Omio response to the EDPB public consultation on Guidelines 03/2026 on web scraping in the context of generative AI

Omio welcomes the opportunity to comment on the draft Guidelines 03/2026 on web scraping in the context of generative AI.

As a digital travel platform operating across Europe, Omio supports a high standard of protection for individuals' personal data. We particularly welcome the Guidelines' emphasis on purpose limitation, data minimisation, transparency and safeguards against the memorisation or reproduction of personal data by generative AI systems.

The final Guidelines should also maintain a proportionate, risk-based approach. This is important to preserve legitimate uses of factual, non-personal information, including public mobility information, while ensuring that the collection and reuse of personal data for AI development is properly controlled.

1. Maintain a clear distinction between personal data and factual public information

The Guidelines should clearly distinguish between scraping that involves personal data and the collection of factual, non-personal information.

In the mobility sector, public information such as routes, timetables, availability and service characteristics can be essential to helping travellers compare and plan journeys. Such information should not be treated as personal data merely because it is collected through automated means or is available on a website that may also contain personal data.

We therefore welcome the statement that the Guidelines address web scraping to the extent that it entails the processing of personal data. The final text should retain this distinction throughout, including when discussing source selection, transparency and technical safeguards. This would avoid uncertainty for organisations that use targeted collection of non-personal information for legitimate commercial or consumer-facing purposes.

The Guidelines should also make clear that they do not determine questions of copyright, database rights, contract or unauthorised access. Those issues may be relevant to a particular activity, but should be assessed under the applicable legal frameworks rather than through GDPR guidance.

2. Support targeted, proportionate collection over indiscriminate crawling

Omio supports the distinction between targeted scraping and unrestricted web crawling. The latter is more likely to result in indiscriminate collection of personal data, including sensitive data, and to create significant challenges for accountability and data minimisation.

The final Guidelines should encourage organisations to define their purpose and collection criteria in advance, identify the categories of data genuinely required, and apply measures to avoid collecting unnecessary personal data. This is preferable to broad collection followed by retrospective attempts to identify and remove personal data.

At the same time, the Guidelines should recognise that a large volume of data does not, in itself, mean that processing is disproportionate. The relevant question is whether the categories of personal data collected are necessary and appropriately limited in relation to a clearly defined purpose. This is particularly important for sectors in which reliable and current factual information is distributed across multiple public sources.

3. Give meaningful weight to website restrictions on AI scraping

Omio strongly supports the draft Guidelines' recognition that website-level restrictions, including authentication requirements, CAPTCHAs, robots.txt and ai.txt files, are relevant when assessing data subjects' reasonable expectations and the fairness of scraping.

These measures are an important way for website operators to communicate that automated collection and reuse of content, including for generative-AI training, is not permitted. Where a website has clearly expressed this position and implemented technical measures to support it, a third party that disregards those measures should face a significantly more demanding assessment of necessity, fairness and legitimate interests.

The final Guidelines could usefully clarify that the evidential weight of such measures will depend on the circumstances. Authentication and CAPTCHAs may provide particularly clear evidence of restricted access. Machine-readable signals, such as robots.txt and ai.txt files, should also be treated as relevant contextual indicators, especially when supported by clear notices or terms of use.

This clarification would help protect individuals whose personal data may be accessible through online services, while providing greater certainty to website operators seeking to set clear rules for the reuse of material available through their services.

4. Transparency requirements should be effective and workable

Omio welcomes the draft Guidelines' emphasis on transparency where individual notice is impossible or would involve disproportionate effort. A public notice explaining the purpose of scraping, legal basis, categories of personal data, source categories and available rights is an important safeguard.

We support the recommendation that controllers provide information about source domains, collection periods and, where relevant, the characteristics of the crawler. However, the final Guidelines should preserve flexibility in how this information is provided. Requiring a complete, continuously updated list of every URL collected could be impracticable for large-scale, dynamic datasets and may not meaningfully assist individuals.

The appropriate standard should be that controllers provide information that is sufficiently specific, accessible and searchable to allow individuals to understand whether their data may have been collected and to exercise their rights effectively. Domain-level information, clear descriptions of source categories, collection periods and a workable contact mechanism will often be more useful than exhaustive technical disclosure.

5. Safeguards should be assessed in a technology-neutral and outcome-focused way

The draft Guidelines rightly identify technical and organisational safeguards that can reduce the impact of web scraping, including source exclusion, data filtering, pseudonymisation, deletion of unnecessary data, deduplication and measures to reduce memorisation and regurgitation.

The final Guidelines should retain an outcome-focused approach. Controllers should be required to demonstrate that the measures selected are effective and proportionate to the risks of the specific processing, rather than being expected to deploy a fixed set of technologies in every case.

This is especially important in relation to special categories of data. Omio agrees that organisations should not intentionally collect special-category data for AI training without a valid Article 9 GDPR condition. Where such data are inadvertently collected despite appropriate preventive measures, the Guidelines should focus on rapid deletion, effective output controls and regular testing of the safeguards applied.

Conclusion

Omio supports the EDPB's objective of ensuring that generative-AI development does not rely on indiscriminate collection and reuse of personal data. The final Guidelines

should provide robust protection for individuals while preserving a clear, proportionate distinction between personal data processing and legitimate uses of factual, non-personal information.

We encourage the EDPB to retain the risk-based approach in the draft and to clarify the importance of website-level restrictions on AI scraping. This would strengthen individuals' control over their data, provide more certainty for digital services, and support responsible innovation in Europe.