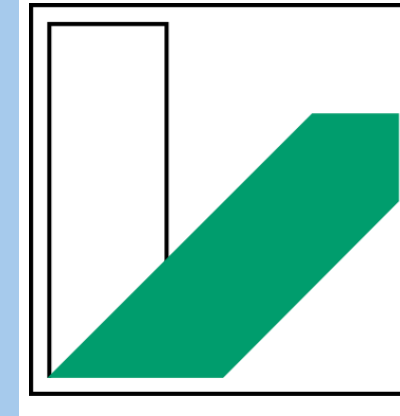


Data protection in the process of training AI models and collecting training data; A legal evaluation

Max Grebe, Nils Henkel



UNIVERSITÄT
BAYREUTH

Seminar Critical Data
Studies

Introduction

The focus of our research is the effectiveness of the existing legal frameworks (especially of the GDPR and the EU AI Act) regarding data protection of consumers personal data in the process of training AI models. We would like to do this by using the recent change to Meta's privacy policy regarding the use of public posts for the purpose of training AI tools as an example/case study. Several Data protection experts criticize the **missing necessity** of the processing and a violation of **Art. 9 GDPR**. Thereby we want to identify potential legal loopholes and possible ways to close them and to reform the legal regulations.

Legislation/Guidelines



1. Relevant Provisions:

GDPR:

Art. 5 -> contains important principles relating to processing of personal data, e.g.:

- Transparency of data processing
- Data minimisation
- Responsibility of the controller to comply with the principles (accountability)

Art. 6 I -> is the most relevant provision regarding the lawfulness of data processing; processing is only lawful if it fulfils the exceptions a) - f)

- **lit. a)** consent to the processing
- **lit. f)** processing is necessary for **legitimate interests** pursued by the controller, except the interests and fundamental rights and freedoms of the data subject override the interests

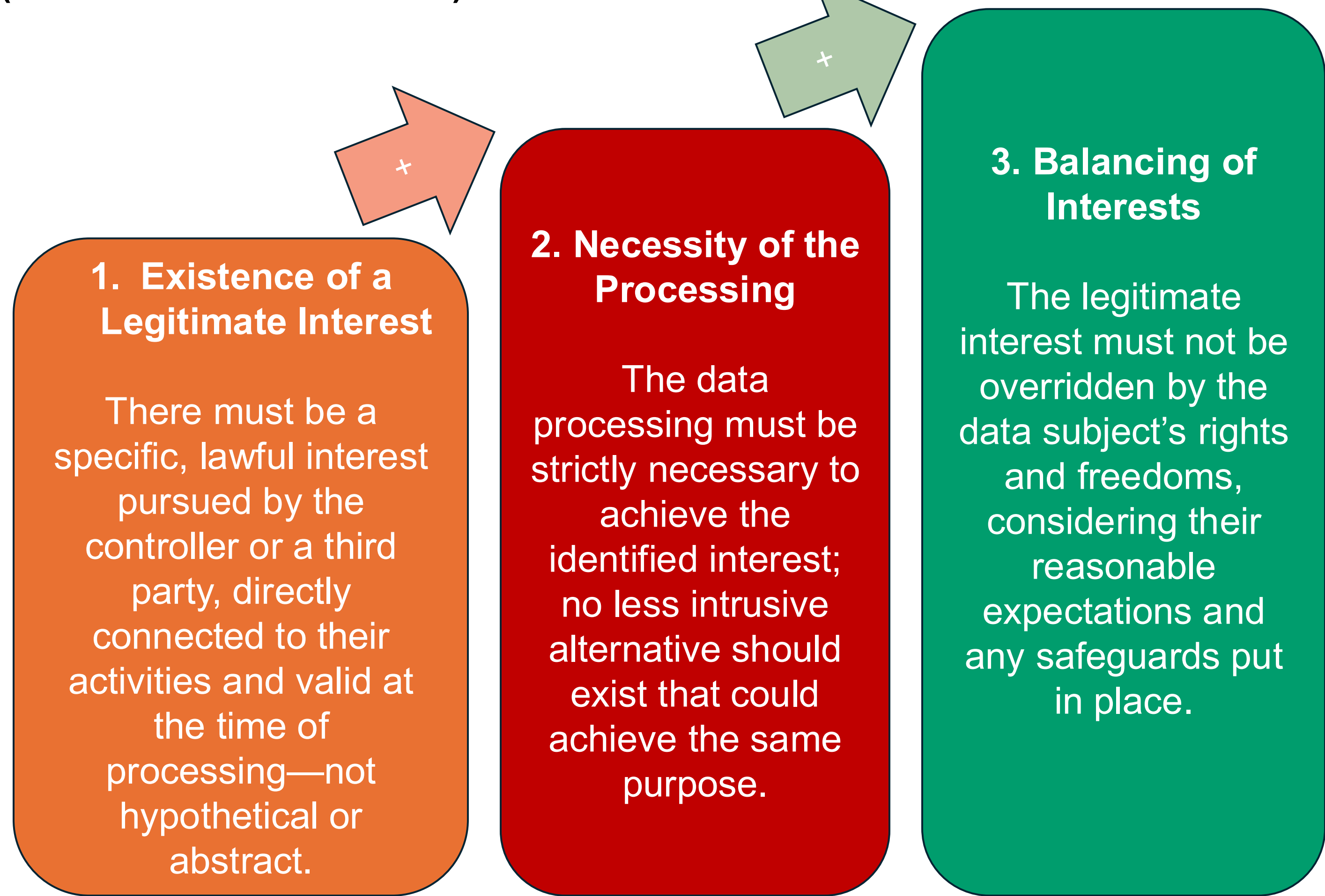
Art. 9 -> **principle (Art. 9 I):** processing of personal data revealing information that is especially worth protecting (e.g. political opinions, religious beliefs, health data, sexual orientation) is prohibited
exception (Art. 9 II a-j)): e.g. Art 9 II e) -> if the relevant data was manifestly made public by the data subject

Recital 38 -> Children merit specific protection with regard to their personal data

EU AI Act:

Art. 53 -> detailed summary of the content used for the purpose of training AI models must be published by the controller in accordance with the AI Office's guidelines

2. EDPB opinion on legitimate interest in the context of AI training (based on Art. 64 II GDPR)



Practice

1. Recent controversial example:

Meta's recent privacy policy change regarding the collection of training data for Meta AI tools:

state of affairs:

- updated policy includes the use of public posts and metadata of facebook and Instagram profiles for the purpose of training Meta AI tools
- only applies to adult users -> cf. recital 38 GDPR
- claimed legal justification: Art. 6 I f) GDPR (legitimate interest)
- training with user's personal data started on May 27 2025

! criticism:

- **Missing necessity** and adequacy of the processing (cf. EDPB opinion)
- **Violation of Art. 9 GDPR**



OLG Köln (preliminary proceedings):

- Processing is necessary and doesn't violate Art. 9
- Less intrusive alternative to anonymize the data is disproportionately complex and weakens training success
- Exception of Art. 9 II e) GDPR -> But: **What about third person's data?**



- According to OLG Köln: Application requirement of the third party-data subject in order to exercise the rights under Art. 9
- **This may make sense in the context of unauthorised social media posts, but not in the case of training sets for AI models. In this case, the person concerned is almost never aware that their critical data has been used.**

2. Identified loopholes in the legislation and the EDPB opinion

- **lack of clear requirements regarding measures to identify/filter out critical data in terms of Art. 9 GDPR (especially from third persons)**
- **identification of minors**
 - although meta restricts the processing of data to adult data subjects -> What measures to verify the age of a user are prescribed? -> no special requirements within the EU AI Act and the GDPR
- **Indeterminacy of the term "legitimate interests"** -> Failure to create regulatory examples through the AI Act that specify exemplary legitimate interests in terms of the training of AI models

Conclusion

- legislation is generally sufficient
- legal specification of measures for the protection of data within the meaning of Art. 9 GDPR regarding third parties is needed
 - EU AI Act should give some regulatory examples that specify important legal terms in the context of training Ai models
 - the courts draw invalid analogies