Generative KI in Unternehmen

Dr. Sven Schmeier, DFKI GmbH
schmeier@dfki.de

1. Was bedeutet das?

- **Was bedeutet das?**
- Was ist eine KI Strategie?
- Make or Buy?

Hat generative KI ein disruptives Potential wie das Internet?

1 Internet

Kosten der Distribution → 0
(von Daten, Inhalten, Informationen)

2 Generative KI

Kosten der Generierung → 0
(von Daten, Inhalten, Informationen)

Wo kann man generative KI bereits einsetzen?

Text- und Bildgenerierung

Programmieren

Video

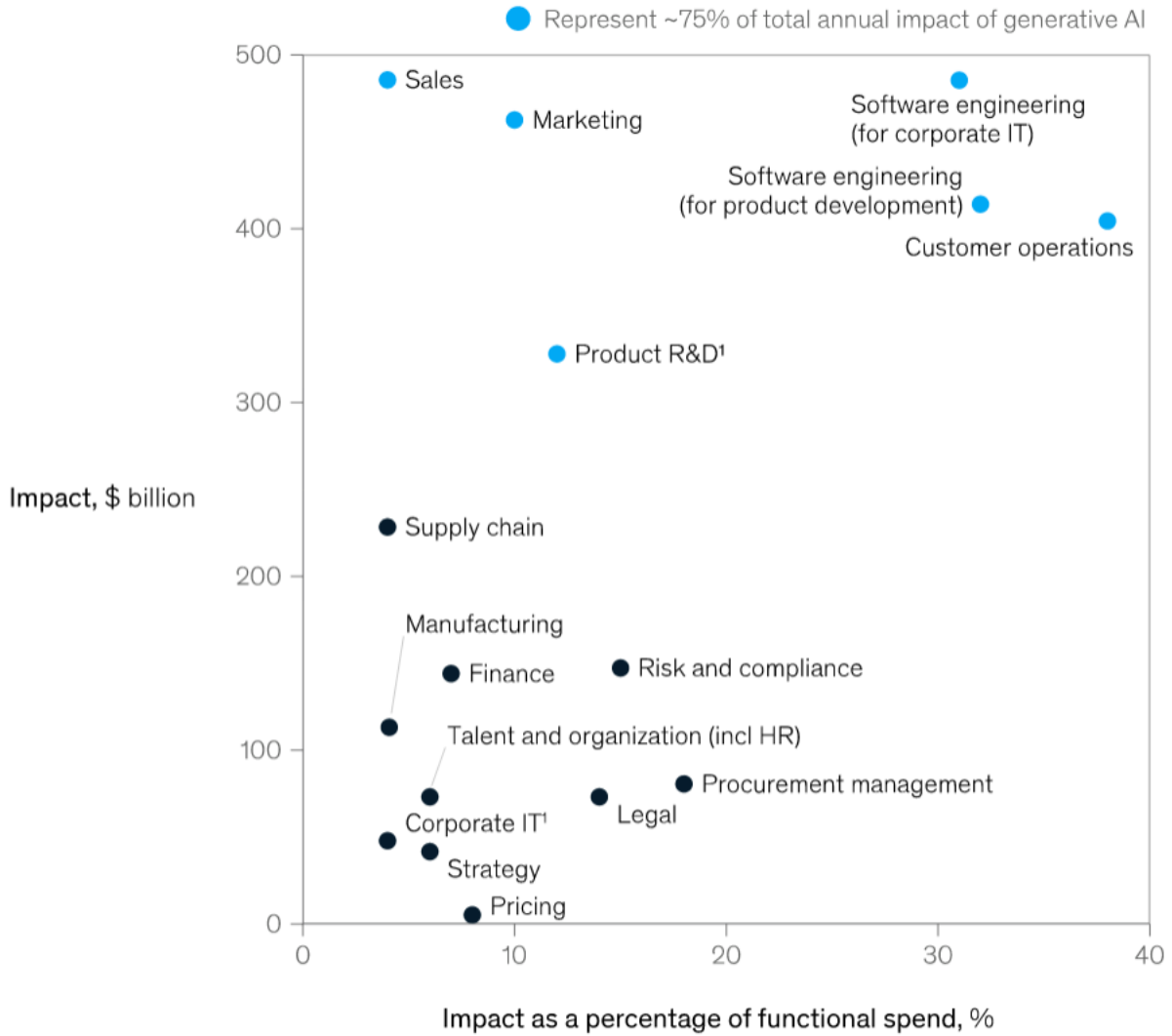
Tongenerierung

Kundenservice

Forschung
Gesundheitswesen
Finanzwesen
Bildung
...

McKinsey Studie: Das Potential fokussiert sich noch auf bestimmte Funktionen

Using generative AI in just a few functions could drive most of the technology's impact across potential corporate use cases.

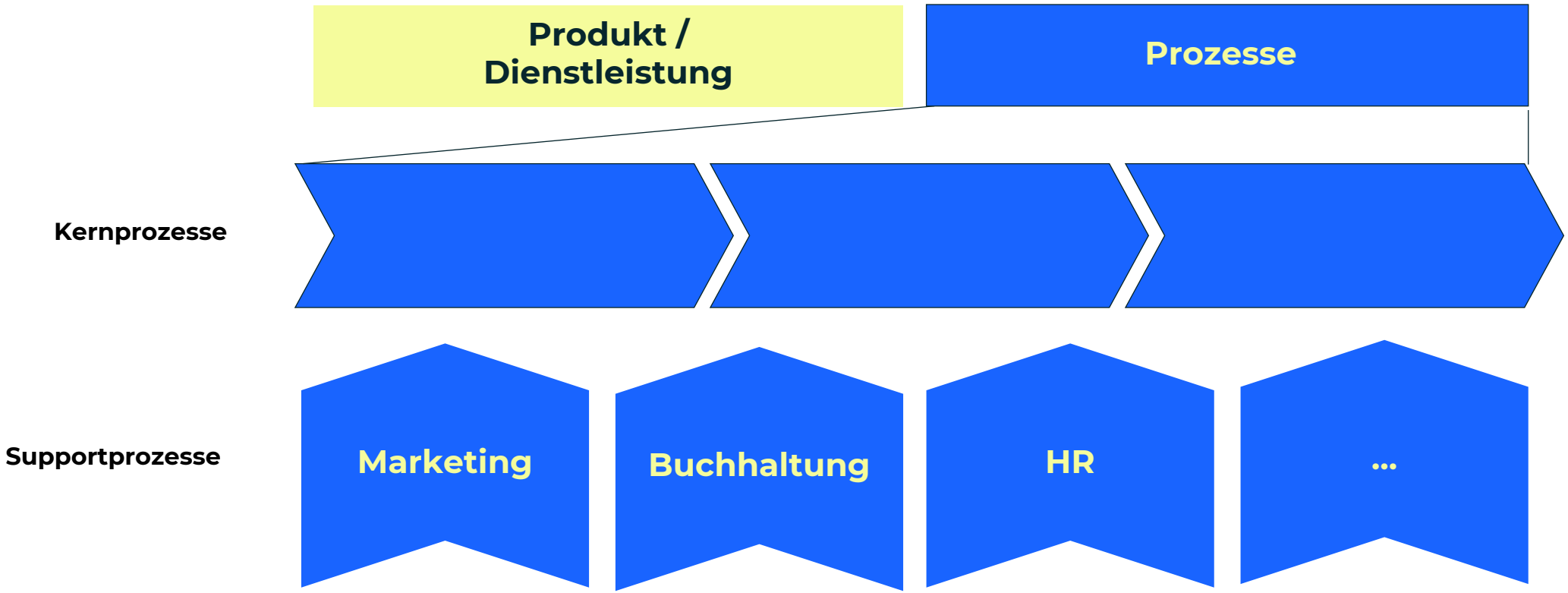


Note: Impact is averaged.
[1]Excluding software engineering.
Source: Comparative Industry Service (CIS), IHS Markit; Oxford Economics; McKinsey Corporate and Business Functions database; McKinsey Manufacturing and Supply Chain 360; McKinsey Sales Navigator; Ignite, a McKinsey database; McKinsey analysis

McKinsey & Company

Welchen Einfluss hat generative KI auf meine Branche und mein Unternehmen?

Ist mein Produkt beeinflusst? Welche Prozesse können von KI erledigt oder gestützt werden?



McKinsey Studie: Größter Einfluss von KI in Tech, Banking und Pharma

Exhibit 4

Generative AI use cases will have different impacts on business functions across industries.

< Prev 01 — 02 Next >

Generative AI productivity impact by business functions[1]

Low impact — High impact

	Total, % of industry revenue	Total, $ billion	Marketing and sales	Customer operations	Product and R&D	Software engineering	Supply chain and operations	Risk and legal	Strategy and finance	Corporate IT[2]	Talent and organization
			760–1,200	340–470	230–420	580–1,200	280–530	180–260	120–260	40–50	60–90
High tech	4.8–9.3	240–460									
Banking	2.8–4.7	200–340									
Pharmaceuticals and medical products	2.6–4.5	60–110									
Education	2.2–4.0	120–230									
Telecommunications	2.3–3.7	60–100									
Healthcare	1.8–3.2	150–260									
Insurance	1.8–2.8	50–70									
Media and entertainment	1.8–3.1	60–110									
Advanced manufacturing[3]	1.4–2.4	170–290									
Consumer packaged goods	1.4–2.3	160–270									
Advanced electronics and semiconductors	1.3–2.3	100–170									

Note: Figures may not sum to 100%, because of rounding. [1]Excludes implementation costs (eg, training, licenses). [2]Excluding software engineering. [3]Includes aerospace, defense, and auto manufacturing. [4]Including auto retail.
Source: Comparative Industry Service (CIS), IHS Markit; Oxford Economics; McKinsey Corporate and Business Functions database; McKinsey Manufacturing and Supply Chain 360; McKinsey Sales Navigator; Ignite, a McKinsey database; McKinsey analysis

Exhibit 4

Generative AI use cases will have different impacts on business functions across industries.

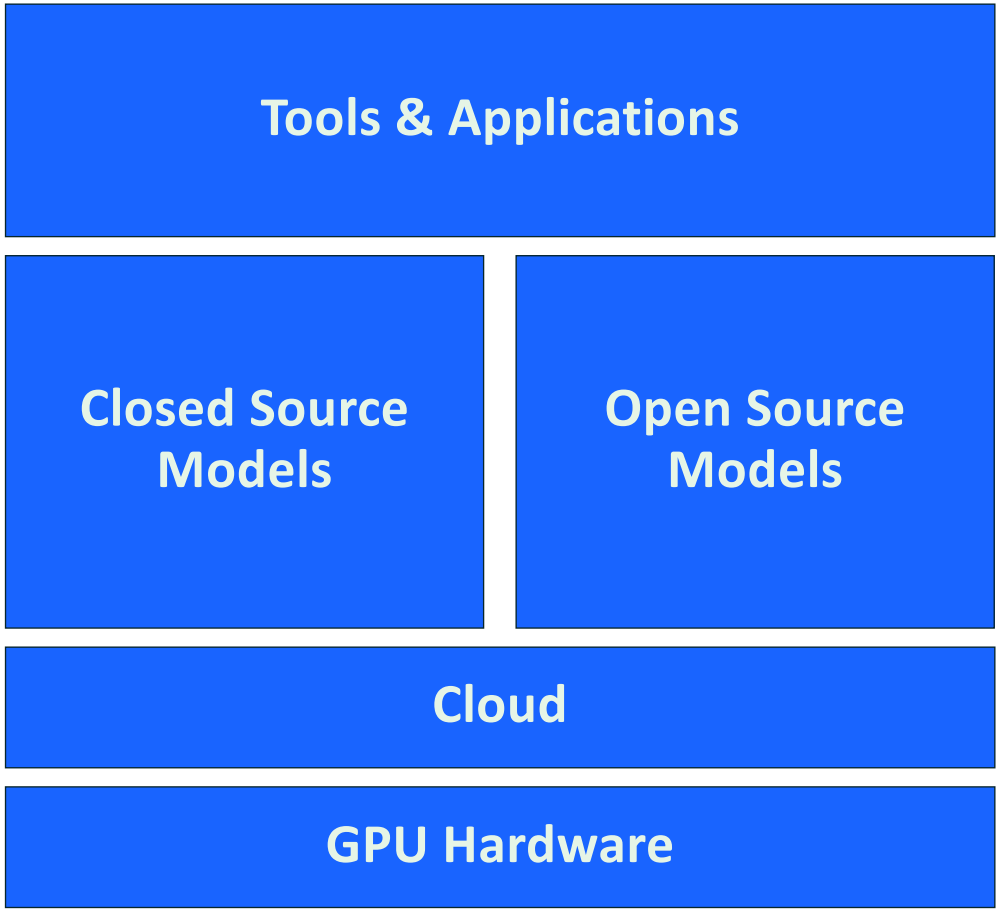
Generative AI productivity impact by business functions[1]

Low impact — High impact

	Total, % of industry revenue	Total, $ billion	Marketing and sales	Customer operations	Product and R&D	Software engineering	Supply chain and operations	Risk and legal	Strategy and finance	Corporate IT[2]	Talent and organization
			760–1,200	340–470	230–420	580–1,200	280–530	180–260	120–260	40–50	60–90
Travel, transport, and logistics	1.2–2.0	180–300									
Retail[4]	1.2–1.9	240–390									
Real estate	1.0–1.7	110–180									
Energy	1.0–1.6	150–240									
Administrative and professional services	0.9–1.4	150–250									
Chemical	0.8–1.3	80–140									
Basic materials	0.7–1.2	120–200									
Construction	0.7–1.2	90–150									
Agriculture	0.6–1.0	40–70									
Public and social sector	0.5–0.9	70–110									
		2,600–4,400									

Note: Figures may not sum to 100%, because of rounding. [1]Excludes implementation costs (eg, training, licenses). [2]Excluding software engineering. [3]Includes aerospace, defense, and auto manufacturing. [4]Including auto retail.
Source: Comparative Industry Service (CIS), IHS Markit; Oxford Economics; McKinsey Corporate and Business Functions database; McKinsey Manufacturing and Supply Chain 360; McKinsey Sales Navigator; Ignite, a McKinsey database; McKinsey analysis

Das Ecosystem hinter dem Hype



- Gibt es einen Überblick, welche Tools es gibt? Welche bieten einen Mehrwert?
- „The Winner takes it all“? Dominanz weniger großer Modelle, die alles können?
- Haben kleine spezialisierte Modelle eine Zukunft?
- Big Tech vs. Open Source? Wer ist vorne?
- „GPU rich“ vs. „GPU poor“?

Tools: „There is an AI for that“

Tools & Applications

Closed Source Models

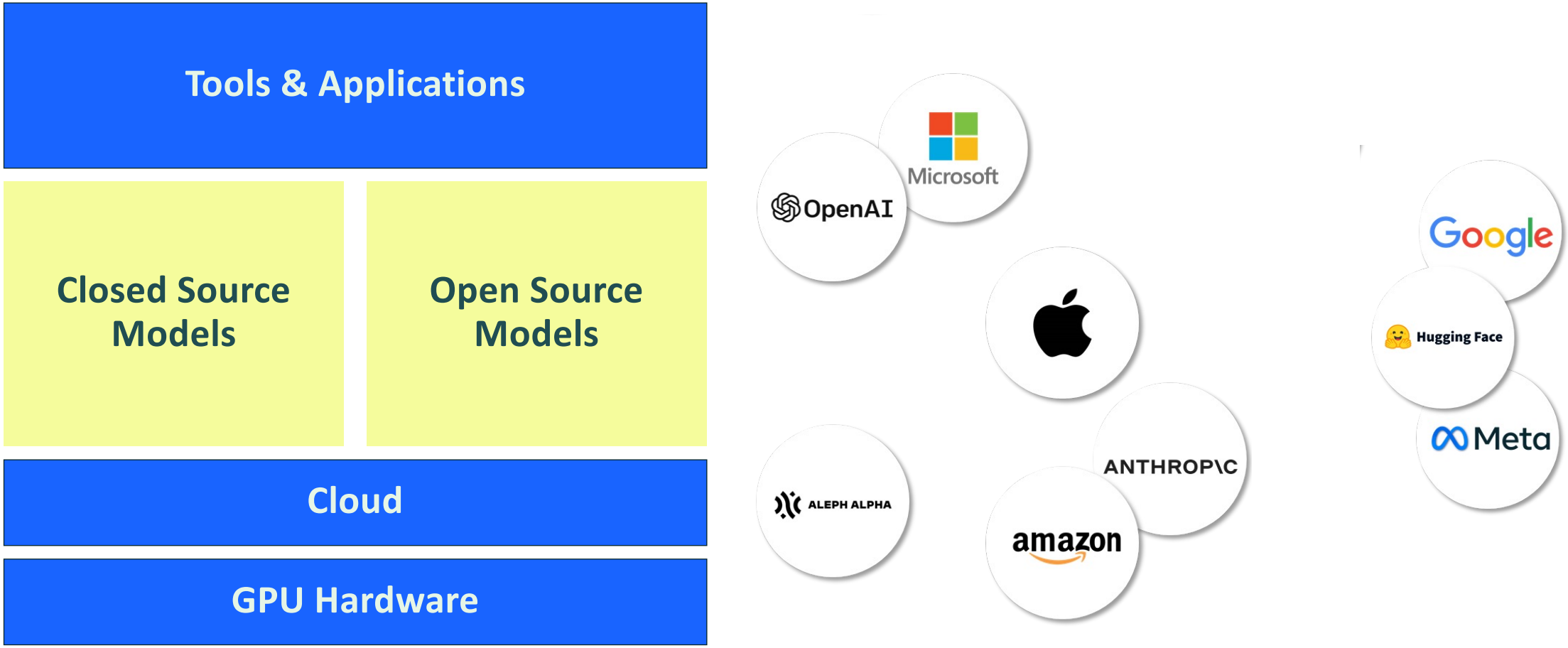
Open Source Models

Cloud

GPU Hardware

Tools, die aktuelle Probleme von LLMs lösen, werden durch neue Versionen teilweise obsolet

Closed vs. Open Source: The winner takes it all?



Closed vs. Open Source: Die Größe ist wichtig, es gibt aber Alternativen

Model Name	Win Rate	Length
GPT-4 Turbo	97.70%	2049
XwinLM 70b V0.1	95.57%	1775
PairRM+Tulu 2+DPO 70B (best-of-16)	95.40%	1607
GPT-4	95.28%	1365
Tulu 2+DPO 70B	95.03%	1418
Yi 34B Chat	94.08%	2123
PairRM+Zephyr 7B Beta (best-of-16)	93.41%	1487
LLaMA2 Chat 70B	92.66%	1790
UltraLM 13B V2.0 (best-of-16)	92.30%	1720
XwinLM 13b V0.1	91.76%	1894
UltraLM 13B (best-of-16)	91.54%	1980
Claude 2	91.36%	1069
PairRM+Tulu 2+DPO 13B (best-of-16)	91.06%	1454
Cohere Command	90.62%	1983
Zephyr 7B Beta	90.60%	1444
OpenChat V3.1 13B	89.49%	1484
ChatGPT	89.37%	827

Closed vs. Open Source: Und welches Modell nehme ich?

Vorteile großer (geschlossener) Modelle

- Größe, und flexibel einsetzbar
- Menge der Trainingsdaten
- Verknüpfung zu großen Suchindizes bzw. Knowledge Graphen
- Sehr leicht zugänglich
- Resultate sehr gut

Vorteile kleinerer (Open Source) Modelle

- Geringere Kosten
- IP bleibt in House
- Datenschutz
- Fachspezifisches Vokabular
- Intransparenz (Bias)
- Anpassbar

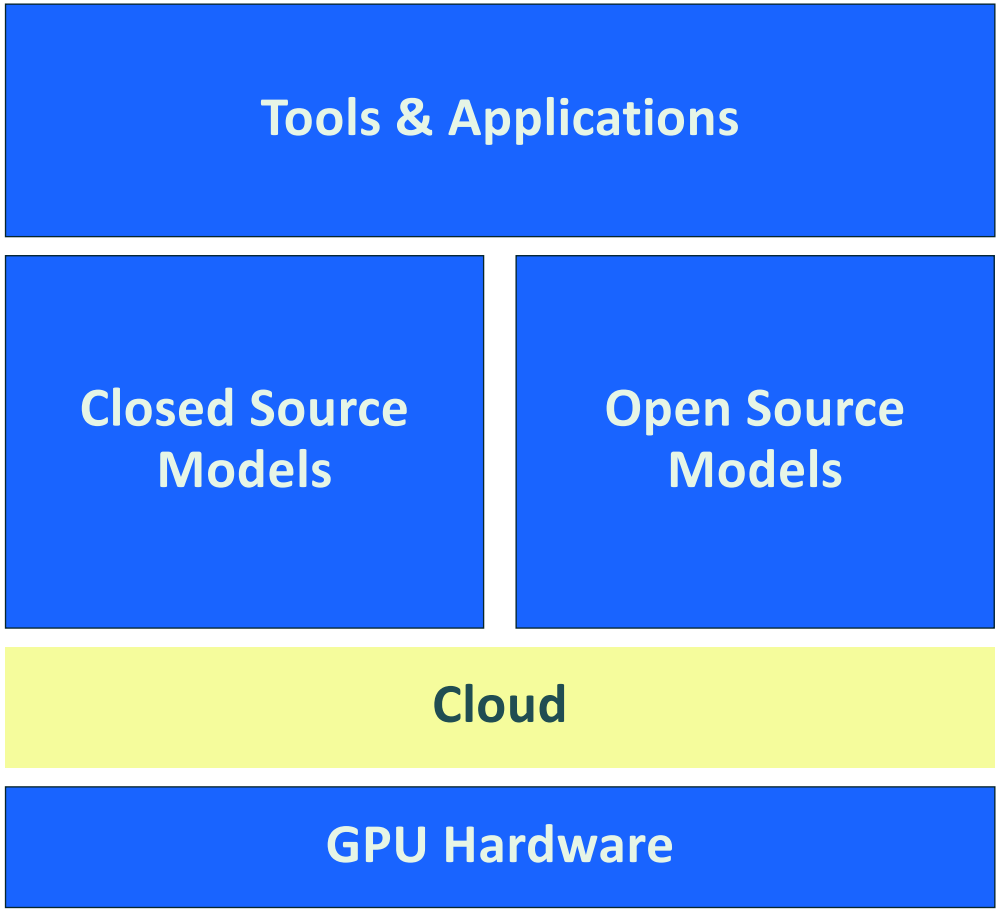
Kleine Modelle sind die Zukunft …

Grundregel:

Testen und mit dem besten Modell starten

Und wenn es funktioniert nach kleineren Alternativen oder Tools suchen

Cloud: Die Big Player



aws

Azure

Google Cloud

Alibaba Cloud

HUAWEI

GPU: GPU poor vs. GPU rich

Tools & Applications

Closed Source Models

Open Source Models

Cloud

GPU Hardware

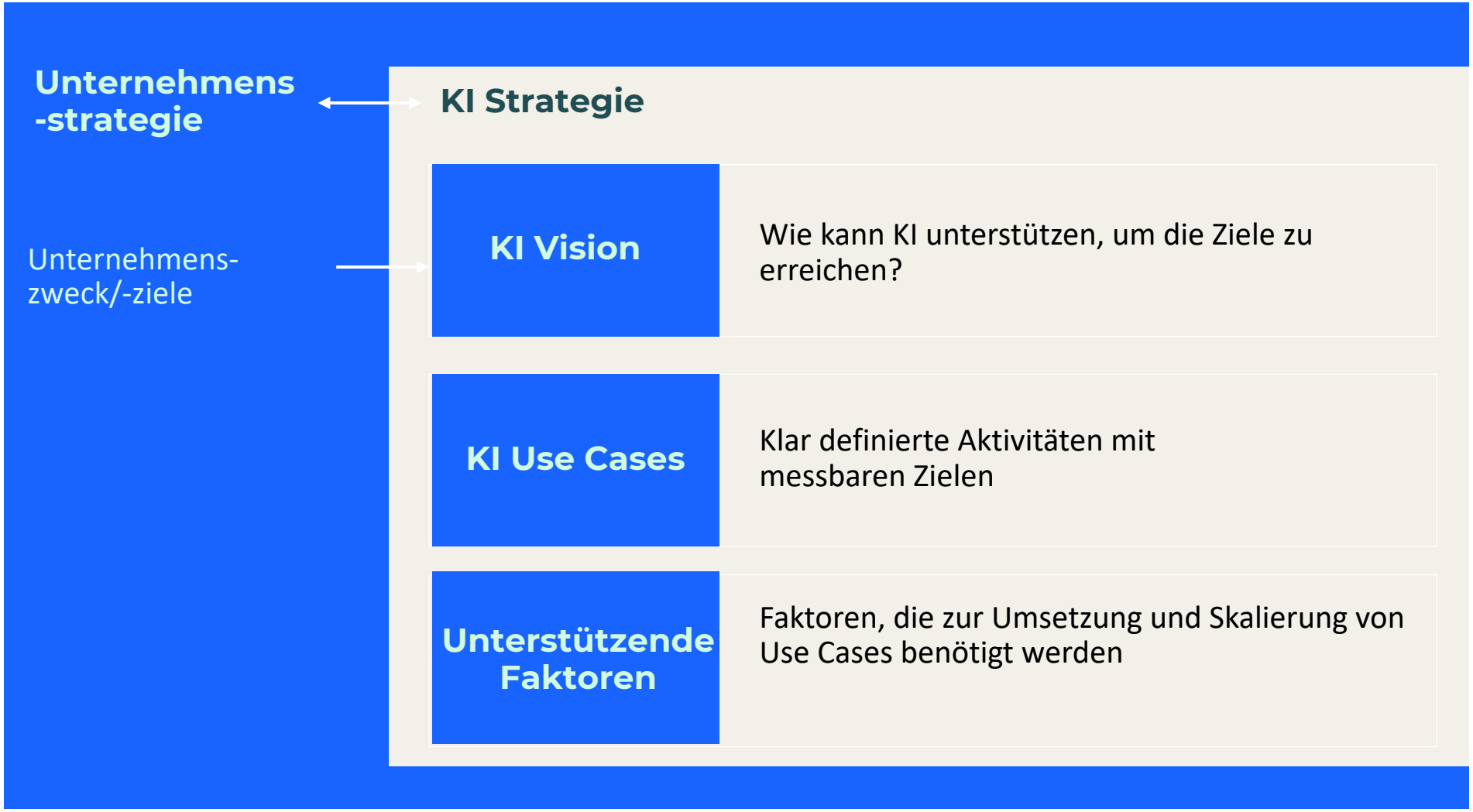
2. Was ist eine KI Strategie?

- Was bedeutet das?
- **Was ist eine KI Strategie?**
- Make or Buy?

„Understanding what AI can do and how it fits into your strategy is the beginning, not the end, of that process.“

Andrew Ng, Professor Stanford University

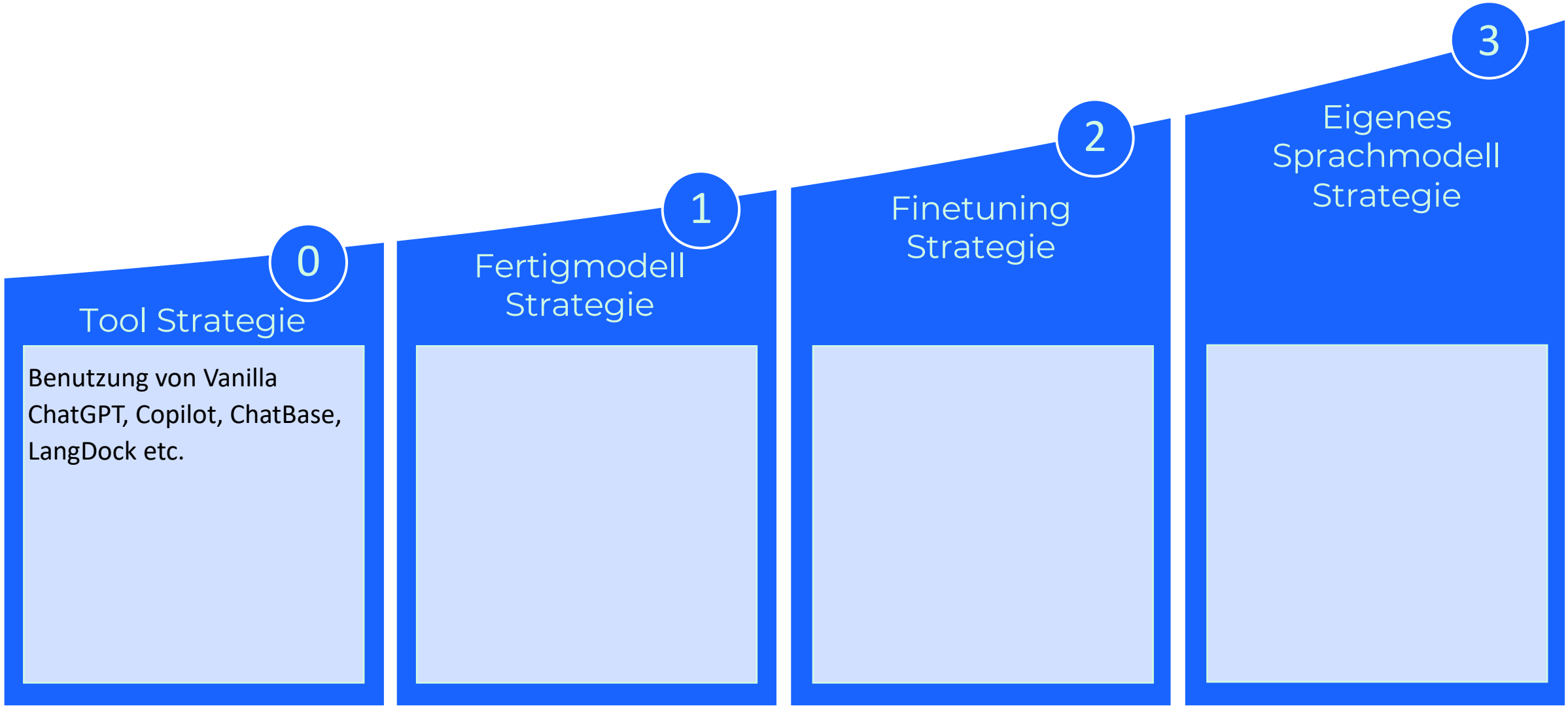
Die KI Strategie im Unternehmenskontext

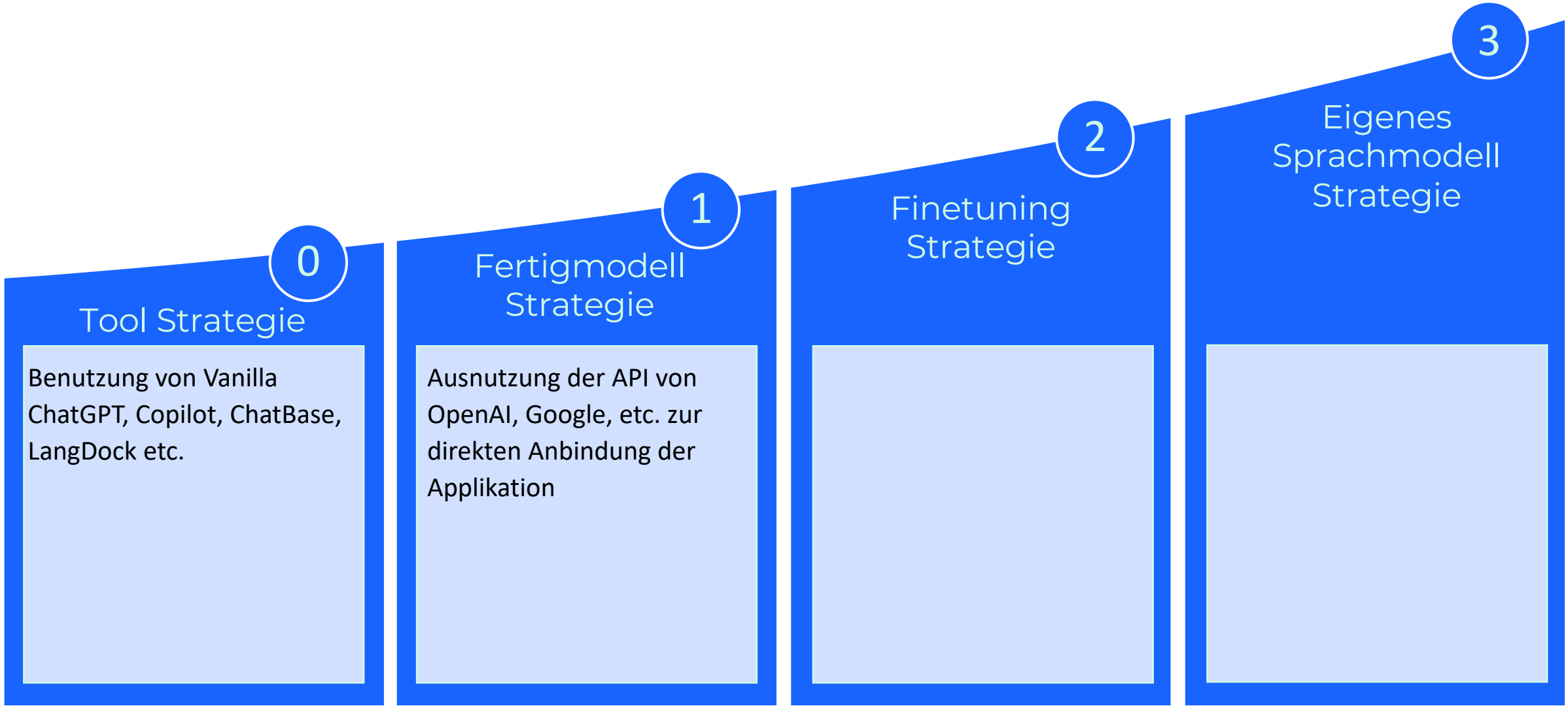


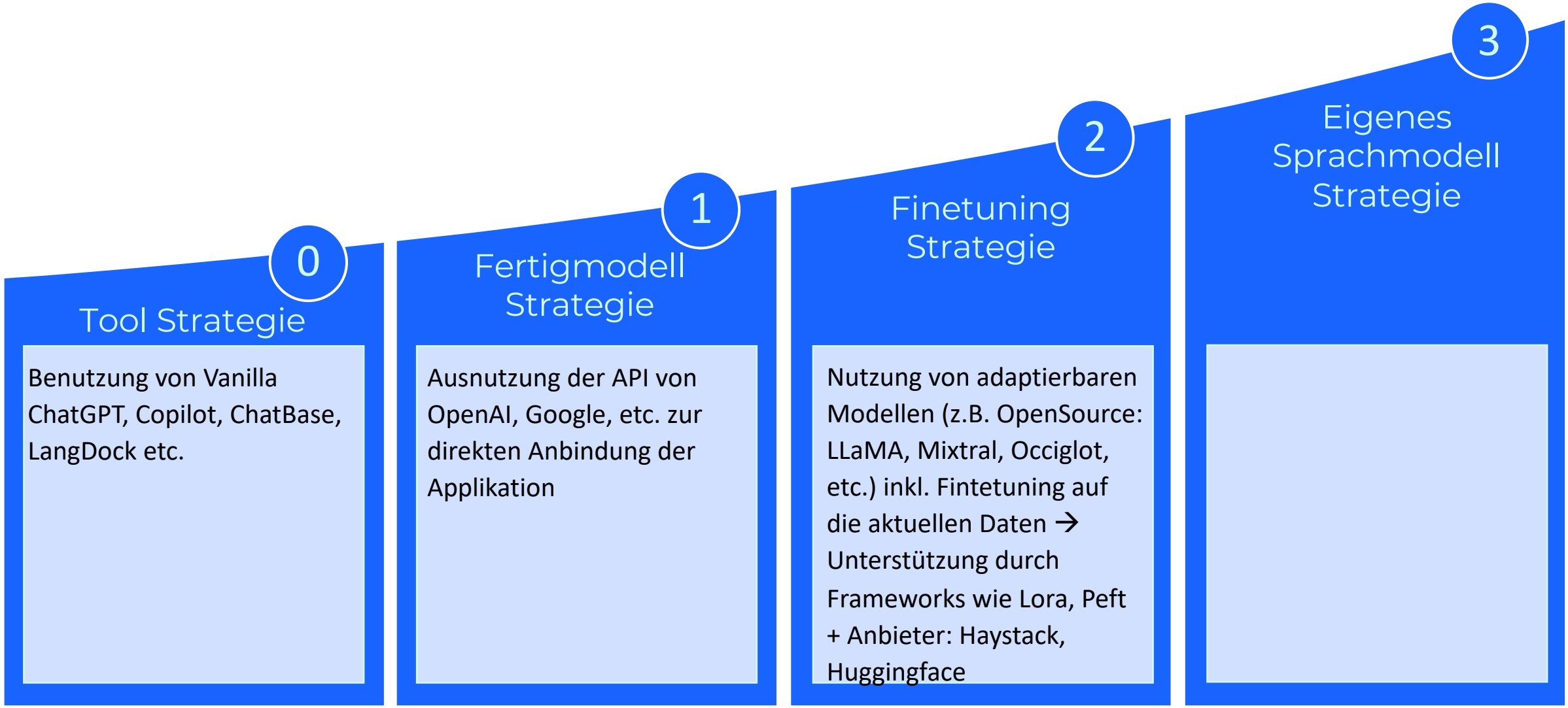
Die Bestandteile einer KI Strategie

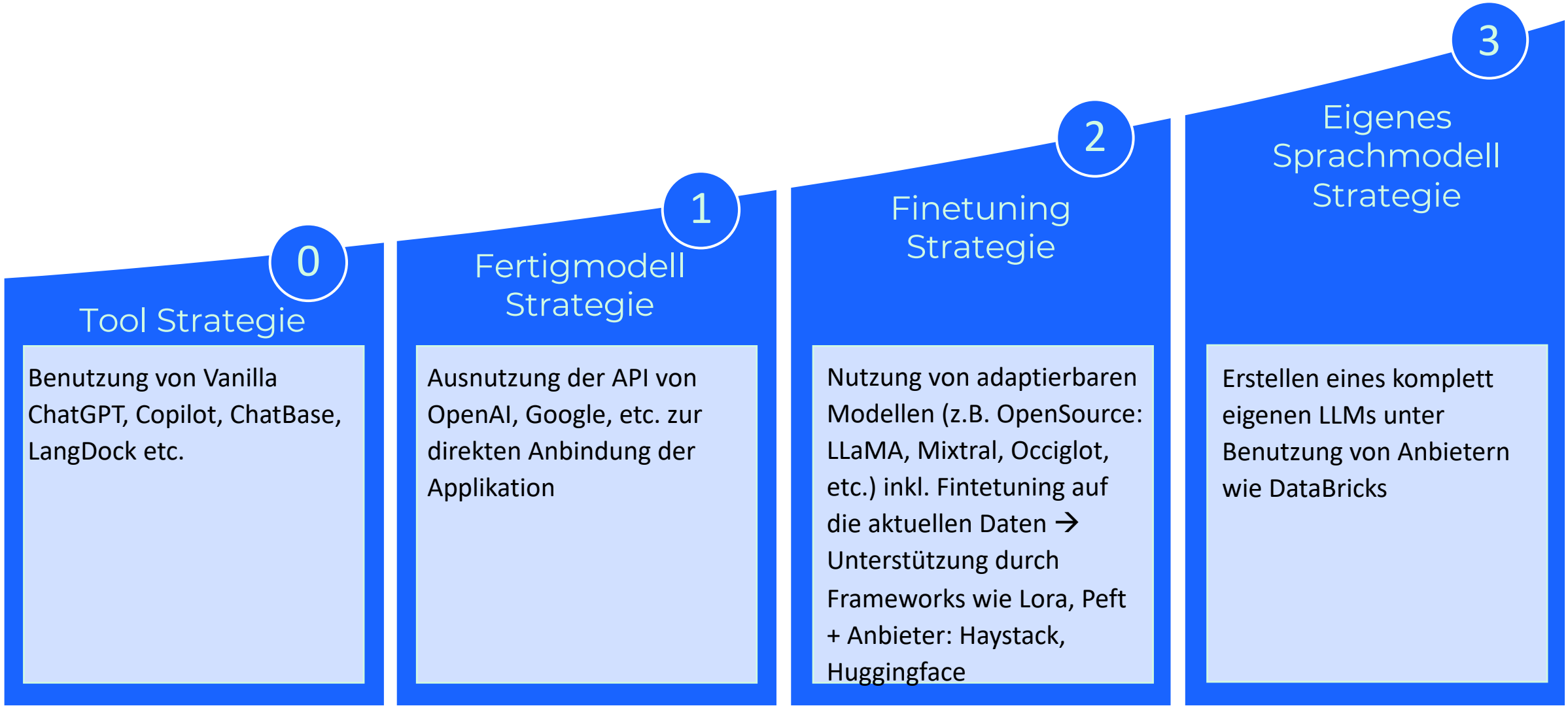
3. Make or Buy?

- Was bedeutet das?
- Was ist eine KI Strategie?
- Make or Buy?


0
Tool Strategie
Benutzung von Vanilla ChatGPT, Copilot, ChatBase, LangDock etc.
1
Fertigmodell Strategie
2
Finetuning Strategie
3
Eigenes Sprachmodell Strategie


0
Tool Strategie
Benutzung von Vanilla ChatGPT, Copilot, ChatBase, LangDock etc.
1
Fertigmodell Strategie
Ausnutzung der API von OpenAI, Google, etc. zur direkten Anbindung der Applikation
2
Finetuning Strategie
3
Eigenes Sprachmodell Strategie


0
Tool Strategie
Benutzung von Vanilla ChatGPT, Copilot, ChatBase, LangDock etc.
1
Fertigmodell Strategie
Ausnutzung der API von OpenAI, Google, etc. zur direkten Anbindung der Applikation
2
Finetuning Strategie
Nutzung von adaptierbaren Modellen (z.B. OpenSource: LLaMA, Mixtral, Occiglot, etc.) inkl. Fintetuning auf die aktuellen Daten → Unterstützung durch Frameworks wie Lora, Peft + Anbieter: Haystack, Huggingface
3
Eigenes Sprachmodell Strategie


0
Tool Strategie
Benutzung von Vanilla ChatGPT, Copilot, ChatBase, LangDock etc.
1
Fertigmodell Strategie
Ausnutzung der API von OpenAI, Google, etc. zur direkten Anbindung der Applikation
2
Finetuning Strategie
Nutzung von adaptierbaren Modellen (z.B. OpenSource: LLaMA, Mixtral, Occiglot, etc.) inkl. Fintetuning auf die aktuellen Daten → Unterstützung durch Frameworks wie Lora, Peft + Anbieter: Haystack, Huggingface
3
Eigenes Sprachmodell Strategie
Erstellen eines komplett eigenen LLMs unter Benutzung von Anbietern wie DataBricks

Projektmanagement

Der CRISP-DM Standard

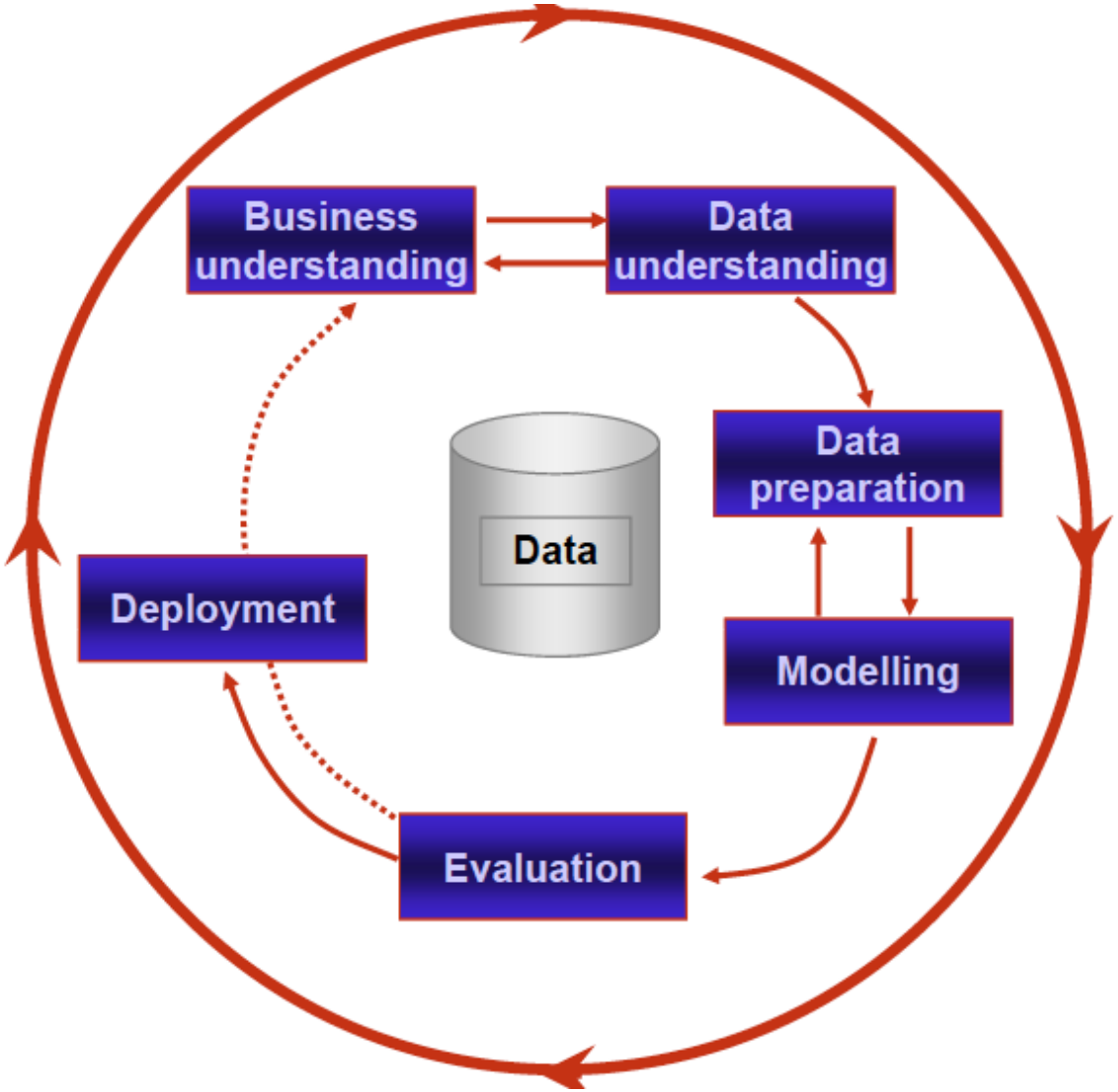
CRISP-DM: Cross-Industry Standard Process for Data Mining

Entwickelt im Rahmen eines EU-Projekts von 1996-99

Partner: DaimlerChrysler (Deutschland), NCR Systems Copenhagen (USA,Dänemark), OHRA Bank Groep B.V. (Niederlande)
SPSS Inc. (USA)

Für Informationen zu CRISP-DM siehe http://www.crisp-dm.org

Das CRISP-DM Referenzmodell



Business Understanding

Initiale Phase
Inhalt:

- Verstehen der Projektziele und Anforderungen aus Unternehmenssicht
- Transformation dieses Wissens in eine Data-Mining-Problemdefinition und Erstellung eines vorläufigen Projektplan

Data Understanding

Beginnt mit einer ersten Datenerfassung
Fährt fort mit Aktivitäten, die darauf abzielen:

- Sich mit den Daten vertraut machen
- Erkennen von Datenqualitätsproblemen
- Entdecken von ersten Erkenntnissen über die Daten
- Erkennen interessanter Teilmengen zur Bildung von Hypothesen für versteckte Informationen

Data Preparation

- Umfasst alle Aktivitäten zur Erstellung des endgültigen Datensatzes (Daten, die in das/die Modellierungswerkzeug(e) eingespeist werden) aus den ursprünglichen Rohdaten
- Datenvorbereitungsaufgaben werden häufig mehrfach und nicht in einer vorgeschriebenen Reihenfolge durchgeführt
- Zu den Aufgaben gehören die Auswahl von Tabellen, Datensätzen und Attributen sowie die Transformation und Bereinigung von Daten für Modellierungswerkzeuge

Modelling

- Verschiedene Modellierungstechniken werden ausgewählt und angewendet, und ihre Parameter werden auf optimale Werte kalibriert
- Typischerweise gibt es mehrere Techniken für denselben Data-Mining-Problemtyp
- Einige Techniken haben spezifische Anforderungen an die Form der Daten, daher ist oft ein Rücksprung in die Phase der Data Preparation erforderlich

Evaluation

- In diesem Stadium wurde ein Modell (oder Modelle) erstellt, das aus Sicht der Datenanalyse eine hohe Qualität zu haben scheint.
- Bevor Sie mit der endgültigen Implementierung des Modells fortfahren, ist es wichtig, das Modell gründlicher zu bewerten und die Schritte zu überprüfen, die zur Erstellung des Modells ausgeführt wurden, um sicherzustellen, dass es die Geschäftsziele richtig erreicht
- Ein Hauptziel ist es, festzustellen, ob es ein wichtiges Geschäftsproblem gibt, das nicht ausreichend berücksichtigt wurde
- Am Ende dieser Phase sollte eine Entscheidung über die Verwendung der Data-Mining-Ergebnisse getroffen werden

Deployment

- Die Erstellung des Modells ist im Allgemeinen nicht das Ende des Projekts
- Selbst wenn der Zweck des Modells darin besteht, das Wissen über die Daten zu erweitern, muss das gewonnene Wissen so organisiert und präsentiert werden, dass der Kunde es nutzen kann
- Abhängig von den Anforderungen kann das Deployment so einfach sein wie die Erstellung eines Berichts oder so komplex wie die Implementierung eines wiederholbaren Data-Mining-Prozesses
- In vielen Fällen wird der Kunde, nicht der Datenanalyst, die Implementierungsschritte durchführen.
- Aber auch wenn der Datenanalyst die Implementierung nicht selbst durchführt, ist es wichtig, dass der Kunde im Voraus weiß, welche Aktionen durchgeführt werden müssen, um die erstellten Modelle tatsächlich nutzen zu können

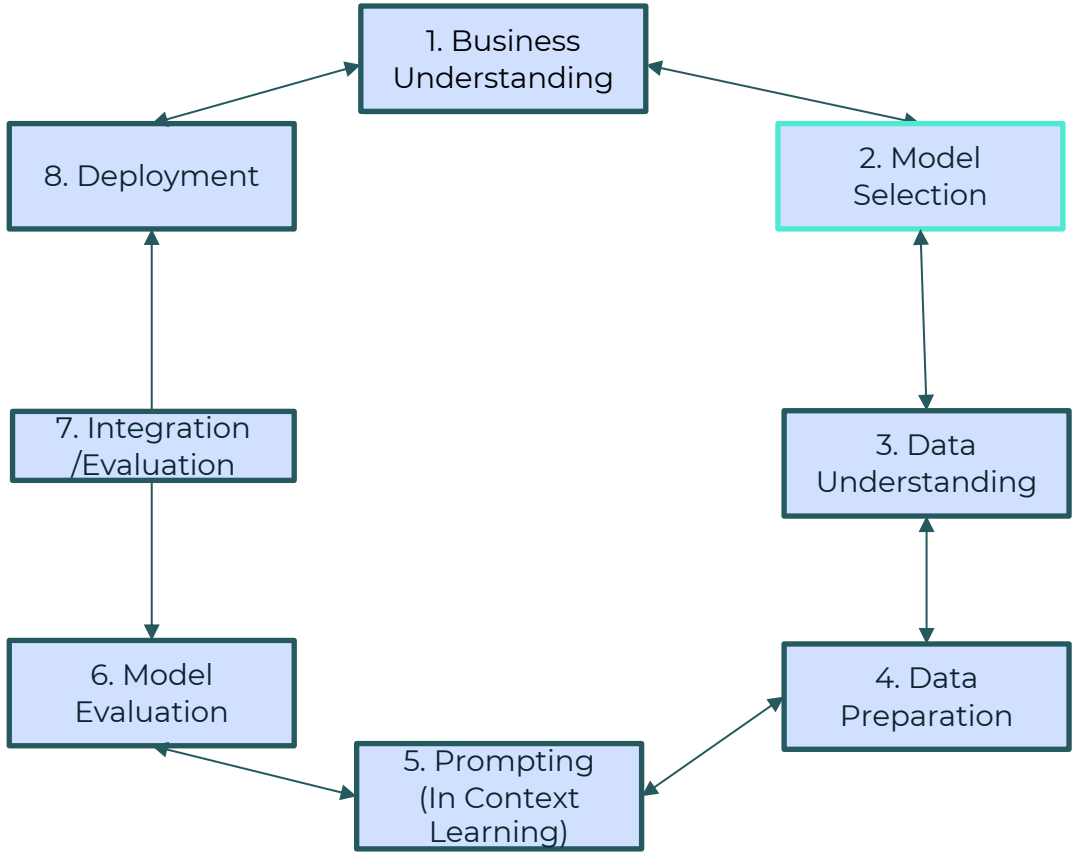
Projektmanagement bei generativer KI

- Das Projektmanagement bei generativer KI unterscheidet sich grundsätzlich vom CRISP DM Prozess
- Wir haben üblicherweise bereits ein Foundation Modell → oft ist keine Datenarbeit erforderlich

- Das Projektmanagement hängt zudem von der KI Strategie des Unternehmens ab:

1. Nutzung von kostenpflichtigen LLMs
2. Nutzen von Open Source LLMs + Finetuning
3. Training eines LLM vom Scratch

Nutzung kostenpflichtiger LLMs



Modellauswahl tritt an Stelle der Daten als zentrales Element.

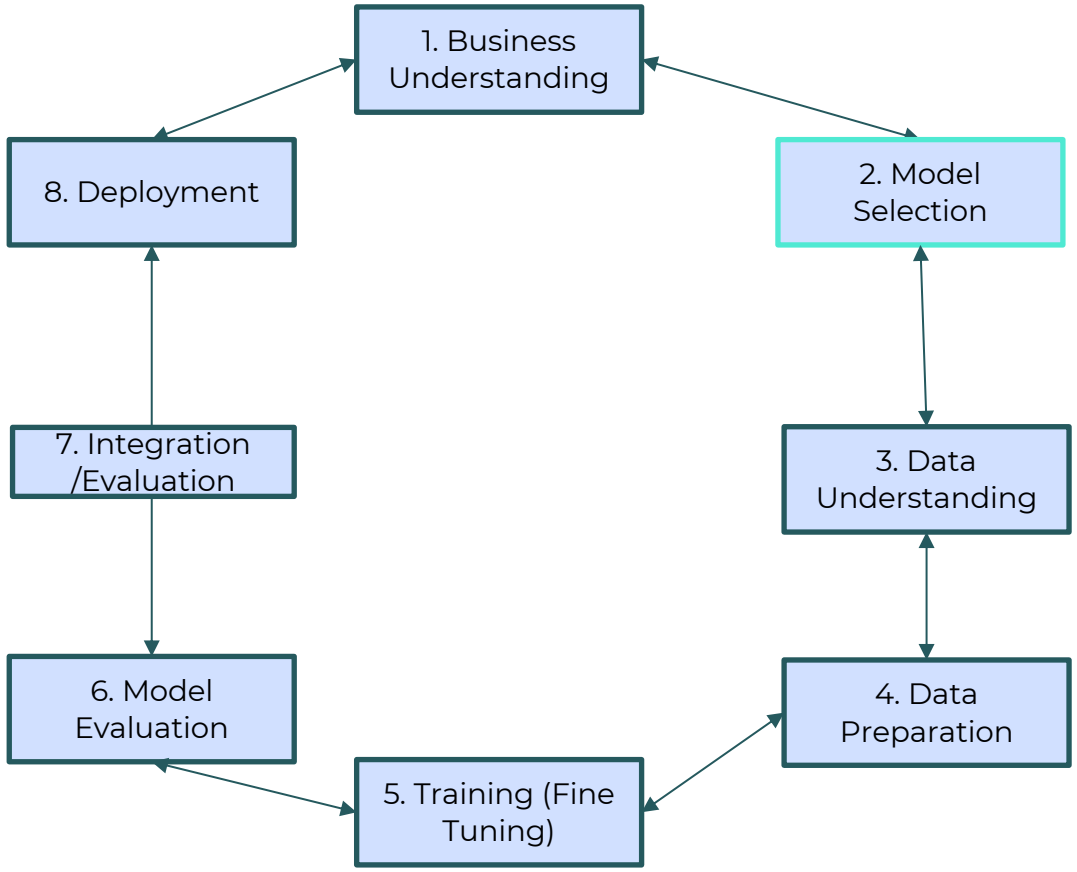
3. und 4. Data Understandlung und Preparation

- Wenn die API Calls bzw. Prompts mit Kontext angereichert werden müssen, z.B. Zusammenfassung wie diese 3 Beispiele (Few Shot Learning)
- ABER deutliche geringerer Datenaufwand

5. Statt des Trainings: wird „In Context gelernt“ also die zusätzliche Anweisungen und Beispiele verwendet

7. Zwischenschritt Integration und Evaluation der Integrierten Lösung

Nutzung Open Source LLMs



Modellauswahl tritt an Stelle der Daten als zentrales Element.

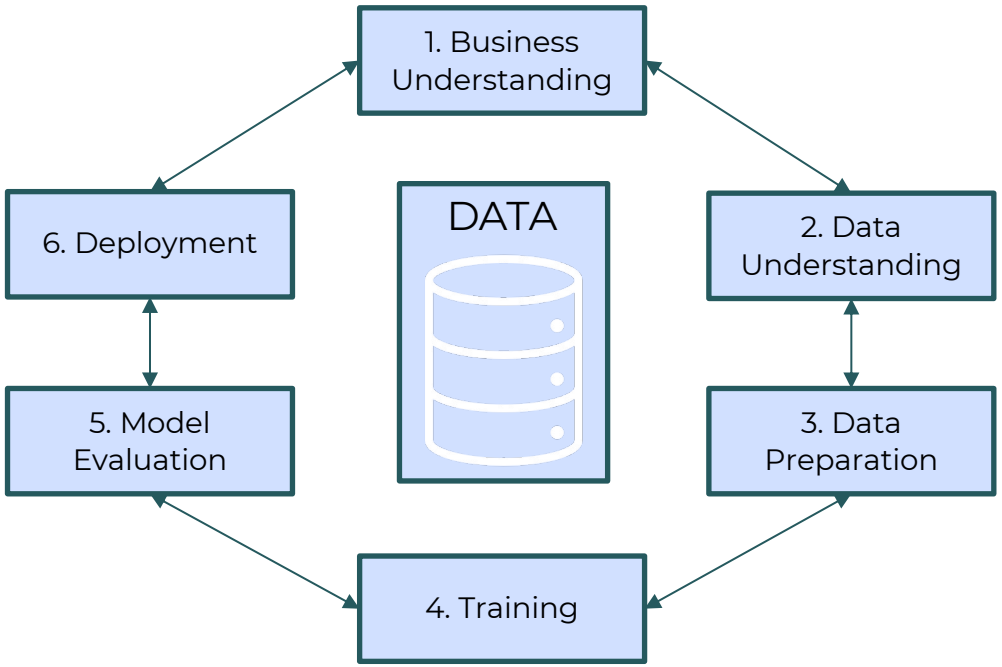
3. und 4. Data Understandlung und Preparation

- **Wenn die API Calls bzw. Prompts mit Kontext angereichert werden müssen, z.B. Zusammenfassung wie diese 3 Beispiele (Few Shot Learning)**
- **ABER deutliche geringerer Datenaufwand**

5. Neben den In Context Learning muss das Modell evtl. Fingetuned werden

7. Zwischenschritt Integration und Evaluation der Integrierten Lösung

Training eines LLM vom Scratch



Fast keine Änderung gegenüber dem klassischen Projektmanagement, aber:

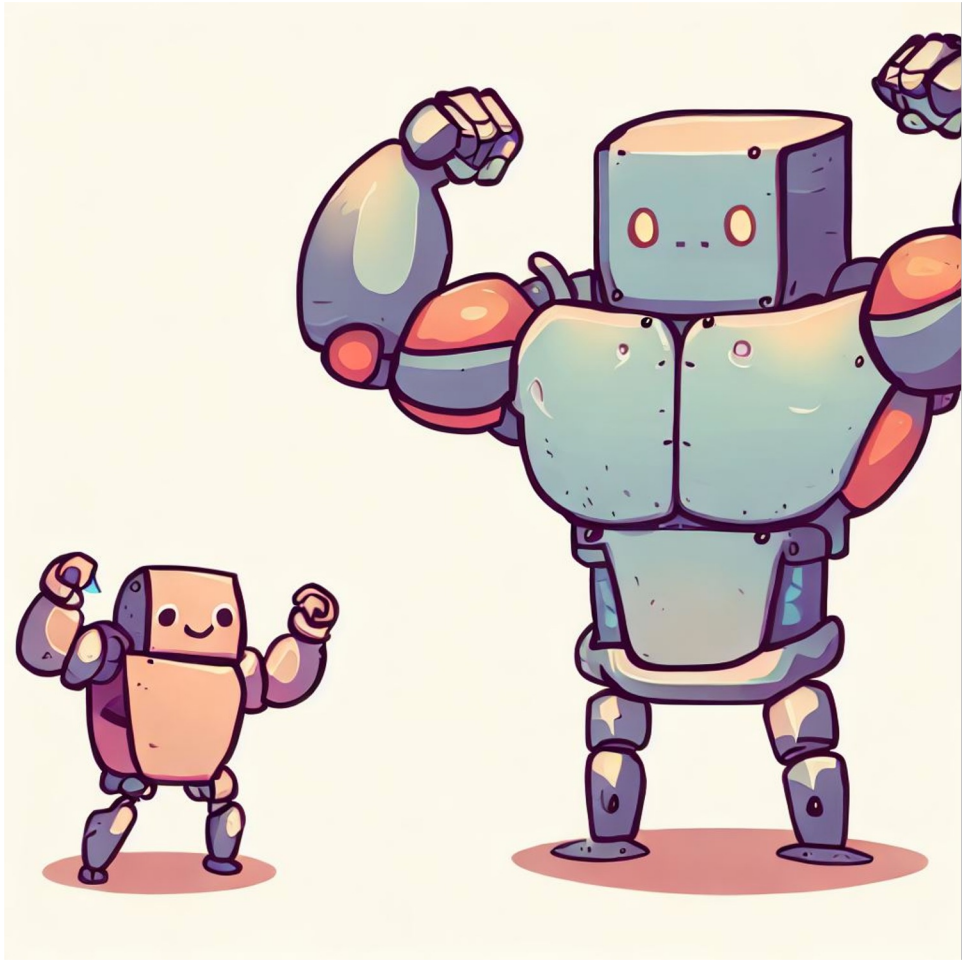
- **Projektrunde für das Sprachmodell (+ Finetuning)**

Ist das neue Modell fertig, wird abhängig vom Anwendungsfall eine zusätzlicher Projektablauf – Analog der Strategie 1 nötig.

Evaluation: Vom Bauchgefühl zur produktiven Anwendung

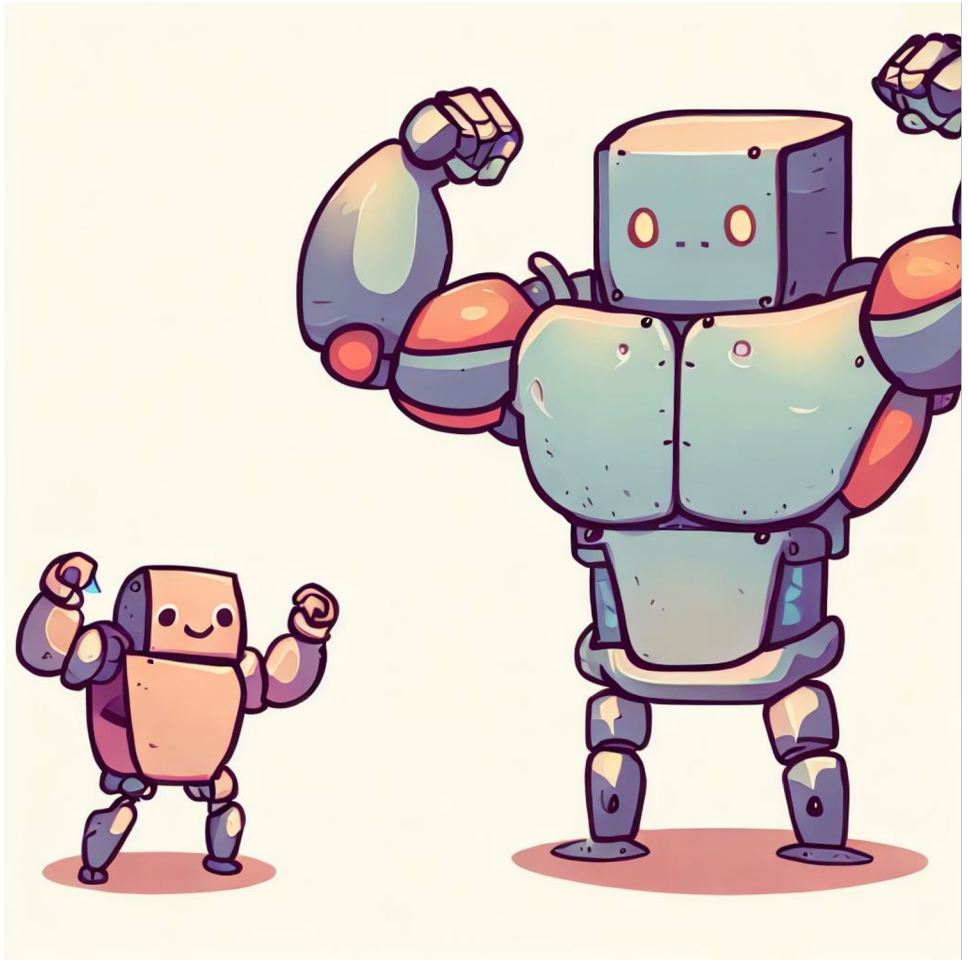
Dimensionen von Qualität - Leistungsfähigkeit

- Abhängig von Task und Benchmark
- Wichtige Tasks
  - Information Extraction
  - Text Classification
  - Question Answering
    - Weltwissen
    - Dokumentenbasiert
  - Summarization
  - Textgenerierung (Nach Anweisungen)
  - Code Generierung
  - Deutsch / Englisch / Multilingual
  - Sensibleness /
  - Specificity / Interestingness



Dimensionen von Qualität - Leistungsfähigkeit

- Abhängig von Task und Benchmark
- Benchmarks
  - Big-Bench
  - Helm



Big Bench

Keyword	Number of tasks	Description
traditional NLP tasks		
contextual question-answering	22	identifying the meaning of a particular word/sentence in a passage
context-free question answering	24	responses rely on model's knowledge base, but not on context provided during query time
reading comprehension	36	a superset of contextual question-answering, measuring the degree to which a model understands the content of a text block
conversational question answering	1	a superset of reading comprehension, measuring the degree to which a model understands the content of a text block and a conversation
summarization	8	involves summarizing a block of text
paraphrase	14	express the same meaning using different words
text simplification	1	express the same meaning using simpler vocabulary
word sense disambiguation	11	identifying the meaning of a word based upon the context it appears
coreference resolution	1	finding all expressions that refer to the same entity in a text

HELM

61 models

AI21 Labs / J1-Jumbo v1 (178B)
AI21 Labs / J1-Large v1 (7.5B)
AI21 Labs / J1-Grande v1 (17B)
AI21 Labs / J1-Grande v2 beta (17B)
AI21 Labs / Jurassic-2 Jumbo (178B)
AI21 Labs / Jurassic-2 Grande (17B)
AI21 Labs / Jurassic-2 Large (7.5B)
Aleph Alpha / Luminous Base (13B)
Aleph Alpha / Luminous Extended (30B)
Aleph Alpha / Luminous Supreme (70B)
Anthropic / Anthropic-LM v4-s3 (52B)
Anthropic / Anthropic Claude v1.3
Anthropic / Anthropic Claude Instant V1
UC Berkeley / Koala (13B)
BigScience / BLOOM (176B)
BigScience / BLOOMZ (176B)
BigScience / T0pp (11B)
BigCode / SantaCoder (1.1B)
BigCode / StarCoder (15.5B)
Cerebras / Cerebras GPT (6.7B)
Cerebras / Cerebras GPT (13B)

42 scenarios

Question answering

- MMLU
- BoolQ
- NarrativeQA
- NaturalQuestions (closed-book)
- NaturalQuestions (open-book)
- QuAC
- HellaSwag
- OpenbookQA
- TruthfulQA

Information retrieval

- MS MARCO (regular)
- MS MARCO (TREC)

Summarization

- CNN/DailyMail
- XSUM

Sentiment analysis

- IMDB

Toxicity detection

59 metrics

Accuracy

- none
- Quasi-exact match
- F1
- Exact match
- RR@10
- NDCG@10
- ROUGE-2
- Bits/byte
- Exact match (up to specified indicator)
- Absolute difference
- F1 (set match)
- Equivalent
- Equivalent (chain of thought)
- pass@1

Calibration

- Max prob
- 1-bin expected calibration error
- 10-bin expected calibration error
- Selective coverage-accuracy area
- Accuracy at 10% coverage

Vergleich ChatGPT mit dem State-of-the-Art

Tasks	Dataset	Metric	Reference	Fine-Tuned SOTA	Zero-Shot SOTA	ChatGPT
Summarization	CNN/DM	ROUGE-1	Lewis et al. (2020a)	44.47	35.27[7]	35.29
	SAMSum	ROUGE-1	Lewis et al. (2020a)	47.28	-	35.29
MT (XXX→Eng	FLoRes-200 (HRL)	ChrF++	Team et al. (2022)	63.5	-	58.64
	FLoRes-200 (LRL)	ChrF++	Team et al. (2022)	54.9	-	27.75
MT (Eng→XXX)	FLoRes-200 (HRL)	ChrF++	Team et al. (2022)	54.4	-	51.12
	FLoRes-200 (LRL)	ChrF++	Team et al. (2022)	41.9	-	21.57
Sentiment Analysis	NusaX - Eng	Macro F1	Winata et al. (2022)	92.6	61.5	83.24
	NusaX - Ind	Macro F1	Winata et al. (2022)	91.6	59.3	82.13
	NusaX - Jav	Macro F1	Winata et al. (2022)	84.2	55.7	79.64
	NusaX - Bug	Macro F1	Winata et al. (2022)	70.0	55.9	55.84
Question Answering	bAbI task 15	Accuracy	Weston et al. (2016a)	100	-	93.3
	bAbI task 16	Accuracy	Weston et al. (2016a)	100	-	66.7
	EntailmentBank	Accuracy	Clark et al. (2018)	86.5	78.58	93.3
	CLUTRR	Accuracy	Minervini et al. (2020)	95.0	28.6	43.3
	StepGame (k=9)	Accuracy	Mirzaee and Kordjamshidi (2022)	48.4	-	23.3
	StepGame (k=1)	Accuracy	Mirzaee and Kordjamshidi (2022)	98.7	-	63.3
	Pep-3k	AUC	Porada et al. (2021)	67.0	-	93.3
Misinformation Detection	COVID-Social	Accuracy	Lee et al. (2021)	77.7	50.0	73.3
	COVID-Scientific	Accuracy	Lee et al. (2021)	74.7	71.1	92.0
Task-Oriented Dialogue	MultiWOZ2.2	JGA	Zhao et al. (2022)	60.6	46.7	24.4
	MultiWOZ2.2	BLEU	Nekvinda and Dušek (2021)	19.1	-	5.65
	MultiWOZ2.2	Inform Rate	Yang et al. (2021)	95.7	-	71.1
Open-Domain KGD	OpenDialKG	BLEU	Ji et al. (2022c)	20.8	3.1	4.1
	OpenDialKG	ROUGE-L	Ji et al. (2022c)	40.0	29.5	18.6
	OpenDialKG	FeQA	Ji et al. (2022c)	48.0	23.0	15.0

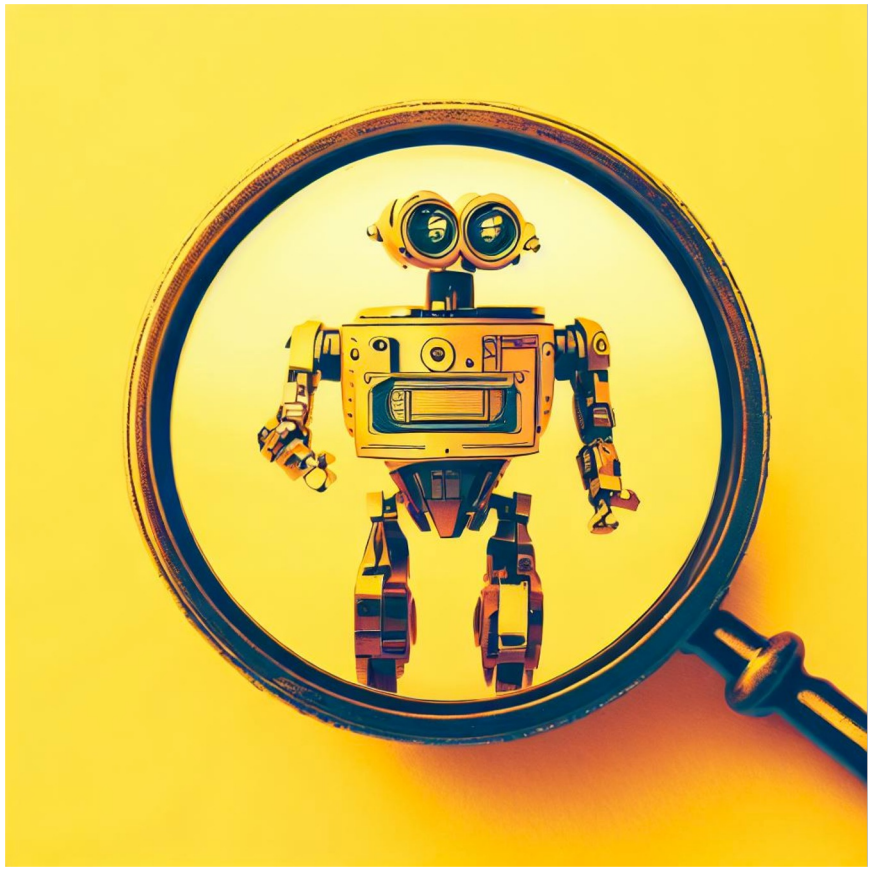
Quelle: Bang et al., 2023

Dimensionen von Qualität - Lizenzen

- Proprietär (OpenAI ChatGPT und GPT4, Aleph Alpha Luminous)

- Open Source
  - A: Nur für wissenschaftliche Zwecke nutzbar (z.B. Alpaka, Vicuna)
  - B: Kommerziell nutzbar ABER (LLama2)
  - C: Uneingeschränkt kommerziell nutzbar (z.B. Falcoon Modelle)

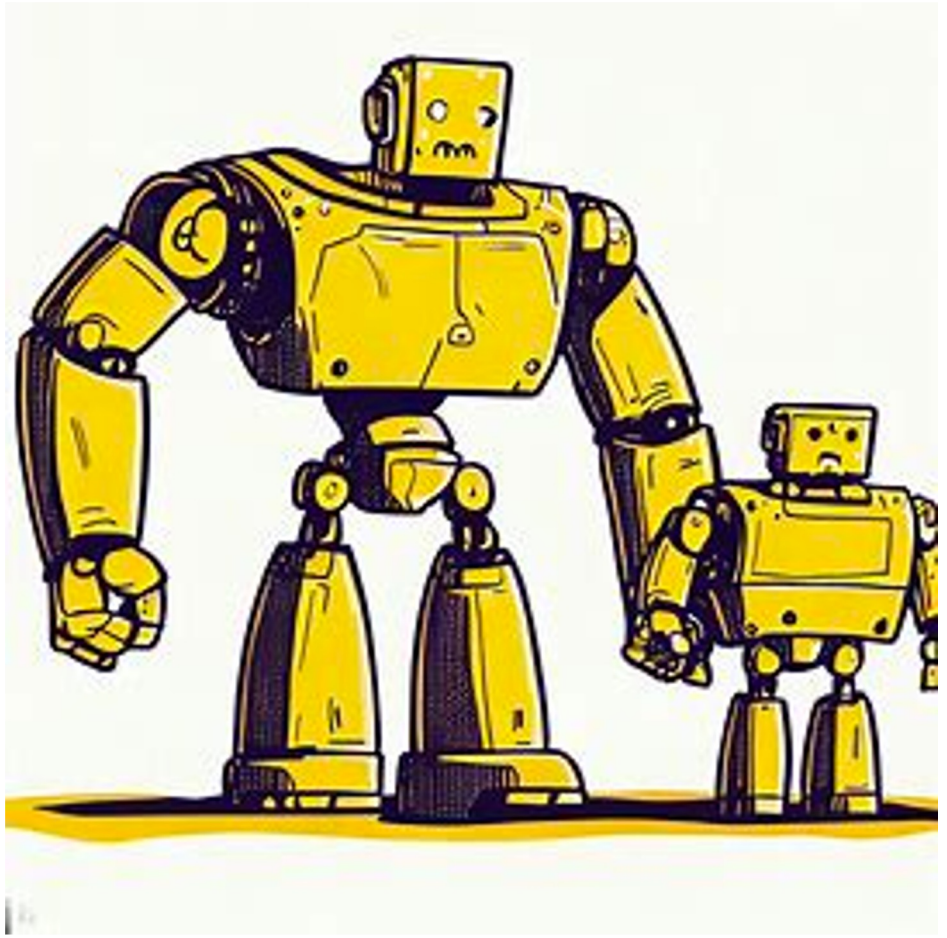
Dimensionen von Qualität - Transparenz



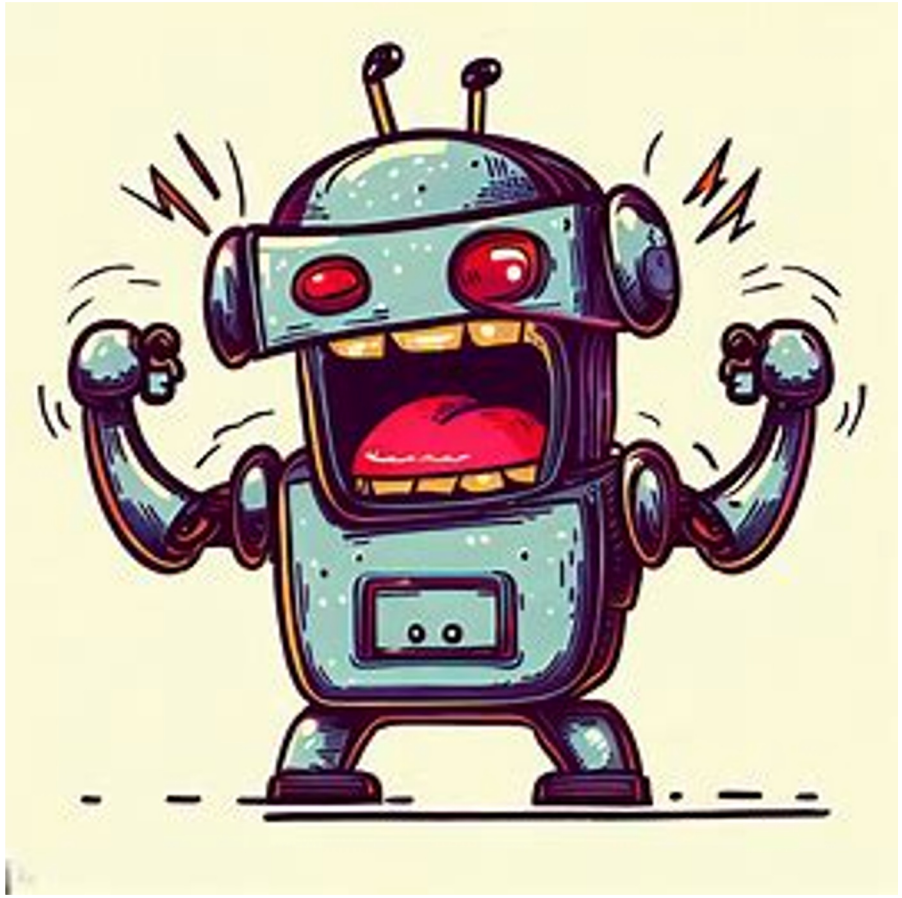
- Proprietäre Modelle haben eine niedrige Transparenz.
- Probleme:
  - Sie ändern sich ohne Ankündigung
  - Es ist unklar mit welchen Zielen sie trainiert wurden

Dimensionen von Qualität - Größe und Aufwand für Deployment

- LLMs sind “Large”. Das neue Falcoon 180b ist 360 GB groß.
- Die 360 GB müssen komplett in den Arbeitsspeicher der Grafikkarte geladen werden. Dazu braucht es noch weiteren GPU Speicher für Inferenz.
- Sharding: Ist das Modell in mehrere kleinere “Shards” aufgeteilt? Wie groß sind die Shards und passen sie zu meiner GPU?



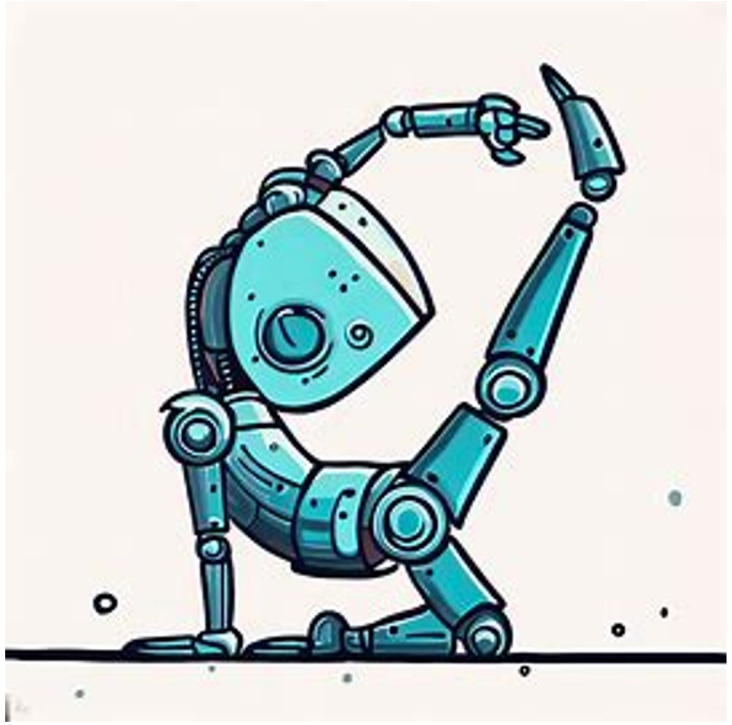
Dimensionen von Qualität - Sicherheit



- Implementiert das Modell Sicherheitsmethoden?
- Wie sind diese implementiert?
- LLMs sprechen frei über jedes Thema und sind niemals zu 100% sicher.

Dimensionen von Qualität: Adaptierbarkeit

- Kommerzielle Modelle muss man meist so nehmen wie sie sind.
- Open Source Modelle sind komplett adaptierbar und damit sehr flexibel.



Open Source Modelle

Leaderboards

- Es exisitieren Evaluationen von Open Source Modellen
  - Hugging Face (https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard?ref=pandia.pro)
  - AlpacaEval (https://tatsu-lab.github.io/alpaca_eval/)
  - Sapling (https://sapling.ai/llm/index)
- Die dahinter liegenden Evaluationen sind jeweils unterschiedlich, aber: Sie sind meistens in Englisch und behandeln viele Disziplinen: Summarisation, Informationsextraktion, Generierung, Recherche, Lösen von Multiple Choice Tests, usw.
- In der Regel ist es wichtig eigene Evaluationen durchzuführen

Rechtliche Themen

Welche Themen muss ich beachten

- GDPR
- AI Act
- Urheberrecht

GDPR

GDPR

- Verordnung der Europäischen Union
- EU-weite Harmonisierung der Regeln für die Verarbeitung personenbezogener Daten
- In Kraft seit 25. Mai 2018

GDPR

Personal Data

Nach der Datenschutz-Grundverordnung ist die Verarbeitung personenbezogener Daten verboten, es sei denn, es liegt eine angemessene Rechtfertigung vor.

Gründe für die Rechtfertigung

- Gültige Einwilligung der betroffenen Person
- Erfüllung eines Vertrags oder Durchführung von vorvertraglichen Maßnahmen.
- Schutz der überwiegenden berechtigten Interessen.
  - Notwendigkeit für die Erfüllung einer rechtlichen Verpflichtung
  - Nur das Recht der EU oder der EU-Mitgliedstaaten ist anwendbar

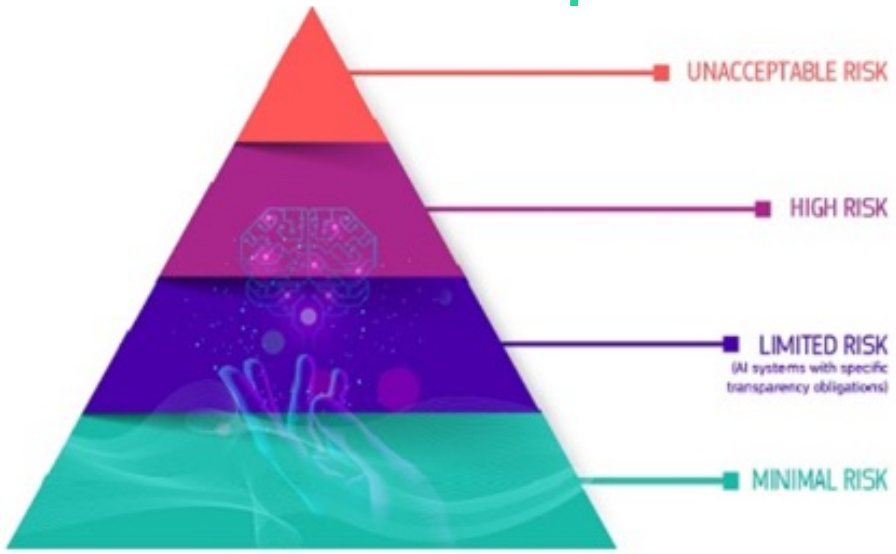
GDPR

Umgang mit personenbezogenen Daten

- Solange die erhobenen Daten nicht personenbezogen sind, können sie in der Regel ohne größere Einschränkungen verwendet werden

- Persönliche Quelldaten entsprechend sichern
- Möglichkeiten
  - Anonymisierung
  - Pseudonymisierung
  - Verschlüsselung

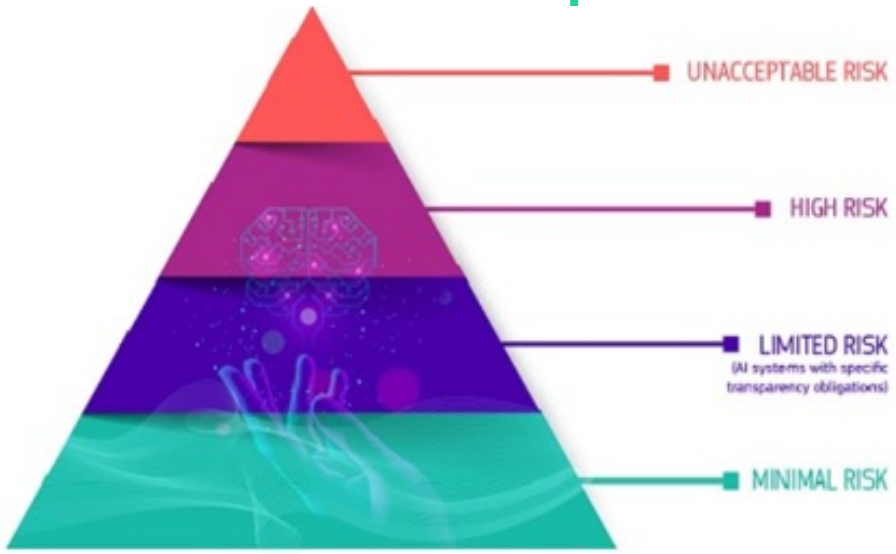
AI Act



AI Act

- Regulates the use of AI in research and business
- Providers and users are affected
- The aim is to categorise AI into risk groups and ban certain AI applications
- Carrying out an assessment to classify AI

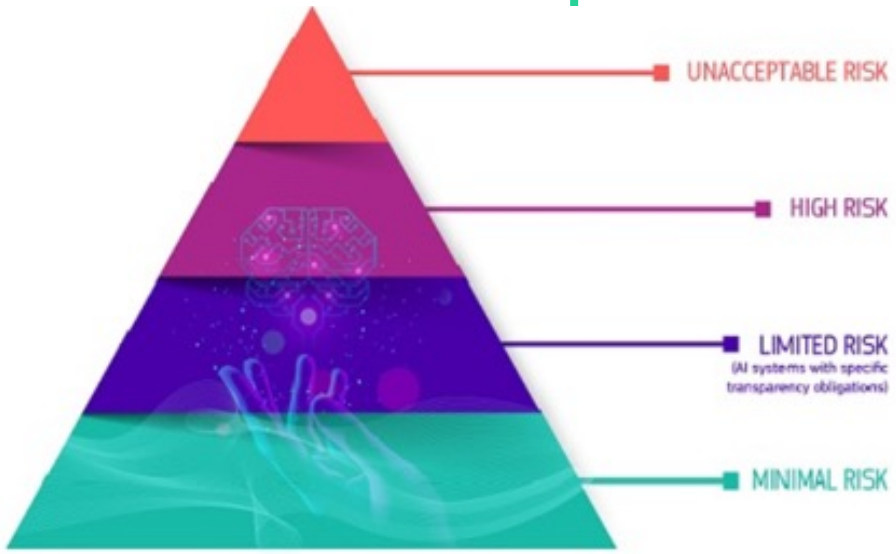
AI Act



AI Act

- Unannehmbares Risiko
- Hohes Risiko
  - Kritische Infrastrukturen (z. B. Verkehr), Sicherheitskomponenten von Produkten
- Begrenztes Risiko
  - KI-Systeme mit spezifischen Transparenzverpflichtungen
- Geringes oder kein Risiko
  - Freie Nutzung (KI-gestützte Videospiele oder Spam-Filter)
  - Die überwiegende Mehrheit der in der EU verwendeten KI-Systeme

AI Act



AI Act Hohes Risiko - Schritte

- Schritt 1:
  - Ein AI-System wird entwickelt
- Schritt 2:
  - Es muss die Konformitätsbewertung durchlaufen und die AI-Anforderungen erfüllen. Bei einigen Systemen ist eine benannte Stelle beteiligt
- Schritt 3:
  - Registrierung von eigenständigen AI-Systemen in einer EU-Datenbank
- Schritt 4:
  - Eine Konformitätsbewertung ist erforderlich. Das AI-System muss das CE-Zeichen tragen.