CEDPO Response to EDPB Consultation Guidelines 032026 on web scraping in the context of generative AI

1. Controller/Joint Controllers/Processor

- Considering the complexity of AI chains, which can involve many parties, further examples of real-life relationships between parties, in terms of respective data protection obligations, (whether they be controller-processor, joint controllers, controller-controller or processor-to-processor), would be helpful.

- It would be helpful to include a distinct paragraph on independent controllers and what such a relationship would look like in the case of web scraping relationships.

- As regards the criteria for determining when a party is a controller or a processor, additional granularity would help to give more clarity. In particular, Section 18 of the draft Guidance, below, could be more detailed:

  'If the scraper creates a training data set on behalf of the AI developer, according to its documented instructions (in particular on how the data set should be constituted with regard to data sources and categories), the scraper may be regarded as a processor.'

2. Transparency

- With respect to Section 28 of the draft Guidelines, below, which concern the test for whether or not a controller can rely on the exception in Article 14(5)(b) GDPR, two additional factors for the balancing exercise could be the context of the data processing and also the feasibility of data subjects being able to subsequently exercise their data subject rights should they become aware that their data is being processed without their knowledge.

  'Where a controller seeks to rely on the exception in Article 14(5)(b) GDPR, it should assess the effort involved for the controller to provide the information to the data subject against the impact and effects on the data subject if they were not provided with the information.11 The balancing exercise should be carried out with regard to the dataset as a whole and not on every single piece of personal data, and should take into account:

- *the number of data subjects,*
- *the age of the data,*
- *any appropriate safeguards adopted.'*

3. Accuracy

- It would be helpful to clarify what responsibility and/or liability, if any, a data scraper may have in circumstances where it has scraped a data set containing personal data and made it publicly available, and where inaccuracies in that data set only become consequential because an unrelated AI developer subsequently uses that data set to train an AI model.

  Further to this point, Section 19 of the draft Guidelines state:

  'In situations where an AI developer trains an AI model using a data set which has already been scraped by another entity (scraper) it is necessary to distinguish between the responsibility for the different data processing. The scraper and AI developer are, in principle, responsible for their own, separate, processing activities, since each of them determines the objectives and the essential means of its own processing. The scraper is not, in principle, responsible for the re-use of the data.'

  With respect to accuracy, then, for example, might there be scenarios where the scraper is in fact responsible for the re-use of the data, even if that re-use is for unrelated purposes, and by an unrelated AI developer.

END