

Tokyo University, 1 April 2026

Copyright vs. AI in the European Union – The Current State of Play

Prof. Dr. Alexander Peukert
Goethe University Frankfurt am Main
a.peukert@jur.uni-frankfurt.de

Overview

- (1) EU copyright law and AI
- (2) EU AI law protecting copyright-related interests
- (3) Relationship between Copyright and AI law
- (4) Reform proposals

The overall legal framework concerning © and genAI

AI Act obligations concerning ©, Art. 53(1)(c) and (d)

Providers of general-purpose AI models (including FOS models) must

- (1) put in place a policy to comply with EU © law
- (2) make publicly available a sufficiently detailed summary about the training content



27 harmonised © laws

© MS1

© MS2

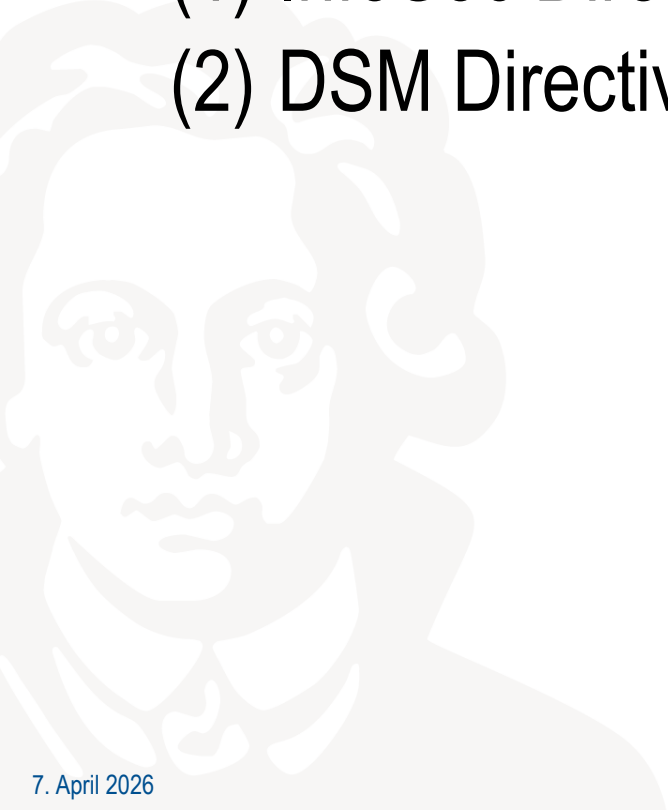
© MS3

... ©
MS27

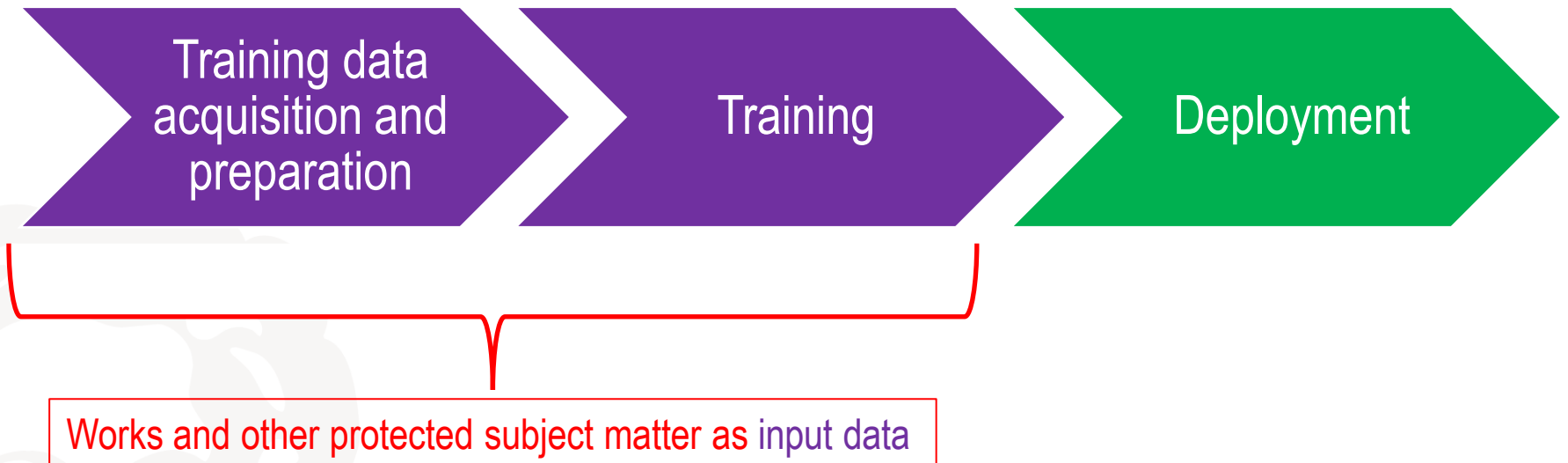
Copyright directives of particular relevance for genAI

(1) InfoSoc Directive 2001/29

(2) DSM Directive 2019/790



Copyright in the GenAI lifecycle



The input phase: Training data acquisition and training

- Is EU copyright law applicable?
 - Strict territoriality: EU copyright law only applies if the reproductions occurred on EU territory
 - Or does EU copyright law also apply to AI training reproductions in third countries if the respective AI is made available on the EU market?

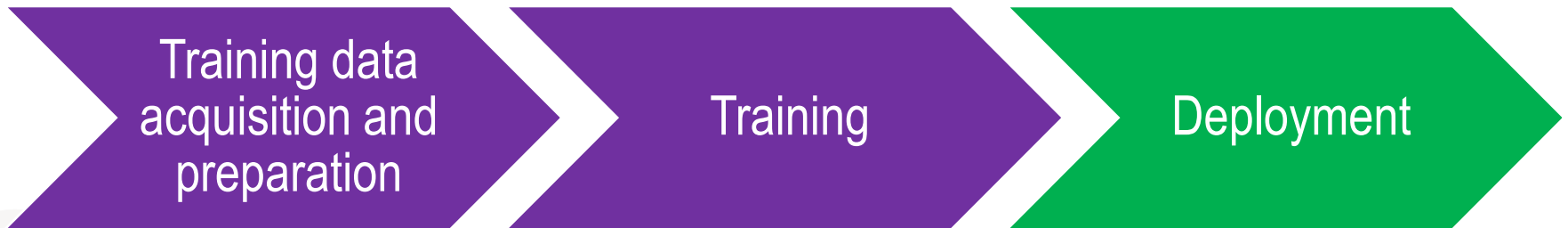
Does genAI training data acquisition and training amount to **text and data mining (TDM)**?

- TDM = “any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations” (Art. 2(2) DSM Directive)
 - Conventional forms of big data analysis, for AI training or other purposes
 - e.g. laion.ai (5,85 billion URLs of images with metadata):
Yes: [Hamburg Higher Regional Court 2025 – LAION](#)
 - GenAI training
 - Cf. question 3 of reference [C-250/25 Like Company](#) concerning press publications

Does genAI training meet the conditions of **Art. 4 DSM** Directive?

- Scope of Application: any TDM
- Conditions
 1. Content must be “lawfully accessible”
 2. The TDM use must not have “been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online”
 3. Training corpus “may be retained for as long as is necessary for the purposes of text and data mining” (Art. 4(2) DSMD).

Copyright in the GenAI lifecycle



Works and other protected subject matter as input data

- (1) Reproduction of protected content „in“ the model/system
- (2) Output may infringe rights

“Memorizations” of protected training data



- Technical background
 - Training data not “stored” in a model
 - Representation of training data in the parameters of a model

Is memorization a „reproduction“ “in” the model?

- Yes: Munich I Regional Court – [GEMA v. OpenAI](#)
- No: High Court London – [Getty v. Stability AI](#);
[Peukert 2026a](#)

Infringements through genAI outputs

- A protected work is recognizable in an AI output
 - Reproduction if the work formed part of the training corpus (+)
 - Communication to the public or only to the user prompting the output?
- Applicable limitations and exceptions?
 - AI providers: quotation right?
 - AI users: private copying, parodies, caricature or pastiche?
- Liability of the AI provider
 - full liability for infringing output (Munich Regional Court [GEMA v. OpenAI](#))
 - or limited liability for insufficient copyright protection measures ([Peukert 2026b](#))?

Overview

(1) EU copyright law and AI

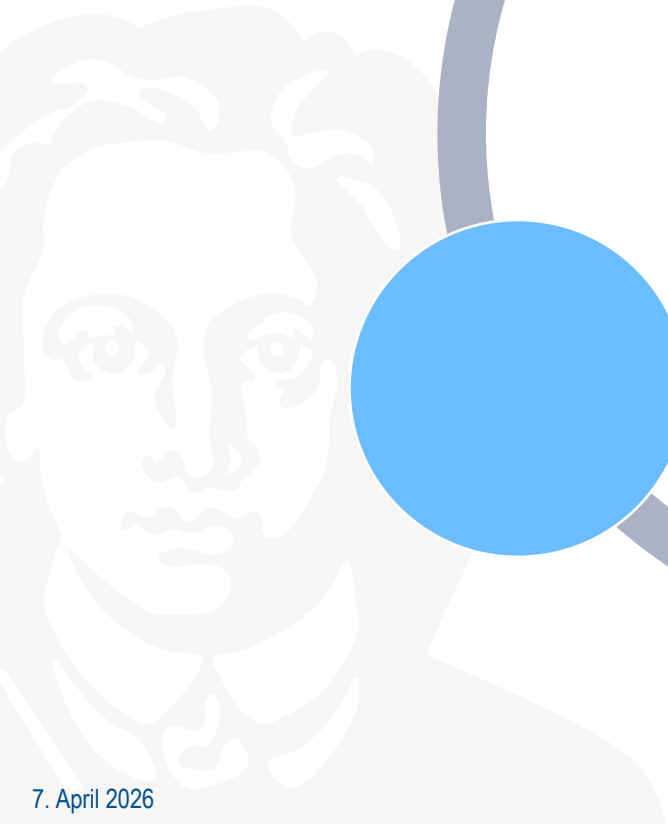
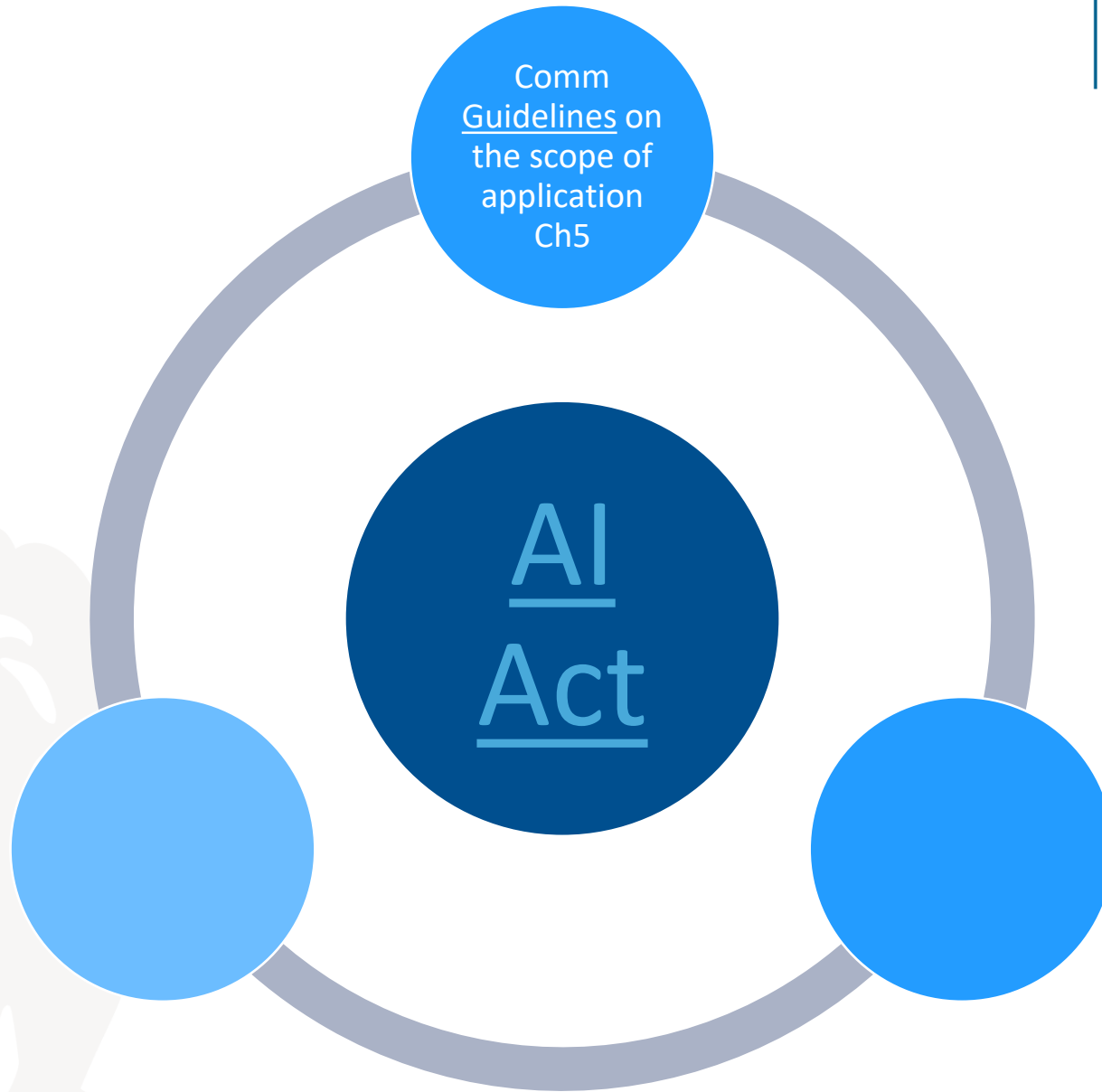
 (2) EU AI law protecting copyright-related interests

(3) Relationship between Copyright and AI law

(4) Reform proposals

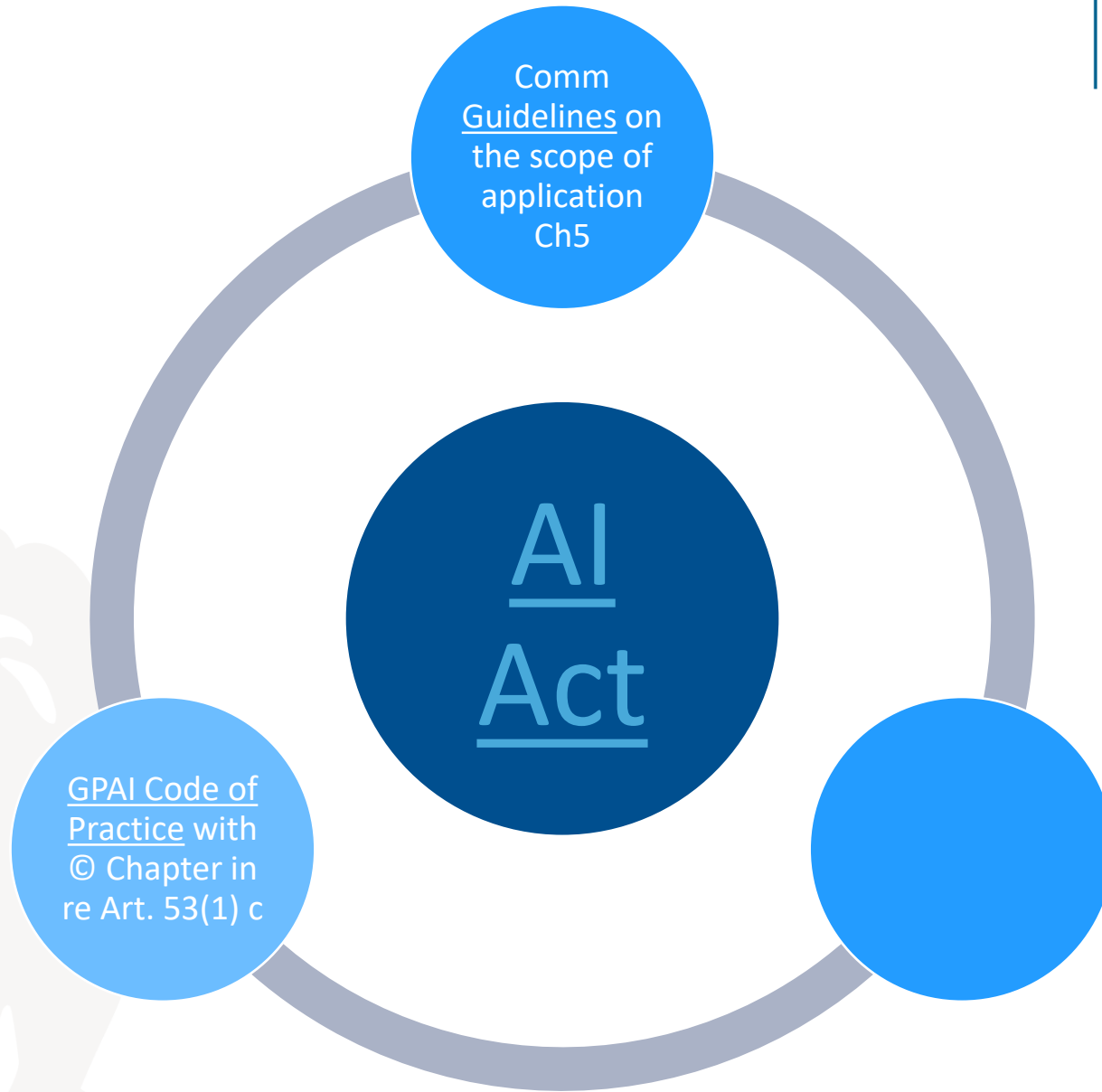
Art. 53(1)(c) and (d) AI Act

- Providers of general-purpose AI models shall:
 - (c) put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;
 - (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.

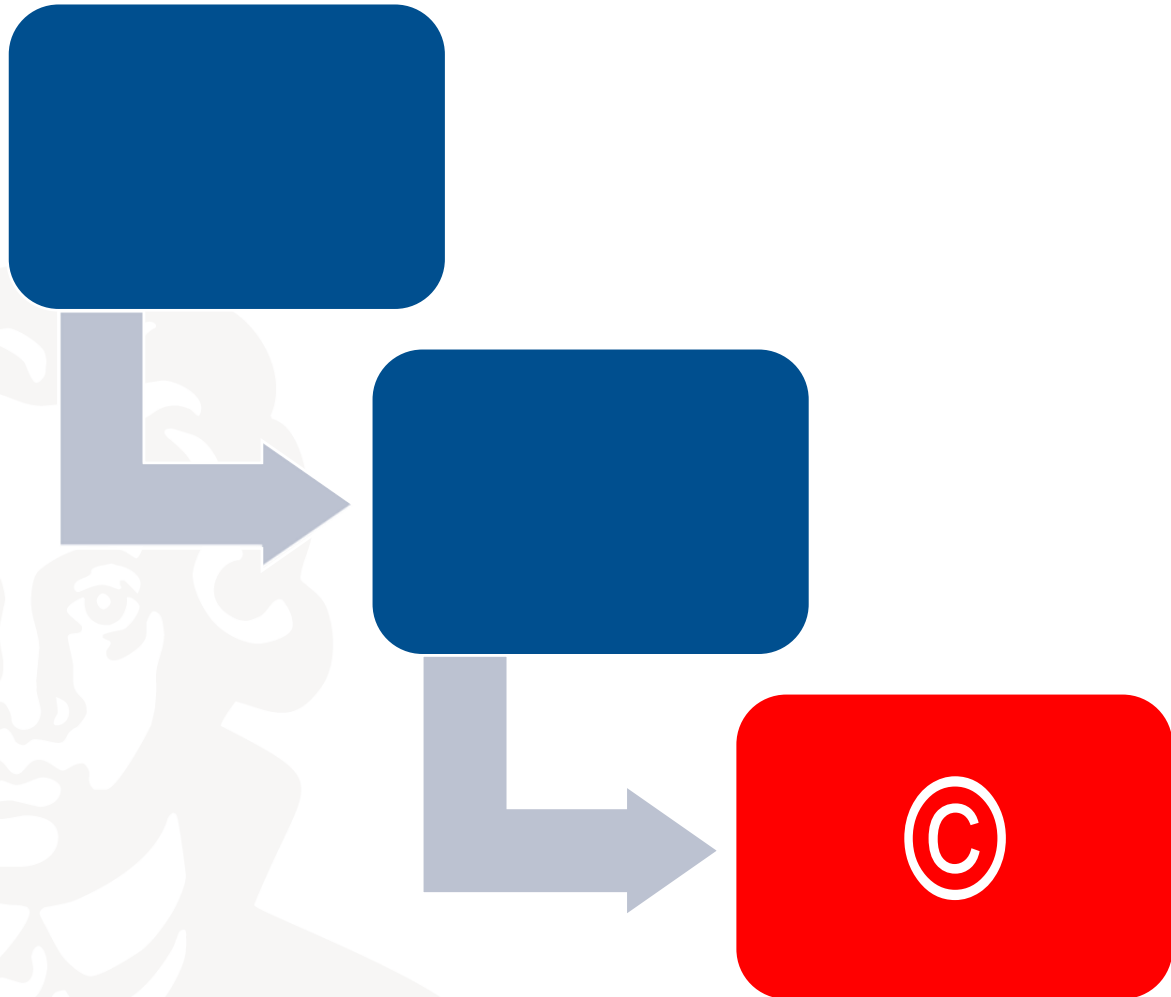


Scope of application of Art. 53(1)(c) and (d) AI Act

- General-purpose AI models (Art. 3(63) AI Act)
- All providers who develop a GPAIM and place it on the EU market under their own name or TM (Art. 3(3) AI Act)
 - Irrespective of whether those providers are established or located within the Union or in a third country (Art. 2(1)(a) AI Act)
 - “regardless of the jurisdiction in which the copyright-relevant acts underpinning the training ... take place” (recital 106 s. 3 AI Act)



© and AI law are linked in a threefold compliance structure



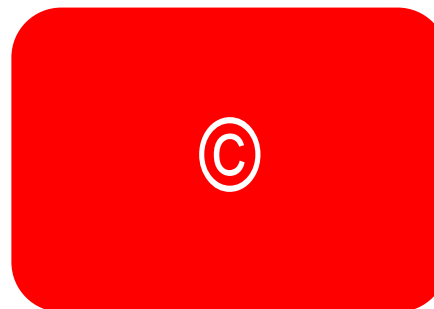
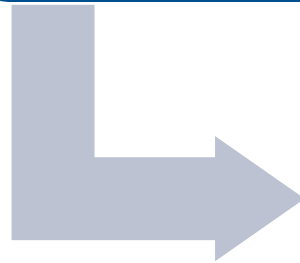
- Do not use without prior authorization, unless you are covered by I&E

© and AI law are linked in a threefold compliance structure



Art.
53(1)(c)

- Comply with EU © law = do not infringe!



- Do not use without prior authorization, unless you are covered by I&e

© and AI law are linked in a threefold compliance structure

CoP

- Comply with Art. 53(1)(c)!

Have a © compliance policy!

Art.
53(1)(c)

- Comply with EU © law = do not infringe!

©

- Do not use without prior authorization, unless you are covered by I&e

Introductory, overarching Measure 1.1:
Draw up, keep up-to-date and implement a copyright policy

Measure 1.2:
Lawful
access

Measure 1.3:
TDM opt-out

Measure 1.4:
Output
infringements

Closing, overarching Measure 1.5:
Set up a point of contact and enable complaints

Core commitments:

1. Have a © policy, assign responsibility for implementation and oversight
2. Comply with lawful access requirement
 - Do not circumvent access controls
 - Blacklist piracy websites recognized as such by EU/EEA Member State authorities (cf. [CJEU UPC Telekabel](#))
3. Identify and comply with TDM rights reservations
 - Comply with robots.txt and subsequent versions
 - Comply with future opt-out standards (international or EU)
 - Provide real-time, public transparency about crawlers and opt-out compliance
4. Mitigate the risk of output infringements
 - Implement appropriate and proportionate technical safeguards to mitigate the risk of output infringements
 - Prohibit copyright-infringing uses in your policies
5. Have a point of contact and handle complaints about Code compliance

Encouragements for voluntary action:

- (1) Publish a summary of the © policy
- (2) Support international and EU opt-out standardisation processes
 - Internet Engineering Task Force, robots.txt update
 - EU process started with a [consultation 2025](#)
- (3) Avoid adverse effects of AI opt-outs on the findability of the content in your search engine

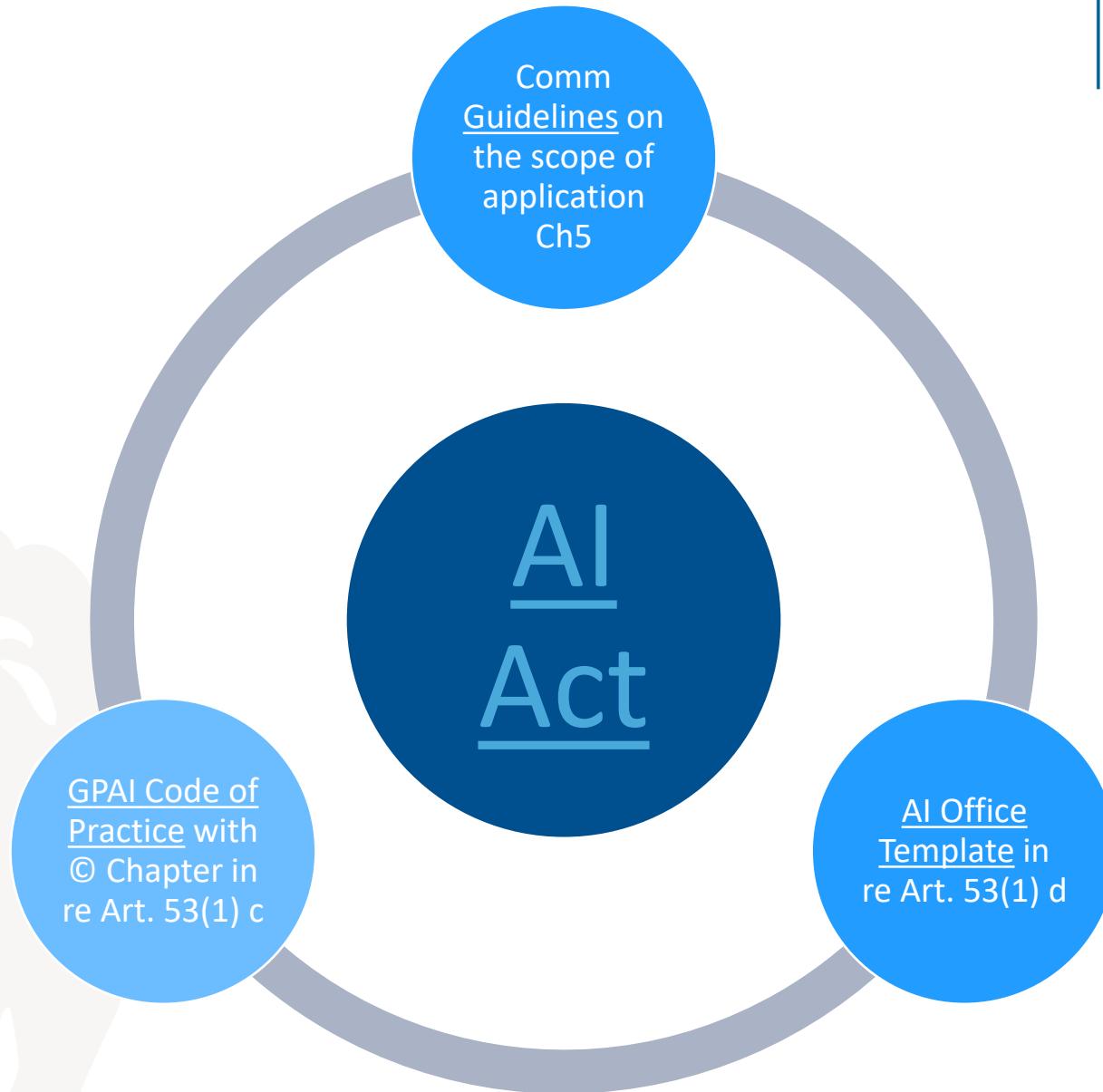
On the one hand,
the GPAI **Code** is a
voluntary tool to
demonstrate
compliance with
Art. 53(1)(c) AI Act

Code Signatories:

- Accessible
- AI Alignment Solutions
- Aleph Alpha
- Almawave
- Amazon
- Anthropic
- Black Forest Labs
- Bria AI
- Cohere
- Cyber Institute
- Domyn
- Dweve
- Euc Inovação Portugal
- Fastweb
- Google
- Humane Technology
- IBM
- Lawise
- LINAGORA
- Microsoft
- Mistral AI
- Open Hippo
- OpenAI
- Pleias
- re-inventa
- ServiceNow
- Virtuo Turing
- WRITER

On the other hand, **non-Signatories have to comply with the AI Act** in any case:

- report their © measures to the AI Office and explain how these measures ensure compliance, for instance by comparing their measures with the Code (gap analysis)
- may be subject to a larger number of requests for information and for access
- have to comply with the training data summary obligation under Art. 53(1)(d)



Main information to be provided in the training data template:

- Model-related i.a.
 - name, dates of market placement and end of training data collection
- Training data-related i.a.
 - List of large publicly available datasets (> 3% the all publicly available datasets used for training for each modality, eg CommonCrawl)
 - Top 10% of domains crawled
- Other information
 - Name of crawler and respected opt-out protocols
 - Measures taken to avoid or remove illegal content from the training data (such as blacklists, keywords, and model-based classifiers)

Overview

(1) EU copyright law and AI

(2) EU AI law protecting copyright-related interests

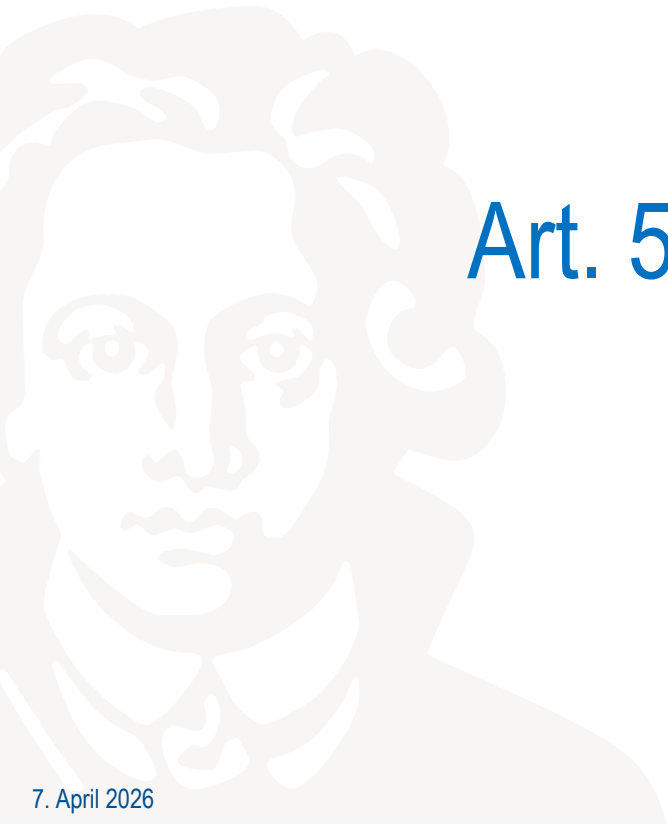
 (3) Relationship between Copyright and AI law

(4) Reform proposals

Copyright law

≠

Art. 53(1)(c) and (d) AI Act



But common aims of **copyright** and **AI law**

high level of
© protection



promotion of
innovative AI

fair balance

 consistent interpretation of
copyright and **AI law**

- **Input phase**: AI training can qualify as permissible TDM
- **Deployment phase**: AI providers have to take effective and appropriate copyright-protective measures
- An AI model is not an infringing copy of protected training content

Overview

(1) EU copyright law and AI

(2) EU AI law protecting copyright-related interests

(3) Relationship between Copyright and AI law

 (4) Reform proposals

Outlook: reform proposals

- The black box problem
 - “the failure by AI providers ... to provide complete transparency should give rise to a rebuttable presumption that any relevant copyright-protected work or other protected subject matter has been used for the purposes of training, inferencing or retrieval-augmented generation” ([European Parliament Resolution 2026](#))
- Remuneration
 - Extended collective licensing agreements per sector with opt-out possibility ([European Parliament Resolution 2026](#))
 - Statutory remuneration claim