

**Question for written answer E-002709/2024
to the Commission**

Rule 144

Andreas Schwab (PPE), Brando Benifei (S&D)

Subject: Openness in AI and transparency of training data

Transparency of the data used to train general-purpose artificial intelligence (GPAI) models is a prerequisite for understanding how they work and for improving accountability. Yet, opacity in training data creates barriers to market entry and hinders innovation. By requiring summaries of training data to be made publicly available, the AI Act¹ allows smaller entities and new entrants to not only comply with regulatory requirements and manage the risks, but also increase their competitiveness.

Article 53(1) of the AI Act requires providers of GPAI models to make available information and detailed summaries about the data used for training, testing and validating GPAI models. Ensuring the effective implementation of these provisions is critical to fulfilling the AI Act's broader objectives, and the future Code of Practice prepared by the AI Office will play a crucial role in achieving this aim.

1. How does the Commission plan to ensure that the training data and the summaries required under Article 53 are sufficiently detailed/granular to provide meaningful transparency?
2. What measures will the Commission take to ensure that trade secrets are not used as a blanket justification for withholding information about training data, particularly where there are legitimate public interests, such as privacy protection or fair competition, in making this information available?

Submitted: 2.12.2024

¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>).