

Estonian Ministry of Justice and Digital Affairs
Liisa-Ly Pakosta
Minister of Justice and Digital Affairs
Suur-Ameerika 1, Tallinn 10122
Estonia

Open data of the Estonian language and META

Brussels, 10 February 2025

Dear Madam Minister Liisa-Ly Pakosta,

News Media Europe took note of the Estonian Ministry of Justice and Digital Affairs' decision to provide Meta, the parent company of Facebook, Instagram and WhatsApp, with the open data of the Estonian language corpus¹, containing almost 4 billion words. We urge the Ministry to reconsider its decision, or at least suspend it until the Ministry has agreed with rightsholders, including news publishers, on the appropriate framework for the use of their content by Meta under this deal.

News Media Europe recalls that any media content – including the content archives of publications – is an important asset for publishers and that copyright and related rights do apply. While we wholeheartedly subscribe to the notion that the sustainability of the Estonian language and culture needs to be safeguarded in the development of large language models, the reality of the relevant legal frameworks and the protections offered therein to rightsholders, cannot and should not be ignored.

Doing so undermines the legal rights and the financial sustainability of editorial media, who contribute uniquely to the security of Estonia and the EU by building information resilience against foreign interference and manipulation, while underpinning the democratic processes.

In connection with this, we ask for information:

- What data does the "Open Data of the Estonian Language Corpus" contain, and more specifically, does it contain copyrighted content of media companies (including members of the Estonian Association of Media Companies), to what extent and in what form?
- On what legal basis (incl. in the sense of AutÕS and IKS) was the content in the above-mentioned dataset transferred to Meta (or, if this transfer has not yet taken place, on what legal basis is it planned/intended to transfer this content)?

¹ <https://www.justdigi.ee/en/news/meta-interested-using-open-data-estonian-language-corpus>

NEWS MEDIA EUROPE

- Under what conditions was it agreed/planned to agree with Meta on the use of language corpus content and what safeguards were taken by the Ministry and/or other parties to the agreement to prevent unlawful use of data (incl. in application outputs)?
- What steps has the Ministry taken to protect the rights of Estonian authors and owners of related rights in the content of the database described above, to obtain their permission and to pay them fair remuneration? If the content of media companies was transferred/given to Meta (and another developer of a large language model) without permission and free of charge, how is it intended to remedy the damage caused by this offence?

Thank you, Minister, for your kind consideration, and I remain at your disposal for any enquiries you might have.

Yours sincerely,

A handwritten signature in black ink on a light background. The signature starts with a large 'W' followed by a series of connected loops and a long horizontal stroke that tapers off to the right.

Wout van Wijk
Executive Director

News Media Europe
Square de Meeus 35
1000 Brussels
Belgium
wout.vanwijk@newsmediaeurope.eu
+32 473685864