

AI and public opinion-forming

Filter bubbles and countermeasures

Prof. Dr. Bernd Holznagel, LL.M.



Negative Effects of AI in Public Discourse

- Online communication is increasingly characterized by **disinformation, fake news and hate speech**.
 - Facts and Opinions are not clearly distinguishable,
 - Uncertainties arise regarding the credibility of sources and evaluations.
- **Amplification through algorithm-based** information distribution:
 - Favoring divisive content, reinforcing filter bubbles and echo chambers,
 - Through bots and other automated dissemination methods.

Traditional Approach

- **Hate speech and disinformation** might be **illegal** under national criminal law.
- Under DSA the enforcement of national criminal law has been strengthened:
 - **notice- and-take-down-approach** (Art. 16 DSA), designate a legal representative.
- Shortcomings: Still hard to enforce national criminal law in this field.
- Recently AI-tools (like the **Kivi-software of media regulators**) are used to **detect possible hate speech cases** on the web. Usually few policemen are searching for illegal content.

New Instrument: Risk Prevention Measures

- As a flexible instrument aimed at protecting a wide variety of legal interests, the DSA contains additional obligations for **managing systemic risks** (Art. 34 ff. DSA).
 - Rules apply exclusively to very large online platforms such as Facebook, YouTube.
 - To this end, online platforms and search engines are to identify, analyze, assess and **report** on all material systemic risks on an annual basis.
- Risk assessment refers to the **dissemination of illegal content**, but also to the negative effects on civic discourse, electoral processes, freedom and pluralism of the media.

Risk Prevention Measures

- Risk assessment **coupled with risk mitigation (Art. 35 DSA)**:
 - Obligation to take appropriate, proportionate and effective risk mitigation measures tailored to the particular systemic risks identified.
 - May result in **new design specifications, e.g. using detection software** like “Kivi”
 - Compliance is verified once a year in an independent audit; non-compliance: auditor issue a negative note (Art. 37 DSA).
 - Sanctions of up to 6% of the annual turnover may be imposed (Art. 52 DSA).

Conclusion

- New European Platform Regulation (DSA and AIA) focuses on risk management. Other new instruments are transparency obligations and design specifications.
- Can the law effectively counter the dangers to the democratic public sphere?
 - Risk management tools need regulators who balance interests involved and find proportionate (design) obligations.
 - Law can lessen the negative effects of using AI in the field of communication. It cannot change the causes for the use of hate speech, disinformation etc. like the underlying business models.

Thank you for your attention!



Institut für Informations-, Telekommunikations- und Medienrecht (ITM)

an der Westfälischen Wilhelms-Universität Münster

Abteilung II

Prof. Dr. Bernd Holznagel LL.M.

Leonardo-Campus 9

D- 48149 Münster

Ansprechpartner:

Geschäftsführer Jan Kalbhenn, LL.M

Tel: +(49) 251 – 83 386 42

E-Mail: kalbhenn@uni-muenster.de