VILNIAUS STR. 31
01402 VILNIUS
LITHUANIA

T. +370 5 231 2211
E. INFO@TRINITI.LT
WWW.TRINITI.LT

E. DELIVERY BOX
ADDRESS
306233023

TRINITI
TALLINN RIGA VILNIUS

To the European Data Protection Board

REGARDING COMMENTS ON GUIDELINES 03/2026 ON WEB SCRAPING IN THE CONTEXT OF GENERATIVE AI

Lithuania, 2026-10-30

Following the publication by the European Data Protection Board (hereinafter referred to as the EDPB) of the Guidelines 03/2026 on web scraping in the context of generative AI (hereinafter referred to as the Guidelines), we would like to express our support for the EDPB's efforts to provide greater legal certainty regarding the application of the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (hereinafter referred to as the GDPR) to web scraping in the context of generative AI. We believe that the Guidelines provide useful clarification on a number of complex issues, including data minimisation, transparency, legitimate interests and the processing of special categories of personal data.

At the same time, based on our practical experience advising organisations of different sizes on compliance with GDPR, including organisations developing and deploying AI-based solutions, we would like to submit the following observations and recommendations for consideration. Attention is given to the practical implementation of the Guidelines by small and medium-sized enterprises (SMEs).

1. A CLEARER PROPORTIONALITY AND RISK-BASED APPROACH FOR SMES IS NEEDED

Several measures described in the Guidelines may require significant technical, organisational and financial resources, including detailed data mapping, source-level filtering, continuous monitoring of scraped sources, validation of datasets, mechanisms facilitating the exercise of data subject rights, and measures preventing memorisation or extraction of personal data from AI models. While these measures may be readily available to large AI developers, their cumulative implementation may be particularly challenging for SMEs, which frequently rely on third-party tools and infrastructure and may not have dedicated privacy, AI governance or engineering teams. We would therefore suggest making clearer that the implementation of safeguards should follow a risk-based and proportionate approach, taking into account, inter alia, the scale and purpose of the processing, the categories of data concerned, the risks to individuals, the technical possibilities available to the controller and the costs of implementation. We would also encourage the EDPB to include at least one practical example showing how an SME developing or fine-tuning a limited-purpose AI system could implement the Guidelines proportionately.

2. SCOPE OF PRELIMINARY DATA MAPPING AND FILTERING

Paragraphs 36–38 recommend defining collection criteria, undertaking preliminary data mapping and applying preventive filtering measures. In practice, however, it may be impossible to identify all personal data before collection, particularly where websites contain dynamic or user-generated content. For example, an SME collecting technical documentation and support discussions may be able to identify relevant websites and exclude certain categories of sources, but it cannot reasonably predict every individual contribution or every item of personal data that may appear within those sources.

The Guidelines themselves recognise that controllers may have limited knowledge of the precise personal data contained in scraped datasets. We would therefore suggest clarifying that preliminary mapping may, where appropriate, be performed at the level of website categories, sources, expected data types or

content categories, rather than requiring an exhaustive ex ante assessment of every individual webpage or data item. The Guidelines could also expressly recognise that preventive measures may be combined with filtering during and after collection, where complete pre-collection filtering is technically impracticable.

Similarly, where the Guidelines recommend maintaining information on individual domains and URLs of scraped webpages in a searchable format, the EDPB could clarify that source-, domain- or dataset-level records may be sufficient where maintaining individual URL-level records would involve disproportionate technical or administrative effort. This is particularly relevant for SMEs relying on third-party scraping tools or infrastructure which may not support granular pre-collection filtering. In such cases, equivalent safeguards, including post-collection filtering, deletion before training or targeted review of higher-risk sources, should be expressly recognised.

3. EFFECTIVENESS OF SAFEGUARDS CONCERNING INCIDENTAL PROCESSING OF SPECIAL CATEGORIES OF PERSONAL DATA

Paragraphs 68 and 73–77 address situations in which special categories of personal data may be incidentally collected in the course of web scraping. This issue is particularly important because, in practice, even carefully designed scraping processes may incidentally collect special category data that could not reasonably have been identified in advance. For example, an SME developing a specialised language model using technical forum discussions may unintentionally collect references to identifiable individuals' health conditions. The Guidelines appropriately emphasise the need for technical and organisational safeguards; however, even with appropriate filtering, deletion and model testing, it may not always be technically possible to guarantee that special category data will never be collected, memorised or subsequently extracted. We would therefore suggest that the EDPB clarify what level of effectiveness, testing and documentation is considered sufficient where the risk cannot be eliminated. In particular, it would be useful to explain whether compliance should be assessed on the basis of the reasonable effectiveness of safeguards, taking into account the state of the art, rather than on the basis of an expectation of absolute prevention.

Further clarification would also be useful regarding the allocation of responsibilities where different organisations are involved in the AI lifecycle, particularly between the developer of the model and an independent deployer. Such clarification should, of course, not imply that incidental processing is automatically exempt from Article 9 GDPR, but it would provide greater legal certainty as to what controllers are expected to demonstrate in practice.

4. DATA ACCURACY AND VALIDATION REQUIREMENTS

The Guidelines recommend scraping from reliable sources, timestamping data and validating data before using them for AI training. We support this approach; however, further clarification would be useful regarding the expected level of validation. For large datasets, verifying the factual accuracy of individual data points would often be technically and economically impossible, particularly for SMEs. The level of validation should therefore depend on the intended use of the AI system and the potential impact of inaccurate personal data on individuals. We would suggest expressly recognising risk-based validation methods, including statistical sampling, automated anomaly detection and targeted validation of higher-risk data.

5. ASSESSMENT OF ALTERNATIVES UNDER ARTICLE 6(1)(F) GDPR

The Guidelines indicate that, when assessing necessity under Article 6(1)(f) GDPR, controllers should consider whether less intrusive alternatives may achieve the same objective, including narrower datasets, synthetic data or other technical approaches. For SMEs developing new AI solutions, however, it may not always be possible to determine in advance whether such alternatives will provide an equivalent level of

effectiveness. In many cases, this can only be established through experimentation. We would therefore suggest clarifying that controllers should not be required to demonstrate that every conceivable alternative is ineffective. A documented assessment of reasonably available and technically viable alternatives, taking into account the state of the art and the specific use case, should be sufficient.

6. EXERCISE OF DATA SUBJECT RIGHTS AND OPT-OUT MECHANISMS

The Guidelines encourage measures facilitating the exercise of individuals' rights, including mechanisms allowing individuals to object to the collection of their data and the creation of opt-out lists. We support measures that give data subjects meaningful control over their data; however, building sophisticated individual-level suppression mechanisms may be technically complex for SMEs, particularly where there is no prior relationship with the data subject. The EDPB could therefore provide examples of proportionate mechanisms, such as a centralised online objection mechanism, source-specific exclusion requests or recognised machine-readable signals. The Guidelines should also clarify the extent to which such mechanisms are expected from controllers whose datasets contain large amounts of unstructured web content and where identification of a specific individual within the dataset may itself require disproportionate effort.

7. GREATER CLARITY ON ACCEPTABLE SAFEGUARDS RATHER THAN ABSOLUTE OUTCOMES

More generally, the Guidelines contain a number of references to measures intended to prevent collection, memorisation, extraction or dissemination of personal data. In practice, AI systems cannot always provide an absolute technical guarantee that such outcomes will never occur. Requiring an absolute outcome could create considerable uncertainty and may disproportionately affect smaller market participants. We would therefore encourage the EDPB to clarify throughout the Guidelines that the relevant standard should be the implementation of appropriate, effective and proportionate technical and organisational measures, assessed in light of the risks, the state of the art and the circumstances of the specific processing. Such clarification would preserve a high level of protection for data subjects while providing controllers, particularly SMEs, with a more practicable and legally predictable compliance framework.

Best regards,

APB TRINITI

Lithuania