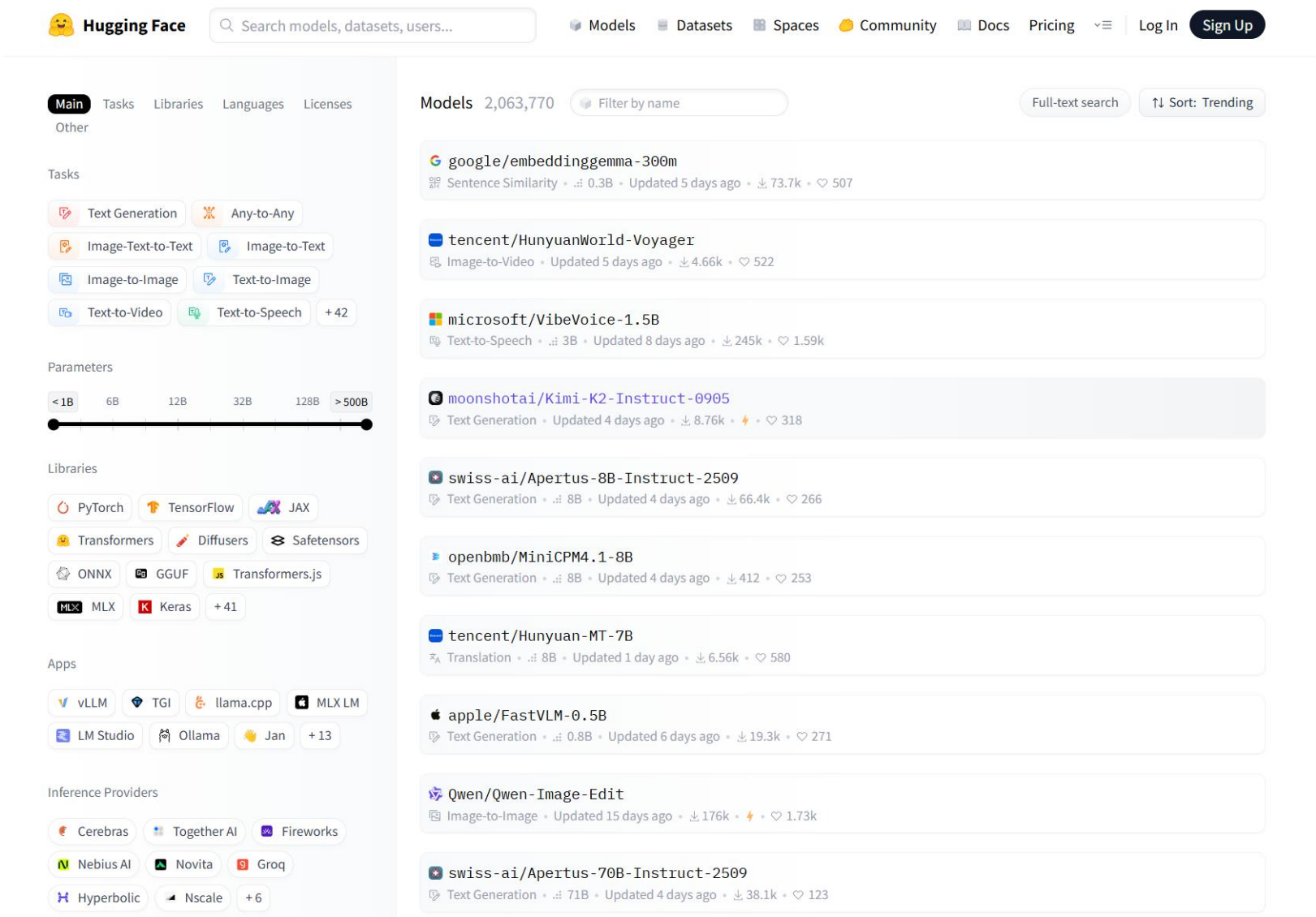
Deutsche Stiftung für Recht und Informatik

Modifikation von GPAI-Modellen und ihre Rechtsfolgen

Eine Fallstudie

Tim Schmetzer
Osborne Clarke

Herbstakademie 2025



https://huggingface.co/models

Anbieter (Art. 3 Nr. 3 KI-VO)

Eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle,

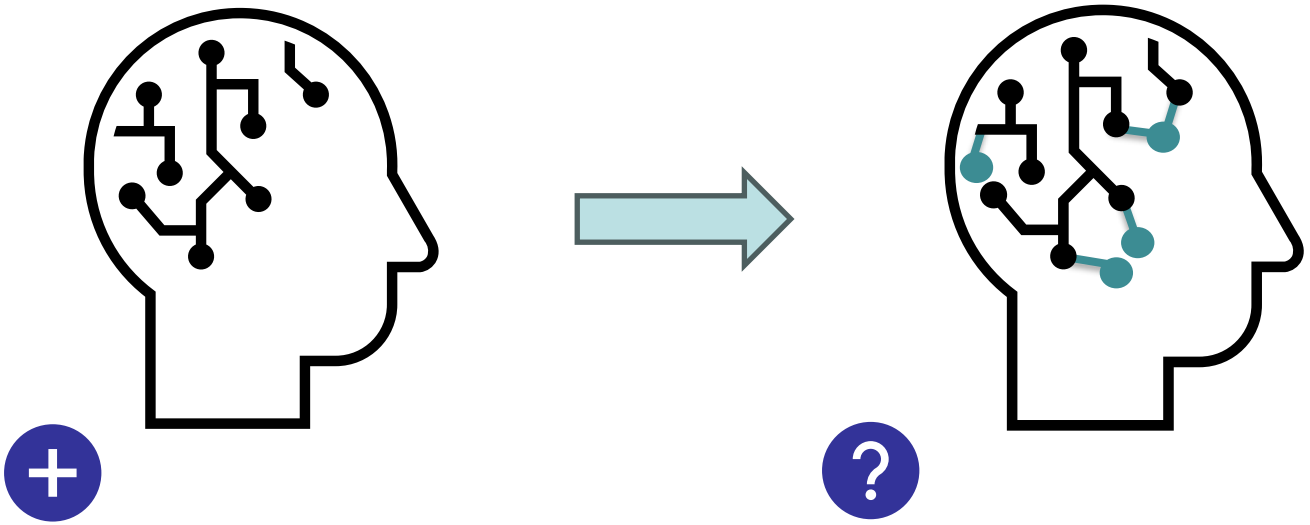
die ein KI-System oder ein KI-Modell mit allgemeinem Verwendungszweck **entwickelt oder entwickeln lässt**

und es unter ihrem eigenen Namen oder ihrer Handelsmarke **in Verkehr bringt oder**

ODER

das **KI-System** unter ihrem eigenen Namen oder ihrer Handelsmarke **in Betrieb nimmt**, sei es entgeltlich oder unentgeltlich;

Anbieter von GPAI-Modellen: Entwickeln oder entwickeln lassen

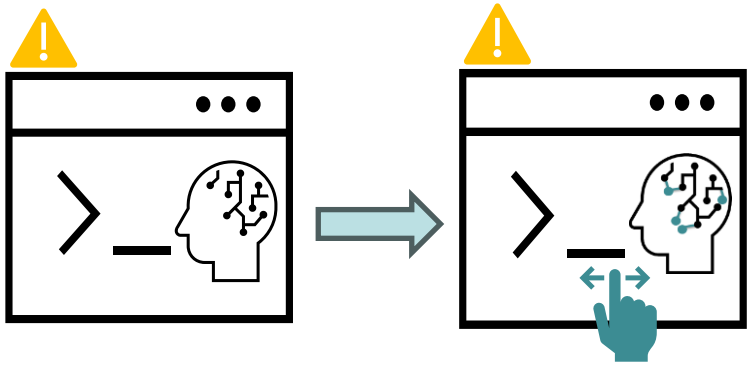


Entwickeln (lassen) von Grund auf

Modifikation (downstream) von bereits in Verkehr gebrachtem GPAI-Modell

Signifikante Veränderung des GPAI-Modells

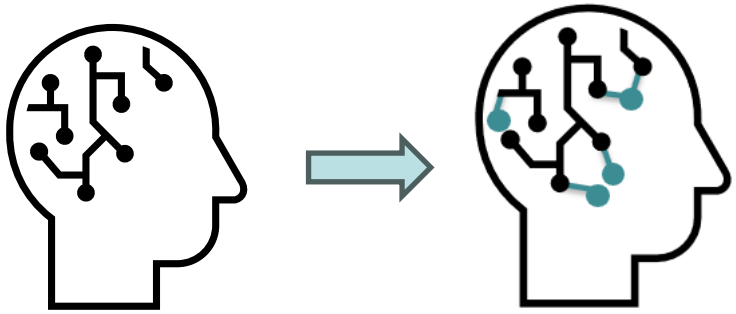
Hochrisiko-KI-Systeme: “Wesentliche Veränderung” (Art. 25 Abs. 1 lit. b KI-VO)



Entwickeln (lassen) von Grund auf

Anbieterfiktion bei **wesentlicher Veränderung** bereits in Verkehr gebrachter KI-Systeme (Art. 25 Abs. 1 lit. b KI-VO)

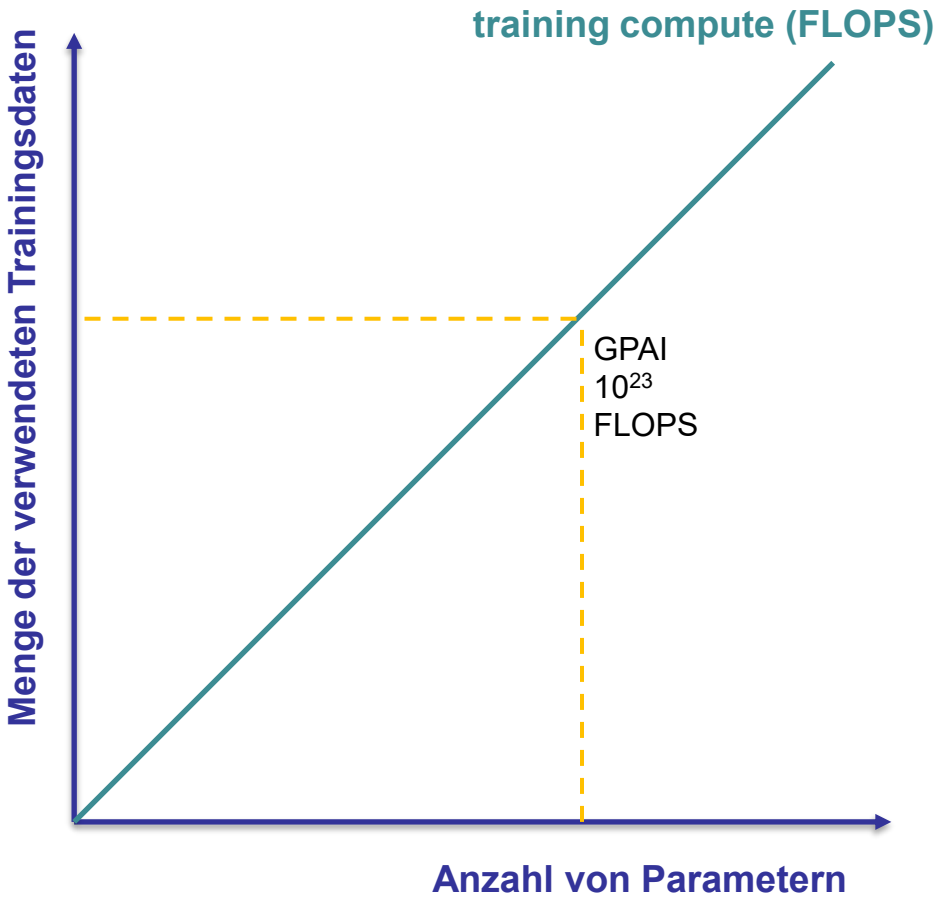
GPAI-Modelle: Signifikante Änderung (vgl. Rn. 62 GPAI-Guideline)



Downstream-Modifikation als “Entwickeln oder entwickeln lassen”?

NEU

GPAI-Guideline der Kommission: *Training Compute (1)*

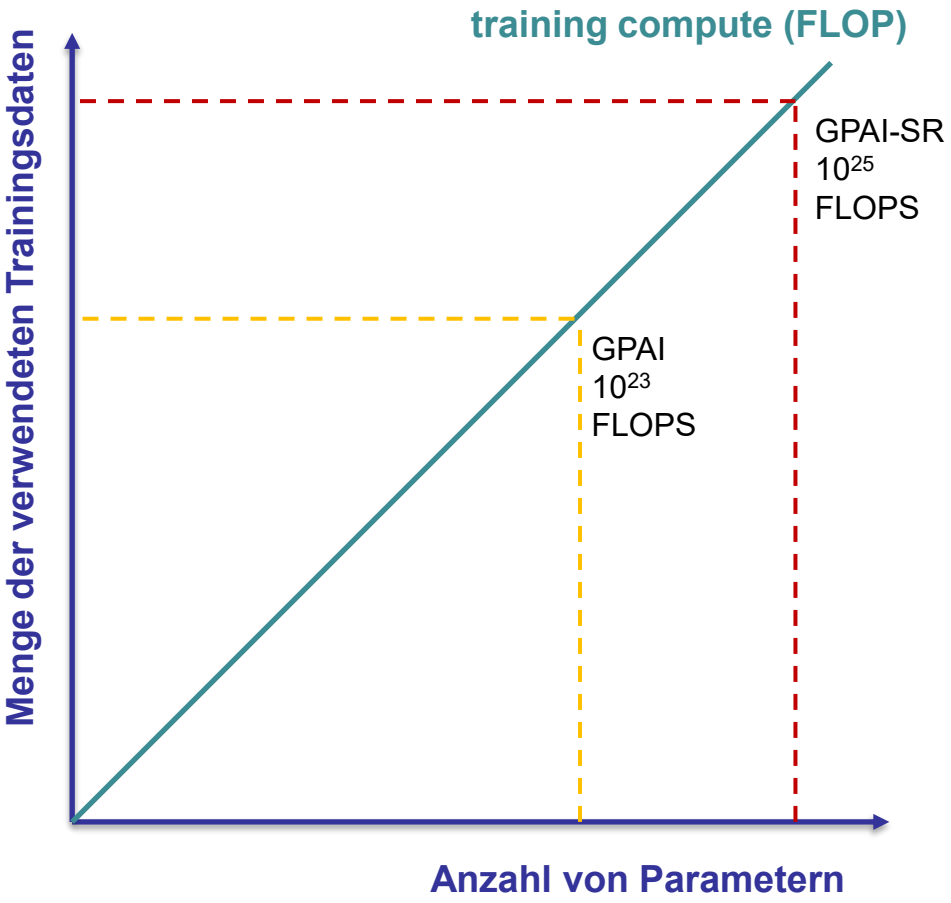


- Messbares Kriterium zur Einordnung als **GPAI-Modell**: *training compute* in *Floating point operations per second (FLOPS)*
- Widerlegliche Vermutung, dass ein KI-Modell ein GPAI-Modell ist, wenn es:
  1. Sprache (Text oder Audio), Text-to-image oder Text-to-Video generieren kann

  UND

  2. Mit mehr als 10^{23} FLOPS trainiert wurde

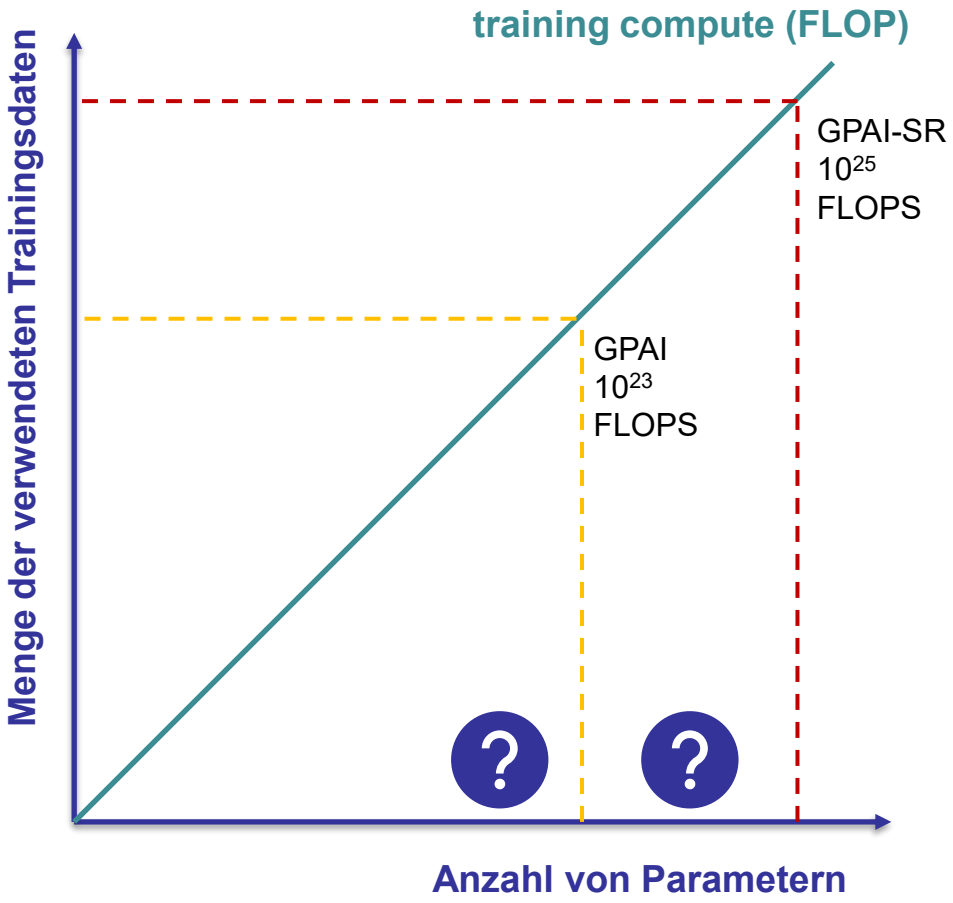
GPAI-Guideline der Kommission: *Training Compute (2)*



- Widerlegliche Vermutung, dass ein GPAI-Modell ein **GPAI-SR-Modell** ist, wenn es: mit mehr als 10^{25} FLOPS trainiert wurde (Art. 51 Abs. 2 KI-VO)

NEU

GPAI-Guideline der Kommission: *Training Compute (3)*



Signifikante Änderung durch Downstream-Modifier

- Grundsatz: 1/3 der tatsächlich aufgewendeten Rechenkapazität zum Training des originalen Modells (widerlegliche Vermutung)
- **Falls <u>nicht ermittelbar</u>:**
  - 10^{23} FLOPS für GPAI-Modelle
  - 10^{25} FLOPS für GPAI-SR-Modelle

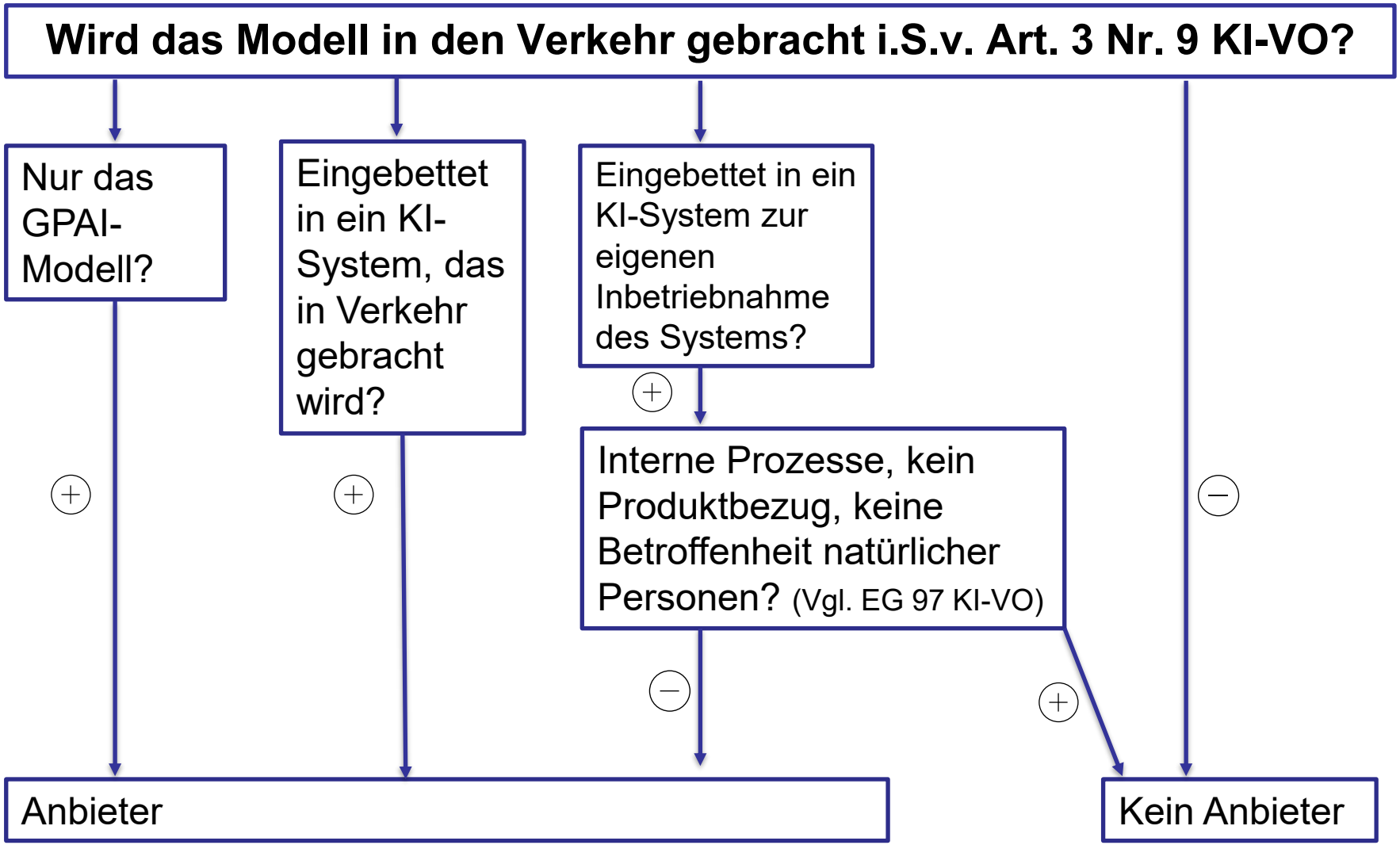
?

Prüfungsschema: Entwickeln oder entwickeln lassen

1. Liegt eine signifikante Änderung vor?

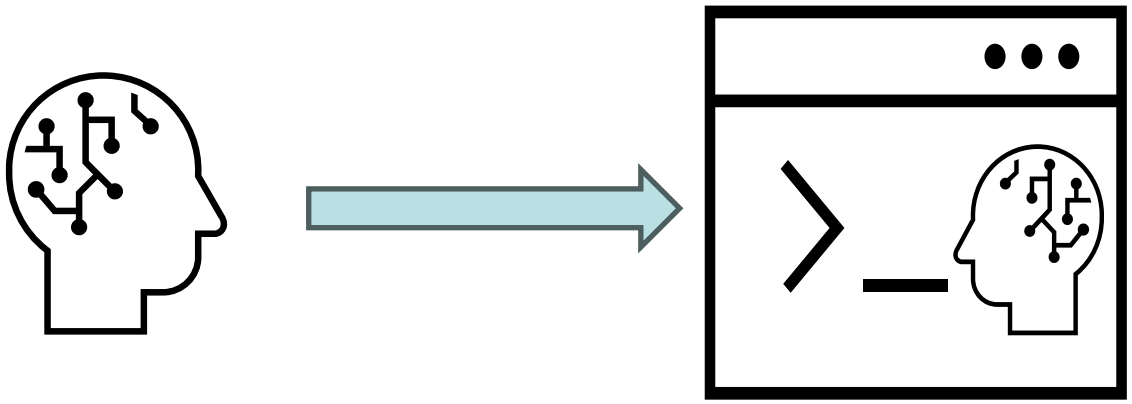
2. Weist das modifizierte Modell weiterhin einen allgemeinen Verwendungszweck i.S.d. Art. 3 Nr. 63 KI-VO auf?

3. Wird das Modell in Verkehr gebracht i.S.v. Art. 3 Nr. 9 KI-VO?


Wird das Modell in den Verkehr gebracht i.S.v. Art. 3 Nr. 9 KI-VO?
Nur das GPAI-Modell?
Eingebettet in ein KI-System, das in Verkehr gebracht wird?
Eingebettet in ein KI-System zur eigenen Inbetriebnahme des Systems?
Interne Prozesse, kein Produktbezug, keine Betroffenheit natürlicher Personen? (Vgl. EG 97 KI-VO)
Anbieter
Kein Anbieter

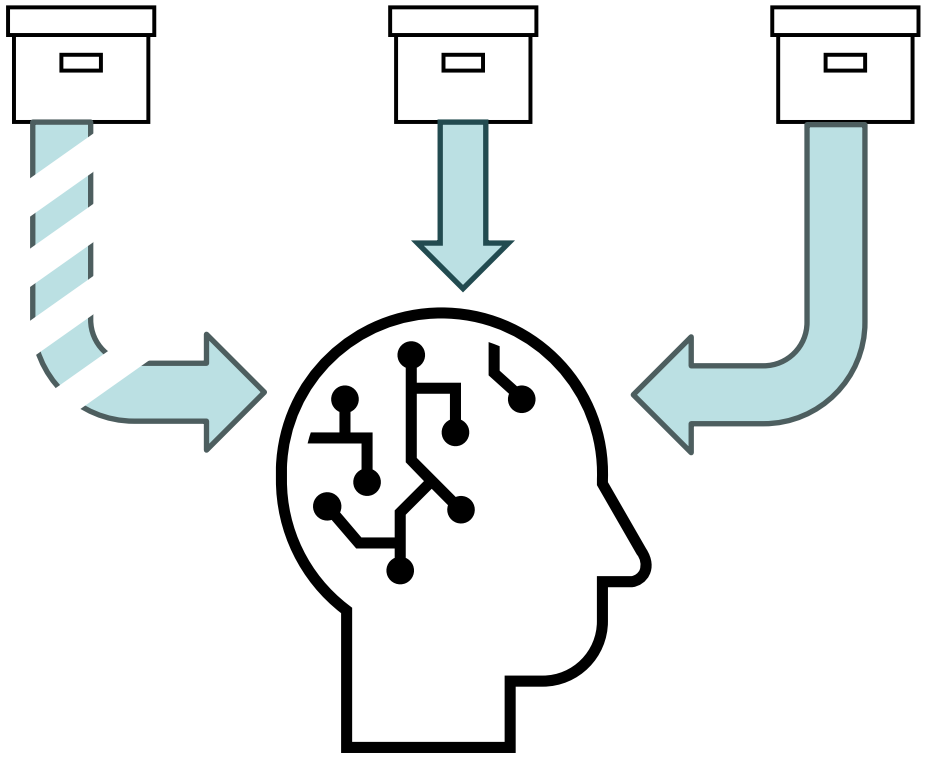
Fallgruppen

1. Integration eines nicht modifizierten GPAI-Modells in ein KI-System

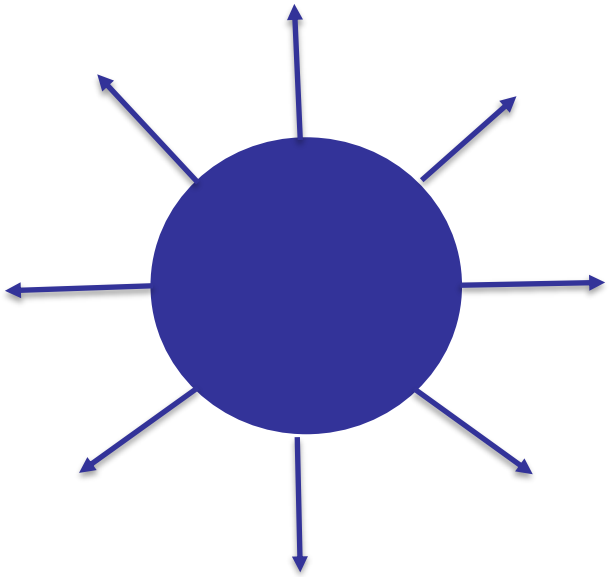


- Grds. **keine Anbieterrolle des Downstream-Modifiers**
- Aus ErwG 97 KI-VO folgt nichts anderes
  - Durch Auslegungsgrundsatz in ErwG 97 soll nur Schutzlücke in der KI-Wertschöpfungskette geschlossen werden (kein im Verkehr befindliches GPAI-Modell ohne Anbieter)
  - Vorliegend bleibt der ursprüngliche GPAI-Anbieter weiterhin Anbieter (dasselbe GPAI-Modell) → keine Schutzlücke in der KI-Wertschöpfungskette

2. Veränderung von Datenquellen

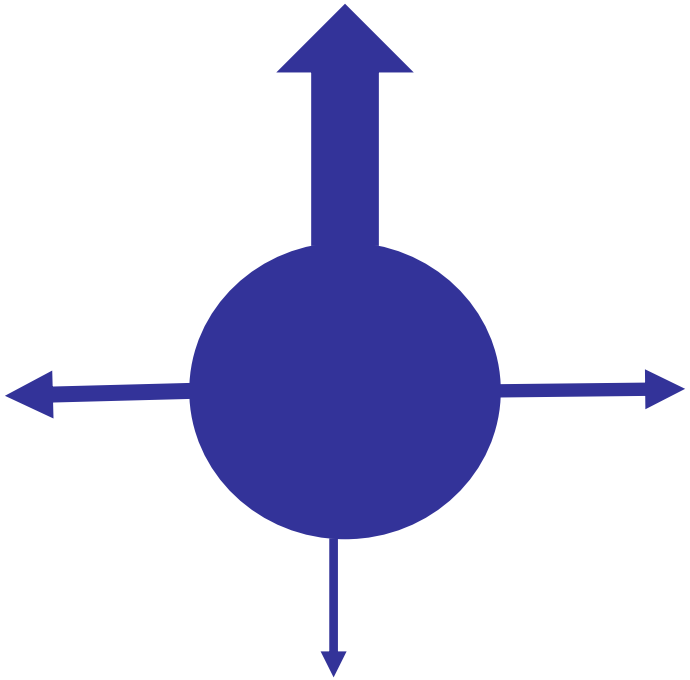


3. Fine-Tuning



Ursprüngliches GPAI-Modell

Training für allgemeine Aufgaben

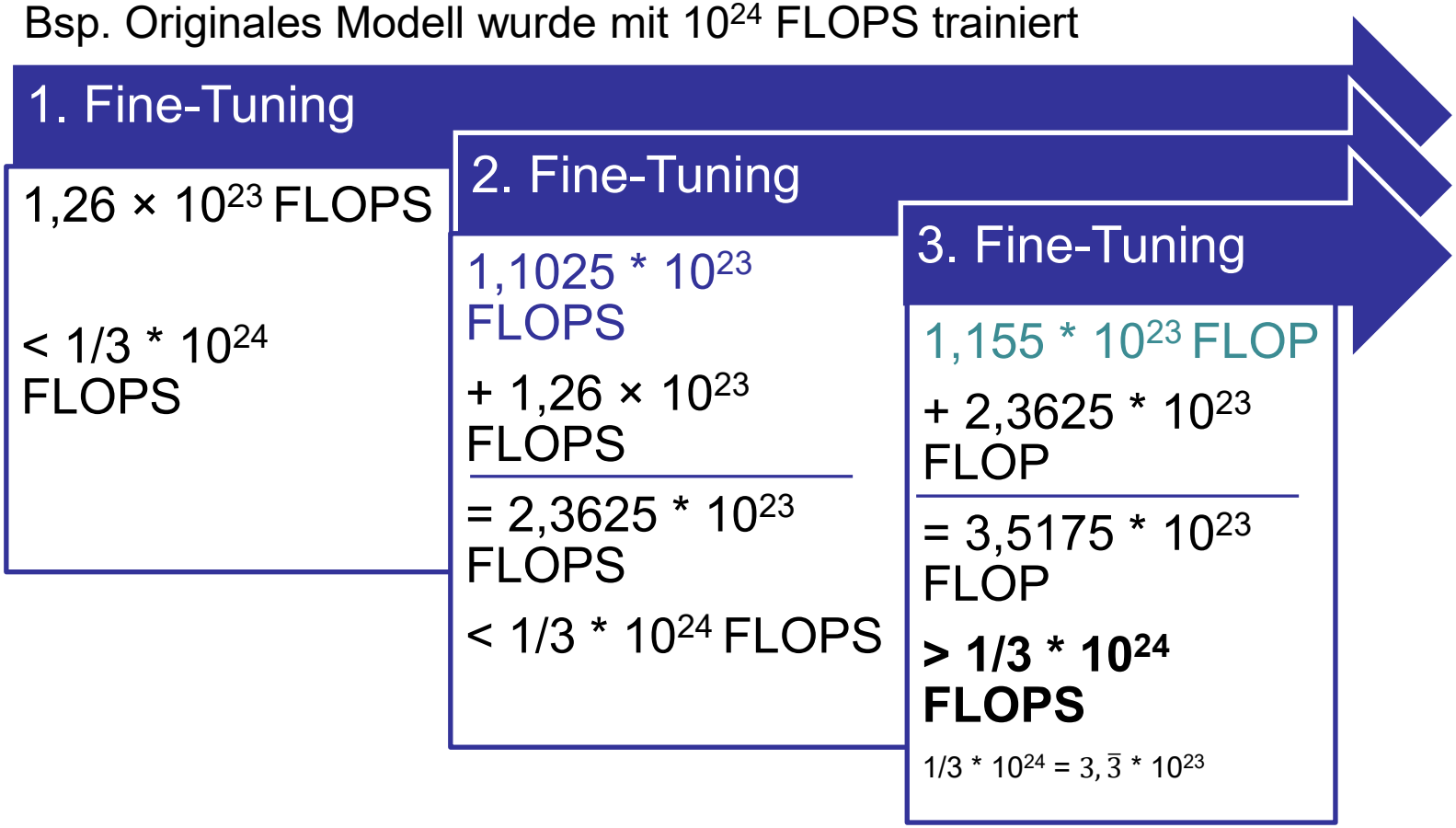


GPAI-Modell mit Fine-Tuning

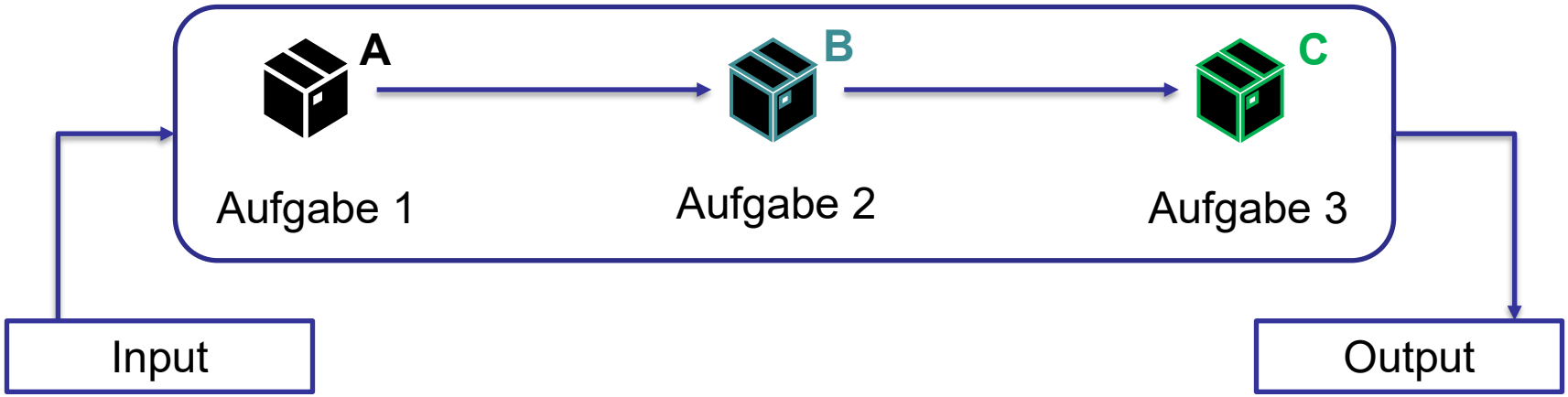
Training für spezifische Aufgaben

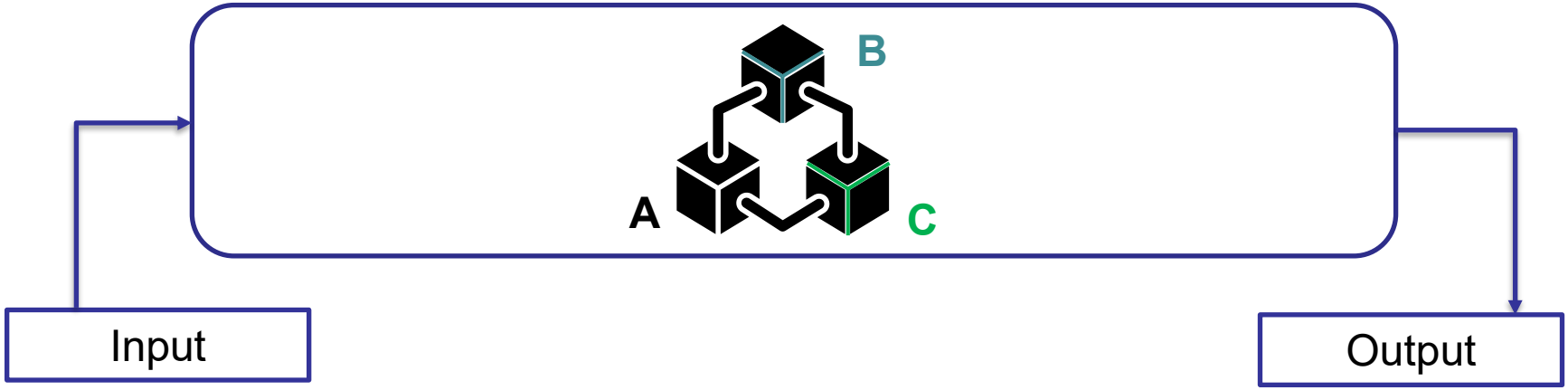
3. Fine-Tuning

Bsp. Originales Modell wurde mit 10^{24} FLOPS trainiert



4. Model Merging und Model Fusion

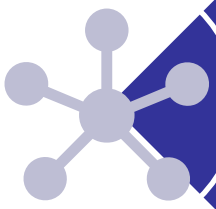




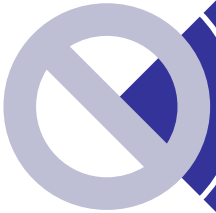
Fazit und Ausblick

Kritik am Kriterium *training compute*

Zweifelhafter langfristiger Bestand des Kriteriums



Zusammenhang zwischen der zum Training verwendeten Rechenleistung und den Fähigkeiten des KI-Modells zweifelhaft



Nur zur Bewertung von (unsupervised) Fine-Tuning geeignet

Questions & Answers