

Data Protection and Large Language Models

05.10.2023 3 min read

Authors

Felipe Palhares

Related areas

Blockchain and Innovation

Topics

Artificial Intelligence

When the topic is artificial intelligence systems, and specifically large language models (LLMs), the first name that comes to mind is ChatGPT.

ChatGPT captured the world's attention. It holds the record as the fastest-growing consumer application ever, amassing 100 million active users in the two months following its launch. The application's attraction was not limited to those interested in testing its abilities: data protection authorities around the world have also been keeping a close on its data protection practices.

The stance taken by regulatory agencies with respect to ChatGPT, and allegations of potential violations of data protection legislation by that service, can offer valuable insights for providers of similar services, or other artificial intelligence solutions that process personal data in a manner comparable to ChatGPT.

On March 20, 2023, Garante, Italy's data protection authority, issued an emergency decision suspending the processing of personal data of users located in Italy by ChatGPT. In its order, Garante cite inadequate information given to data subjectson how OpenAI processes information, and the potential lack of a legal basis for mass data collection and processing for the purpose of training the system's algorithms.

Garante revisited is order on April 28, 2023, after OpenAI made changes to improve compliance with the GDPR, including providing additional information on the data processing, adopting measures to allow users to exercise opposition and exclusion rights, and including age checks to restrict use of the service by individual under 13 years. Nonetheless, the Italian authority's investigation into ChatGPT is ongoing.

After Germany's and Spain's data protection authorities announced that they too would investigate ChatGPT's practices, on April 13, 2023, the European Data Protection Board, which is responsible for ensuring consistent application of the GDPR, announced the creation of a task force to assess OpenAI's practices and to share information on possible enforcement actions by data protection authorities.

In May 2023, Professor Luca Belli published an article pointing to various violations of Brazil's General Data Protection Law (LGPD) by ChatGPT, which led to the filing of a petition against OpenAI with the ANPD, Brazil's National Data Protection Authority. The suspected violations include inadequate information, the lack of a legal basis for data processing, failure to respect data subjects' rights under the LGPD, and failure to appoint a Personal Data Protection Officer. Although there have been reports in the media that the ANPD has commenced an investigation into the complaints against OpenAI, no official mention of the investigation can be found on the ANPD's website.

More recently, in August 2023, a Polish researcher filed a complaint with the local data protection authority, alleging that OpenAI systematically violates the GDPR by disregarding data subjects' requests and directly infringing all transparency obligations.

These cases are only the tip of the iceberg. To date, no data protection authority has taken a definitive position on ChatGPT's personal data processing practices, but it is clear that there are serious questions as to the legality of its conduct.

In particular, recent cases should draw the attention of AI system developers to three essential issues: (i) transparency is fundamental and should be considered at all times; (ii) data subjects' rights must be respected, especially when they make requests for personal data correction and exclusion; and (iii) suppliers should keep data protection on their radar from the conception of any new product or service, in compliance with the principle of privacy by design.

› PROFESSIONALS

Professionals who combine legal insight and market perspective



Felipe Palhares