POSITION PAPER

ESBG response to the **EDPB Call for Feedback on Web Scrapping in the context of Generative AI Guidelines**

ESBG (European Savings and Retail Banking Group)

Rue Marie-Thérèse, 11 - B-1000 Brussels

October 2026

ESBG

The European Savings and Retail Banking Group welcomes the European Data Protection Board's (EDPB) efforts to provide greater legal certainty through its draft Guidelines on Web Scraping in the Context of Generative AI. These Guidelines address increasingly important questions arising from the growing use of data-driven technologies, AI systems, and data-sharing ecosystems. At the same time, ESBG believes that several aspects of the draft Guidelines would benefit from further clarification to ensure consistent application and practical implementation. In particular, the draft Guidelines leave a number of key legal, operational, and accountability issues insufficiently addressed, including for example the application of data subjects' rights, purpose limitation, international data transfers, AI-specific risks, and documentation requirements. Greater guidance in these areas would strengthen legal certainty for controllers while maintaining a high level of protection for individuals. The following identify the main shortcomings and gaps that ESBG believes should be addressed to provide effective and efficient guidance on web scraping in the context of generative AI.

03/2026 Guidelines on web scraping in the context of generative AI

I. Additional Clarification

Inadequate Treatment of the DPIA Requirement (Article 35 GDPR)

The DPIA is mentioned only in passing as a possible "appropriate measure" for protecting data subjects' rights alongside anonymisation, pseudonymisation, and security measures. This underplays its significance considerably. Web scraping for AI training almost certainly meets several of the Article 35(3) GDPR criteria that trigger a mandatory DPIA, in particular: large-scale processing of special categories of data (Article 35(3)(b)) and systematic monitoring of publicly accessible areas (Article 35(3)(c)). The guidelines should have included a dedicated analysis concluding that a DPIA will typically be mandatory and setting out what such a DPIA should cover in this specific context.

Lack of Thresholds in the Balancing test

The guidelines acknowledge that where data subjects' interests override the legitimate interests being pursued and sufficient mitigating measures are not taken, the processing cannot be based on Article 6(1)(f) GDPR, but they do not articulate any concrete threshold or scenario in which the balance would definitively tip against the controller regardless of mitigating measures. The list of mitigating measures is presented as open-ended and almost always available, which risks creating the misleading impression that legitimate interest can always be rescued by adopting some measure or another. Clearer guidance on when the scale of processing or the sensitivity of data makes Article 6(1)(f) unavailable in principle, irrespective of mitigating measures, would strengthen the legal clarity of the guidelines considerably.

Accountability Principle (Article 5(2) GDPR) Acknowledged but Not Addressed

The guidelines acknowledge that it is not obvious how to meet the principle of accountability laid down in Article 5(2) GDPR when processing large datasets for AI development, while simultaneously stating that compliance must be met by the controller. However, having identified this difficulty, the guidelines provide no concrete guidance on how controllers should document and demonstrate their compliance — for example, the requirements of Article 30 GDPR (records of processing activities), internal data protection policies, audit trails of scraping operations, or documentation of the balancing test. Simply acknowledging the problem without providing solutions is a shortcoming.

Incomplete Coverage of the Data Broker Scenario

The guidelines explicitly state that they will not address the role of data brokers who scrape data and hold datasets available for other organisations without training AI models themselves. This is an important exclusion that leaves a significant compliance gap. In practice, many AI developers obtain pre-scraped datasets from third parties, and the conditions under which they may lawfully re-use such data, including chain-of-compliance questions, the applicability of Article 14 to downstream controllers, and liability for original scraping conducted without a lawful basis, are not addressed. The responsibility allocation section between scrapers and AI developers (which distinguishes between separate and joint processing activities) does not provide sufficient certainty and clarification without guidance on due diligence obligations for re-used datasets.

Absence of Purpose Limitation Analysis Specific to the Scraping Context

The guidelines address purpose limitation by cross-referencing EDPB Opinion 28/2024 without engaging substantively with the specific purpose limitation issues that arise in the web scraping context. Despite it being common industry practice, the Guidelines lack any analysis of whether personal data scraped for training one AI model can lawfully be re-used for training a different model (or fine-tuning), re-used by a different controller, or used for purposes beyond AI training (such as analytics or targeted advertising).

II. Gaps in the addressed GDPR topics

Data Subjects' Rights (Articles 15–22 GDPR)

The document acknowledges that compliance with GDPR requirements is challenging when web scraping is used to collect personal data, yet it provides no dedicated section on how controllers should operationalise data subjects' individual rights, most critically:

- The right to erasure ("right to be forgotten", Article 17 GDPR): The guidelines acknowledge that once the model is trained, personal data cannot easily be deleted from a model, yet they offer no substantive guidance on how a controller should handle an erasure request in this circumstance. This is a glaring omission given that it is one of the most practically significant tensions in AI development.

- The right of access (Article 15 GDPR): No guidance is provided on how a controller can respond to access requests when it may have limited knowledge of which data subjects' information is in the dataset.
- The right to object (Article 21 GDPR): While opt-out mechanisms are mentioned as a mitigating measure, the guidelines do not engage with the standalone and unconditional right to object to processing based on legitimate interest. The guidelines refer to "a discretionary and prior right to object" and creating an opt-out list, but this characterisation as "discretionary" understates the mandatory nature of Article 21. The right to object is not optional for the controller, upon a valid objection, the controller must cease processing unless it can demonstrate compelling legitimate grounds that override the data subject's interests.

Fairness Principle (Article 5(1)(a) GDPR)

The guidelines indeed identify fairness as a relevant principle but provide no dedicated substantive analysis of what fairness means in the context of web scraping for AI training. Web scraping inherently involves processing data in ways that are unknown to data subjects and that may reflect or amplify societal biases. The fairness dimension, i.e. whether the processing produces effects that are unjust, discriminatory, or contrary to reasonable expectations, warranted its own section alongside the other principles. The guidelines address transparency and data minimisation in depth but leave fairness almost entirely unexplored.

Storage Limitation Principle (Article 5(1)(e) GDPR)

The guidelines list several GDPR principles as relevant, including purpose limitation, data minimisation, fairness, and accuracy, but make no mention of storage limitation. This is a significant gap, web scraping for AI training involves retaining large volumes of personal data over extended periods (during data collection, cleaning, structuring, model training, and fine-tuning), yet the guidelines provide no analysis of how long scraped personal data may lawfully be retained, what retention schedules should look like, or when deletion obligations arise at the various stages of the AI development pipeline.

Criminal Convictions and Offences Data (Article 10 GDPR)

The guidelines briefly mention data relating to criminal convictions and offences as enjoying special protection under Article 10 GDPR in the context of the balancing test, but provide no substantive guidance on how Article 10 applies in the web scraping context. Court judgments, news reports of criminal proceedings, and publicly available criminal records are precisely the kind of data that may be scraped at scale, yet the guidelines provide no analysis equivalent to what they provide for Article 9 special categories.

Data Protection by Design and by Default (Article 25 GDPR)

The guidelines contain no reference to Article 25 GDPR, which requires controllers to implement data protection by design and by default. Since

untargeted scraping can result in the exploration of a large portion of the internet and increases the risk that the controller has limited knowledge of what personal data are collected and processed, guidance on embedding privacy-protective features into the architecture of the scraping process from the outset, rather than retrospectively applying filters, is essential.

Security Obligations (Article 32 GDPR)

Security is mentioned only briefly as one mitigating measure among many. Implementing technical and organisational measures to ensure a high level of security is listed as an "appropriate measure" in the context of transparency, but there is no dedicated analysis of Article 32 GDPR obligations. Given the volume and sensitivity of the data involved, and the specific risks of data breaches, model inversion attacks, and membership inference attacks, a substantive discussion of security requirements was warranted. The guidelines acknowledge that AI models could be attacked and that the information memorised by the model could be used for arbitrary purposes, yet do not translate this acknowledgement into concrete Article 32 guidance.

International Data Transfers (Chapter V GDPR)

The guidelines do not provide any information on international data transfers. This is a material gap for several reasons. Web scraping frequently involves collecting data from servers outside the EEA, scraping is often conducted by non-EEA entities targeting EEA data subjects, and AI training infrastructure (cloud computing, GPU clusters) is commonly hosted outside the EEA. Despite being directly triggered by the typical operational realities of large-scale AI development, the requirements of Chapter V GDPR (adequacy decisions, Standard Contractual Clauses, Binding Corporate Rules, and transfer impact assessments) are not mentioned.

About ESBG (European Savings and Retail Banking Group)

ESBG represents the locally focused European banking sector, helping 32 savings and retail banks in 27 European countries strengthen their unique approach that focuses on providing service to local communities and boosting SMEs. Advocating for a proportionate approach to banking rules, ESBG unites at EU level some 859 banks, which together employ 620,000 people driven to innovate at 37,000 branches. ESBG members have total assets of € 6,35 trillion, provide € 3,72 trillion in loans to customers, and serve 163 million Europeans seeking retail banking services. ESBG members commit to further unleash the promise of sustainable and responsible 21st century banking.

Our transparency ID is 8765978796-80.

European Savings and Retail Banking Group – aisbl
Rue Marie-Thérèse, 11 ▪ B-1000 Brussels ▪ Tel : +32 2 211 11 11 ▪ Fax : +32 2 211 11 99
Info@wsbi-esbg.org ▪ www.wsbi-esbg.org

Published by ESBG. April 2026.