educa

Erlachstrasse 21
3012 Bern

+41 31 300 55 00
info@educa.ch

(Bildungs-)Daten als Grundlage für KI-Systeme

Dr. Edouard Lamboray

bit media e-Learning Tagung | 06.06.2024

Digitaler Bildungsraum Schweiz
Espace numérique suisse de formation
Spazio formativo digitale svizzero
Spazi da furmazium digital svizzer
Swiss digital education space

AUFTRAG

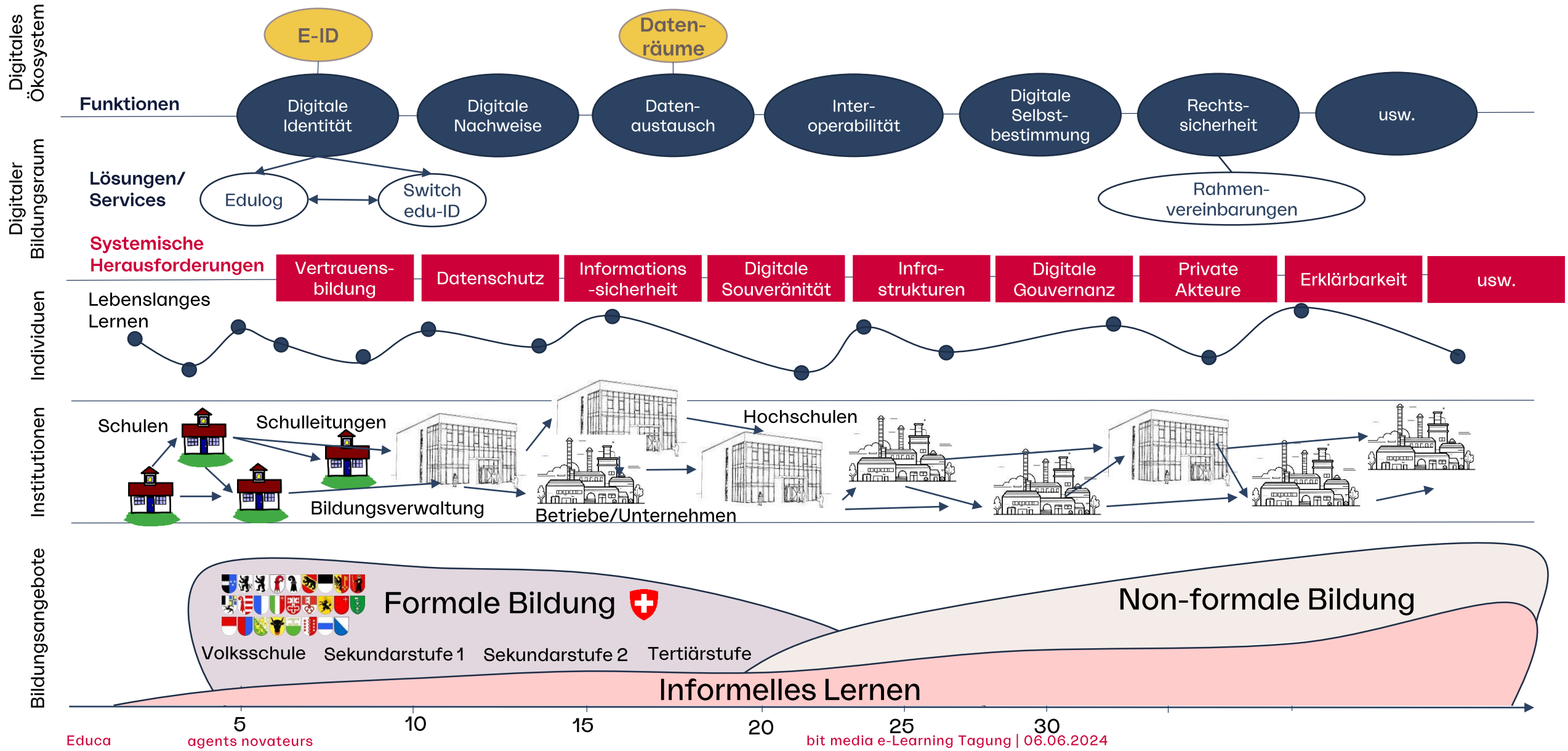
Qualitäts-
entwicklung im
Bildungssystem

«Educa untersucht technologische Entwicklungen und verbindet sie mit der Qualitätsentwicklung der obligatorischen Schule (Primarstufe und Sekundarstufe I), der beruflichen Grundbildung, der Gymnasien und der Fachmittelschulen (Sekundarstufe II). Sie schafft schweizweit Grundlagen für den digitalen Bildungsraum Schweiz.»

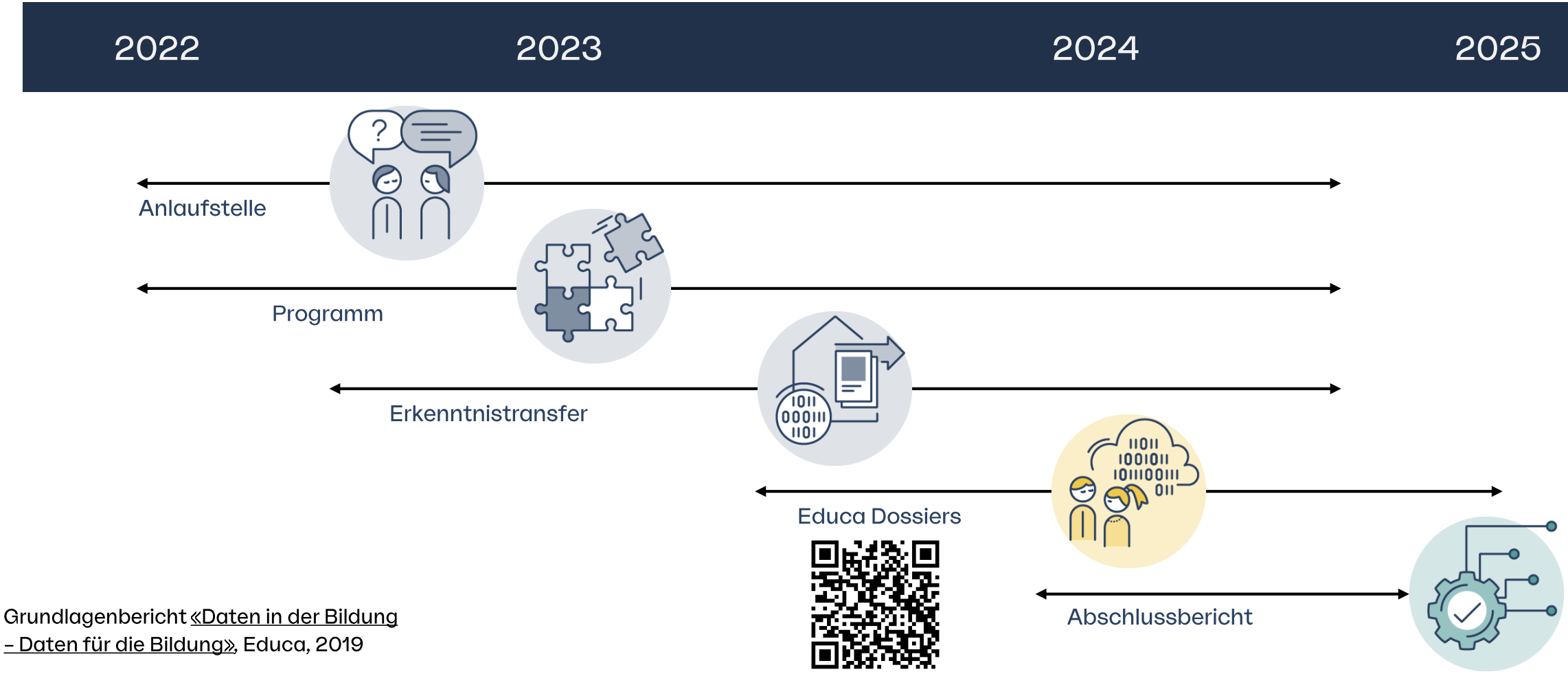
Statut der Fachagentur

Technologische
Entwicklungen

SYSTEMARCHITEKTUR

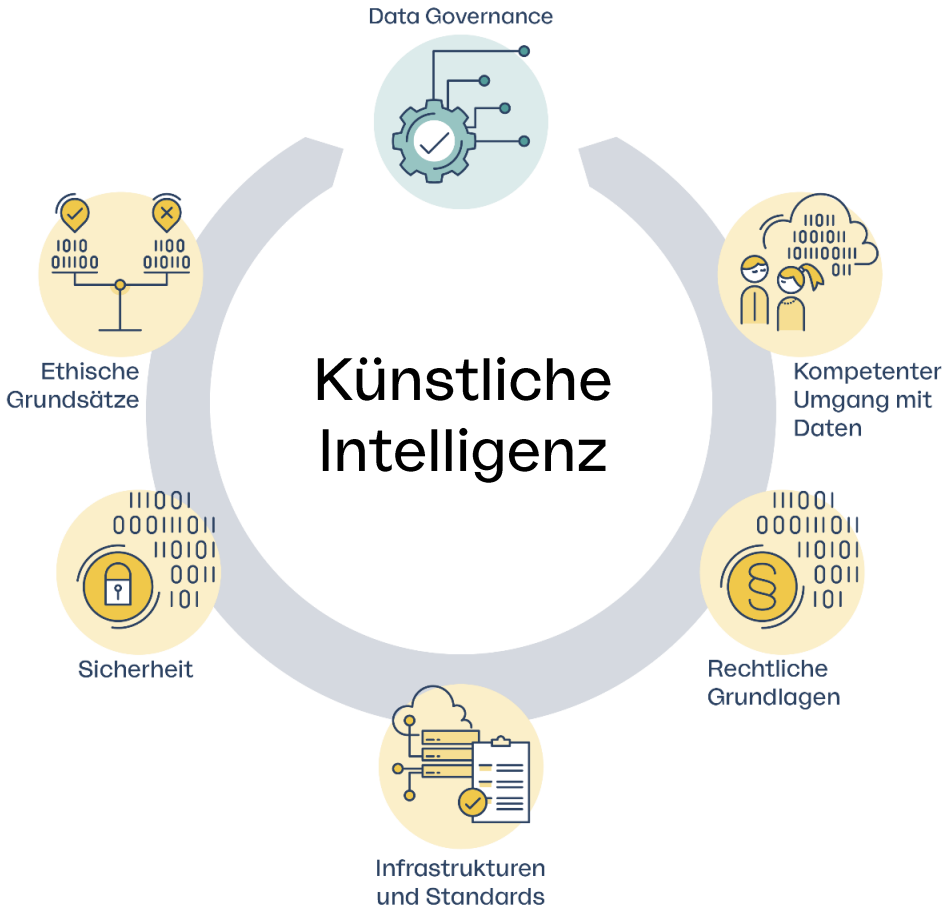


ENTWICKLUNG EINER DATENNUTZUNGSPOLITIK FÜR DEN BILDUNGSRAUM SCHWEIZ



Grundlagenbericht «Daten in der Bildung – Daten für die Bildung», Educa, 2019

Daten und Künstliche Intelligenz



Im DNP Programm:

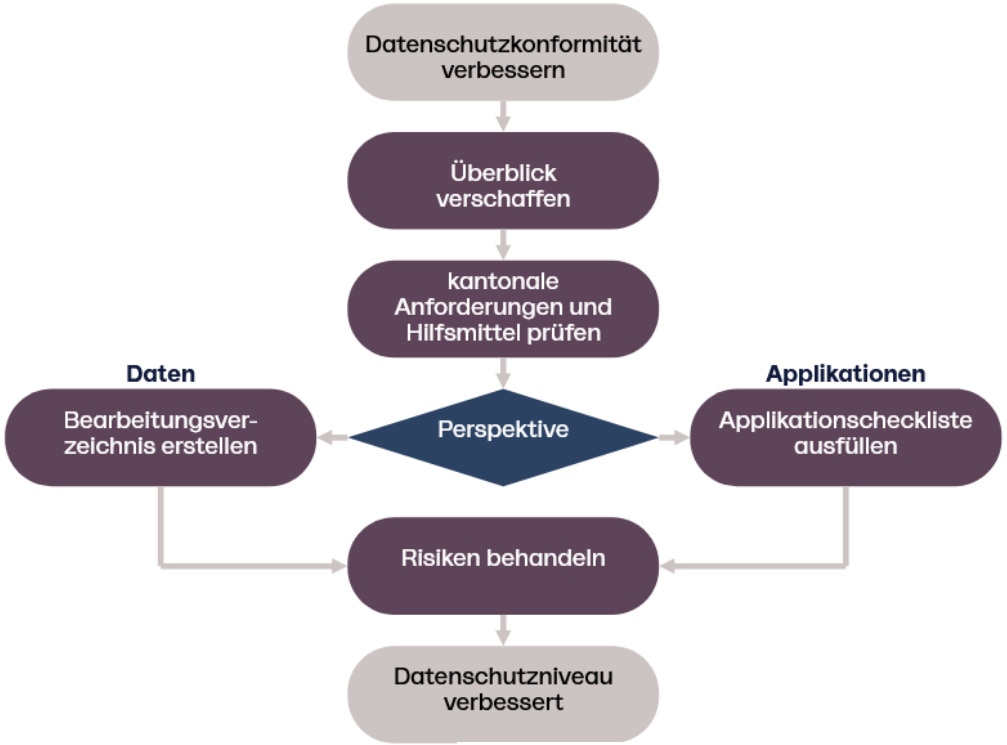
- Was sind die Voraussetzungen für die sozial und pädagogisch nachhaltige Gestaltung von algorithmischen Systemen
- Ausgestaltung von Testumgebungen zur Förderung verantwortungsvoller Innovation

- Datennutzung in KI Systemen
  - Daten- und Modelllebenszyklus
  - Betriebsmodelle von KI Systemen

educa.ch/educa24

- Regulierung von KI
  - Auslegeordnung, horizontal und sektoriell

Informationssicherheit und Datenschutz



NEU ab Mitte Juni
Educa Dossier:
Datenschutzkonforme Schule

Im DNP Programm:

- Datenschutzkonforme Schule
  - Unterstützende Instrumente für die Schulleitungen

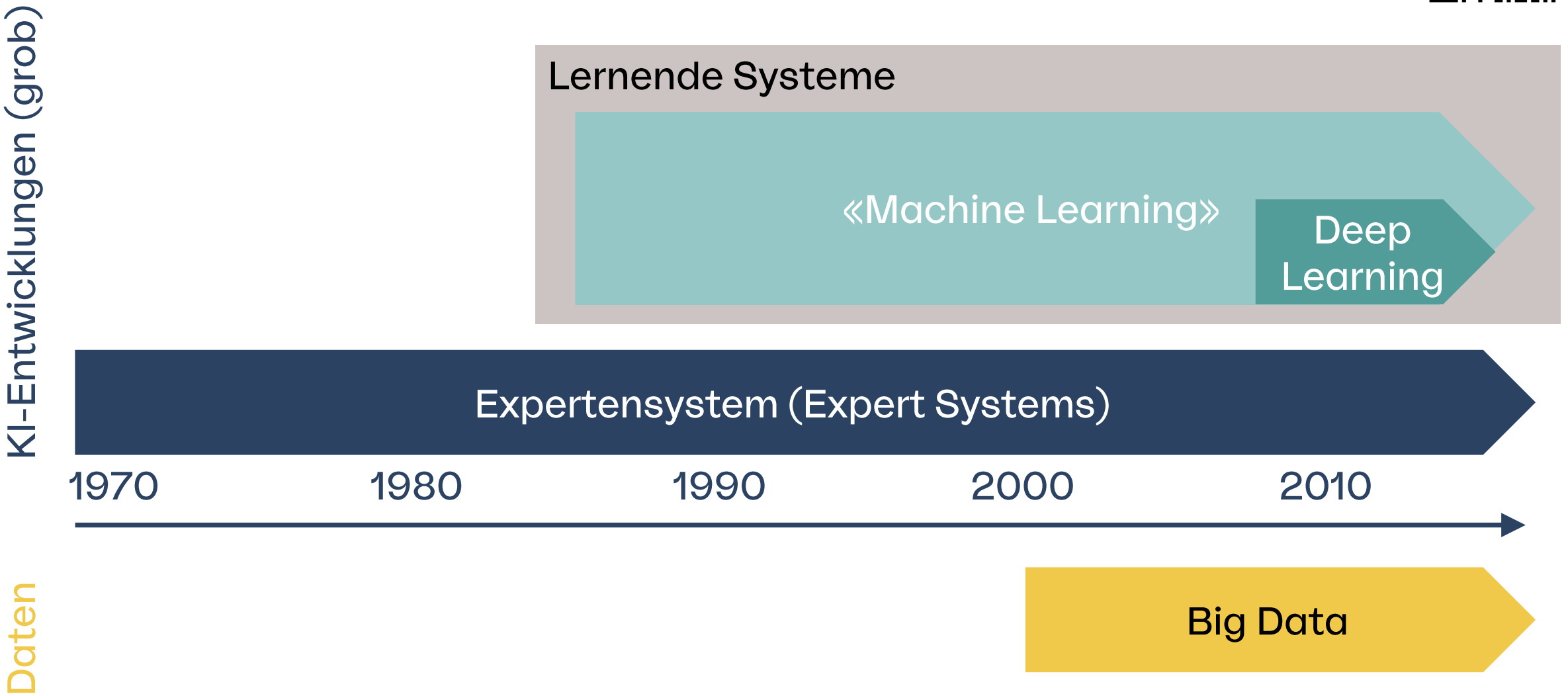
- Datenmanagement-Kompetenzen von Schulen
  - Standortbestimmung und Perspektiven für Weiterentwicklung

Educa Navigator

- Orientierung im Markt digitaler Werkzeuge

KI-SYSTEME ÜBER DIE JAHRZEHNTE

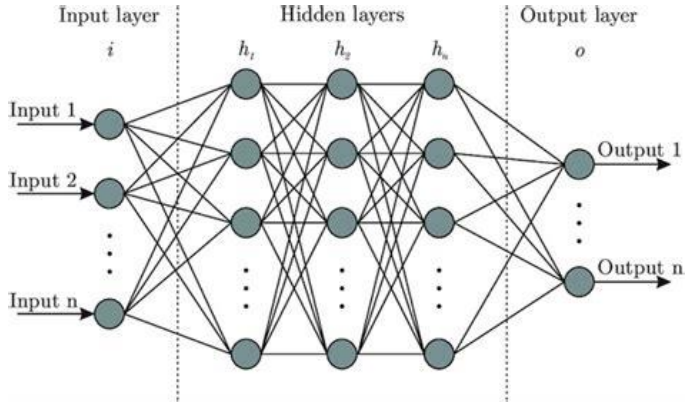
Warum wir den Begriff "Künstliche Intelligenz" vermeiden sollten, Beat Döbeli Honegger



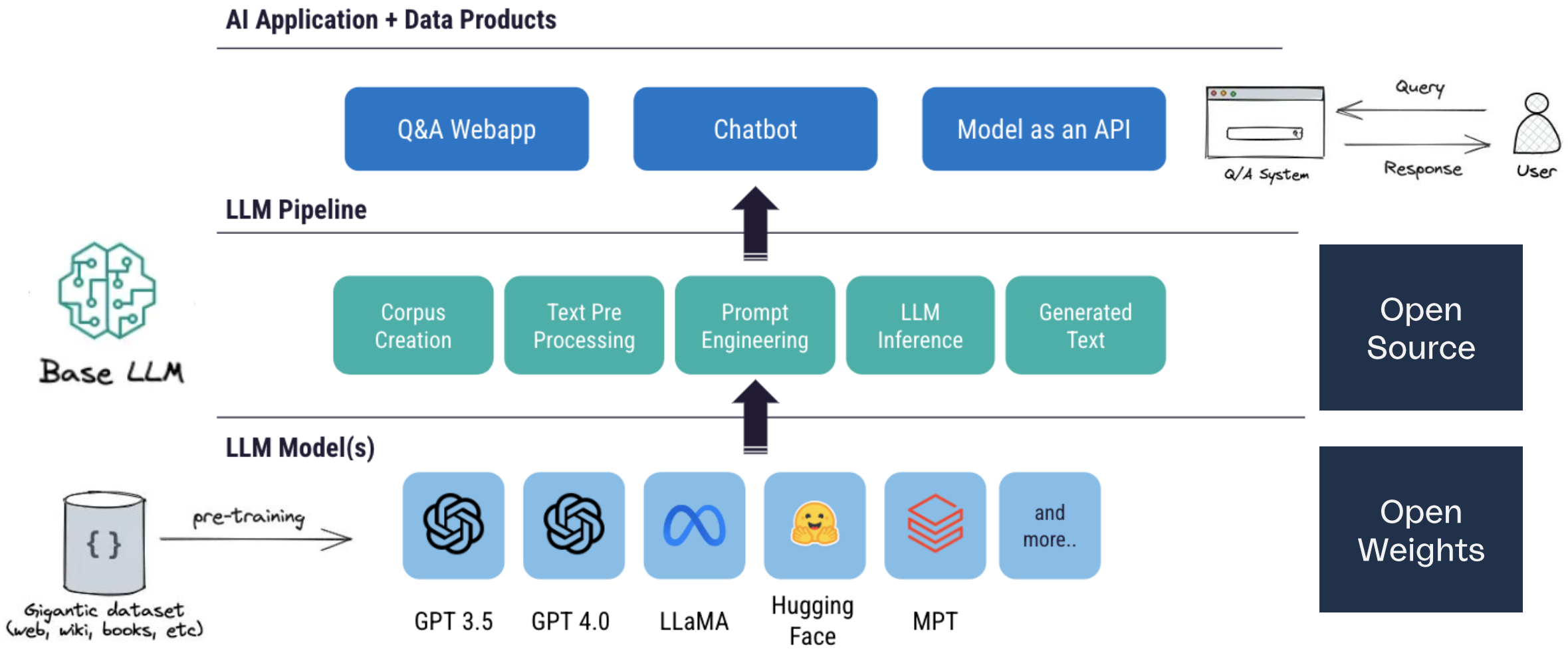
Klassische Algorithmen


```python
def dijkstra(graph,src,dest,visited=[],distances={},predecessors={}):
    """ calculates a shortest path tree routed in src
    """
    if src not in graph:
        raise TypeError('The root of the shortest path tree cannot be found')
    if dest not in graph:
        raise TypeError('The target of the shortest path cannot be found')
    if src == dest:
        path=[]
        pred=dest
        while pred != None:
            path.append(pred)
            pred=predecessors.get(pred,None)
        readable=path[0]
        for index in range(1,len(path)): readable = path[index]+'--->'+readable
        print('shortest path - array: '+str(path))
        print("path: "+readable+",   cost="+str(distances[dest]))
    else :
        if not visited:
            distances[src]=0
        for neighbor in graph[src] :
            if neighbor not in visited:
                new_distance = distances[src] + graph[src][neighbor]
                if new_distance < distances.get(neighbor,float('inf')):
                    distances[neighbor] = new_distance
                    predecessors[neighbor] = src
        visited.append(src)
        unvisited={}
        for k in graph:
            if k not in visited:
                unvisited[k] = distances.get(k,float('inf'))
        x=min(unvisited, key=unvisited.get)
        dijkstra(graph,x,dest,visited,distances,predecessors)
if __name__ == "__main__":
    graph = {'A1': {'B1': 1, 'B2': 1.41, 'A2': 1},
             'A2': {'B1': 1.41, 'B2': 1, 'B3': 1.41, 'A1': 1, 'A3': 1},
             ..........................................................
    dijkstra(graph,'A1','H9')
    shortest path - array: ['H9', 'G8', 'F8', 'E7', 'D6', 'D5', 'C4', 'C3', 'B2', 'A
path: A1--->B2--->C3--->C4--->D5--->D6--->E7--->F8--->G8--->H9,   cost=11.4599999999
```

Künstliche Intelligenz



APPLIKATIONSSCHICHTEN EINER LLM ANWENDUNG



Bildquellen: https://blog.dataiku.com/llm-in-production und https://towardsdatascience.com/

OFFEN UND TRANSPARENT?

Open Source	MPT
Open Weights	LLaMA
Closed	GPT 3.5, GPT 4.0

https://promptengineering.org/

Foundation Model Transparency Index Scores by Major Dimensions of Transparency, 2023

Source: 2023 Foundation Model Transparency Index

Major Dimensions of Transparency	Meta Llama 2	BigScience BLOOMZ	OpenAI GPT-4	stability.ai Stable Diffusion 2	Google PaLM 2	ANTHROPIC Claude 2	cohere Command	AI21 labs Jurassic-2	Inflection Inflection-1	amazon Titan Text	Average
Data	40%	60%	20%	40%	20%	0%	20%	0%	0%	0%	20%
Labor	29%	86%	14%	14%	0%	29%	0%	0%	0%	0%	17%
Compute	57%	14%	14%	57%	14%	0%	14%	0%	0%	0%	17%
Methods	75%	100%	50%	100%	75%	75%	0%	0%	0%	0%	48%
Model Basics	100%	100%	50%	83%	67%	67%	50%	33%	50%	33%	63%
Model Access	100%	100%	67%	100%	33%	33%	67%	33%	0%	33%	57%
Capabilities	60%	80%	100%	40%	80%	80%	60%	60%	40%	20%	62%
Risks	57%	0%	57%	14%	29%	29%	29%	29%	0%	0%	24%
Mitigations	60%	0%	60%	0%	40%	40%	20%	0%	20%	20%	26%
Distribution	71%	71%	57%	71%	71%	57%	57%	43%	43%	43%	59%
Usage Policy	40%	20%	80%	40%	60%	60%	40%	20%	60%	20%	44%
Feedback	33%	33%	33%	33%	33%	33%	33%	33%	33%	0%	30%
Impact	14%	14%	14%	14%	14%	0%	14%	14%	14%	0%	11%
Average	**57%**	**52%**	**47%**	**47%**	**41%**	**39%**	**31%**	**20%**	**20%**	**13%**	

https://crfm.stanford.edu/fmti/

ecosystem graphs

Foundation models define an emerging ecosystem of datasets, models, and applications. As their footprint grows, so does their societal impact.

Ecosystem graphs is a framework to document the foundation models ecosystem, namely both the assets (datasets, models, and applications) and their relationships. Using it, one can answer questions such as: What are the latest foundation models? Who builds them and where are they used downstream? What are the general trends over time? We hope that ecosystem graphs will be a useful resource for researchers, application developers, policymakers, and the public to better understand the foundation models ecosystem.

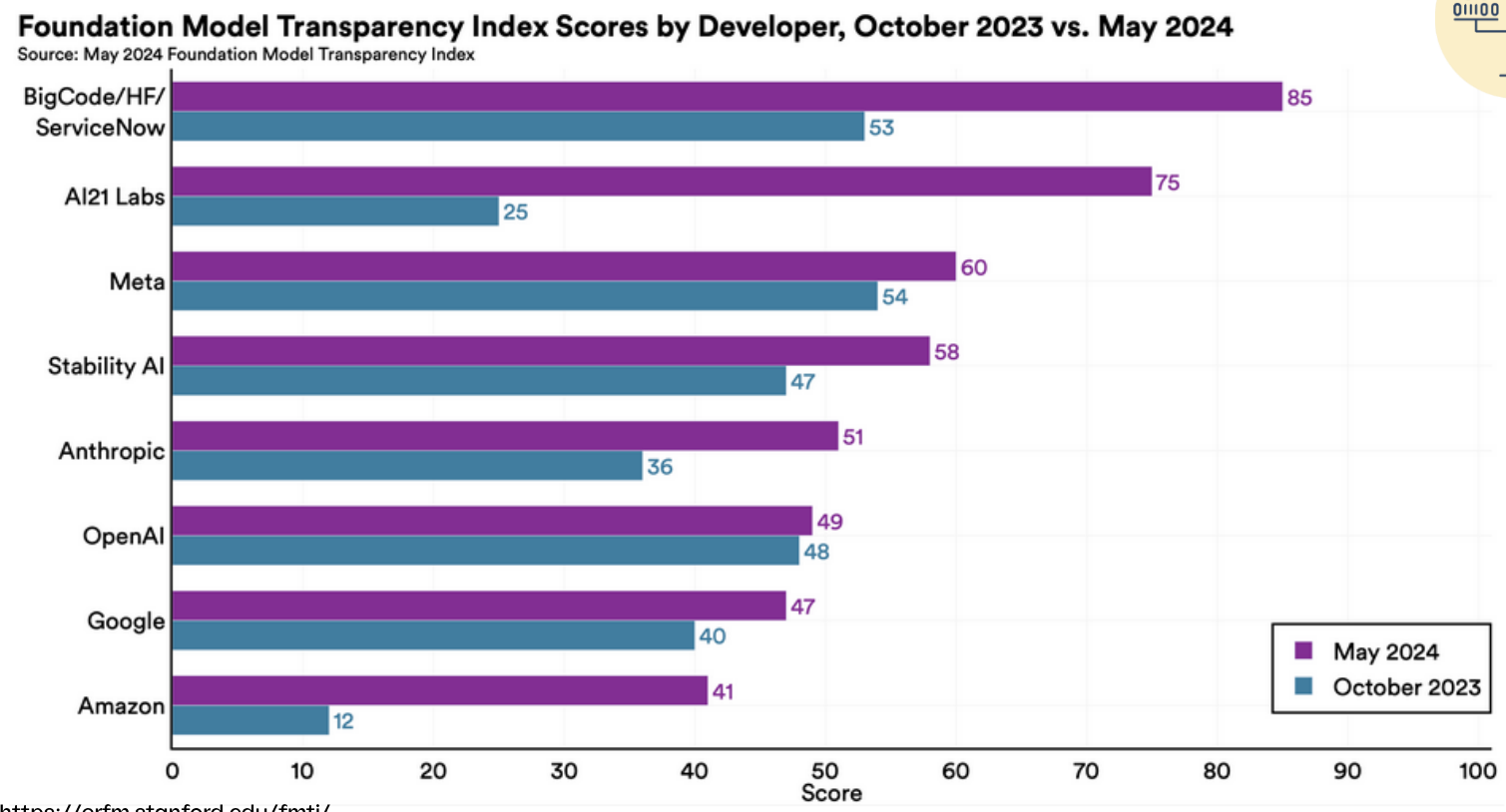
https://crfm.stanford.edu/ecosystem-graphs/index.html

Type	Name	Organization	Created date	Size	Modality (In; Out)	Access	License	Dependencies
model	Med-Gemini	Google	Apr 29, 2024	unknown	image, text; text	closed	unknown	Gemini | MultiMedBench
dataset	CosmicMan-HQ 1.0	Shanghai AI Laboratory	Apr 28, 2024	6 million images	image	open	unknown	
model	OpenELM	Apple	Apr 24, 2024	3B parameters	text; text	open	Apple	RefinedWeb | The Pile | RedPajama-Data | Dolma | CoreNet library
model	Phi-3 Mini	Microsoft	Apr 23, 2024	3.8B parameters	text; text	open	MIT	
model	Aurora-M	Tokyo Institute of Technology, MIT-IBM Watson Lab, Sapienza University of Rome	Apr 23, 2024	15B parameters	text; text	open	unknown	StarCoderPlus

OFFEN UND TRANSPARENT?

Open Source	MPT
Open Weights	LLaMA
Closed	GPT 3.5, GPT 4.0

https://promptengineering.org/



https://crfm.stanford.edu/fmti/

ecosystem graphs

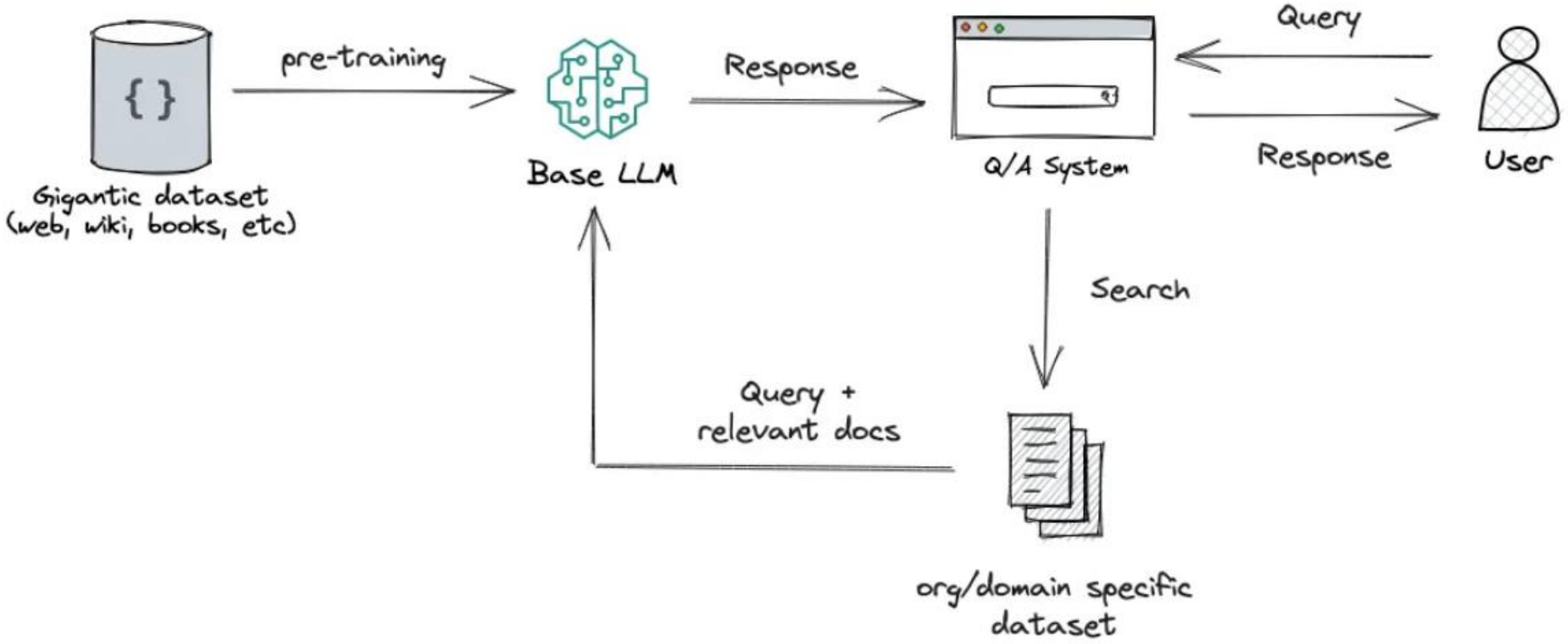
Foundation models define an emerging ecosystem of datasets, models, and applications. As their footprint grows, so does their **societal impact**.

Ecosystem graphs is a framework to document the foundation models ecosystem, namely both the assets (datasets, models, and applications) and their relationships. Using it, one can answer questions such as: What are the latest foundation models? Who builds them and where are they used downstream? What are the general trends over time? We hope that ecosystem graphs will be a useful resource for researchers, application developers, policymakers, and the public to better understand the foundation models ecosystem.

https://crfm.stanford.edu/ecosystem-graphs/index.html

Type	Name	Organization	Created date	Size	Modality (In; Out)	Access	License	Dependencies
model	Med-Gemini	Google	Apr 29, 2024	unknown	image, text; text	closed	unknown	Gemini | MultiMedBench
dataset	CosmicMan-HQ 1.0	Shanghai AI Laboratory	Apr 28, 2024	6 million images	image	open	unknown	
model	OpenELM	Apple	Apr 24, 2024	3B parameters	text; text	open	Apple	RefinedWeb | The Pile | RedPajama-Data | Dolma | CoreNet library
model	Phi-3 Mini	Microsoft	Apr 23, 2024	3.8B parameters	text; text	open	MIT	
model	Aurora-M	Tokyo Institute of Technology, MIT-IBM Watson Lab, Sapienza University of Rome	Apr 23, 2024	15B parameters	text; text	open	unknown	StarCoderPlus

Retrieval Augmented Generation (RAG)



Bildquellen: https://towardsdatascience.com/

educa 24

18.9.2024

Daten als Grundlage für KI-Systeme

agents novateurs

educa.ch/educa24