

Berlin, 20. März 2026

Stellungnahme

der Gesellschaft für Informatik e.V. (GI)

zu Fragen der EU-Kommission zu Wirksamkeit  
und Auswirkungen der DSM-Richtlinie in den  
Mitgliedstaaten



Das Bundesministerium der Justiz und für Verbraucherschutz (BMJV) führt eine öffentliche Konsultation im Zusammenhang mit der durch die EU-Kommission vorzunehmende Evaluierung der EU-Richtlinie über das Urheberrecht im digitalen Binnenmarkt (Richtlinie (EU) 2019/790 – „DSM- Richtlinie“) durch. Die Gesellschaft für Informatik e.V. (GI) nimmt zum Aspekt des Text und Data Mining (TDM) zum Zwecke der wissenschaftlichen Forschung (Artikel 3) wie folgt Stellung.

*Frage 4: Does the national implementation of the TDM exception introduced in Article 3 facilitate cross-border research collaboration, and partnerships between research organisations and industry? If any, please indicate any legal or practical barriers encountered in this context.*

**Antwort:**

The transposition of Article 3 of the DSM Directive into German law is generally suitable for enabling the computer-aided analysis of large datasets. With regard to cross-border research initiatives however, we observe issues due to varying national interpretations of the European legal framework, which are slowing down computer science research.

The problem does not lie within the German norm itself, but rather in the lack of harmonization regarding the sharing of research materials. In some member states, it is not possible to share datasets, corpora, or interim results with partners. This divergence in the implementation of Article 3 creates legal uncertainty in cross-border research and prevents the formation of European research consortia. As long as the exchange of research data across national borders fails due to differing national statutory exceptions, the scientific added value of Article 3 will remain below its full potential.

*Frage 5: In your view, to what extent has the TDM exception introduced in Article 3 been effective in your Member State in improving the ability of research organisations and cultural heritage institutions to carry out text and data mining for scientific research, including in the context of AI development? In your response, please consider whether the implementation has increased legal certainty for eligible institutions, improved practical access to content for TDM, reduced technical or contractual obstacles to conducting TDM, and led to any increase or change in TDM activities.*

**Antwort:**

While the transposition of Article 3 of the DSM Directive into German law has theoretically bolstered the framework for TDM, the practical implementation within the R&D sector is hampered by systemic legal fragmentation. For research organizations, the delta between legislative intent and operational reality remains significant,

particularly concerning the development and dissemination of Large Language Models (LLMs) and other foundational AI architectures.

From a computational perspective, the efficacy of Large Language Models (LLMs) and foundational AI is strictly dependent on the volume and diversity of the underlying training data. Data scraping is not merely an auxiliary process; it is the primary ingestion layer for modern AI pipelines. Currently, the distinction between “lawful access” and “technical accessibility” creates a bottleneck. For research to be reproducible (a core tenant of Open Science), scientists require stable, high-throughput access to web-scale data. Furthermore, the technical implementation of machine-readable opt-outs (Article 4 DSM) remains poorly standardized. This creates high-latency pre-processing requirements for researchers who must filter massive datasets to ensure compliance, often at the cost of data representative-ness.

The European Commission’s commitment to Open Science – specifically the principle of “as open as possible, as closed as necessary” – is currently in direct conflict with the restrictive interpretation of TDM exceptions. Under programs like Horizon Europe, researchers are mandated to provide open access to their outputs. However, the risk of copyright infringement during the ingestion of scraped data often prevents the public release of the resulting “Open Weights” or datasets. Open Science requires the sharing of not just results, but the methodologies and data subsets used. Without a harmonized sharing right across the EU, the technical interoperability of research data remains fragmented.

To bridge the gap between AI engineering and European law, we propose the following:

- **Standardization of Technical Exceptions:** The legislator should provide technical guidelines on how scraping for research purposes can bypass anti-scraping protocols when the objective is non-commercial scientific analysis.
- **The Open Science Safe Harbor:** Explicitly protect researchers who publish model parameters and training methodologies in accordance with Open Science mandates. If a researcher scrapes data for a public-interest project, the subsequent Model Release should be protected from derivative-work claims.
- **Clarification on “Training as TDM”:** Confirm that the transient technical copies created during a scraping/crawling process for AI training fall squarely under the Article 3 *lex specialis*, regardless of whether the final model is capable of generating similar content to the input.



GESELLSCHAFT  
FÜR INFORMATIK

## **Kontakt**

Nikolas Becker  
Leiter Politik & Wissenschaft  
Gesellschaft für Informatik e.V.  
E-Mail: [Nikolas.Becker@gi.de](mailto:Nikolas.Becker@gi.de)

## **Über die Gesellschaft für Informatik e.V. (GI)**

Die Gesellschaft für Informatik e.V. (GI) ist die größte Fachgesellschaft für Informatik im deutschsprachigen Raum. Seit 1969 vertritt sie die Interessen der Informatikerinnen und Informatiker in Wissenschaft, Gesellschaft und Politik und setzt sich für eine gemeinwohlorientierte Digitalisierung ein. Mit 14 Fachbereichen, über 30 aktiven Regionalgruppen und unzähligen Fachgruppen ist die GI Plattform und Sprachrohr für alle Disziplinen in der Informatik. Weitere Informationen finden Sie unter [www.gi.de](http://www.gi.de).