



When do LLMs process personal data?

On 15 July 2024, the Hamburg Commissioner for Data protection and freedom of information (the HCDP) released a discussion paper on Large Language Models (LLMs) and Personal Data (the Discussion Paper). The HCDP distinguishes between storage, training and use of a LLM and provide practical examples on when personal data protection should be ensured.

STORAGE OF PERSONAL DATA

The HCDP considers that a LLM, in view of the way it functions and CJEU case law, does not store personal data. The HCDP explains that LLMs do not reproduce words or full text, but fragment words into small pieces (tokens) and create relationships between those tokens. Accordingly, the HCDP is of the view that tokens do not constitute personal data and do not contain elements related to an identified or identifiable individual as per CJEU case law. Additionally, LLMs never fully reproduce input data provided: they generate new content based on the tokens and relationships established.

Based on those considerations, the HCDP states that the storage of LLMs models does not constitute a processing of personal data under GDPR. This means that data subjects cannot exercise their rights against LLMs, but should exercise them against the input or output, in particular if a LLM is integrated into another AI system.

TRAINING AND USE OF LLM

The training of a LLM on personal data however seems to constitute processing under GDPR as per the HCDP. In this case, GDPR will apply to the training. But the HCDP provides an interesting consideration for companies integrating third parties LLMs into an AI system: those users of the LLM will not be liable for a breach of the LLM during the training phase. Only the developers of the LLM are responsible for GDPR breaches occurring during the training. The users will nonetheless have to clarify responsibilities with the developers/providers, make sure data subjects can exercise their rights towards input and output and take protection measures against attacks and unlawful extraction.

The HCDP also provides some requirements for fine tuning and simple use of third parties LLMs. With respect to fine tuning of a LLM, the entity fine tuning the LLM should ensure no or as little personal data as possible is used. If personal data are used, fine tuning requires a legal basis and data subjects should be able to exercise their rights towards the entity fine tuning the model.

For the use of a third party LLM, beyond ensuring that data subjects can exercise their rights regarding the input and output, the users should implement the safeguards against extraction and attacks provided by the developer/provider and take their own additional measures.

CONCLUSION

The Discussion Paper seems to provide a pragmatic approach for companies using LLMs. It remains to be seen if this position will be followed by other data protection authorities or even courts in Europe.

The Discussion Paper is available [here](#) in English.

Apollolaan 15, 1077 AB Amsterdam, PO Box 75440, Amsterdam, 1070 Ak, The Netherlands | Tel +31 (0)20 674 1000 | Fax +31 (0)20 674 1111

A&O Shearman is an international legal practice with nearly 4,000 lawyers, including some 800 partners, working in 29 countries worldwide. A current list of A&O Shearman offices is available at aoshearman.com/en/global-coverage.

A&O Shearman means Allen Overy Shearman Sterling LLP and/or its affiliated undertakings. Allen Overy Shearman Sterling LLP is a limited liability partnership registered in England and Wales with registered number OC306763. Allen Overy Shearman Sterling LLP is authorised and regulated by the Solicitors Regulation Authority of England and Wales (SRA number 401323).

The term partner is used to refer to a member of Allen Overy Shearman Sterling LLP or an employee or consultant with equivalent standing and qualifications or an individual with equivalent status in one of Allen Overy Shearman Sterling LLP's affiliated undertakings. A list of the members of Allen Overy Shearman Sterling LLP and of the non-members who are designated as partners is open to inspection at our registered office at One Bishops Square, London E1 6AD.

A&O Shearman was formed on 1 May, 2024 by the combination of Shearman & Sterling LLP and Allen & Overy LLP and their respective affiliates (the legacy firms). This content may include or reflect material generated and matters undertaken by one or more of the legacy firms rather than A&O Shearman.

© Allen Overy Shearman Sterling LLP 2024. This document is for general information purposes only and is not intended to provide legal or other professional advice.