necesario en relación con los fines para los

que son tratados (STJUE de 12 de septiembre del 2024, C-17/22 y C-18/22).

Sobre esta base, el tribunal alemán considera, en el juicio sumario propio de los procedimientos de medidas cautelares, que el tratamiento de los datos personales incluidos en las redes sociales es necesario para el entrenamiento de la inteligencia artificial y respeta el referido principio de minimización de los datos personales. A su juicio, no existen opciones menos invasivas que sean adecuadas para satisfacer adecuadamente el interés legítimo perseguido, considerando que la anonimización de los datos sería impracticable (*unpraktikabel*) y que el uso de los llamados *Flywheel-Daten* (referentes al comportamiento del usuario, visitas, clics, etc.) reduciría drásticamente el volumen de los datos utilizables y no se ajustaría a la necesidad de entrenar los modelos generativos de inteligencia artificial con «grandes cantidades de texto, imágenes, videos y otros datos» (necesidad reconocida explícitamente en el considerando 105 del Reglamento sobre inteligencia artificial).

- c) Finalmente, el Tribunal Superior Regional de Colonia también considera cumplido el requisito de que no prevalezcan los intereses o los derechos y libertades fundamentales del interesado en la protección de los datos sobre el interés legítimo del responsable del tratamiento o de un

tercero. Se basa para ello, entre otros argumentos, en el hecho de que los interesados podían esperar razonablemente dicho tratamiento al recopilar los datos;

El tribunal alemán considera que el tratamiento de los datos es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento

en que sólo se usan para el entrenamiento de la inteligencia artificial los datos de las redes que figuran en abierto o con carácter público (y que, por tanto, también se puede localizar a través de motores de búsqueda), y en el hecho de que el demandado ha puesto en práctica numerosas de las medidas recomendadas por el Comité Europeo de Protección de Datos —en su ya citado Dictamen 28/2024, apartado 96 y siguientes— para mitigar la interferencia con los derechos de los interesados.

Entre esas medidas que se adoptaron en el caso concreto se encuentra la supresión de determinados datos (eliminación de nombres completos, direcciones de correo electrónico, números de teléfono, números de identificación nacional, ID de usuario, números de tarjetas de crédito/débito, números de cuenta bancaria, números de matrícula de vehículos, direcciones IP y direcciones

postales) y su recopilación de forma no estructurada. El tribunal es consciente de que, pese a todo, esto no implica una anonimización de los datos, pues se mantienen las caras de personas identificables en la fotografías, pero el riesgo de identificación se reduce considerablemente con la supresión de los referidos datos. Y también es relevante la medida consistente en la posibilidad de que los

usuarios eviten la inclusión de sus datos en el conjunto de datos de entrenamiento, eliminado el carácter público de su cuenta o de sus publicaciones en las redes sociales, incluso con carácter retroactivo.

- 3.3. Pese a todo lo anterior, en las cuentas de las redes sociales implicadas en el caso, los usuarios introducen a veces datos personales de terceras personas, como fotografías en las que aparecen esos terceros. Y esas otras personas no pueden modificar el carácter público de las cuentas ni hacer uso de la posibilidad, ofrecida por el titular de las redes a los titulares de las cuentas, de oponerse al tratamiento de los datos con fines de entrenamiento de la inteligencia artificial.

No obstante, el Tribunal Superior Regional de Colonia considera que esta situación no afecta a la licitud del tratamiento de los datos con base en el artículo 6.1f del Reglamento General de protección de datos. A su juicio, no prevalecen los intereses de estos

terceros porque el eventual perjuicio para ellos se produciría si el sistema de inteligencia artificial pudiese llegar a ofrecer como resultado, ante una determinada pregunta, datos personales de esos terceros. Y, según el tribunal alemán, ese riesgo es muy reducido, pues el sistema de inteligencia artificial no se entrena repetidamente sobre el conjunto de datos en cuestión, lo que hace que exista una alta probabilidad de que la información sobre un individuo concreto se pierda en la masa ingente de datos y, en consecuencia, no aparezca inalterada dicha información

en un resultado o *output* ofrecido por el sistema de inteligencia artificial.

4. Conclusión

Estamos, como se puede comprobar, ante una resolución judicial importante que afronta un tema de máxima importancia y en la que se facilita la utilización de los datos personales para el entrenamiento de los sistemas de inteligencia artificial. Habrá que estar atento, en todo caso, al desarrollo posterior del asunto, toda vez que la resolución se adopta en el marco de la solicitud de medidas cautelares.