Rechtsfragen Künstlicher Intelligenz

Dr. Jonas Siglmüller
Denkfabrik Legal Tech

24.11.2023

Technische Beschaffenheit

Haftung

Leistungsmetriken

Vertrauchlichkeit

Trainingsdaten

Regulierung

KI-Output

Erklärbarkeit

Maintenance


Ressourcen-
anforderungen
Latenzzeit
Technische
Beschaffenheit
Skalierbarkeit
Inferenzzeit
FLOPS
Beweislast
Informationszugang
durch KI-Haftungs-RL
Haftung
Richtigkeit
Vertrauen
Precision
Recall
Leistungsmetriken
F1-Score
Genauigkeit
Mean Squared Error
AGG
Datenlizenz
Trainingsumgebung
Datenqualität
Datenquellen
Monitoring
Verschlüsselung
Personenbezogene
Daten
Trainingsdaten
Aufbewahrung
Vertrauchlichkeit
AI Act-E
CRA / NIS2
DSA
DSGVO
Auskunftsansprüche
Urheberrechtliche
Schranken
(insb. § 44b, 60d UrhG)
Regulierung
on-prem?
Trainingslizenz?
GeschGehG
Data Act
Aufsichtsrecht
Zulassungsrecht
Personenbezogene Daten
Urheberrechtsverletzungen
Visualisierung
(z.B. Heatmaps)
Feature Importance
KI-Output
Model Drift
Aktualität
Erklärbarkeit
Richtigkeit
Entstehung von
Urheberrecht
Maintenance
SHAP
Partial Dependence Plots
RGL +
AVV / Joint Control
Verbesserung
Anomalieerkennung

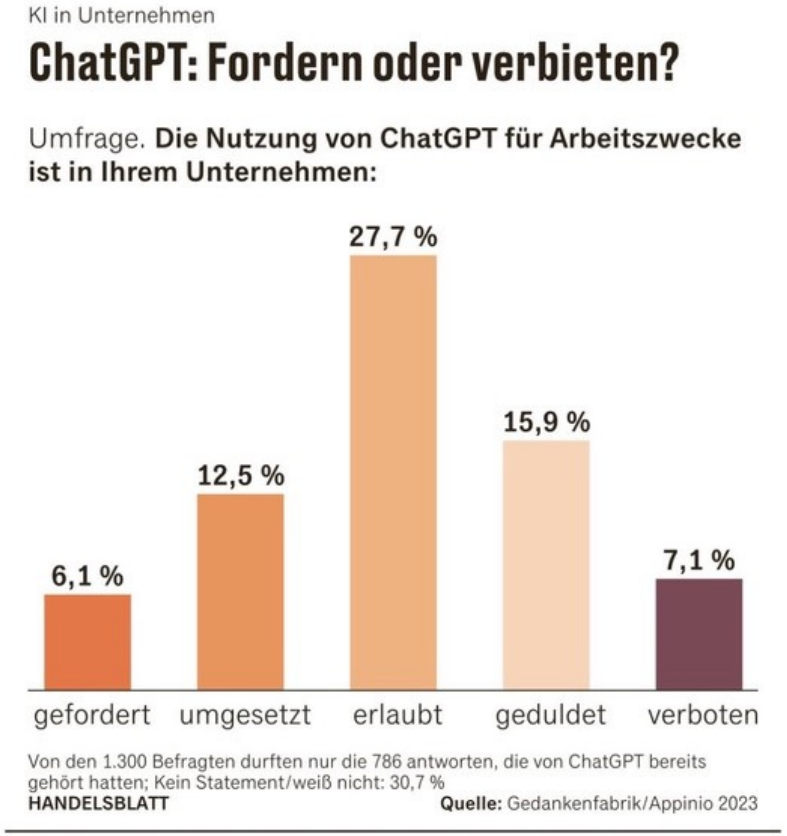
Vertraulichkeit

LLMs für Forschung & Entwicklung

Case Study 1

Unternehmen verbieten ChatGPT aus Sorge vor dem Verlust von Geschäftsgeheimnissen

- Samsung
- JP Morgan
- Apple
- Goldman Sachs
- Citigroup
- Bank of America



§ 2 Nr. 1 GeschGehG

Im Sinne dieses Gesetzes ist Geschäftsgeheimnis eine Information

[...]

b) die Gegenstand von den Umständen nach angemessenen Geheimhaltungsmaßnahmen durch ihren rechtmäßigen Inhaber ist und

[...]

Vertraulichkeit bei OpenAI

Section 3 (c) Terms of Use ChatGPT:

[…] We may use Content from Services other than our API (“Non-API Content”) to help develop and improve our Services. […]

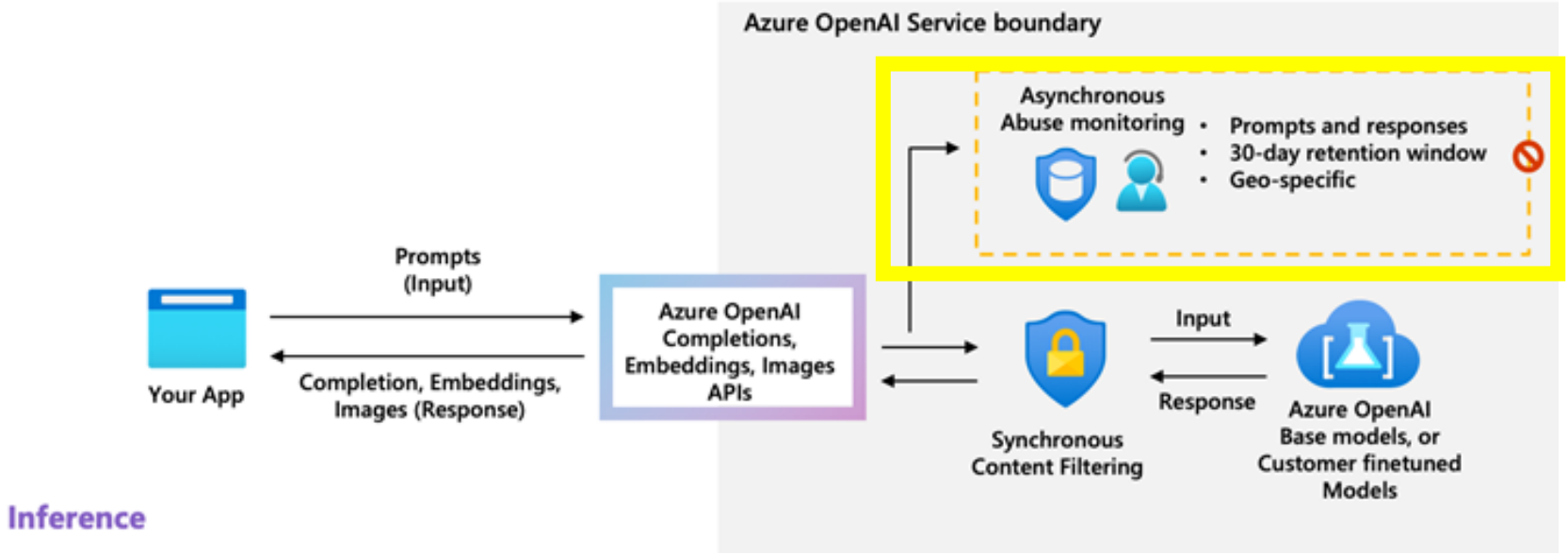
[…] If you do not want your Non-API Content used to improve Services, you can opt out by filling out this form. […]

Microsoft Azure OpenAI

Case Study 1

[…] Your prompts (inputs) and completions (outputs), your embeddings, and your training data:

- are NOT available to other customers.
- are NOT available to OpenAI.
- are NOT used to improve OpenAI models.
- are NOT used to improve any Microsoft or 3rd party products or services. […]



Quelle: https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy

Vertraulichkeit bei KI-Nutzung

Case Study 1

Königsweg

Betrieb auf
eigener IT-Infrastruktur

Plan B

Vermeidung von vertraulichen Informationen

+

Geheimhaltungsvereinbarung mit KI-Anbieter

ITAVA

Betriebsmodelle von ChatGPT

ChatGPT (Plus) (GPT-3.5/4)

- Weiterverarbeitung der Daten
  neu: Opt-Out-Möglichkeit
- Keine ausreichende Vertraulichkeits-vereinbarung
- Verarbeitung in den USA
- h.M. gemeinsame Verantwortlichkeit
- Kostenlos nur GPT-3.5
- Nutzungs-beschränkungen

Neu: ChatGPT Enterprise

- Datenverschlüsselung in rest
- SOC-2 Zertifizierung
- Unternehmensfeatures (Adminkonsole, SSO, Analytics etc.)
- Schneller und längere Texte
- Möglichkeit zum Finetuning des Modells
- Schaffen eigener IP?

ChatGPT API

- Einbettung in eigene Produkte möglich
- Keine ausreichende Vertraulichkeits-vereinbarung
- Keine Weiterverarbeitung der Daten
- OpenAI als Datenverarbeiter

Azure OpenAI

- Server in der EU
- Möglichkeit zum Finetuning des Modells
- Schaffen eigener IP

GPT-2

- Open Source Modell
- Betrieb auf eigener Infrastruktur möglich

Personenbezug

Vertragsverhandlungen in KI-Projekten

Case Study 2

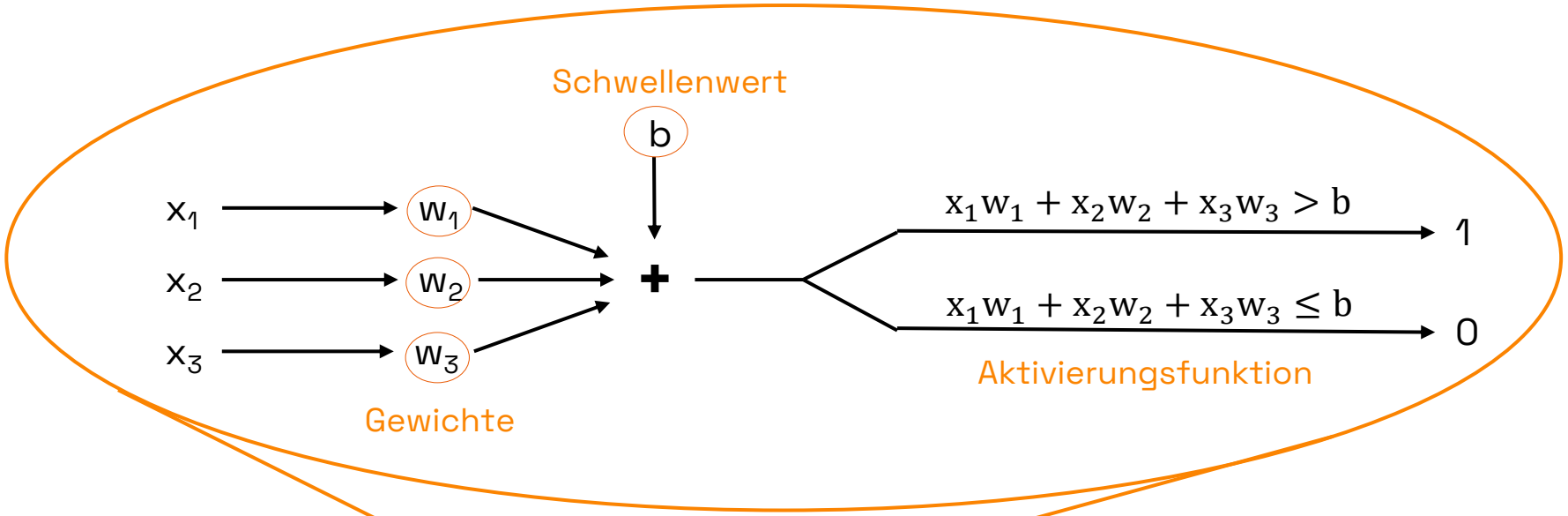
Rechtskonformes System

vs.

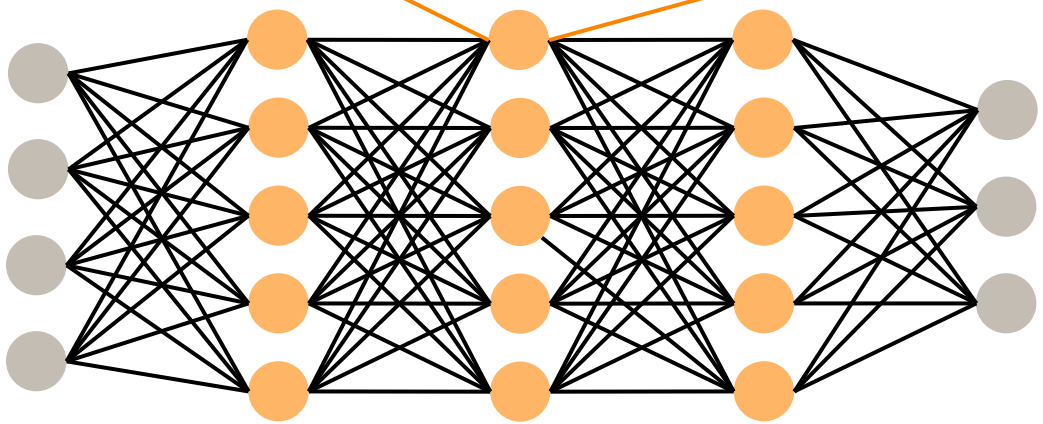
Rechtskonform trainiertes System


Vertrauensleute
BMW Dingolfing
SOLIDARITÄT GEWINNT!
Landshut
Landshut

Künstliches Neuron



Tiefes neuronales Netz



Feature Learning Layers

GPT - 4

1.760.000.000.000

Parameter

120

Layer

4x5+5 5x5+5 5x5+5 5x3+3 = 103 Parameter

Rechtskonformes System

Rechtskonform trainiertes System

Rechtskonformes System

IP-Compliance

Outputfilter bei Stable Diffusion

Case Study 3

LAION-5B

Stable Diffusion

v1.4

Prompt

Output

Extracting Training Data from Diffusion Models

Nicholas Carlini[*1] *Jamie Hayes*[*2] *Milad Nasr*[*1]
Matthew Jagielski[+1] *Vikash Sehwag*[+4] *Florian Tramèr*[+3]
Borja Balle[†2] *Daphne Ippolito*[†1] *Eric Wallace*[†5]
[1]Google [2]DeepMind [3]ETHZ [4]Princeton [5]UC Berkeley
[*]Equal contribution [+]Equal contribution [†]Equal contribution

Figure 3: Examples of the images that we extract from Stable Diffusion v1.4 using random sampling and our membership inference procedure. The top row shows the original images and the bottom row shows our extracted images.

Quelle: https://arxiv.org/pdf/2301.13188.pdf

(Zu) Starke Ähnlichkeit?

Case Study 3

§ 23 (1) Urheberrechtsgesetz

Bearbeitungen oder andere Umgestaltungen eines Werkes, insbesondere auch einer Melodie, dürfen nur mit Zustimmung des Urhebers veröffentlicht oder verwertet werden. **Wahrt das neu geschaffene Werk einen hinreichenden Abstand zum benutzten Werk, so liegt keine Bearbeitung oder Umgestaltung im Sinne des Satzes 1 vor.**

Keine „wiederkennbare Form“

Europäischer Gerichtshof zum Sampling

„Verblassen“ des Ausgangswerks

Bundesgerichtshof zu Porsche 911

Training Set

Caption: Living in the light with Ann Graham Lotz

Generated Image

Prompt: Ann Graham Lotz

Stable Diffusion
v1.4

Prompt

Output

LAION-5B

Filter

Microsoft announces new Copilot Copyright Commitment for customers

Sep 7, 2023 | Brad Smith, Vice Chair and President, Hossein Nowbar, CVP and Chief Legal Officer

To address this customer concern, Microsoft is announcing our new Copilot Copyright Commitment. As customers ask whether they can use Microsoft's Copilot services and the output they generate without worrying about copyright claims, we are providing a straightforward answer: yes, you can, and if you are challenged on copyright grounds, we will assume responsibility for the potential legal risks involved.

IP-Schutz

Schutz der Erzeugnisse von GitHub Copilot

Case Study 4

```
0 references | 0 changes | 0 authors, 0 changes
public static void CreateTables()
{
    using (var context = new TaskContext())
    {
        context.Database.ExecuteSqlRaw("CREATE TABLE tasks (id INT PRIMARY KEY, title VARCHAR(50), priority INT)");
    }
}
}
```

Wem gehört der Code?

- Urheberrechtlich geschützt ist nur die **persönlich geistige Schöpfung** eines Menschen (§§ 2 Abs. 2, 69a Abs. 3 UrhG)

  → Entscheidende Abgrenzung:

KI nur Hilfsmittel

vs.

KI schöpferisch

Use Cases

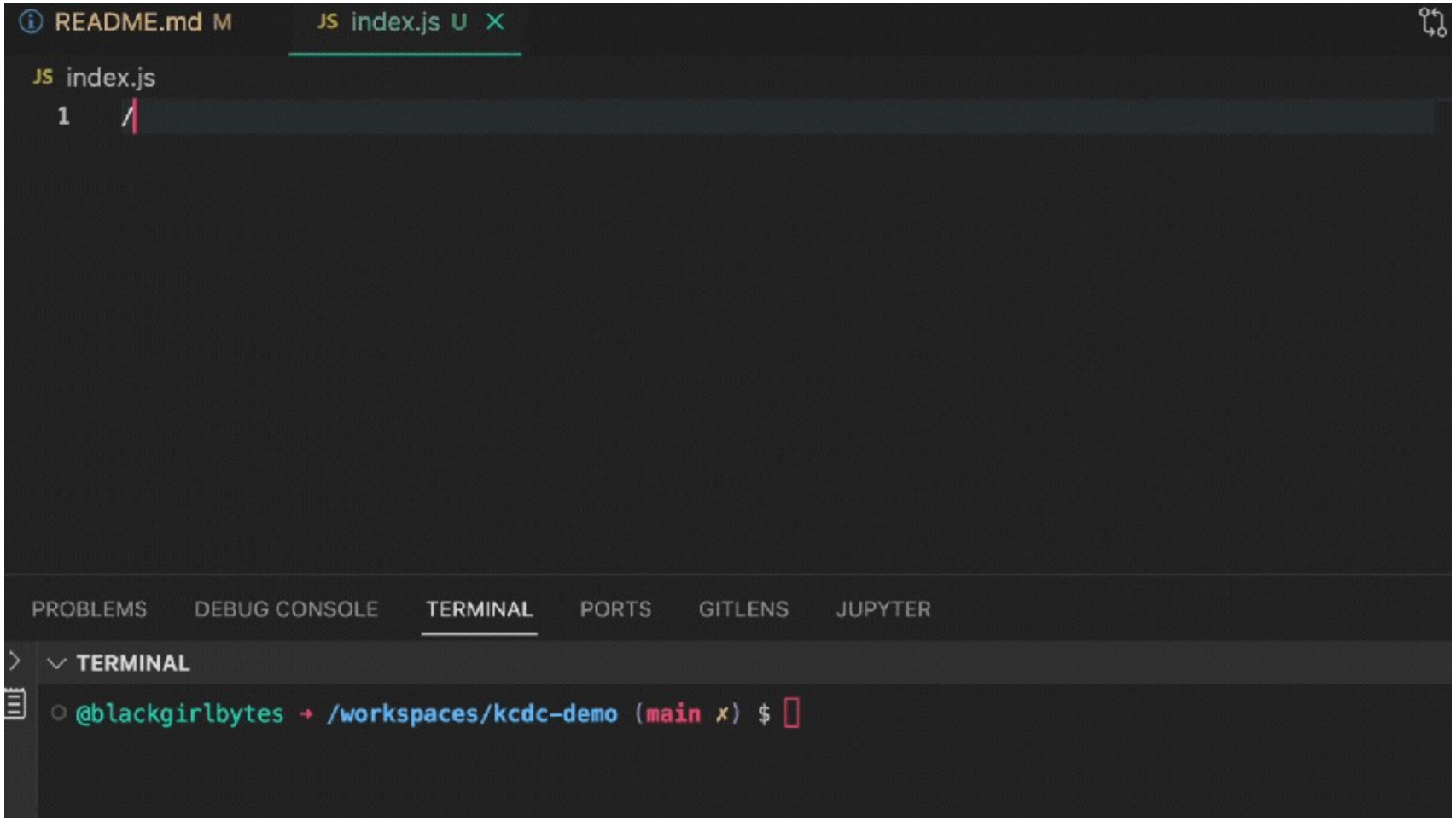
Case Study 4

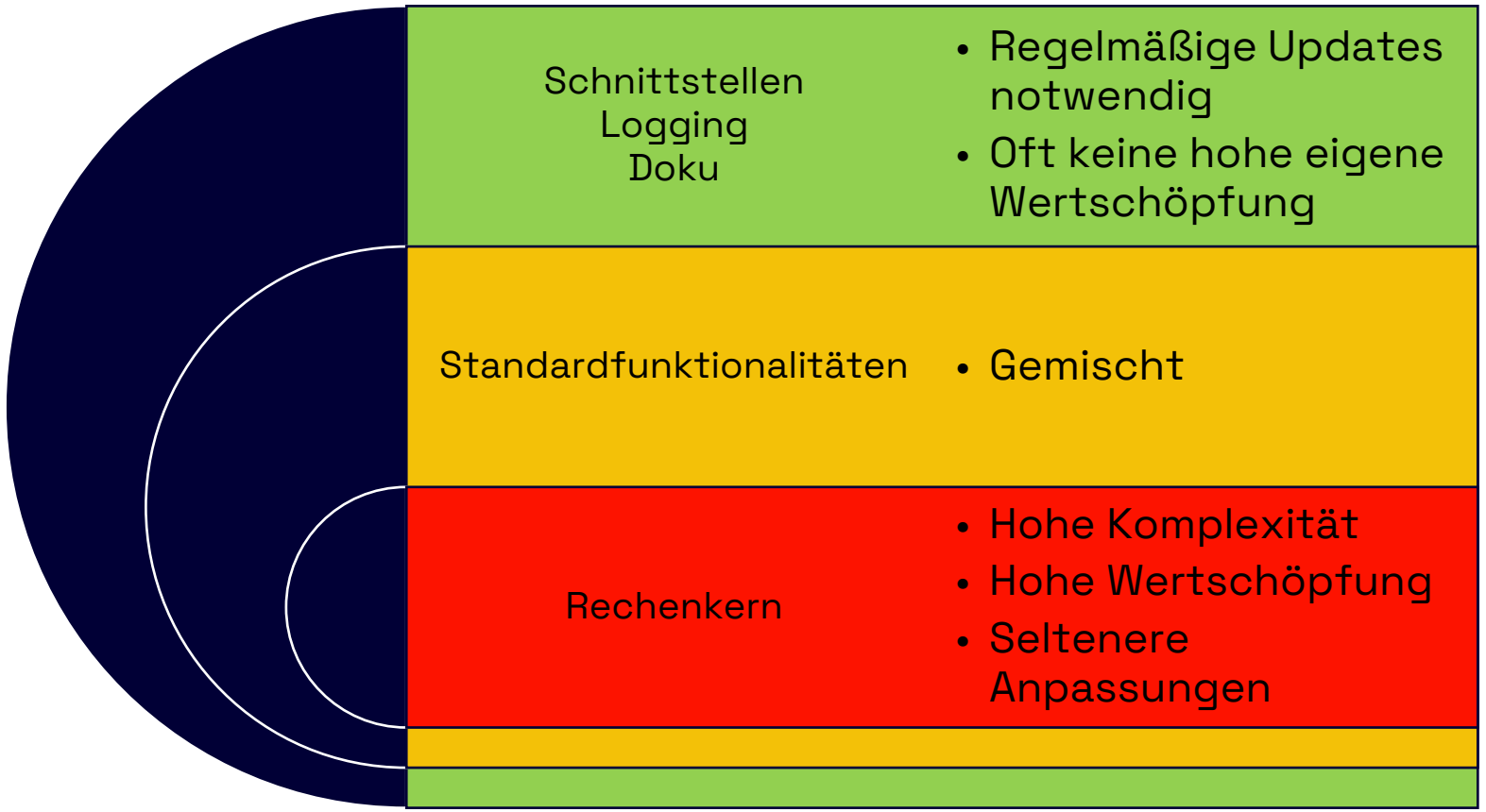
useRowActions.js

```
28      },
29      {
30        label: 'Modify rate/allocation',
31        selected: false,
32        onClick: (e, row) => {
33          const id = row[0]?.value;
34          const record = { ...getRecordById(id) };
35          setRowModal(<ModifyModal booking={record} onClose={e => setRowMod
36        },
37      },
38      {
39        label: 'Delete',
40        selected: false,
41        onClick: (e, row) => {
42          const id = row[0]?.value;
43          const booking = { ...getRecordById(id) };
44          setRowModal(<DeleteModal booking={booking} refreshParent={refresh
45        },
46      },
```

Use Cases

Case Study 4





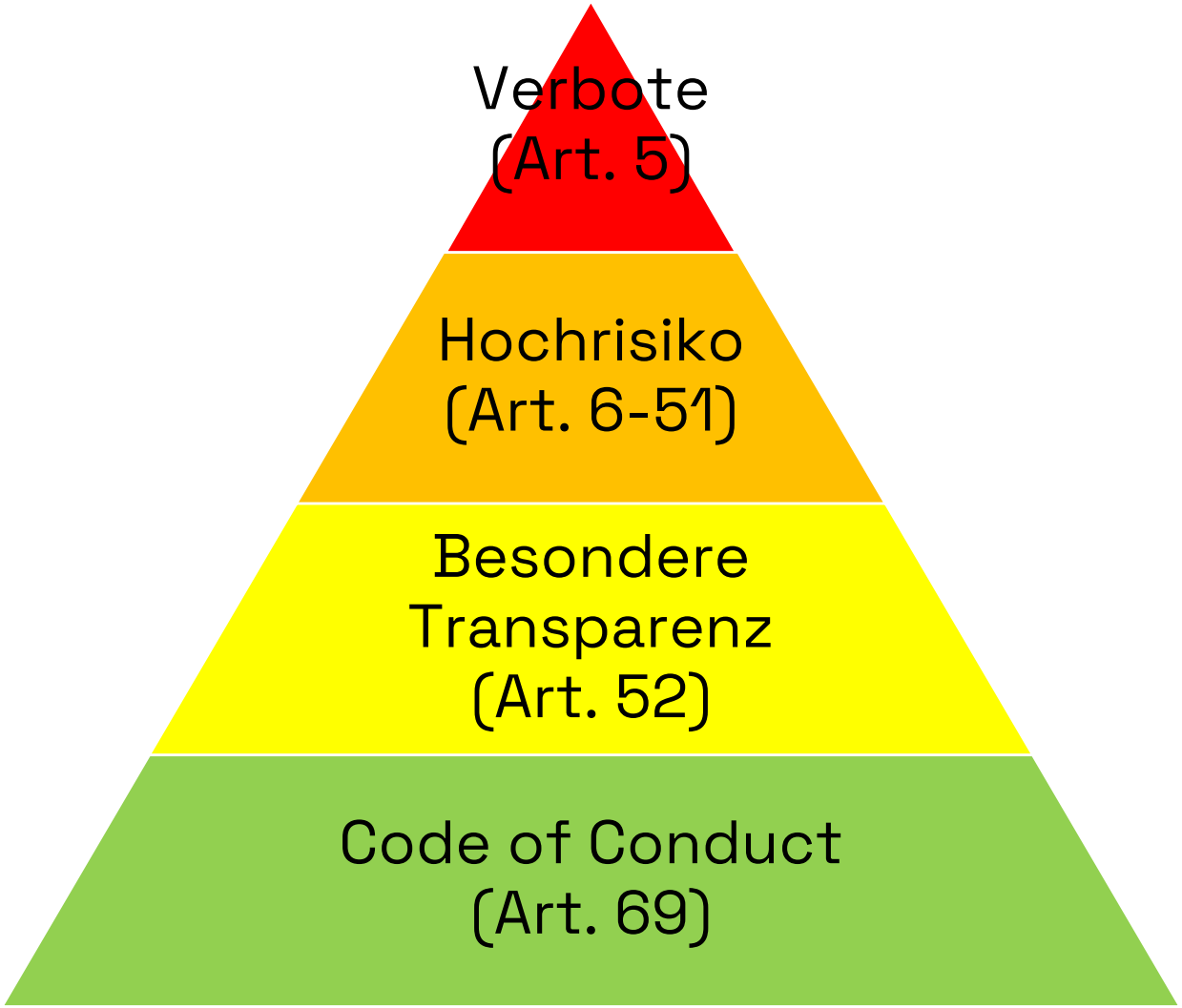
Wichtig:

- Automatisierte Duplikatsprüfung bei GitHub Copilot anschalten
- Klare Unterteilung nach Repositories
- Dokumentation der Copilot Nutzung / Eigenleistung
- Duplikatsprüfung auf Open Source Software (zB auf GitHub) und Foren (zB Stackoverflow)

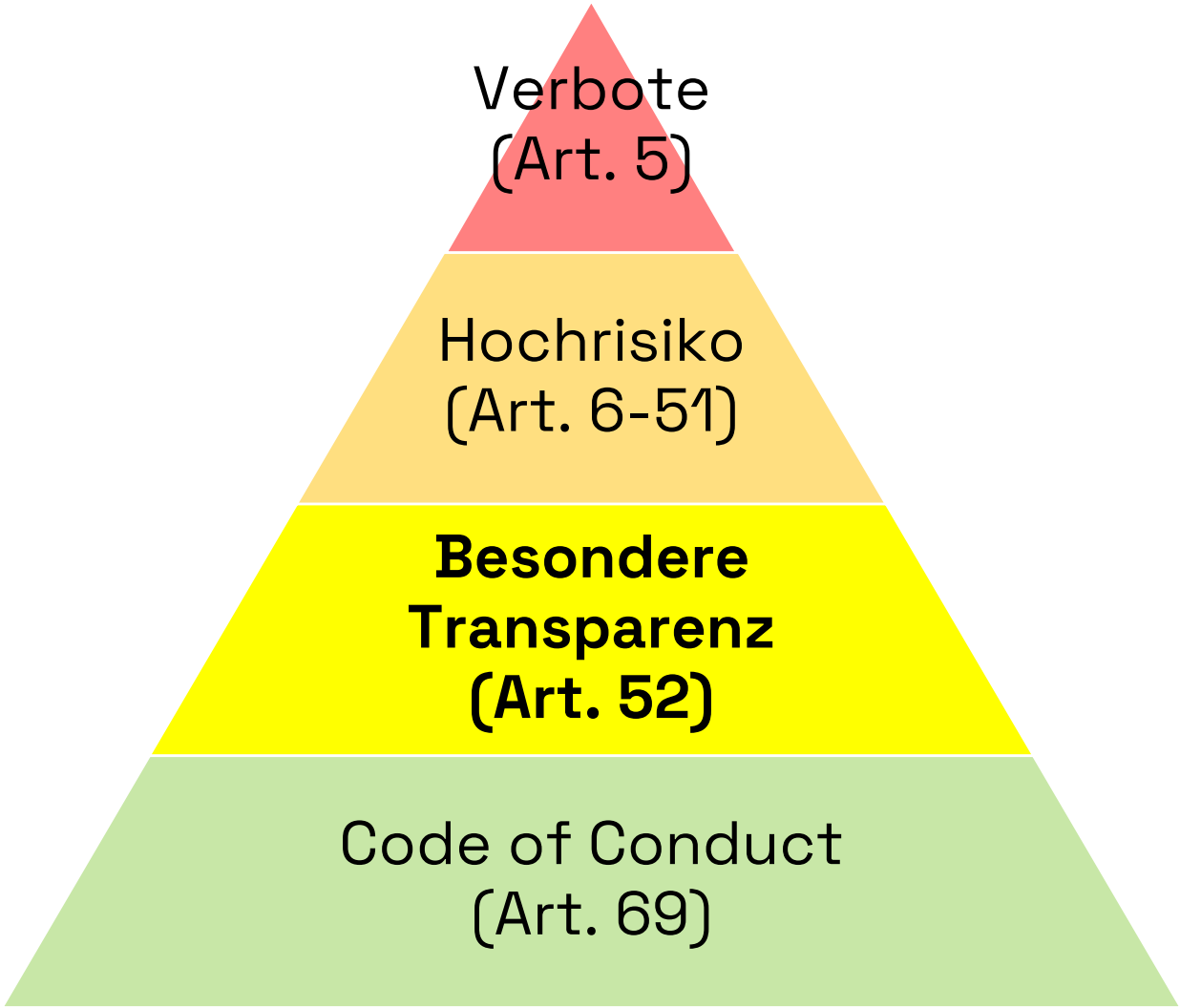
Regulatorik

Update zum AI Act

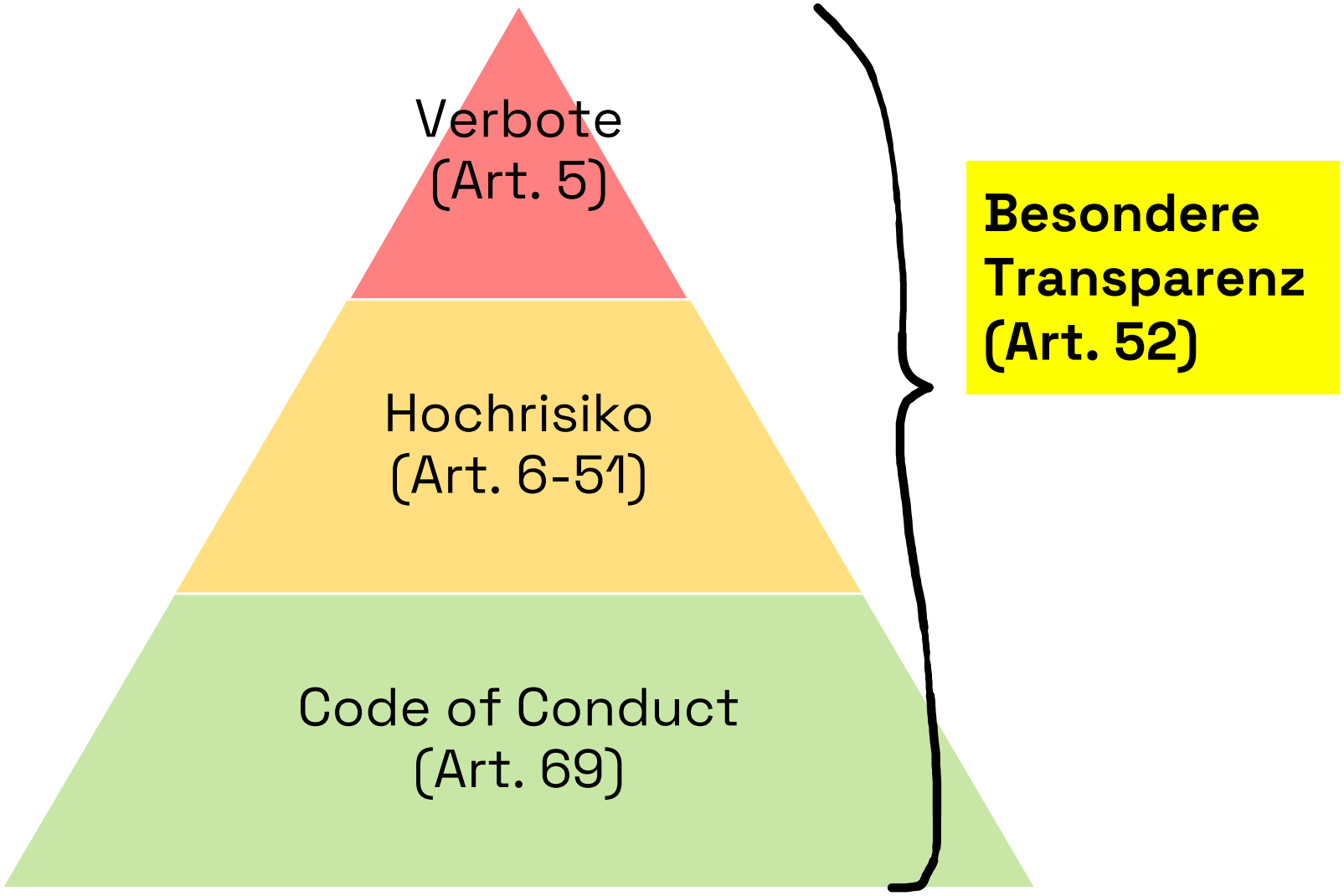
April 2021: Kommissionsentwurf



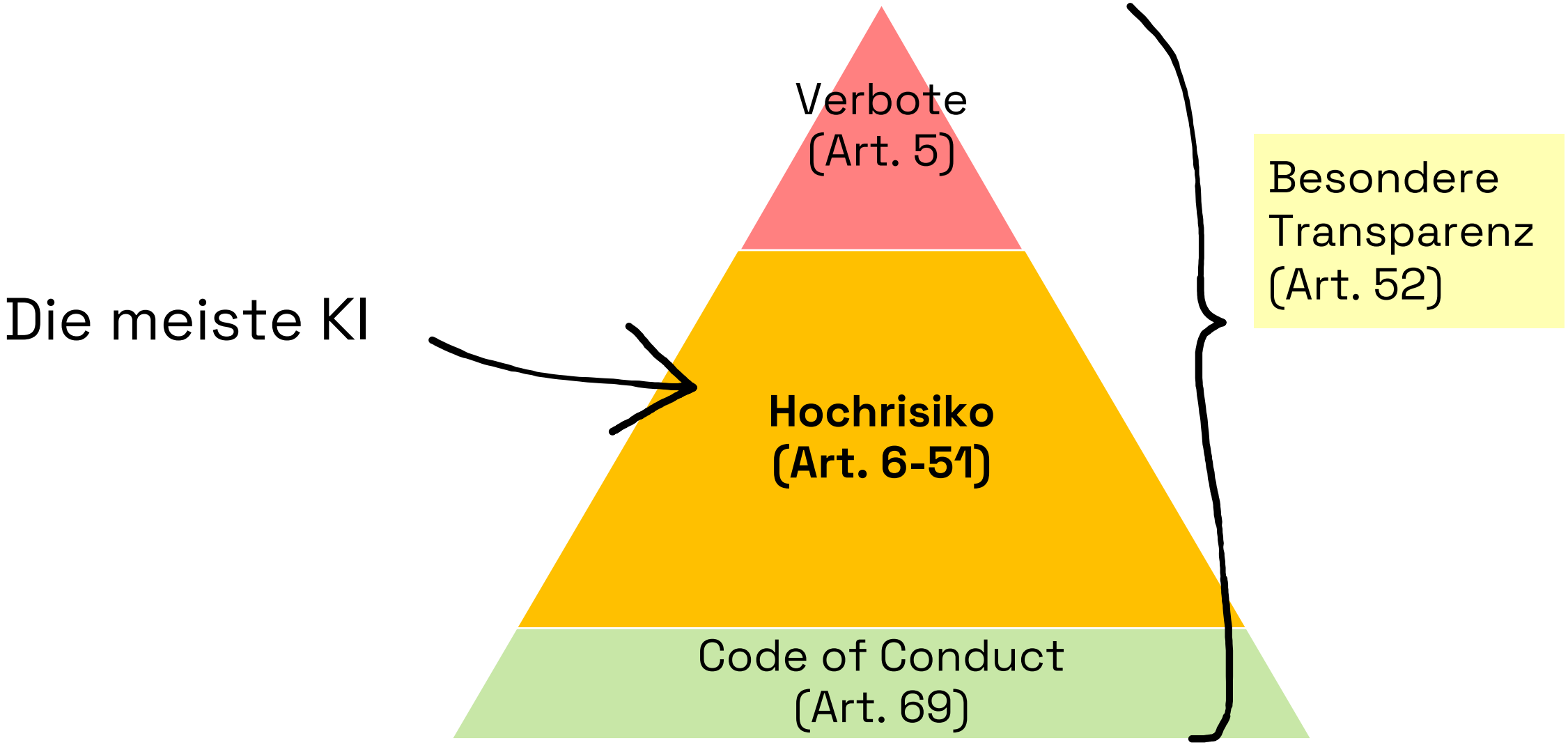
April 2021: Kommissionsentwurf



April 2021: Kommissionsentwurf

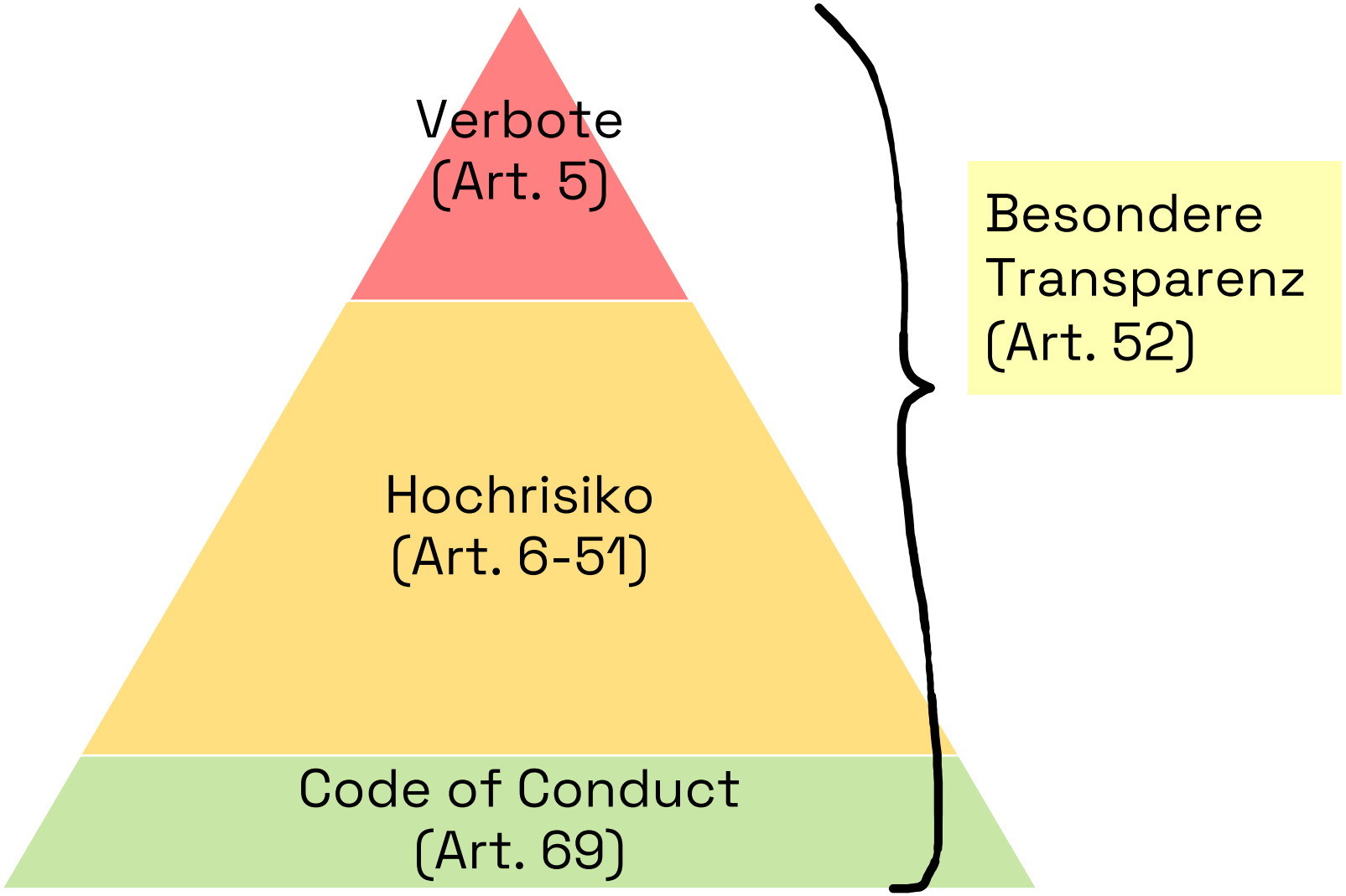


April 2021: Kommissionsentwurf



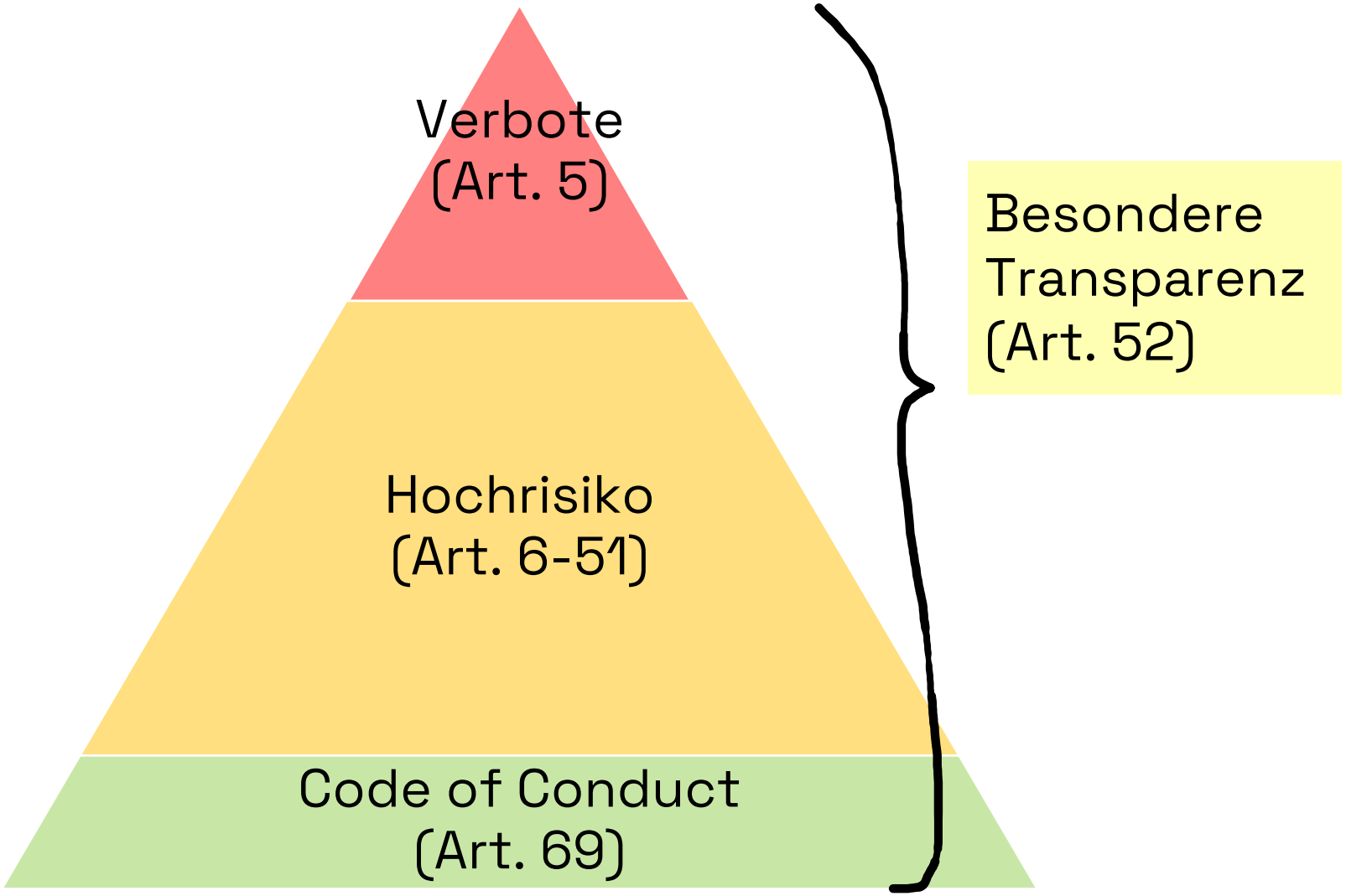
Juni 2023: Beschluss des Parlaments

Foundation Models (Art. 28b)



Aktuell: Trilogphase

Foundation Models (Art. 28b)



Aktuell: Trilogphase

Foundation Models (Art. 28b)

Worüber wird gestritten?

- Allgemeiner Auskunftsanspruch

Art. 28b AI Act ParlE: „Verwendung von urheberrechtlich geschützten Ausbildungsdaten […] öffentlich zugänglich machen.“

Leak: „demonstrate that they [providers] have taken adequate measures to ensure the models are trained in compliance with applicable Union copyright law”

Aktuell: Trilogphase

Foundation Models (Art. 28b)

Worüber wird gestritten?

- 2-Tier-Approach

  Leak Rat:

  ”Providers of very capable foundation models [...] should be subject to additional obligations”

  “a workable mechanism could be to presume a foundation model is ’very capable’ when the threshold of FLOPs is reached”

ITAVA

Aktuell: Trilogphase

Worüber wird gestritten?

Foundation Models (Art. 28b)

‘general-purpose AI model’ means an AI model, including when trained with a large amount of data using self-supervision at scale, that is capable to [competently] perform a wide range of distinctive tasks regardless of the way the model is released on the market. Research, development, and prototyping activities preceding the release on the market are not covered.

‘general-purpose AI system’ means an AI system based on an AI model that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems;

Insight aus Brüssel

Wenn politische Einigung bis 06.12.2023, dann:

> Inkrafttreten voraussichtlich **Anfang 2024**

> Umsetzungsfrist für Unternehmen: 24 bzw. 36 Monate

Wenn keine politische Einigung bis 06.12.2023, dann:

> Verabschiedung in den nächsten 12 Monaten sehr unwahrscheinlich (wegen neuer Ratspräsidentschaft, Europawahl)

> **Neustart** der Trilogverhandlung frühestens im **Oktober 2024**

“Feeding AI systems on the world’s beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy.“

Ruha Benjamin

Herzlichen Dank!

Dr. Jonas Siglmüller
Rechtsanwalt & Softwareentwickler
jonas.siglmueller@aitava.com

Let's connect!