Contribution to guidelines on web scraping in the context of generative AI

We thank you a lot for the possibility of contributing and have the following remarks from a consumer's perspective:

The guidelines on web scraping in the context of generative AI can be read like an instruction for private entities how to scrape or let scrape the web for personal, even sensitive (!) data in the context of training generative AI; even not sensitive data can be cumulated and than become sensitive data. This highly delicate topic which is increasingly important nowadays is dealt with on only 22 pages. The given examples in the guidelines are strikingly obvious. It would be warmly welcomed if the examples are more differentiated to outline the possible limits.
As proposed by the Commission in the digital omnibus regarding the training and operation of AI systems, the processing of web scraping can rely on the legal ground of legitimate interest in line with Art 6 para 1 lit f GDPR respecting i.a. the principles of purpose limitation, transparency, or data minimization. Regarding the special categories of personal data, the guidelines refer to the CJEU-ruling in the case GC and other, C-136/17. If the processing activity of the web scraper is similar to the processing activity of a search engine and the processing only includes incidental and residual processing of special categories of personal data and no intentional processing and the controller implements measures to prevent the dissemination of special categories of personal data, the responsibility and obligations of the processor under the GPDR are restricted to referencing and via verification on the basis of a request by the data subject.
Although the guidelines highlight that there should always be a case by case-analysis it is obvious that the mentioned legal interpretation should be applicable in general.

In detail:
Regarding the principles when processing personal data starting from para 21 the guidelines refer to opinion 28/2024. For a better reading and understanding the main findings and interpretations should be mentioned in these guidelines as well.
For example, the guidelines never mention the right to object pursuant to Art 21 GDPR whereas opinion 28/2024 highlights this data subject right in para 65 of opinion 28/2024. The data subject's right to object pursuant to Art 21 GDPR should be mentioned in connection with the legal ground of legitimate interest.

Para 24ff of the guidelines deals with the principle of transparency and refers to the guidelines on transparency from November 2017. It should be added in the guidelines that Art 14 para 5 lit b GDPR is an exception and should therefore – in line with the CJEU ruling – be interpreted narrowly.

The elaborations in Para 59 ff concerning the consequences for data subjects, namely the difficulties to oppose the scraping of their data, are very important and thus welcomed on the consumers' perspective.