

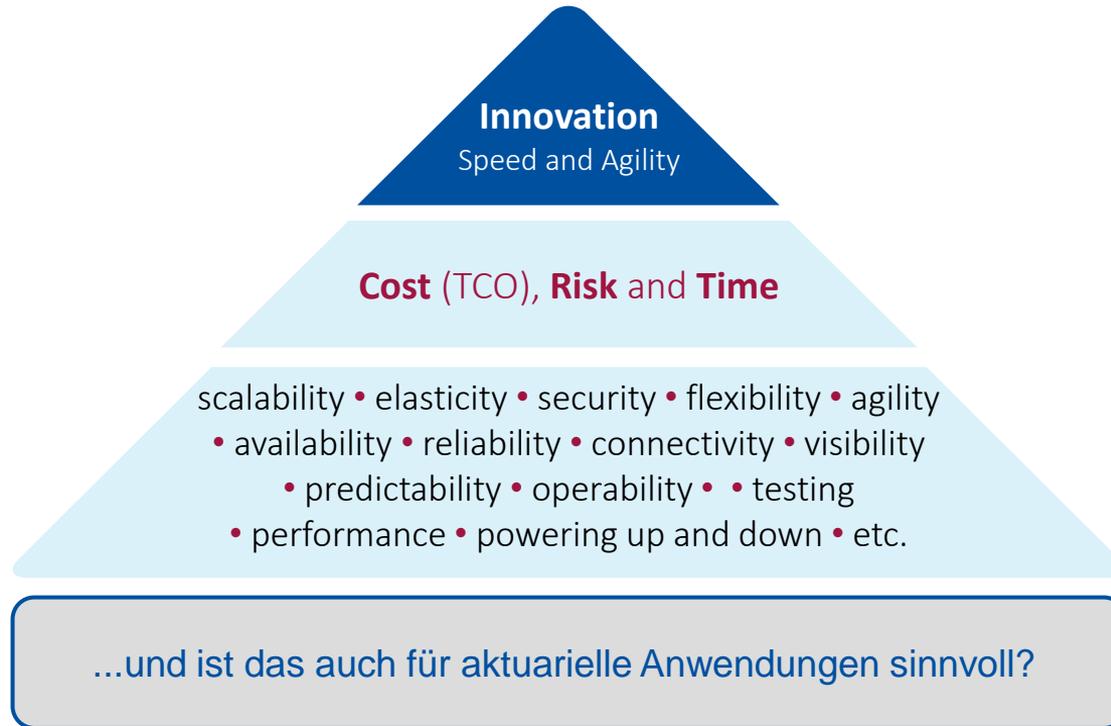
DAV/DGVFM
Herbsttagung
2024

Claudia Schmidt und Georg van Bühren, BELTIOS GmbH

**KI und Cloud meets qx: Aktuarielle
Berechnungen treffen moderne
Technologien – passt das?**

DAV/DGVFM Herbsttagung 2024

Warum ziehen alle plötzlich in die Cloud?

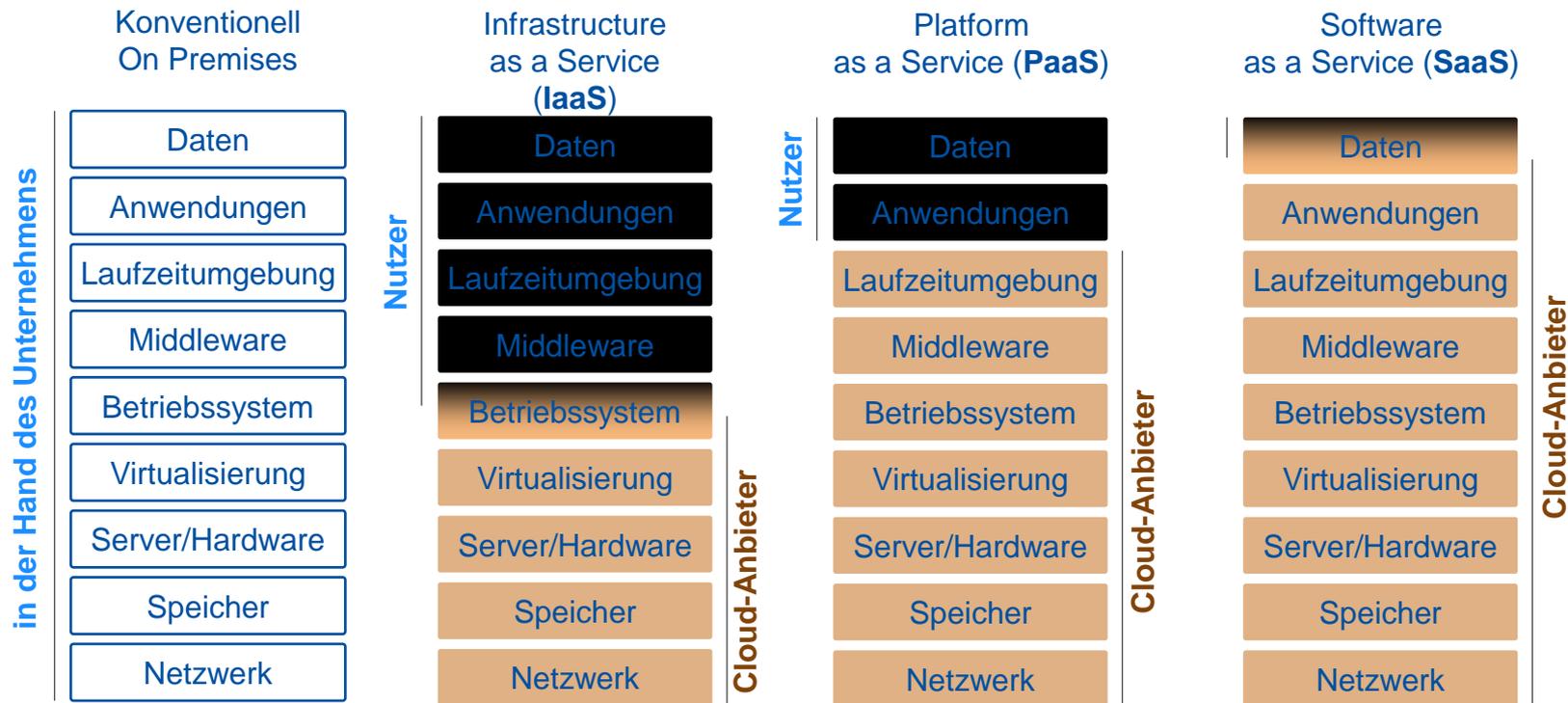


Die Cloud-Transformation besteht nicht nur aus Technik



Verantwortungsbereiche der Cloud-Service-Modelle

sind unterschiedlich verteilt



➔ Für aktuarielle Tätigkeit interessieren: ANWENDUNGEN und DATEN

Vorteile von Anwendungen in der Cloud



✓ **Gemeinsame Nutzung gleicher Anwendungen**

- Unterschiedliche Tools mit gleichen Funktionalitäten können konsolidiert werden
- Kernfunktionalitäten können von diversen Anwendungen genutzt werden



✓ **Effektivität und Modernität der Bereitstellung von (neuen) IT-Services und Anwendungen erhöhen**

- durch klare Kostentransparenz der IT-Komponenten die Nutzen und Konsequenzen abschätzen (z.B. Pay as you go)
- unmittelbarer Einsatz moderner Verfahren und Technologien eröffnen (z.B. DevOps, AI)



✓ **digitale Arbeitsweisen der Mitarbeiter nachhaltig und dauerhaft sicherstellen**

- Fachservices zeitnah implementieren und nutzen
- durch Einsatz moderner und zeitgemäßer Kollaborations-Lösungen immer digitaler werden

Vor allem KI-Modelle lassen sich in der Regel nur über Cloud-Anbieter sinnvoll dauerhaft einsetzen

Datenverwendung & Datenschutz

Datenverwendung

Die Daten werden bestimmungsgemäß verwendet

- ✓ **Vertraglich festgelegte Nutzungsbedingungen, die die Nutzung der Daten zum Training von KIs ausschließen**
- ✓ **Keine Anfertigung von Übersichten der berechneten Tarife und Produkte**
- ✓ **Keine Auswertung der Nutzung auf einzelne Benutzer**
- ✓ **Auswertung von Berechnungen nur zum bestimmungsgemäßen Zweck der Abrechnung für pay-per-use**

Datenschutz

Die Daten werden nicht persistiert

- ✓ **Keine Speicherung oder Verarbeitung von personenbezogenen Vertragsdaten**
- ✓ **Keine Speicherung von Tarif- und Produktdaten**
- ✓ **Zukünftig Unterstützung für homomorphe Verschlüsselung**
- ✓ **State-of-the-art Verschlüsselung bei der Übertragung der Daten**

Die Bereitstellung des actuariellen Rechners kann in unterschiedlichen Varianten erfolgen: **SaaS, Managed Service, on premises.**

Dadurch können zusätzliche **individuelle Datenschutzerfordernungen** erfüllt werden

Compliance mit

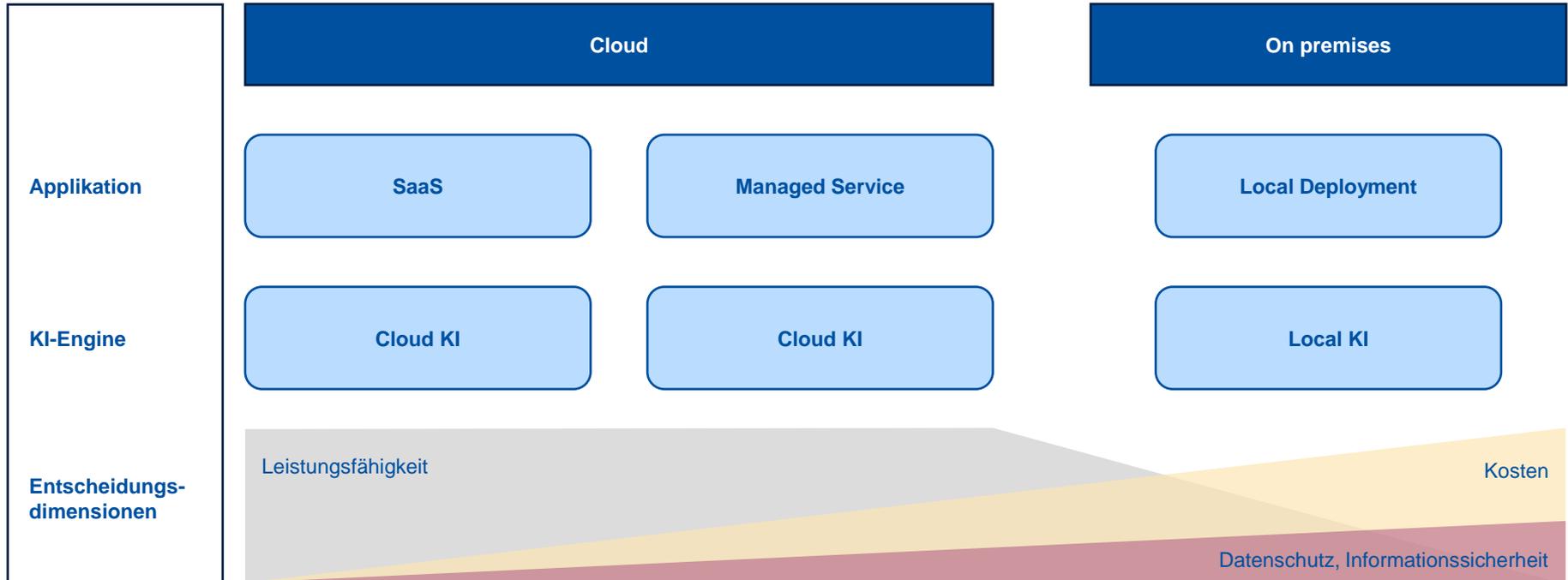
DSGVO

DORA

EU AI Act

ISO27001

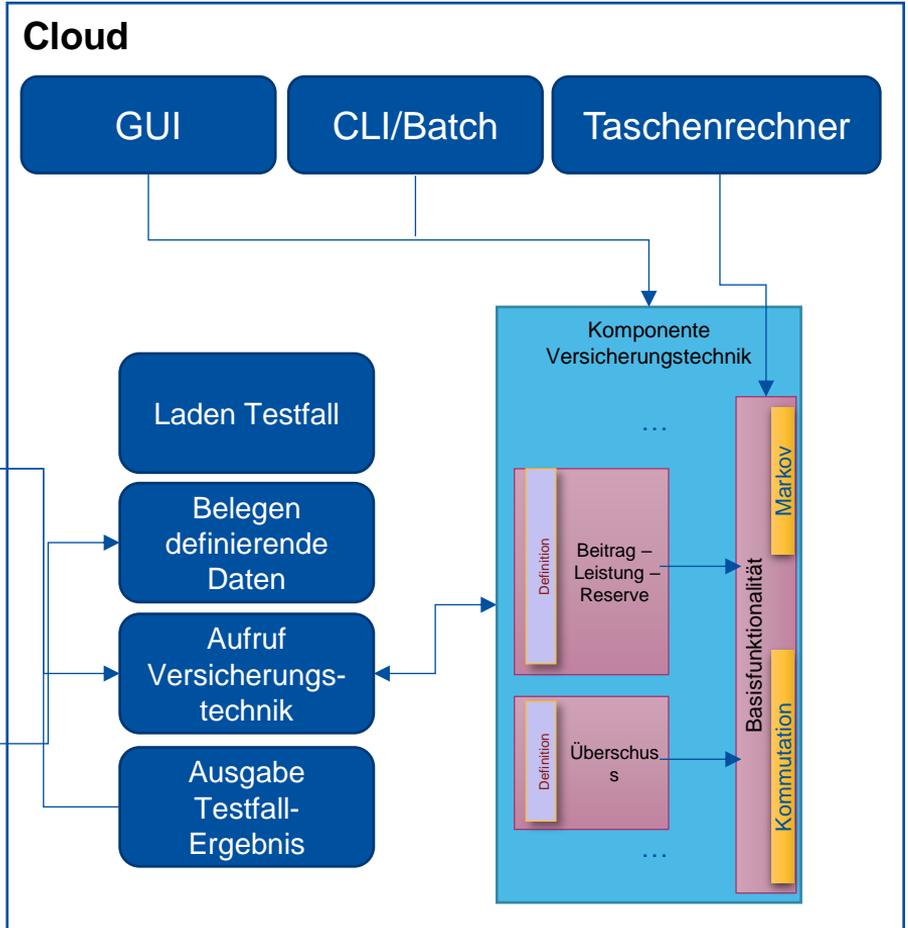
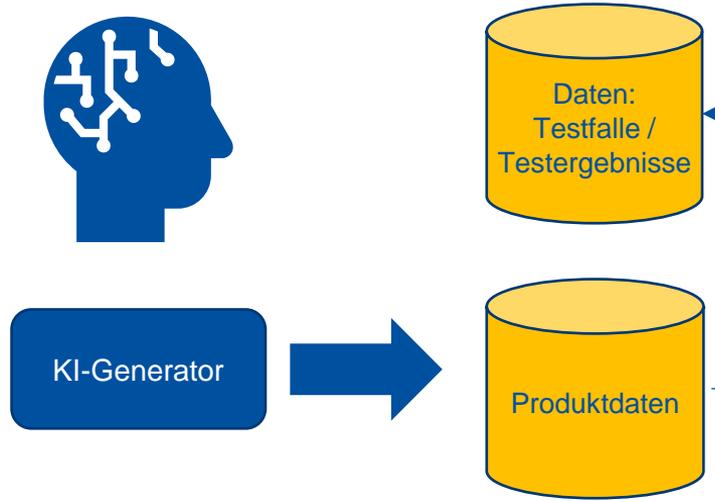
Alternativen für die Bereitstellung



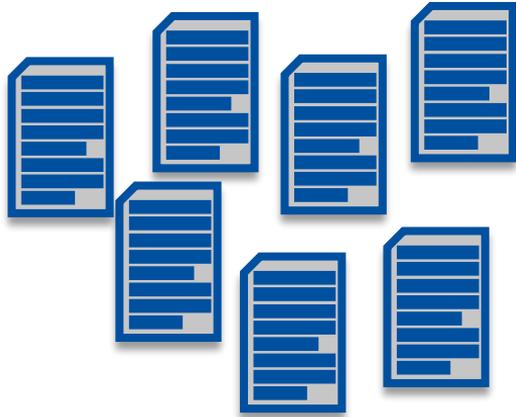
Derzeit ist der actuarielle Rechner bei AWS als SaaS bereitgestellt.

Weitere Optionen auf Anfrage möglich, wie z. B. SaaS auf Microsoft Azure oder bei deutschem/ europäischen Cloud-Anbieter.

Use Case: Aktuarieller Rechner mit KI-Generator für die Produktdaten



Ausgangslage und Anwendergruppe



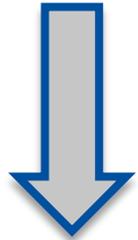
Bestandsmigrationen bei Lebensversicherungen benötigen große Mengen an technisch umgesetzten Tarifen für den Abgleich im Testrechner actualib

→ Ausgangslage sind Geschäftspläne

Versicherungsgesellschaften mit Bestandsmigration (und Neugeschäft)

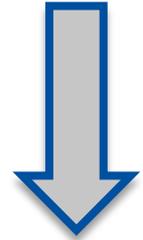


Problem & Pain Points



Manuelle Extraktion und
Verarbeitung ist sehr
zeitintensiv und monoton

Fachexperten/innen werden
dafür benötigt, bieten
allerdings an anderer Stelle
einen größeren Mehrwert



Vorteile durch KI-Unterstützung



Die Implementierungszeit wird deutlich verkürzt.



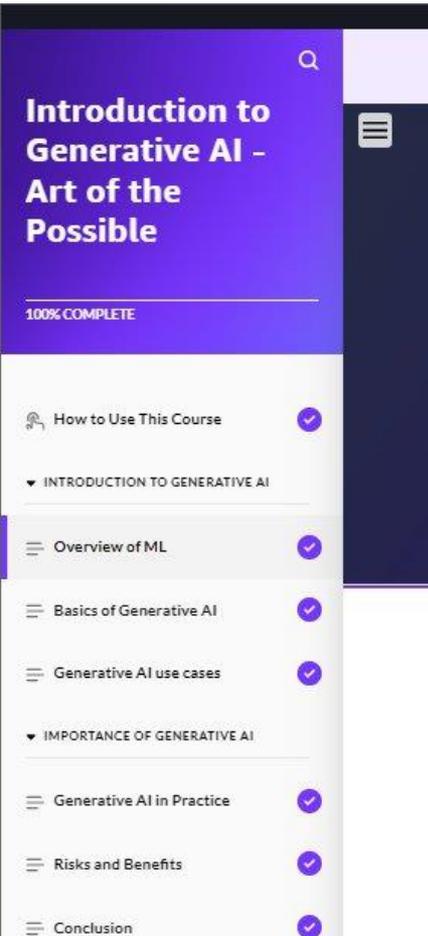
Mitarbeiter können sich auf andere Aufgaben konzentrieren.



Durch die vereinfachte Nutzung kann die Akzeptanz für den Test gesteigert werden.

Der Weg zu unserer 1. nützlichen KI-Anwendung





Foundation

- Generative AI Learning Plan for Developers

Advanced

- AWS PartnerCast - Breaking the Limits with AWS Generative AI – Technical
- AWS PartnerCast - Building Generative AI on AWS: Key Service Features and Demos
- AWS PartnerCast - Building Generative AI Solutions with Best Practices and Design Choices

Specialization

- Amazon CodeWhisperer - Getting Started
- AWS PartnerCast - GenAI Series: Building Generative AI Applications with Hugging Face on AWS
- AWS PartnerCast - Building Generative AI Applications with AI21 on AWS
- Amazon Kendra Getting Started
- AWS PartnerCast - Amazon OpenSearch as Vector Datastore for your Generative AI Applications with Zero ETL from DynamoDB – Technical

The AI Product Canvas



<p>Product</p> <p>actulib Tarifgenerator</p>	<p>Customer</p> <p>Aktuare und Aktuarinnen an der Schnittstelle zwischen Fachbereich und IT, die ausgewählte Tarife mit dem BELTIOS-Testrechner actulib auf versicherungsmathematische Werte testen wollen.</p>	<p>Value proposition</p> <p>Zum Test von versicherungsmathematischen Werten mit actulib müssen ausgewählte Daten aus Textdokumenten extrahiert werden und in ein spezielles json-Format gebracht werden. Durch Einsatz von KI sinkt die Implementierungszeit deutlich.</p>	<p>Channels</p> <p>Dieser Tarifgenerator ist ein weiterer Service für Kunden von actulib und würde zusammen angeboten werden. Dadurch kann es über einen gemeinsamen Kanal vertrieben werden.</p>	<p>Action</p> <p>Für Anwender der actulib würde dieser Service eine Erleichterung in der Verwendung der actulib bieten. Dies hilft sowohl Bestandskunden, könnte aber auch Neukunden einfacher vom Produkt überzeugen.</p>	<p>Next steps (internal)</p> <p>Es muss zunächst das KI-Modell auf Basis der Trainings-Geschäftspläne implementiert werden. Anschließend muss das Modell nach Außen zugänglich gemacht werden.</p>
<p>Analytical approach</p> <p>Grundsätzlich Generative AI, ggf. in Verbindung mit traditionellen Machine Learning-Algorithmen</p>	<p>Jobs and objectives</p> <p>Bspw. im Rahmen von Bestandsmigrationen bei Lebensversicherungen müssen große Mengen an Tarifen technisch umgesetzt werden. Diese Umsetzung basiert auf Geschäftsplänen, die zeitaufwändig analysiert werden müssen.</p>	<p>Gain creators</p> <p>Bislang müssen diese technischen Konfigurationsdateien von BELTIOS-Aktuaren vorgenommen werden. Durch KI-Unterstützung können die Kunden selbst Einstellungen vorgeben.</p>	<p>Interfaces</p> <p>Der Tarifgenerator kann gemeinsam mit dem aktuariellen Testrechner entweder über eine Kommandozeile oder auch über eine Web-GUI angeboten werden.</p>	<p>Downstream commitments</p> <p>Sollte in der Zukunft die Darstellung der Tarifkonfiguration verändert werden, so muss auch die KI dahingehend angepasst werden.</p>	<p>Key Stakeholders</p> <p>Das Testrechner-Team von BELTIOS</p>
<p>AI Risk assessment</p> <p>Es muss sichergestellt werden, dass die von Kunden eingesetzten Geschäftspläne gesichert und vertraulich bleiben.</p>	<p>Challenges and pain-points</p> <p>Die Übertragung von Geschäftsplänen in technische Vorgabedaten ist sehr zeitaufwändig, insbesondere da man sowohl einen "fachlichen" als auch einen "technischen" Aktuar benötigt.</p>	<p>Pain relievers</p> <p>Langwierige Arbeiten und Abstimmungen würden entfallen oder zumindest stark reduziert, insbesondere da diese Tätigkeiten äußerst monoton sind.</p>	<p>Datasets</p> <p>Es liegen anonymisierte Geschäftspläne zum Trainieren der Modelle vor.</p>	<p>Performance</p> <ul style="list-style-type: none"> • Erstens muss das Format der Generierung technisch gültig sein, und • zweitens sollte der Output fachlich korrekt und vollständig sein. (Hier kann gegen bestehende Vertragsumsetzungen getestet werden.) 	<p>Costs of non-delivery</p> <p>Zeitintensive und repetitive Arbeiten müssten weiterhin von Fachexperten ausgeführt werden, die bei anderen Aufgaben einen größeren Mehrwert erzeugen könnten</p>
<p>Cost Structure</p> <p>Lineare Abhängigkeit zur Anzahl der zu erzeugenden Tarife</p>			<p>Impact</p> <p>Das Produkt hat disruptiven Charakter, da es bislang kein vergleichbares Konkurrenz-Produkt am Markt gibt, der Bedarf nach solchen Lösungen jedoch stark nachgefragt wird.</p>		

Ablauf der Verarbeitung

1

Hochladen der Tarifbeschreibungen in PDF-, JPG-, PNG- oder Word-Format

2

Umwandlung der Datei in Textform

3

Extraktion wesentlicher Informationen mittels Generativer KI

4

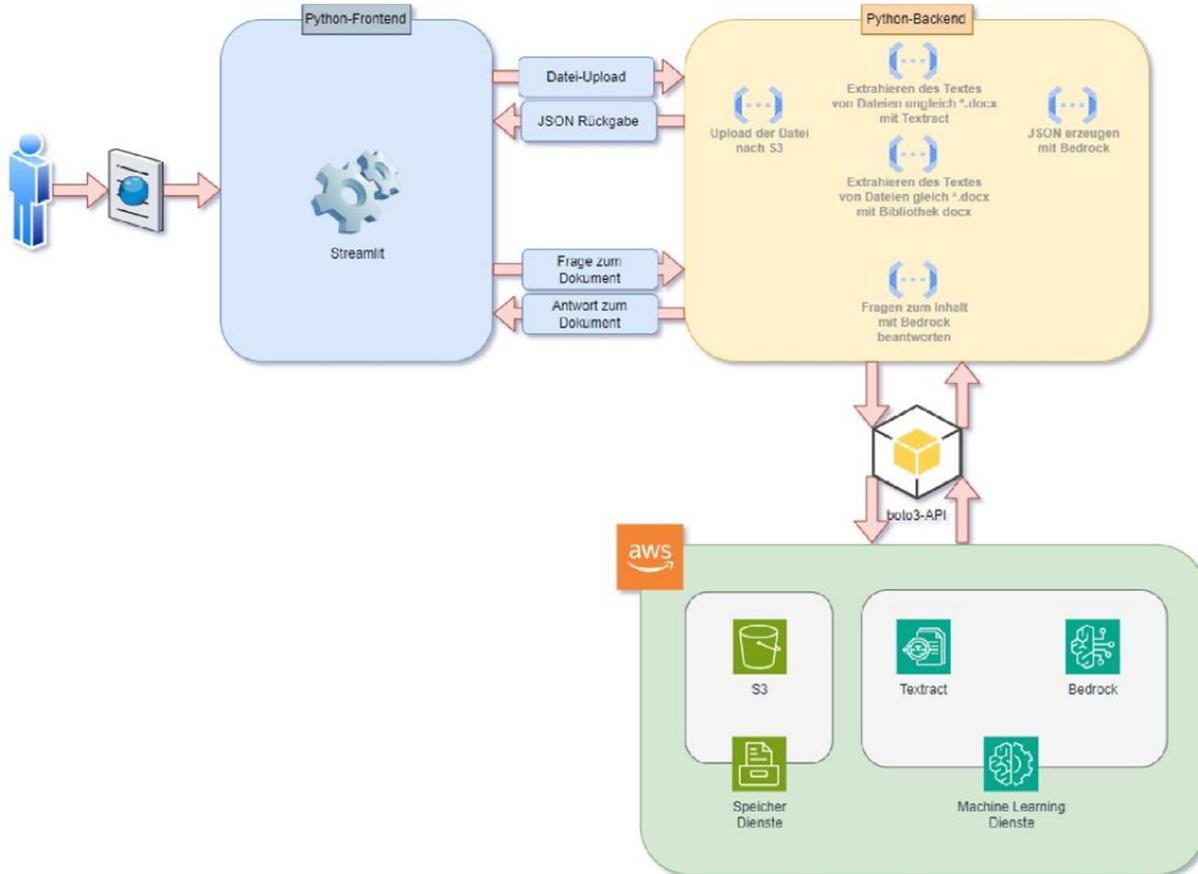
JSON-Datei als Output

5

Interaktion mit Chatbot bei Rückfragen

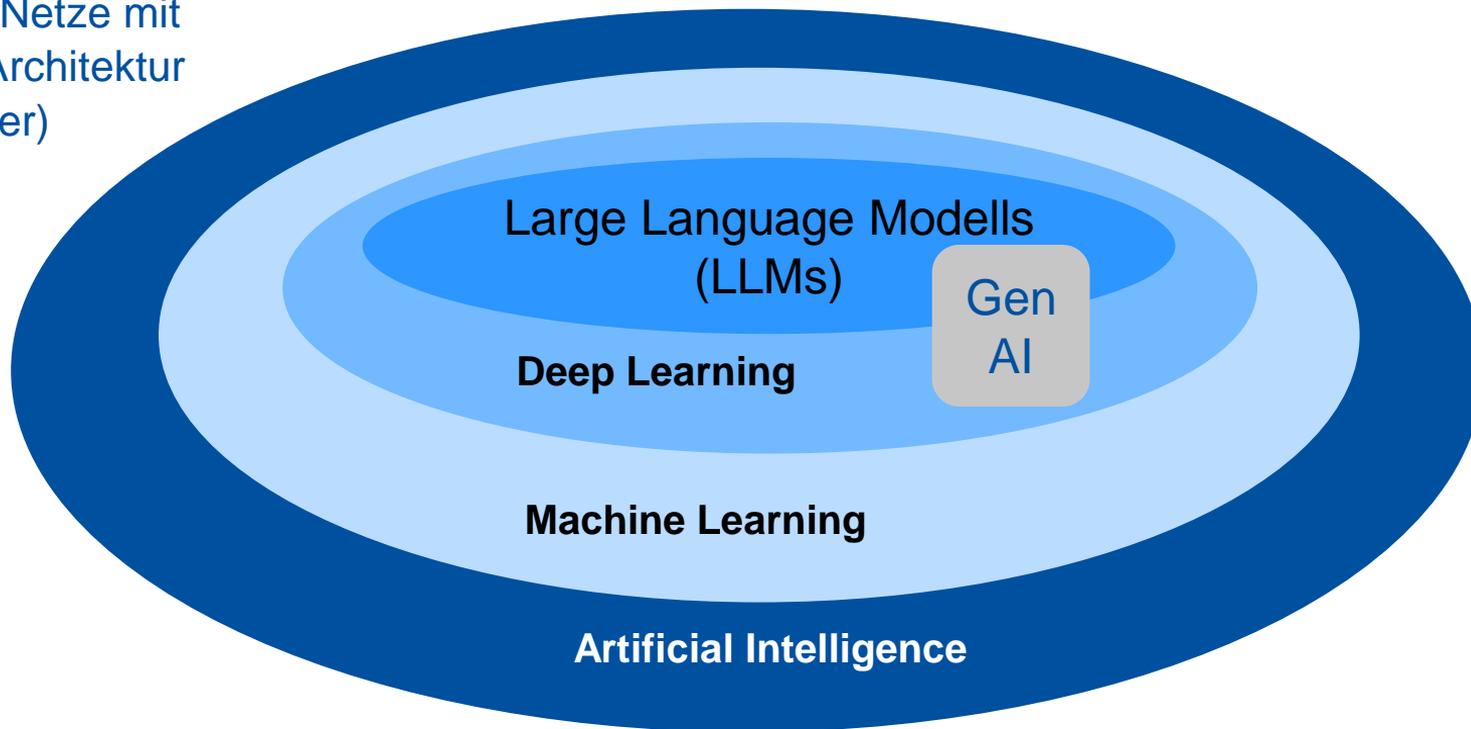


Architektur



Wie sind Large Language Models (LLMs) einzuordnen?

LLMs = riesige vortrainierte
Neuronale Netze mit
spezieller Architektur
(Transformer)



Verfügbare KI-Modelle am Beispiel AWS Bedrock

→ Service für die Nutzung von Basismodellen (Foundation Models) verschiedener Anbieter

Basismodelle

Amazon Bedrock unterstützt Basismodelle von branchenführenden Anbietern. Wählen Sie das Modell, das am besten zum Erreichen Ihrer individuellen Ziele geeignet ist.

**AI21
labs**

Jamba 1.5
Nach AI21 Labs



Titan
Nach Amazon

AI

Claude
Nach Anthropic



Command
Nach Cohere



Llama
Nach Meta



Mistral
Nach Mistral AI



Stable
Diffusion
Nach Stability AI

Beispiele für unterschiedliche API

Anthropic Sonnet 3.5

```
{
  "modelId": "anthropic.claude-3-5-sonnet-20240620-v1:0",
  "contentType": "application/json",
  "accept": "application/json",
  "body": {
    "anthropic_version": "bedrock-2023-05-31",
    "max_tokens": 1000,
    "messages": [
      {
        "role": "user",
        "content": [
          {
            "type": "image",
            "source": {
              "type": "base64",
              "media_type": "image/jpeg",
              "data": "iVBORw..."
            }
          },
          {
            "type": "text",
            "text": "What's in this image?"
          }
        ]
      }
    ]
  }
}
```

Mistral 7B Instruct

```
{
  "modelId": "mistral.mistral-7b-instruct-v0:2",
  "contentType": "application/json",
  "accept": "application/json",
  "body": "{\"prompt\": \"<s>[INST] this is where you place your input text [/INST]\", \"max_tokens\": 200, \"temperature\": 0.5, \"top_p\": 0.9, \"top_k\": 50}"
}
```

Cohere R+

```
{
  "modelId": "cohere.command-r-plus-v1:0",
  "contentType": "application/json",
  "accept": "*/*",
  "body": "{\"chat_history\": [{\"role\": \"USER\", \"message\": \"What is an interesting new role in AI if I don't have an ML background\"}, {\"role\": \"CHATBOT\", \"message\": \"You could explore being a prompt engineer!\"}], \"message\": \"What are some skills I should have\"}"
}
```

Arbeit mit Basismodellen (Foundation Models)

- Hier: In-Context-Learning mit One-Shot Inference
- Einteilung von Input und Output in Tokens
- Daumenregel: 1,3 Tokens je Wort *



Tipps für gute Prompts



Klar das Thema
beschreiben



Ein bestimmtes
Ausgabeformat
vorgeben



Größe der Antwort
festlegen



Explizite Anweisungen
geben



Kreativ sein



Das Modell nach
„Denke Schritt-für-
Schritt“ zur Lösung
kommen lassen



Beschränkungen für
mehr Kontrolle
hinzufügen



XML/HTML-Tags zur
Abgrenzung verwenden



Arbeitsaufträge ans
Ende bei Prompts mit
viel Text stellen



Einfach und knapp sein



Kontext angeben
und Few-Shot-Beispiel
Prompts mitliefern



Angaben,
was das Model tun soll,
wenn es keine
zuverlässige Antwort
geben kann



Negative
Formulierungen
vermeiden



Verweigerung
verwenden, um
Anfragen zu vermeiden,
die das Modell nicht
beantworten soll

**... dann lass
mal schauen!**



**Vielen Dank
für Ihre
Aufmerksamkeit.**

Claudia Schmidt, BELTIOS
Georg van Bühren, BELTIOS

**Besuchen Sie
unsere Webseite und
unseren Testrechner**

www.beltios.com