fortiss

BAVARIAN CENTER FOR SOFTWARE INNOVATION

Technische Hochschule Ingolstadt

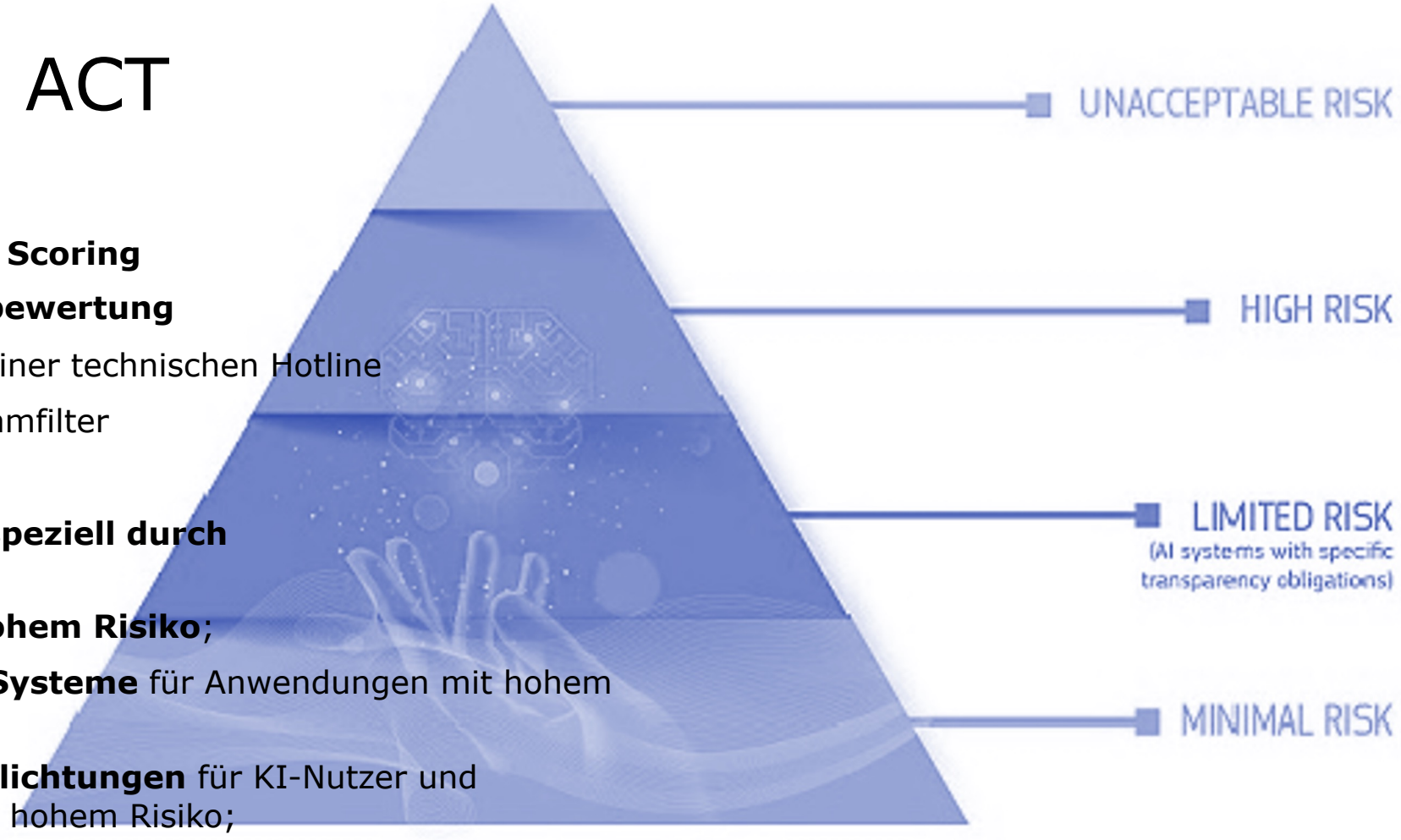
fortiss

Ostbayerische Technische Hochschule Amberg-Weiden

#1: EU AI Act ist ein risikobasierter Ansatz

Einführung EU AI ACT

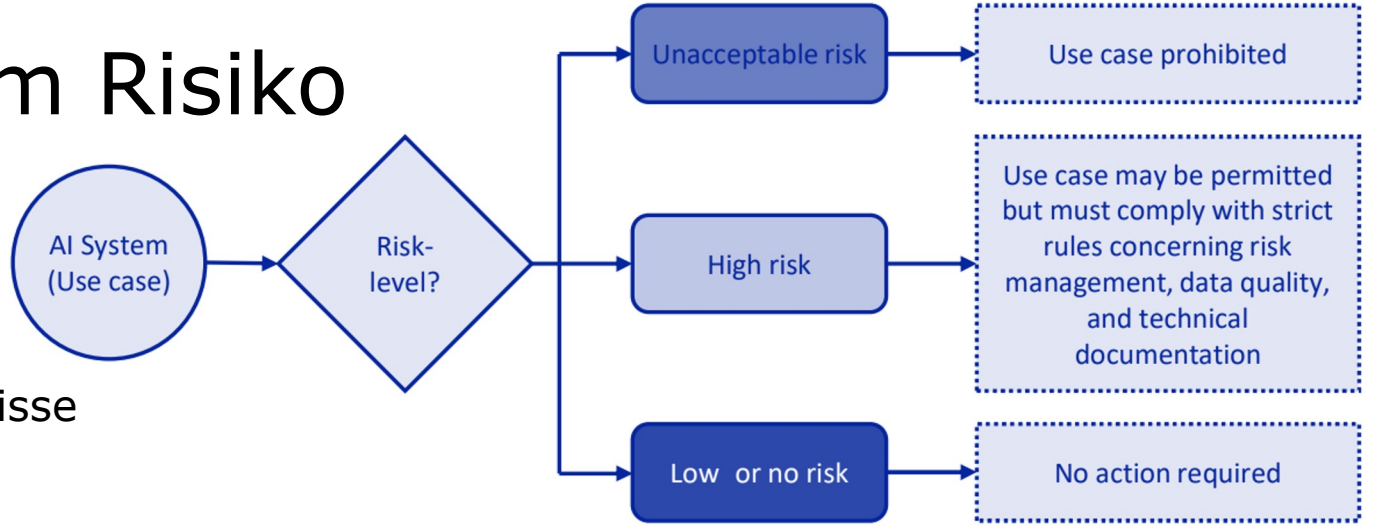
- **Risikobasierter Ansatz**
  - **Inakzeptables Risiko: Social Scoring**
  - **Hohes Risiko: HR-Personenbewertung**
  - Begrenztes Risiko: Chatbot in einer technischen Hotline
  - Minimales oder kein Risiko: Spamfilter
- Zielsetzung
  - Bewältigung von **Risiken**, die **speziell durch KI-Anwendungen** entstehen;
  - eine **Liste der Anträge mit hohem Risiko**;
  - **klare Anforderungen an KI-Systeme** für Anwendungen mit hohem Risiko festlegen;
  - Festlegung **spezifischer Verpflichtungen** für KI-Nutzer und Anbieter von Anwendungen mit hohem Risiko;
  - **Konformitätsbewertung** vorschlagen, bevor das KI-System in Betrieb genommen oder in Verkehr gebracht wird;
  - schlägt die Durchsetzung vor, nachdem ein solches KI-System in Verkehr gebracht wurde;
  - Vorschlag für eine Governance-Struktur auf europäischer und nationaler Ebene.



Quelle: EK
https://ec.europa.eu/information_society/newsroom/image/document/2021-17/pyramid_7F5843E5-9386-8052-931F5C4E98C6E5F2_75757.jpg

Anforderungen bei hohem Risiko

- angemessene Risikobewertungs- und Risikominderungssysteme;
- hohe **Qualität der Datensätze**, die das System füttern, um Risiken und diskriminierende Ergebnisse zu minimieren;
- **Erfassung** der Tätigkeit, um die **Rückverfolgbarkeit** der Ergebnisse zu gewährleisten;
- ausführliche **Unterlagen**, die alle erforderlichen Informationen über das System und seinen Zweck enthalten, damit die **Behörden die Einhaltung der Vorschriften bewerten** können;
- klare und angemessene Informationen für den Nutzer;
- angemessene menschliche Aufsichtsmaßnahmen zur Minimierung des Risikos;
- hohe **Robustheit, Sicherheit und Genauigkeit**.



Quelle: capAI (Floridi et al., 2021)

Beispiele für Anwendungsgebiete:

- kritische Infrastrukturen (z. B. Verkehr)
- Bildungs- oder Berufsausbildung
- Sicherheitskomponenten von Produkten
- Beschäftigung, Management von Arbeitnehmern und Zugang zu selbstständiger Erwerbstätigkeit (z. B. Software zur Auswahl von Lebensläufen)
- wesentliche private und öffentliche Dienstleistungen (z. B. Kreditbewertung)
- Strafverfolgung
- Migrations-, Asyl- und Grenzkontrollenmanagement
- Rechtspflege und demokratische Prozesse

#2: Anforderungen werden gestaffelt verpflichtend

Zeitleiste für das Inkrafttreten der Regelungen

2025

Datum		Verwandte Inhalte des AI-Gesetzes
2. Februar 2025	**Anwendung:** Die Verbote für bestimmte KI-Systeme beginnen zu gelten(Kapitel 1 und Kapitel 2).	Artikel 113 Buchstabe a Erwägungsgrund 179
2. August 2025	**Anwendung:** Die folgenden Regeln beginnen zu gelten: • Benannte Stellen(Kapitel III, Abschnitt 4), • GPAI-Modelle(Kapitel V), • Governance(Kapitel VII), • Vertraulichkeit(Artikel 78) • Sanktionen (Artikel 99 und 100)	Artikel 113 Buchstabe b

2026

Datum		Verwandte Inhalte des AI-Gesetzes
2. August 2026	**Anwendung:** Die übrigen Bestimmungen des AI-Gesetzes finden Anwendung, mit Ausnahme von Artikel 6 Absatz 1.	Artikel 113
2. August 2025	**Anbieter:** Anbieter von GPAI-Modellen, die vor diesem Datum in Verkehr gebracht bzw. in Betrieb genommen wurden, müssen bis zum 2. August 2027 mit dem AI-Gesetz konform sein.	Artikel 111 Absatz 3

+ März 2025: Verhaltenskodex
- Bereitstellung von technischen Unterlagen (z. B. Fähigkeiten und Grenzen)
- Zusammenfassungen der verwendeten Trainingsdaten
- Maßnahmen zur Einhaltung des geltenden Unionsrechts zum Urheberrecht
- Zusätzlich für Für GPAI-Modelle: Modellbewertungen nach dem Stand der Technik, Risikobewertung und –minderung, Meldung schwerwiegender Zwischenfälle, einschließlich Abhilfemaßnahmen, Angemessener Schutz der Cybersicherheit

+ April 2025: Abgabetermin für Ergebnisse aus dem Standardisierungsauftrag C(2023)3215

Quelle: https://artificialintelligenceact.eu/de/implementation-timeline/

Zeitleiste für das Inkrafttreten der Regelungen (2)

2026

Datum		Verwandte Inhalte des AI-Gesetzes
2. August 2026	**Betreiber:** Diese Verordnung gilt für Betreiber von AI-Systemen mit hohem Risiko *(mit Ausnahme der in Artikel 111 Absatz 1 genannten Systeme)*, die vor diesem Datum in Verkehr gebracht/in Betrieb genommen wurden. Dies gilt jedoch nur für Systeme, deren Konstruktion ab diesem Zeitpunkt wesentlich geändert wird.	Artikel 111 Absatz 2

2027

Datum		Verwandte Inhalte des AI-Gesetzes
2. August 2027	**Anwendung**: Artikel 6(1) und die entsprechenden Verpflichtungen in der Verordnung beginnen zu gelten.	Artikel 113
2. August 2027	**Anbieter:** Anbieter von GPAI-Modellen, die vor dem 2. August 2025 in Verkehr gebracht werden, müssen bis zu diesem Datum die erforderlichen Maßnahmen ergriffen haben, um den in dieser Verordnung festgelegten Verpflichtungen nachzukommen.	Artikel 111 Absatz 3
2. August 2027	**IT-Großsysteme:** KI-Systeme, die Bestandteile der in Anhang X aufgeführten IT-Großsysteme sind und vor diesem Datum in Verkehr gebracht/in Betrieb genommen wurden, müssen bis zum 31. Dezember 2030 mit dieser Verordnung in Einklang gebracht werden.	Artikel 111 Absatz 1

Quelle: https://artificialintelligenceact.eu/de/implementation-timeline/

Zeitleiste für das Inkrafttreten der Regelungen (3)

2031

Datum		Verwandte Inhalte des AI-Gesetzes
2. August 2031	**Kommission:** Die Kommission nimmt eine Bewertung der Durchsetzung dieser Verordnung vor und erstattet dem Europäischen Parlament, dem Rat und dem Europäischen Wirtschafts- und Sozialausschuss darüber Bericht.	Artikel 112 Absatz 13

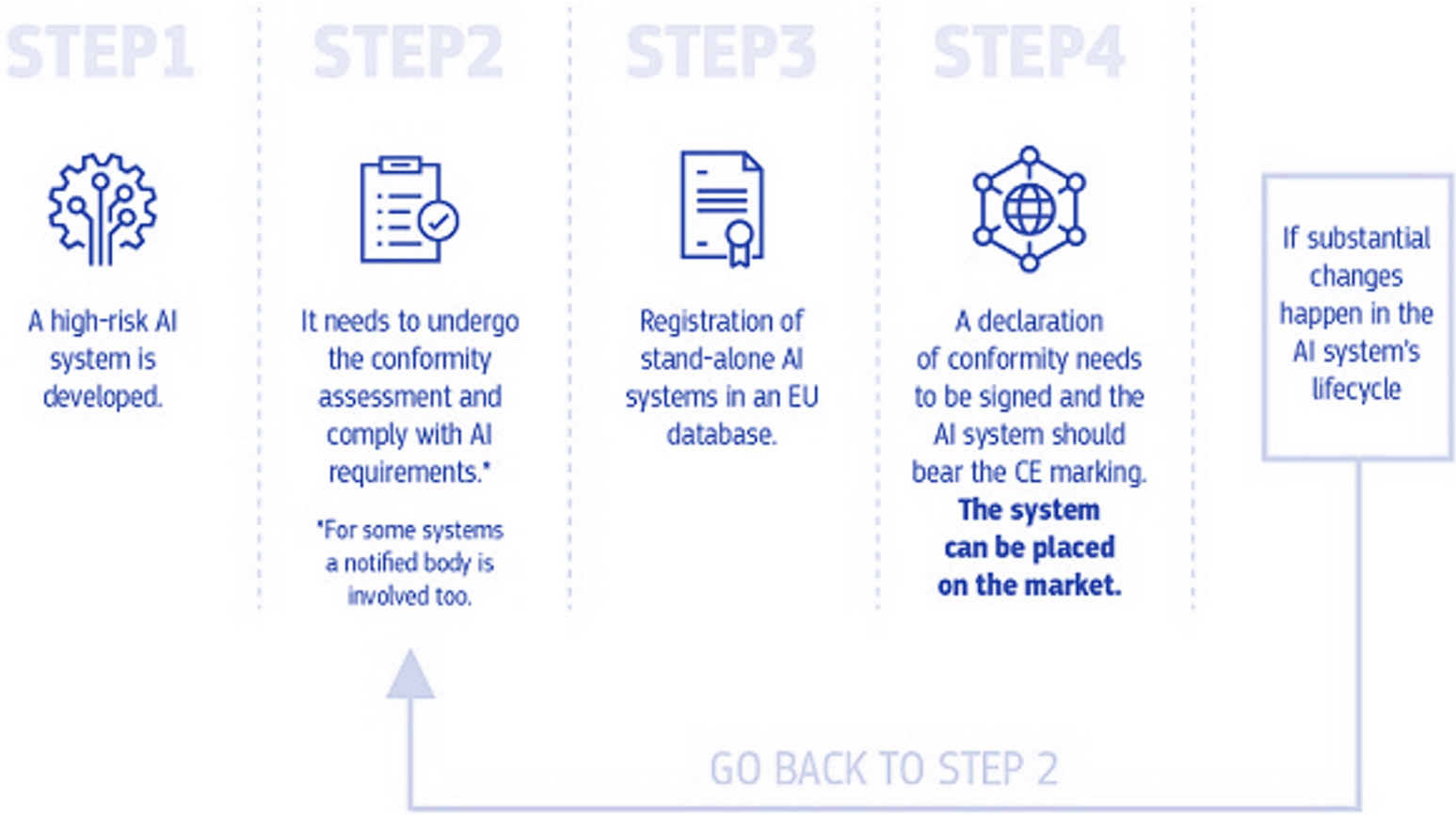
Quelle: https://artificialintelligenceact.eu/de/implementation-timeline/

#3: Idealtypische Umsetzung laut der EK

Wie soll laut EK die Umsetzung in der Praxis aussehen?

- Fokus auf „Endprodukt"
- Anbieter soll die Prüfung durchführen, in manchen Fällen werden „notified bodies" involviert
- Eintragung in EU-Register
- Fortlaufend: bei Änderungen
- Was ist mit Kontext? Fine-tuning? Agentic / Compound AI?
- Kann ein Assessment **nach** der Entwicklung stattfinden?

- aktueller Stand: **nicht umsetzbar**



Quelle: EK https://digital-strategy.ec.europa.eu/de/policies/regulatory-framework-ai

Und was ist mit DSGVO?

Orientierungshilfen-Navigator KI & Datenschutz (ONKIDA)

	A. EDPS Guidelines on generative AI and the EUDPR (2024, PDF)↗ *Datenverarbeitung durch EU-Organe*	B. Report der EDSA Taskforce ChatGPT (2024, PDF)↗	C. DSK: Orientierungshilfe zu KI und Datenschutz (2024, PDF)↗	D. LfDI BW: Rechtsgrundlagen zum Einsatz von KI (2023)	E. BayLDA: Checkliste Datenschutzkonforme KI (2024, PDF)↗	F. Hamburger BfDI: Checkliste zum Einsatz LLM-basierter Chatbots (2023, PDF)↗	G. CNIL: Recommendations on the development of AI systems („How-to sheets") (2024)↗	H. DSB Österreich: FAQ KI und Datenschutz (2024)↗	I. DSK: Positionspapier zu TOM bei Entwicklung und Betrieb von KI-Systemen (2019, PDF)↗
1. Grundsatz der Datenrichtigkeit Art. 5↗ I lit. d) DSGVO	(+) S. 15 f. (Art. 4 I lit. d) VO 2018/1725)	(+) Rn. 29 ff. sowie im Fragebogen im Annex, S. 11	(+/–) Recht auf Berichtigung Rn. 27, Überprüfung der Richtigkeit der Ergebnisse Rn. 64 f.	(–)	(+/–) Recht auf Berichtigung, S. 6, 10	(+/–) Überprüfung der Richtigkeit des Ergebnisses S. 4	(+/–) „data cleaning", „monitoring and updating" Sheet 7	(+)	(–)
2. Grundsatz der Datenminimierung Art. 5↗ I lit. c) DSGVO **Zweckbindungsgrundsatz** Art. 5↗ I lit. b) DSGVO	(+) Datenminimierung: S. 14 (Art. 4 I lit. c) VO 2018/1725) (+/–) Zweckbindung: nur sehr indirekt („consistent with original purpose"), S. 12	(+/–) nur im Rahmen des Fragebogens im Annex, S. 10	(+) Zweckbindung Rn. 1 f.	(+/–) Berücksichtigung Datenminimierung bei Art. 6 I lit. f DSGVO (S. 17) u. § 13 LDSG BW (S. 25) (+) Zweckänderung S. 15	(+/–) Zweckbindung nur eher indirekt S. 6, 8, 11 (Checkliste)	(–)	(+) Sheet 2, Datenminimierung auch Sheet 6, Zweckkompatibilität auch Sheet 4 (2/2)	(+)	(+) Datenminimierung S. 9, 14, 17 (+) Zweckbindung S. 6 f., 7 (Fragebogen), 8, 9, 14, 17
3. Personenbezug Art. 4↗ Nr. 1 DSGVO	(+) S. 7 (Art. 3 Nr. 1 VO 2018/1725)	(–)	(+) Rn. 4 ff., 7 f., 48 ff.	(+) insbes. S. 6	(+) S. 4, 5, 9,10, 11 (Checklisten)	(+) S. 2 f. *vgl. auch Hamburger Thesen zum Personenbezug in Large Language Models↗ v. 15.7.2024*	(+) Introduction	(–)	(+) S. 15 kurzer Satz im Zusammenhang mit Vertraulichkeit beim Training
4. Rechtsgrundlagen für die Datenverarbeitung Art. 6↗ I u. 9↗ II DSGVO	(+) S. 11 ff. (Art. 5 und 10 II VO 2018/1725)	(+) Rn. 13 ff., ebenso im Fragebogen S. 12 f.	(+) Rn. 9 ff. (zudem Verweis auf Positionspapier LfDI BW), Rn. 62 (im Zusammenhang mit sensiblen Daten)	(+) insbes. S. 11 ff.	(+) S. 4 und 9 (Checklisten)	(+) S. 2 (indirekt im Zusammenhang mit Personenbezug) und S. 4 (im Zusammenhang mit Diskriminierung)	(+) Sheet 4 (1/2 und 2/2), Sheet 8 (in consultation)	(+/–) nur allgemeine Bezugnahme	(+/–) vereinzelt kurze Bezugnahmen, dass es einer Rechtsgrundlage bedarf
5. (Mit–)Verantwortlichkeit Art. 26↗ (und 28↗) DSGVO	(+) S. 6	(+/–) Rn. 23 ff. in Zusammenhang mit Fairness-Prinzip, „Abwälzung" der Verantwortlichkeit auf betroffene Personen; im Rahmen des Fragebogens S. 14	(+) Rn. 32 ff.	(+) S. 9 ff.	(+) S. 9	(–)	(+) Sheet 3	(–)	(+/–) indirekt: Klärung der Zugriffsmöglichkeiten von Cloud-Anbietern S. 16; „Rollen- und Berechtigungskonzept" S. 15,18,19
6. Transparenzgebot und Informationspflichten Art. 5↗ I lit. a und 12 ff. DSGVO	(+) S. 17 (Art. 14 VO 2018/1725)	(+) Rn. 27 f., ebenso im Fragebogen S. 13	(+) Rn. 21 ff.	(+) S. 12 (im Zusammenhang mit informierter Einwilligung)	(+) Transparenz S. 7 (als Teil des „Datenschutz-Risikomodells") (+) Infopflichten S. 5 (Checkliste)	(–)	(+) Sheet 2, Dokumentation in Sheet 7	(+)	(+) S. 5, 11 ff., 16 f.
7. Auskunftsanspruch Art. 15↗ DSGVO **Recht auf Löschung** Art. 17↗ DSGVO	(+/–) allgemein Betroffenenrechte S. 22	(+) allgemein Betroffenenrechte Rn. 32 ff.	(+) nur Recht auf Löschung Rn. 26, 28 f.; „weitere Betroffenenechte" Rn. 30	(+) nur Recht auf Löschung S. 12	(+) Auskunftsanspruch S. 5, 10 (Checkliste), Recht auf Löschung S. 6, 10 (Checkliste)	(–)	(–)	(+/–) nur allgemeine Bezugnahme auf Betroffenenrechte	(+) Auskunftsanspruch, S. 7; Betroffenenrechte allgemein S. 18 (in einem Satz)
8. Automatisierte Entscheidungen und Profiling Art. 22↗ DSGVO	(+) S. 18 (Art. 24 VO 2018/1725)	(–)	(+) Rn. 12 ff.	(–)	(–)	(+) S. 22	(–)	(+)	(+) S. 5 (mehr oder weniger), S. 14 (Bezugnahme in einem Satz, indirekt), S. 18
9. Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen Art. 25↗ DSGVO	(+) S. 9 (Art. 27 VO 2018/1725)	(+) Rn. 7 knappe Bezugnahme; Rn. 35 im Zusammenhang mit Betroffenenrechten	(+) Rn. 43	(+/–) S. 7 (Bewertung Personenbezug), S. 18 Fn. 57 (Berücksichtigung bei Art. 6 I lit. f DSGVO)	(+), S. 7 (nur ein knapper Satz)	(–)	(+) Sheet 6 (Vorschrift wird nicht direkt genannt, aber Konzept DPbD wird beschrieben), Sheet 7	(–)	(+) S. 7 (analog?)
10. Datenschutz-Folgenabschätzung Art. 35↗ DSGVO	(+) S. 9 f. (Art. 39 u. 89 VO 2018/1725)	(+/–) nur im Rahmen des Fragebogens im Annex, S. 11	(+) Rn. 38 ff.	(–)	(+) S. 4, 6, 9, 11 (Checklisten), S. 7	(+) S. 2	(+) Sheet 5	(–)	(+) S. 5

Quelle: https://www.baden-wuerttemberg.datenschutz.de/onkida/

Was ist der aktuelle Stand der Standardisierungsarbeitsgruppe?

Project reference	Status	Initial Date	Current Stage	Next Stage	Forecasted voting date
CEN/CLC/TR 17894:2024 (WI=JT021001) Artificial Intelligence - Artificial Intelligence Conformity Assessment	Under Approval	2022-05-18	2024-10-24	2024-11-25	
EN ISO/IEC 22989:2023/prA1 (WI=JT021031) Information technology — Artificial intelligence — Artificial intelligence concepts and terminology — Amendment 1	Under Drafting	2024-05-04	2024-05-04	2024-11-04	2026-06-15
EN ISO/IEC 23053:2023/prA1 (WI=JT021032) Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) — Amendment 1	Under Drafting	2024-05-06	2024-05-06	2024-11-06	2026-06-15
FprCEN/CLC/TR 18145 (WI=JT021010) Environmentally sustainable Artificial Intelligence	Under Approval	2023-01-18	2024-10-17	2025-01-09	
prCEN/CLC/TR XXX (WI=JT021026) Impact assessment in the context of the EU Fundamental Rights	Preliminary		2024-02-16		
prCEN/CLC/TR XXX (WI=JT021009) AI Risks - Check List for AI Risks Management	Preliminary		2022-09-30		
prCEN/CLC/TR XXXX (WI=JT021002) Artificial Intelligence - Overview of AI tasks and functionalities related to natural language processing	Under Drafting	2022-06-29	2023-11-03	2024-03-29	
prCEN/TS (WI=JT021033) Guidance for upskilling organisations on AI ethics and social concerns	Preliminary		2024-06-10		
prCEN/TS (WI=JT021035) Sustainable Artificial Intelligence – Guidelines and metrics for the environmental impact of artificial intelligence systems and services	Preliminary		2024-06-10		
prCEN/TS (WI=JT021034) Guidelines on tools for handling ethical issues in AI system life cycle	Preliminary		2024-06-10		
prEN ISO/IEC 12792 (WI=JT021022) Information technology - Artificial intelligence - Transparency taxonomy of AI systems (ISO/IEC DIS 12792:2024)	Under Approval	2023-08-23	2024-07-16	2025-03-17	2025-03-17
prEN ISO/IEC 23282 (WI=JT021012) Artificial Intelligence - Evaluation methods for accurate natural language processing systems	Under Drafting	2023-11-22	2023-11-22	2024-05-22	2026-01-02
prEN ISO/IEC 24029-2 (WI=JT021015) Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods	Preliminary		2023-01-10		
prEN ISO/IEC 24970 (WI=JT021021) Artificial intelligence — AI system logging	Under Drafting	2024-08-21	2024-08-21	2025-02-21	2026-09-29

- > 30 Standards in Vorbereitung
- Einhaltung des Anwendungsplans ab 2026 unrealistisch
- Ausarbeitung von Prüfverfahren steht noch aus

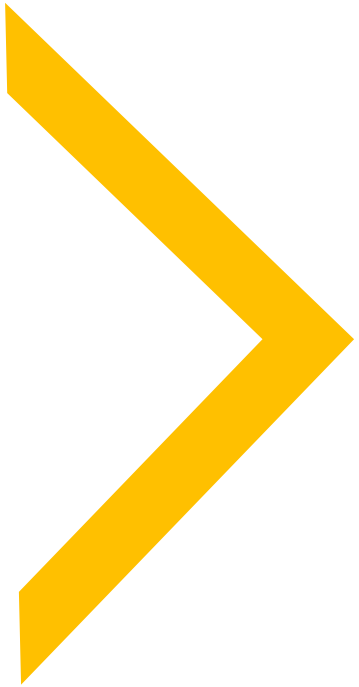
Quelle: CEN/CLC/JTC 21
https://standards.cencenelec.eu/dyn/www/f?p=205:22:0

#4: Umsetzung in der Praxis

EU AIA, auweiia …
Hürde oder Chance?

EU AI Act ist risikobasiert

- Fokus auf "Endprodukt"
- „Irgendwann gibt es konkrete Vorgaben"
- Überprüfung soll durch KI-Anbieter erfolgen (Achtung: Wer ist Anbieter?)
- Betreiber haftet bei Nutzung, durch Anpassung der KI gleichgestellt mit Anbieter



- Wie kann KI mit Blick auf den EU AI Act entwickelt and eingesetzt werden?
- Wie kann KI mit Blick auf den EU AI Act und die Vielzahl von Standards auditiert werden bzw. Konformität sichergestellt werden?

Mitigationsstrategien für KI-Anwender

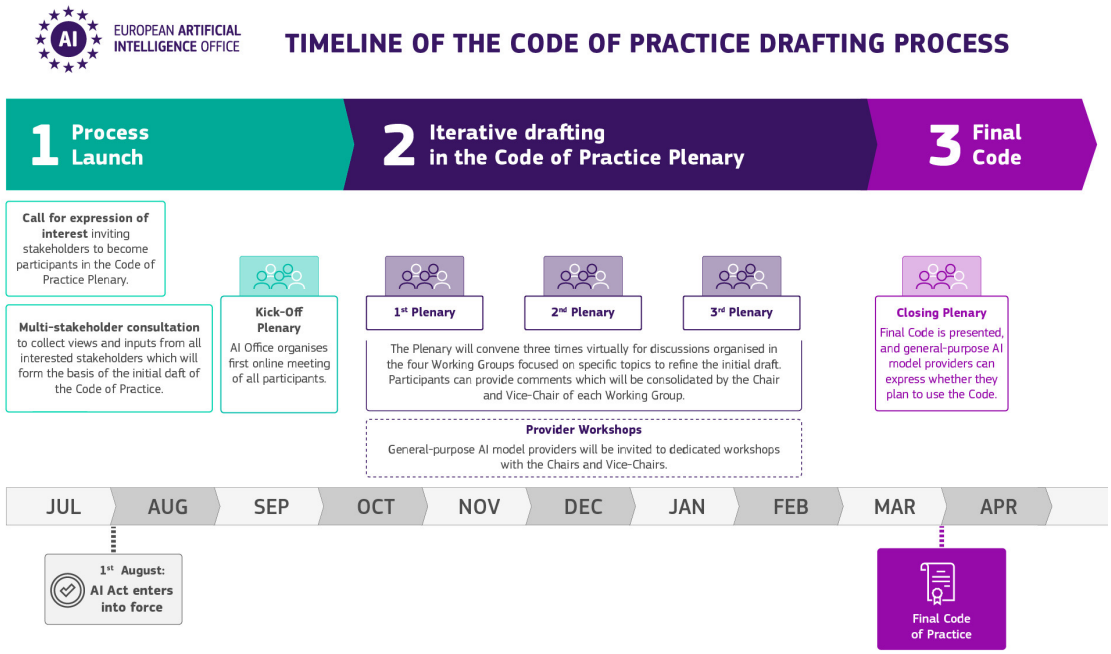
KI nicht nutzen?

Konformitätspflichten reduzieren? (bereits im EU AI Act erfasst)

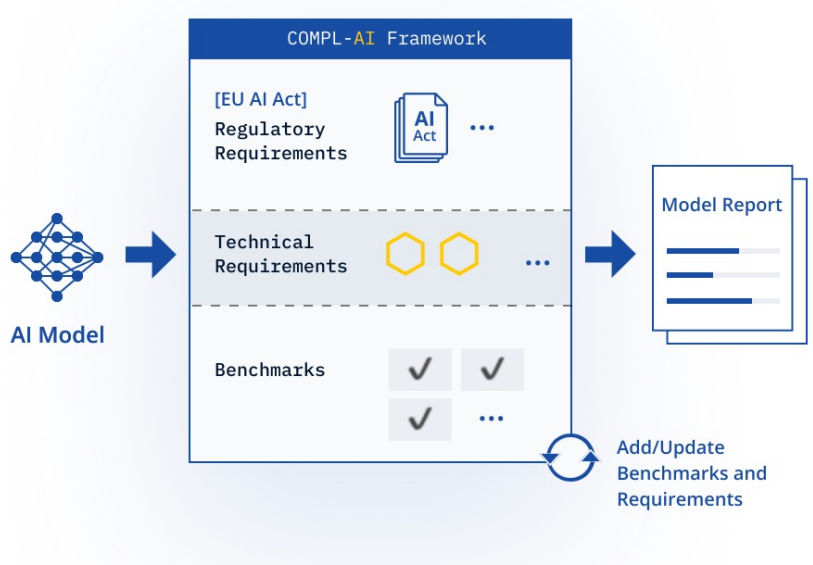
- Einsatzzwecke konkretisieren
- beschränkte verfahrenstechnische Aufgabe
- Ergebnisverbesserung einer menschlichen Tätigkeit
- Erkennen von Entscheidungsmustern
- Durchführung vorbereitender Aufgaben

- Andere zum Anbieter machen
- Anbringen des Namen oder der Marke
- Vornahme wesentlicher Änderungen

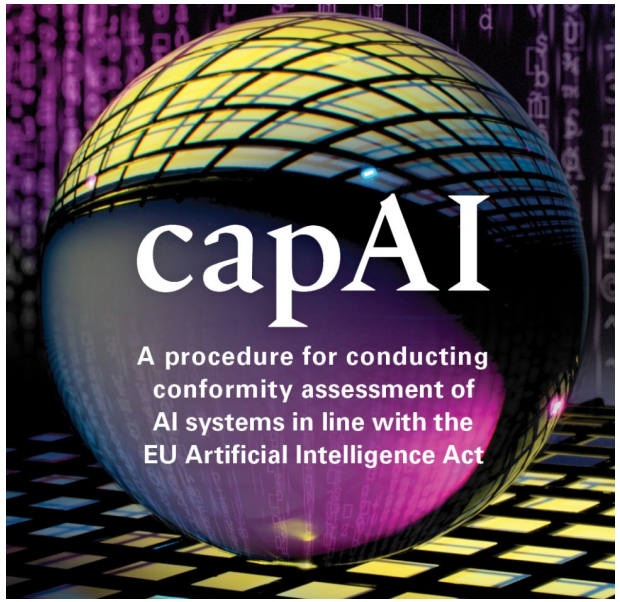
Werkzeuge für die Unterstützung



Quelle:EK https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-code-practice



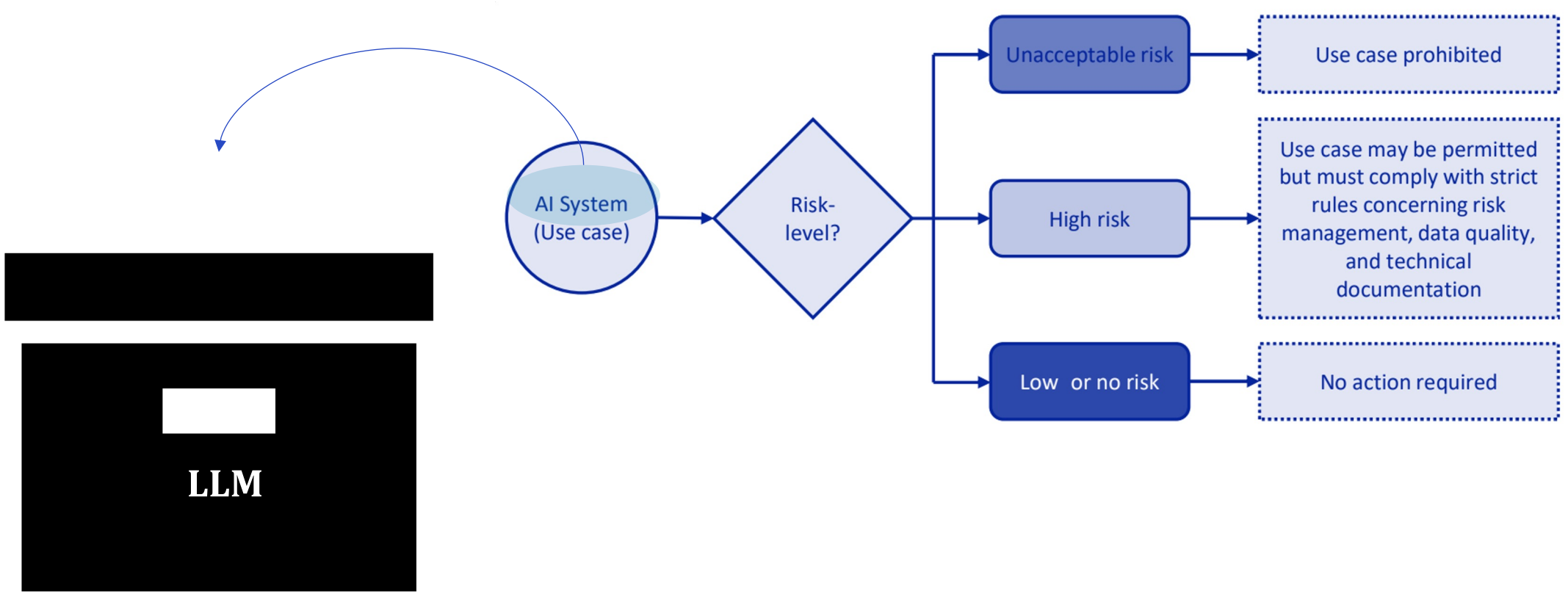
Quelle: https://compl-ai.org



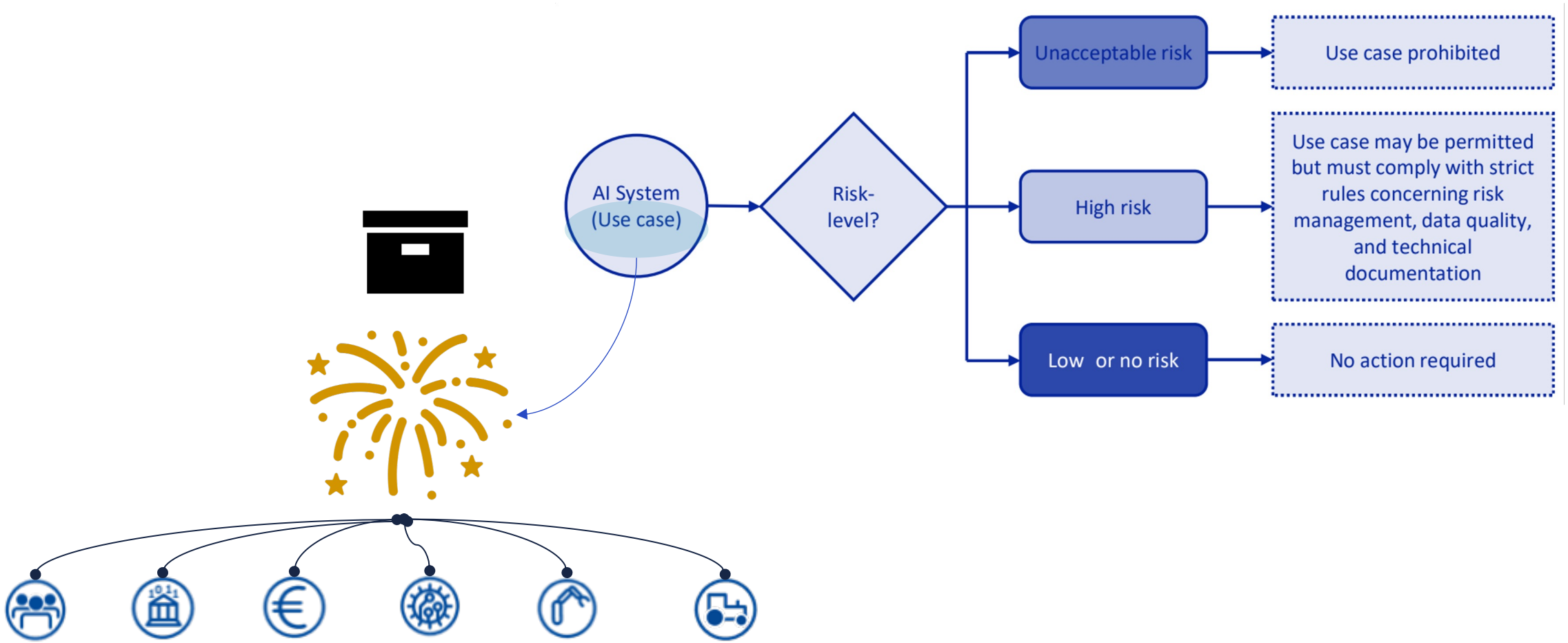
Quelle: https://artificialintelligenceact.eu/assessment/

#5: Software/ AI Engineering als Grundlage für Konformität

EUAIA und GenAI: LLMs sind „Allzweck-Blackboxes“, Robustheit war keine Priorität (z.B. „Jailbreaks“)

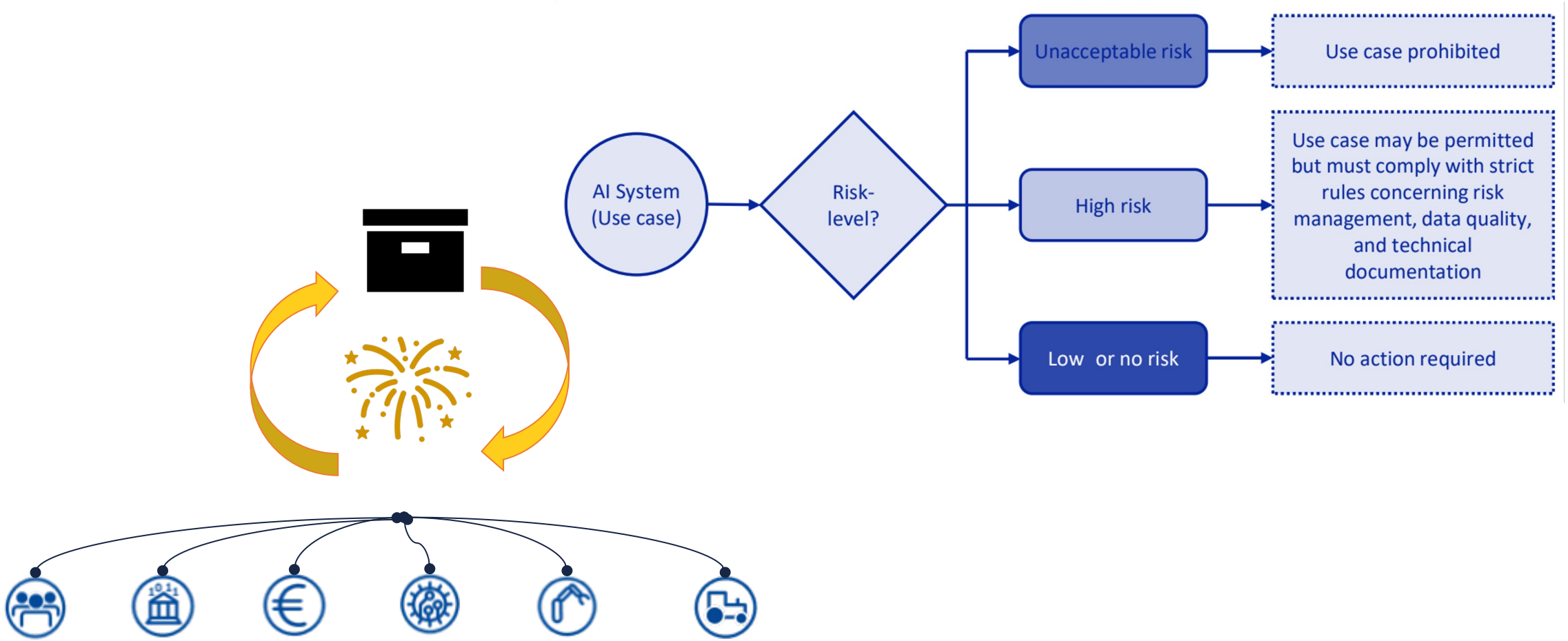


LLMs bringen Risiken in Bezug auf Robustheit und Sicherheit mit, doch entscheidend ist der Anwendungsfall*

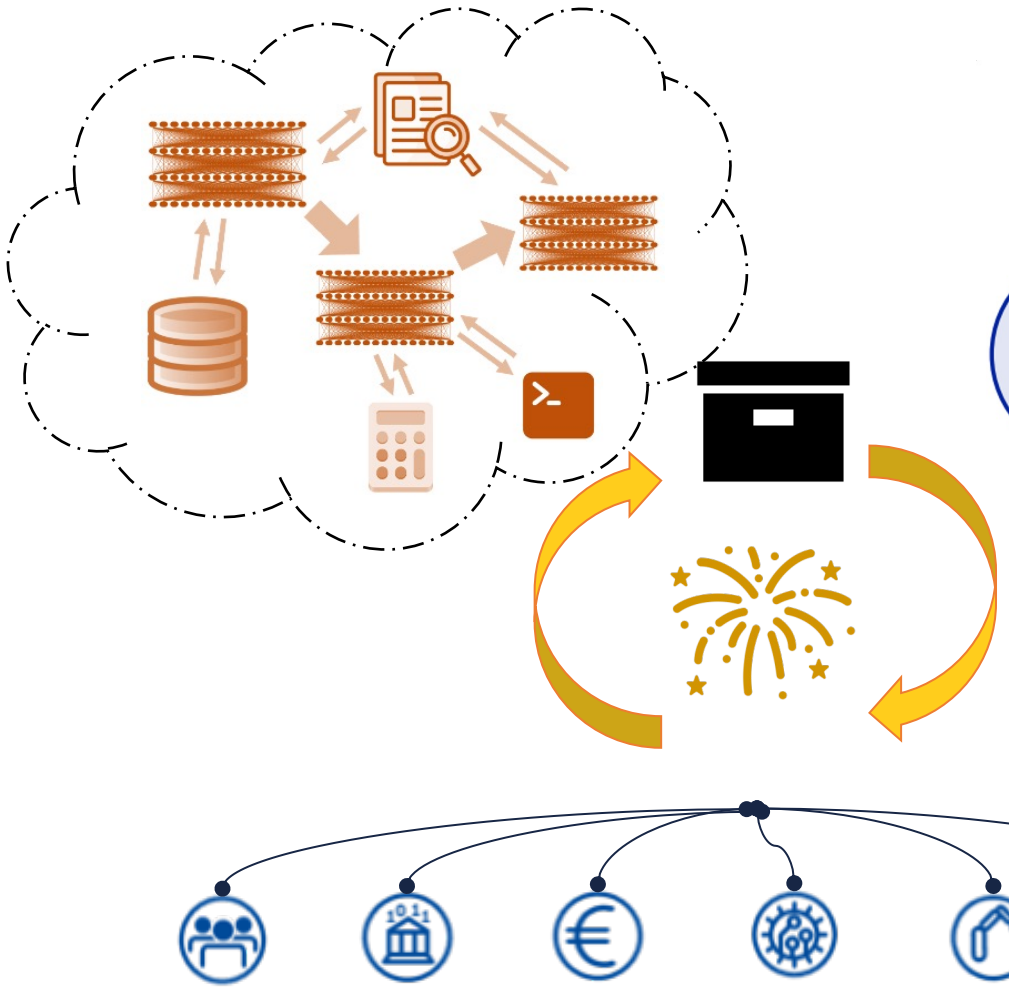


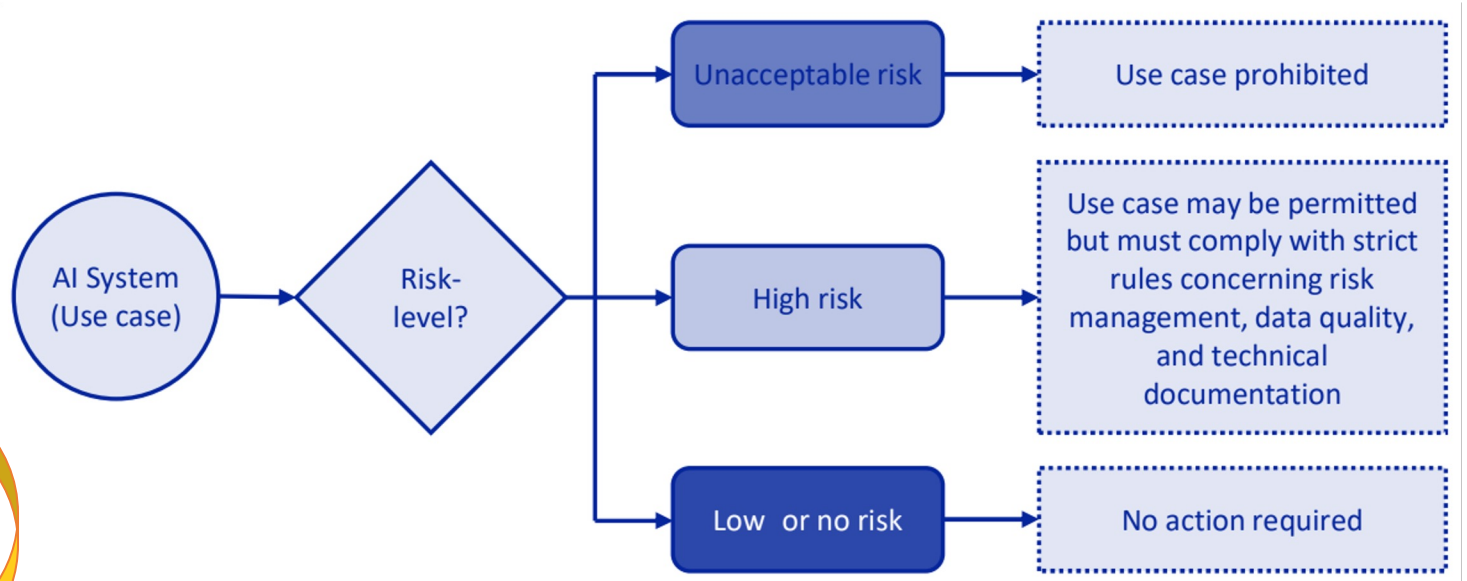
*Anbieter von LLM unterliegen besonderen Konformitätsanforderungen

Ein praktikabler Ansatz ist die Kontrolle und Steuerung von Input und Output

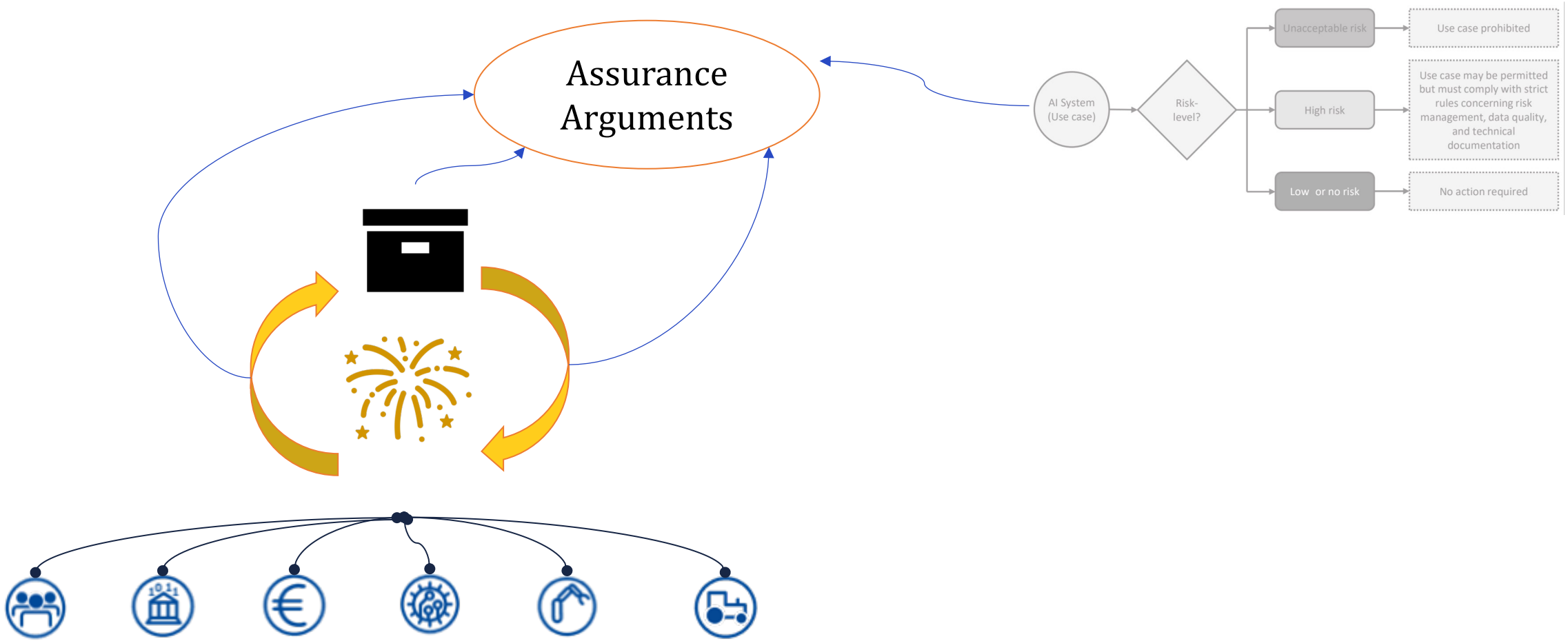


Durch die Evolution von GenAI wird der Ansatz relevanter

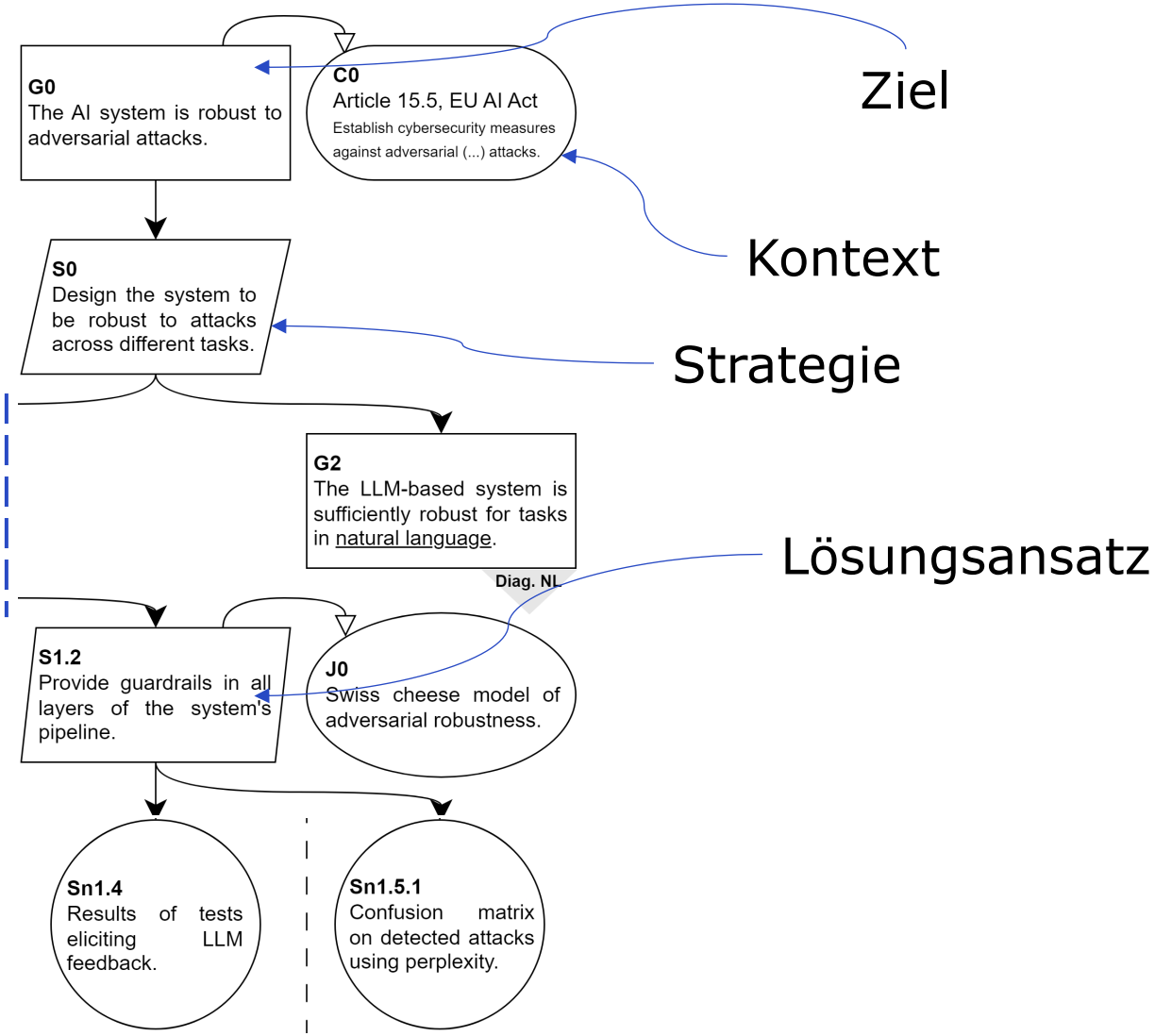




Wir schlagen den Einsatz von Assurance Argumenten für eine fortlaufende Sicherstellung der Konformität vor

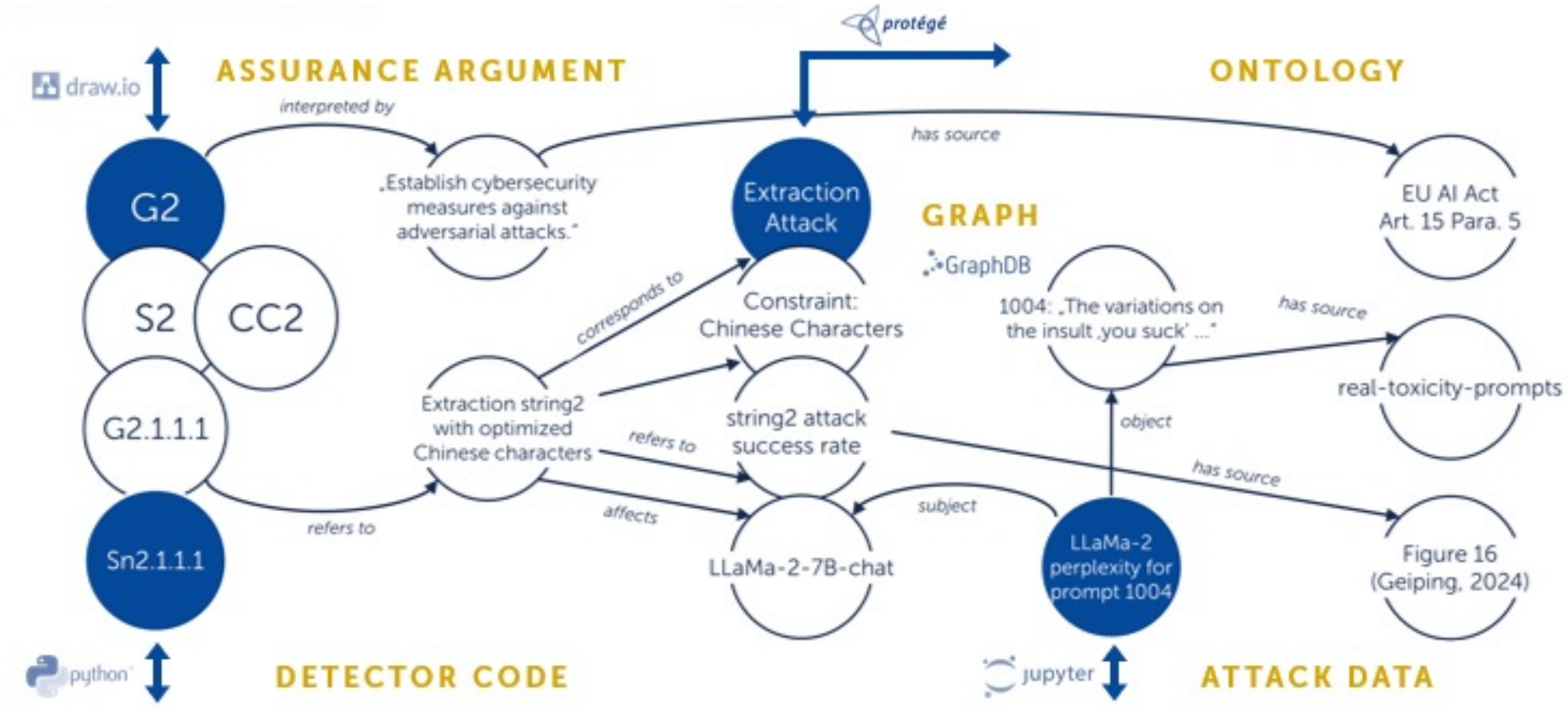


Was ist ein Assurance Argument?



Es sollen möglichst viele Engineering-Entscheidungen transparent gemacht werden

Auf dem Markt verfügbare Werkzeuge helfen im Engineering-Prozess



#1: EU AI Act ist ein risikobasierter Ansatz

#2: Anforderungen werden gestaffelt verpflichtend

#3: Idealtypische Umsetzung laut der EK

#4: Umsetzung in der Praxis

#5: Software/ AI Engineering als Grundlage für Konformität

Vielen Dank!