

Einführung

1. Large Language Models
2. GenAI & RAG
3. Cloud & Container



Andreas Schütz
Sales Engineer



Shubham Sumalya
Sales Engineer



Felix Vetter
Sales Engineer Intern

Meilensteine auf dem Weg zu heutigen LLMs



1940er - Erste Language Models



2010 - Recurrent Neural Networks



2013 - Word2Vec (Word Embeddings)



2017 - Transformer-Architektur



2020 - Instruct-Modelle

Übersetze mir "Herzlich willkommen auf dem InterSystems DACH Symposium" ins Italienische.



"Benvenuti al Simposio InterSystems DACH."



illia.polosukhin@gmail.com

Prompt Engineering



Translate "Bilderrahmen"
into English. Only output
the actual translation.

Prompt

If there is no English version of the
word, ignore it. Translating an
expression ...

Meta Llama3 (8B)

Prompt Engineering



Translate into English.

Input: Cheval

English: Horse

Input: Stuhl

English: Chair

Input: Bilderrahmen

English:

Prompt

Picture frame

Input: Telefon

English: Telephone

Input...

Meta Llama3 (8B)

Chatbot Prompt



You are Bob, a very powerful large language model based chatbot. Answer as helpful and concisely as possible.

User: Suggest me a dinner.

Assistant: Grilled Lemon Herb Chicken with Quinoa Salad and Steamed Broccoli.

User: *<Previous conversation>*

Assistant: *<Previous conversation>*

...

User:

Assistant.

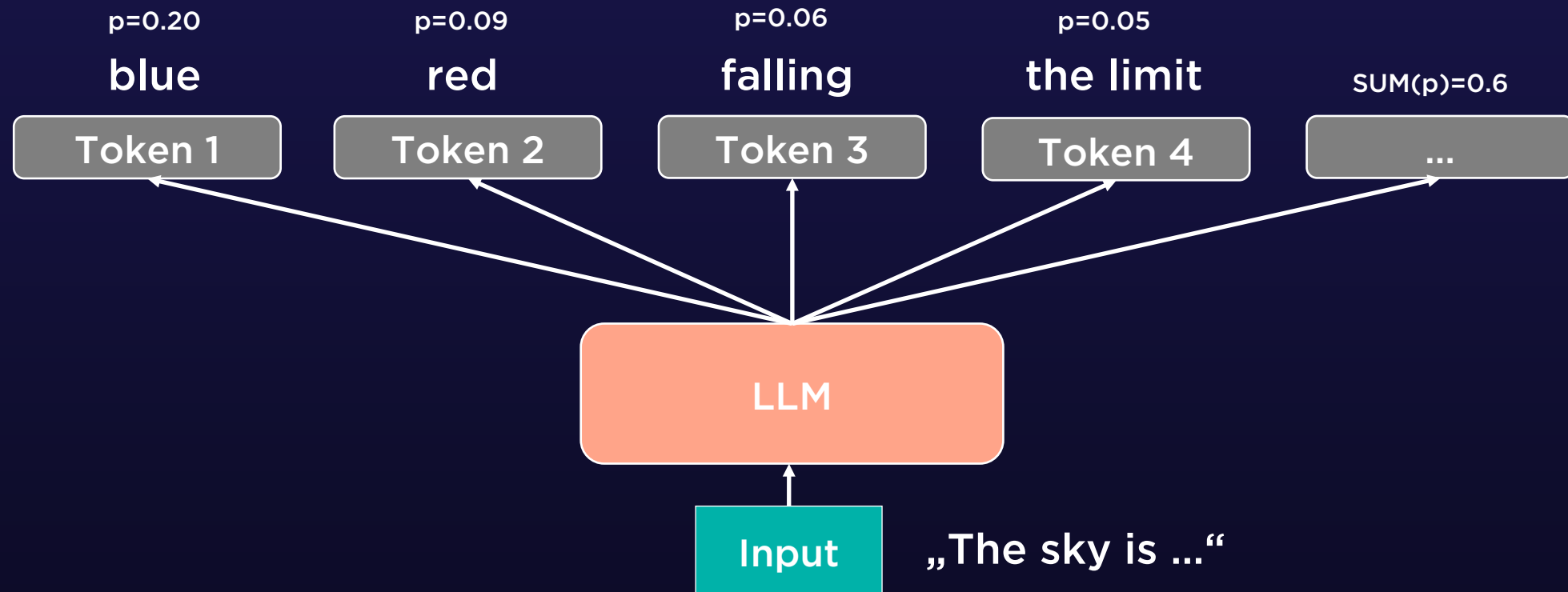
System Message

Example

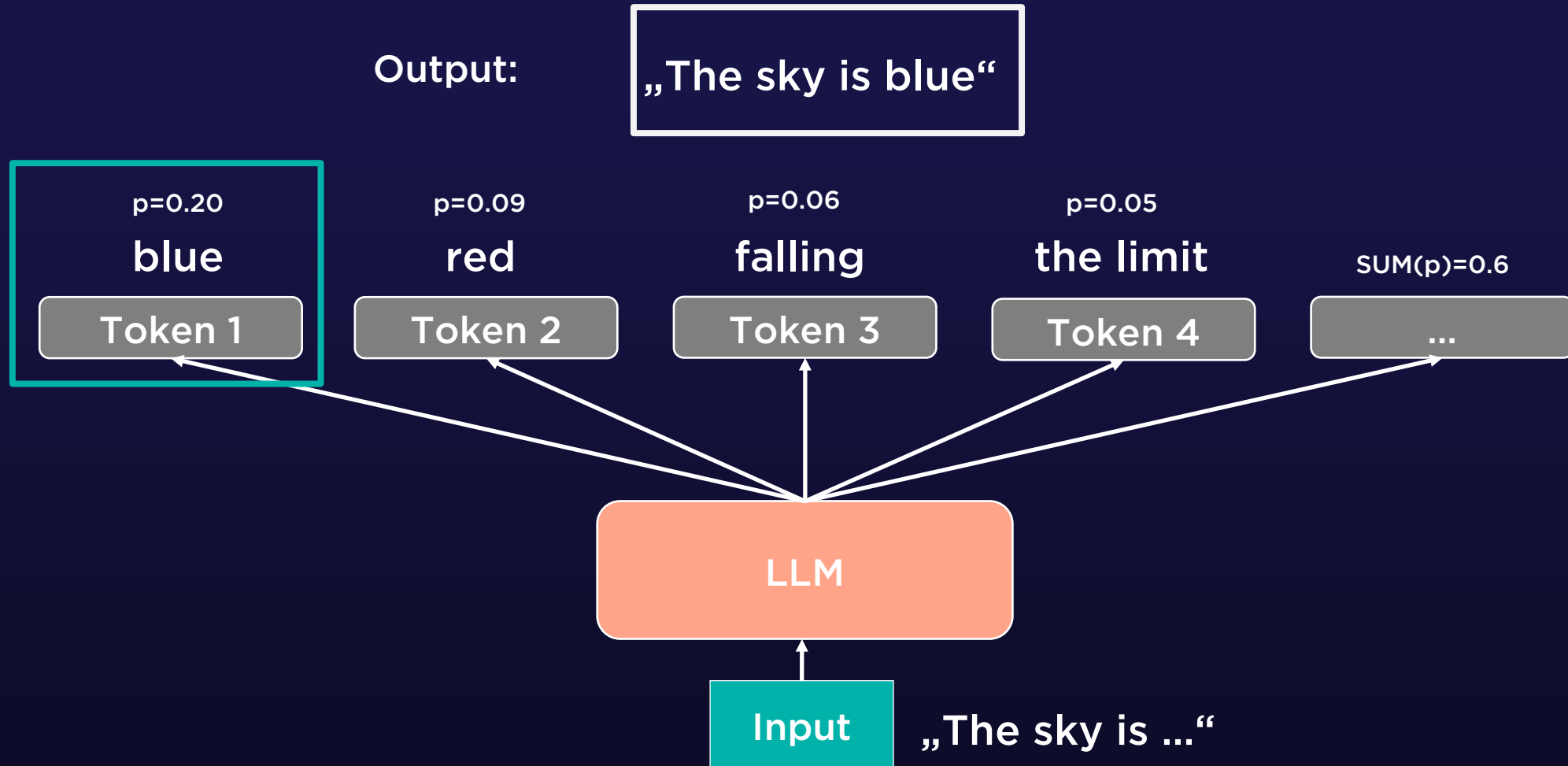
Conversation History

Input

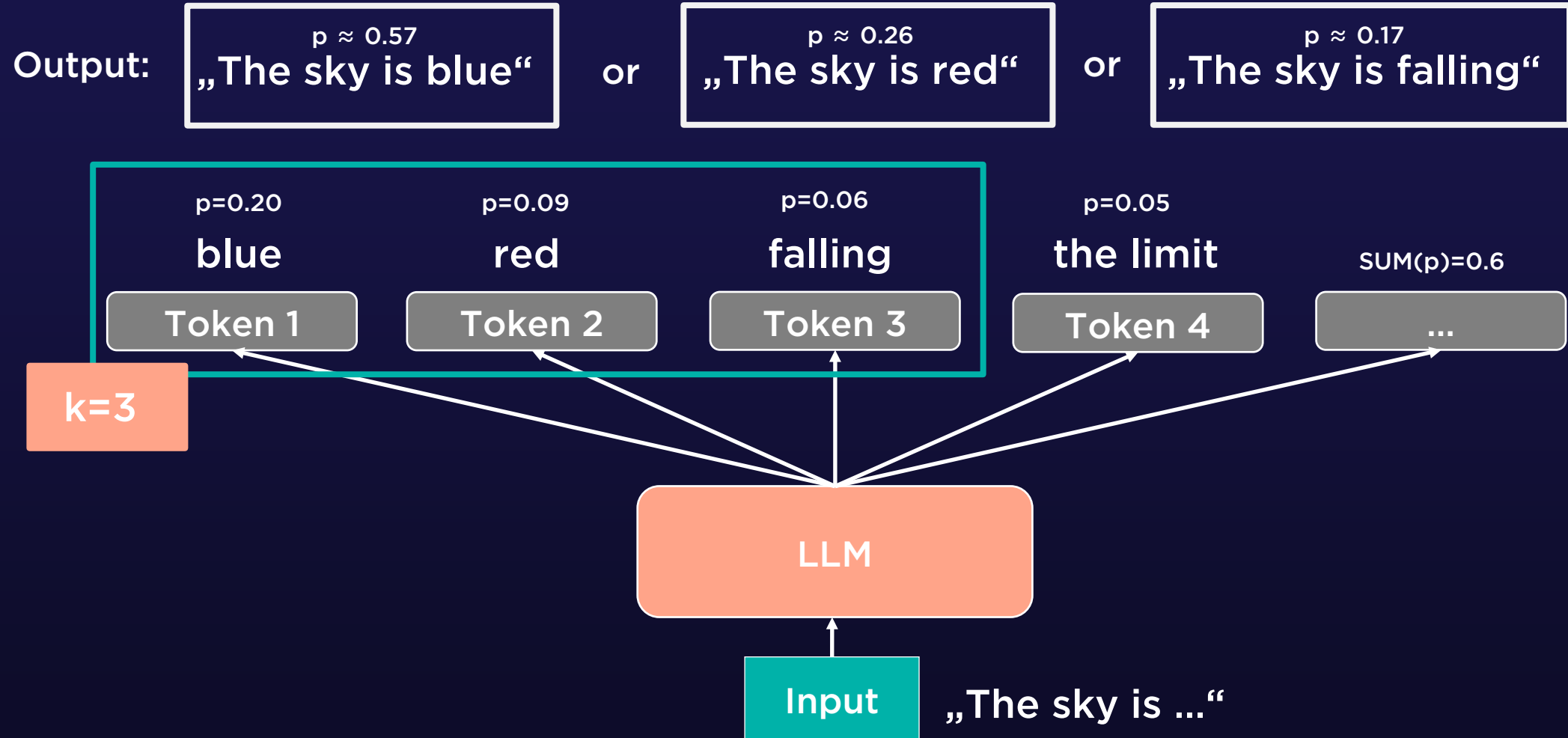
Sampling Strategy



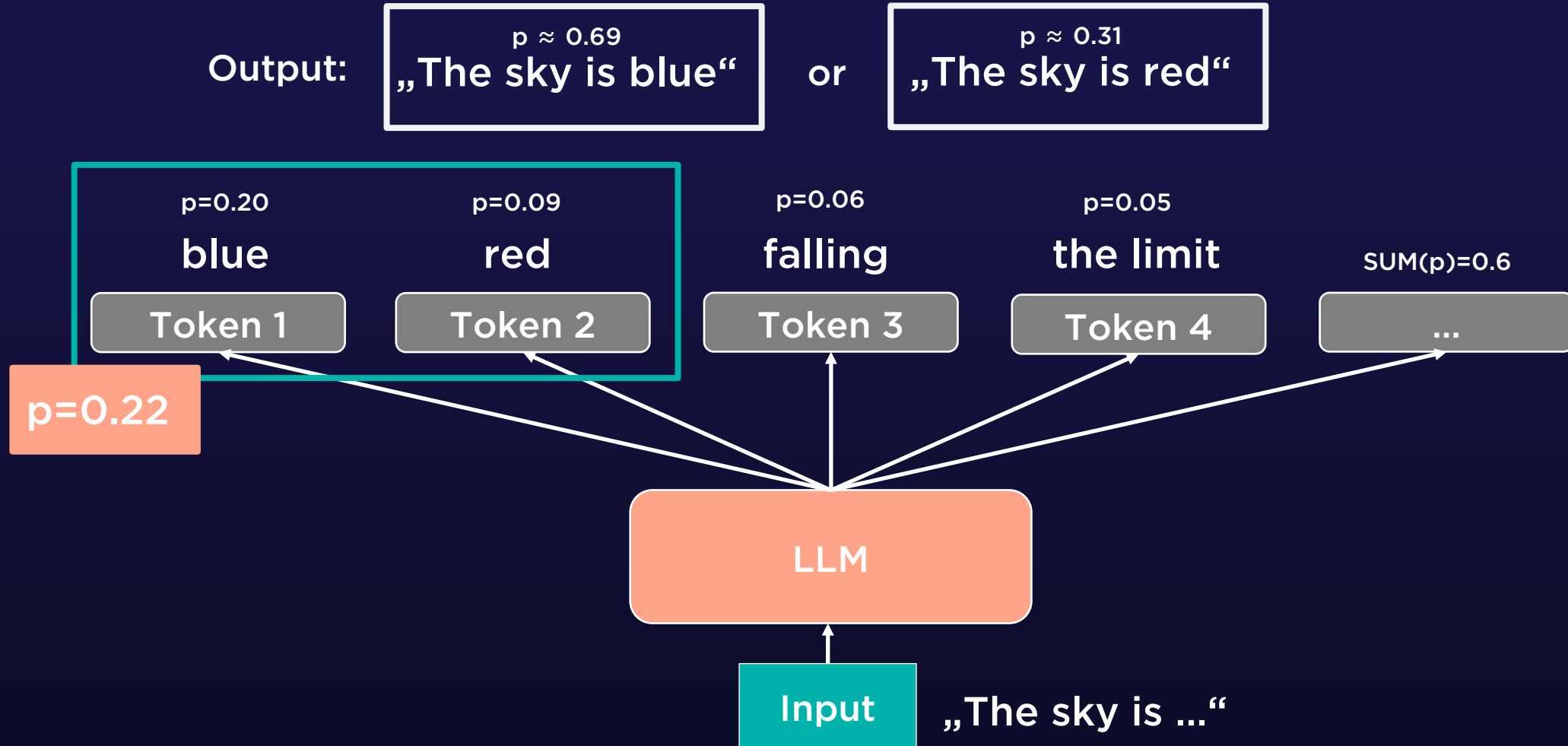
Sampling Strategy - No Sampling



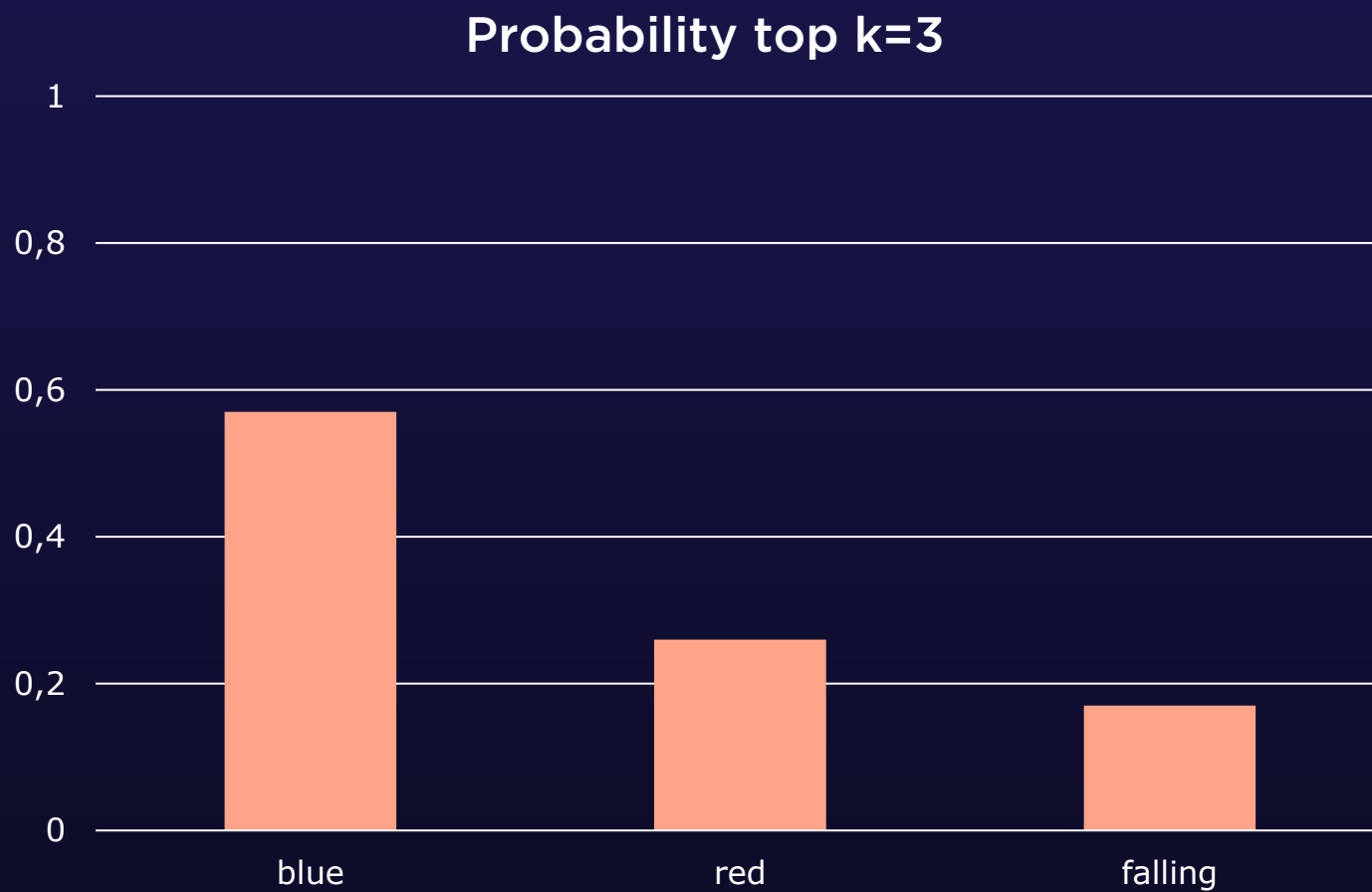
Sampling Strategy - Top k Sampling



Sampling Strategy - Top p Sampling



Sampling Strategy - Temperature



Temperature

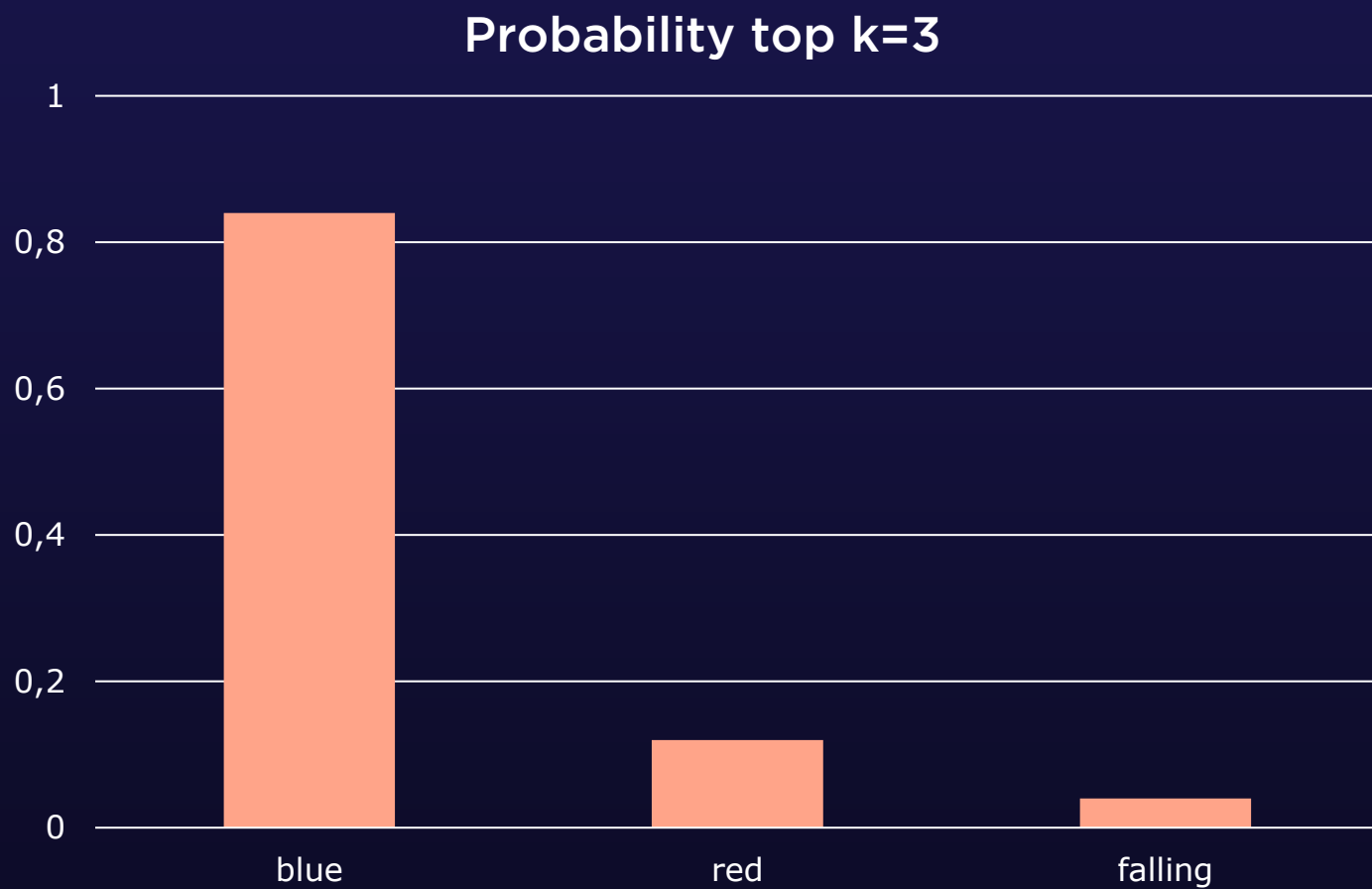
0.1

0.4

1

2

Sampling Strategy - Temperature



Temperature

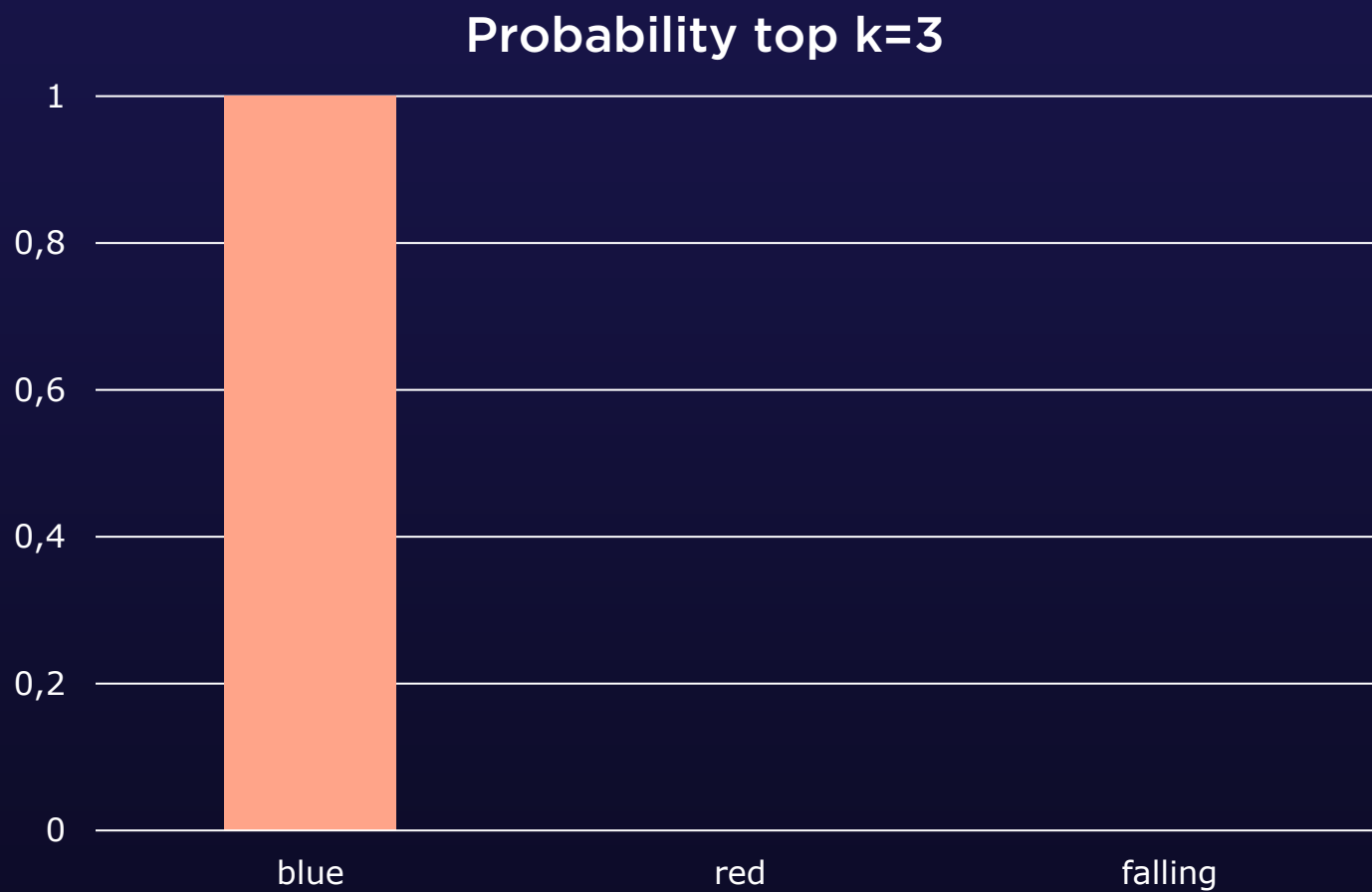
0.1

0.4

1

2

Sampling Strategy - Temperature



Temperature

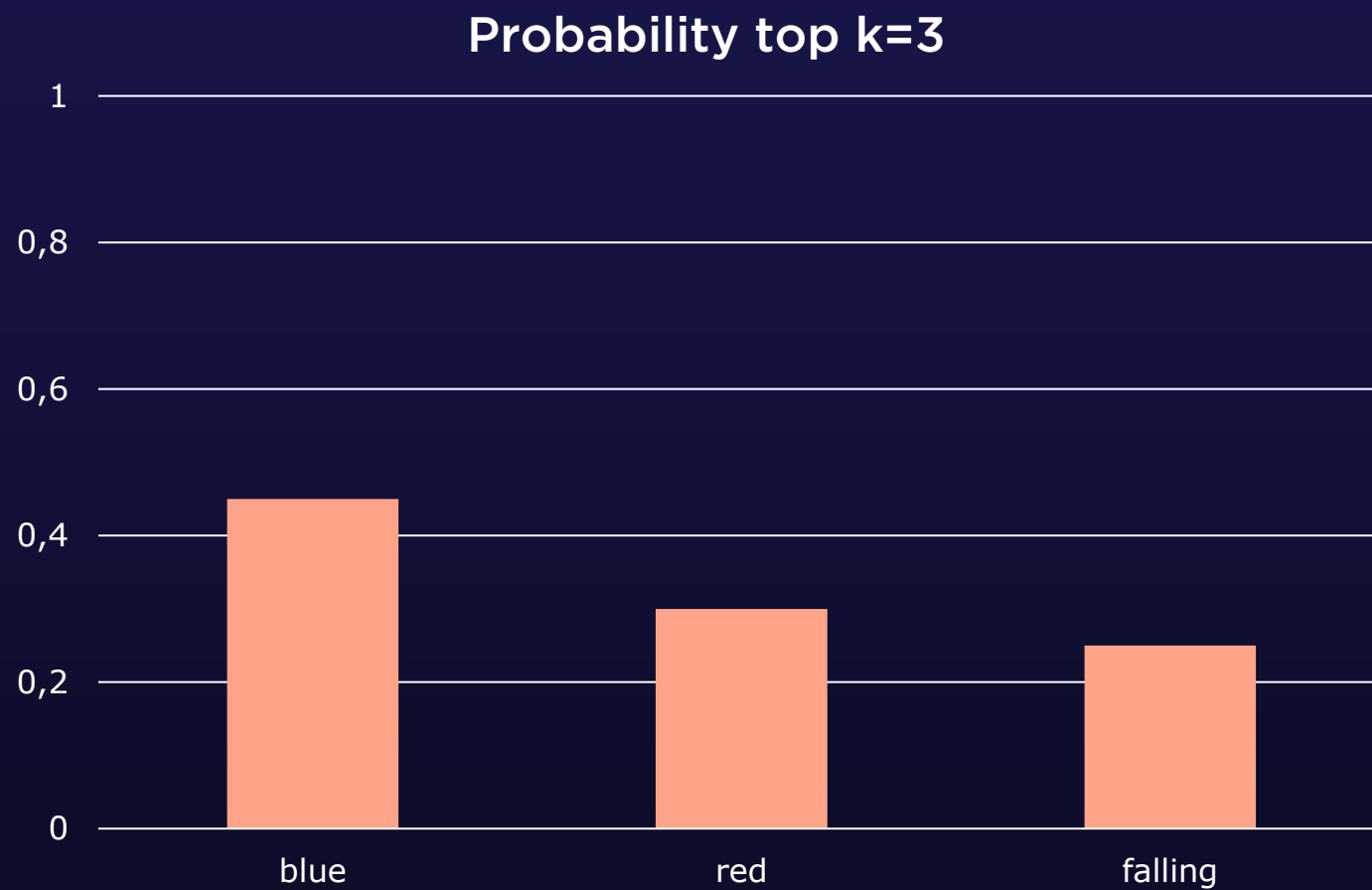
0.1

0.4

1

2

Sampling Strategy - Temperature



Temperature

0.1

0.4

1

2

Endlich schlaue Chatbots! Oder?

Kurze Demo mit Einführung zu GenAI



Felix Vetter
Sales Engineer Intern



Artificial Intelligence

Machine Learning

Image Recognition

Survival Analysis

...

Tabular ML

Classification

Regression

TimeSeries

...

Deep Learning

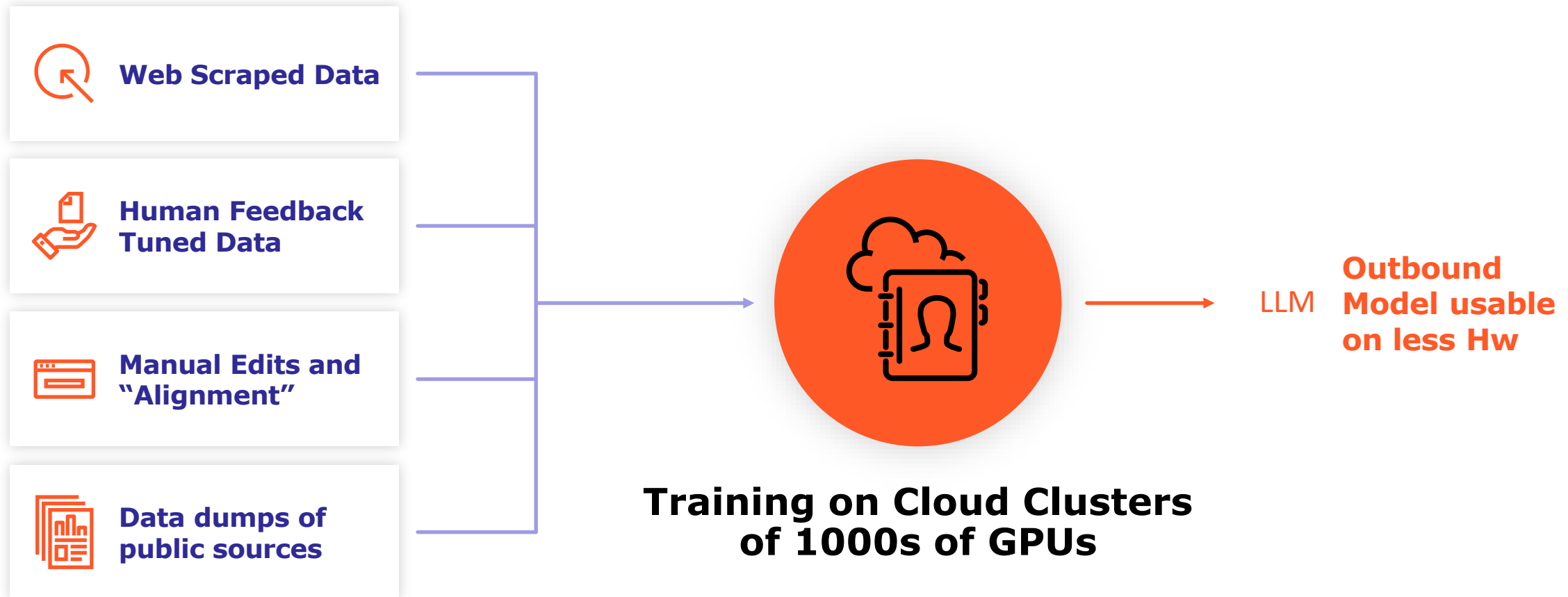
Generative AI

Large Language Models

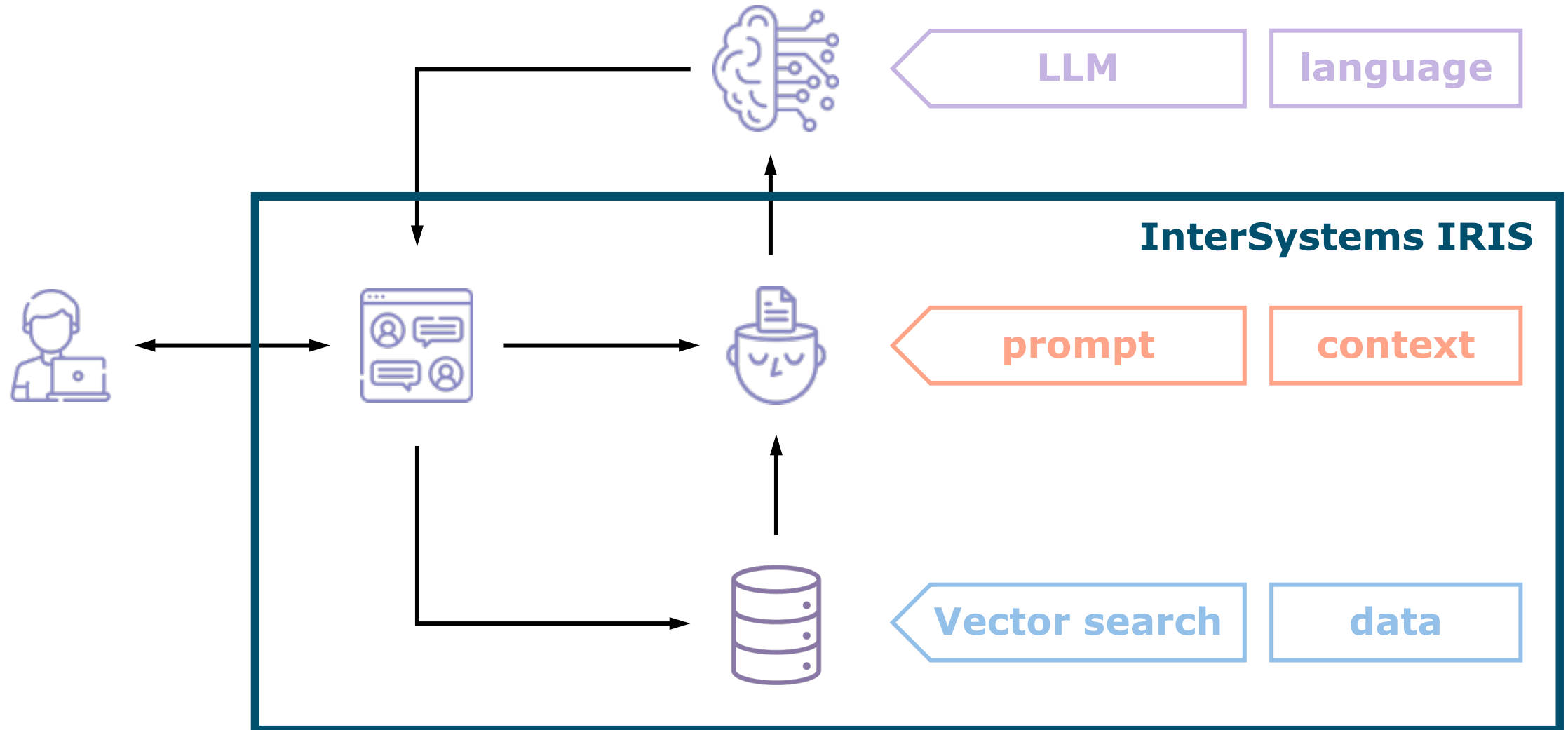
Vector Search

...

LLMs werden mit einer VIELFALT von Quellen trainiert



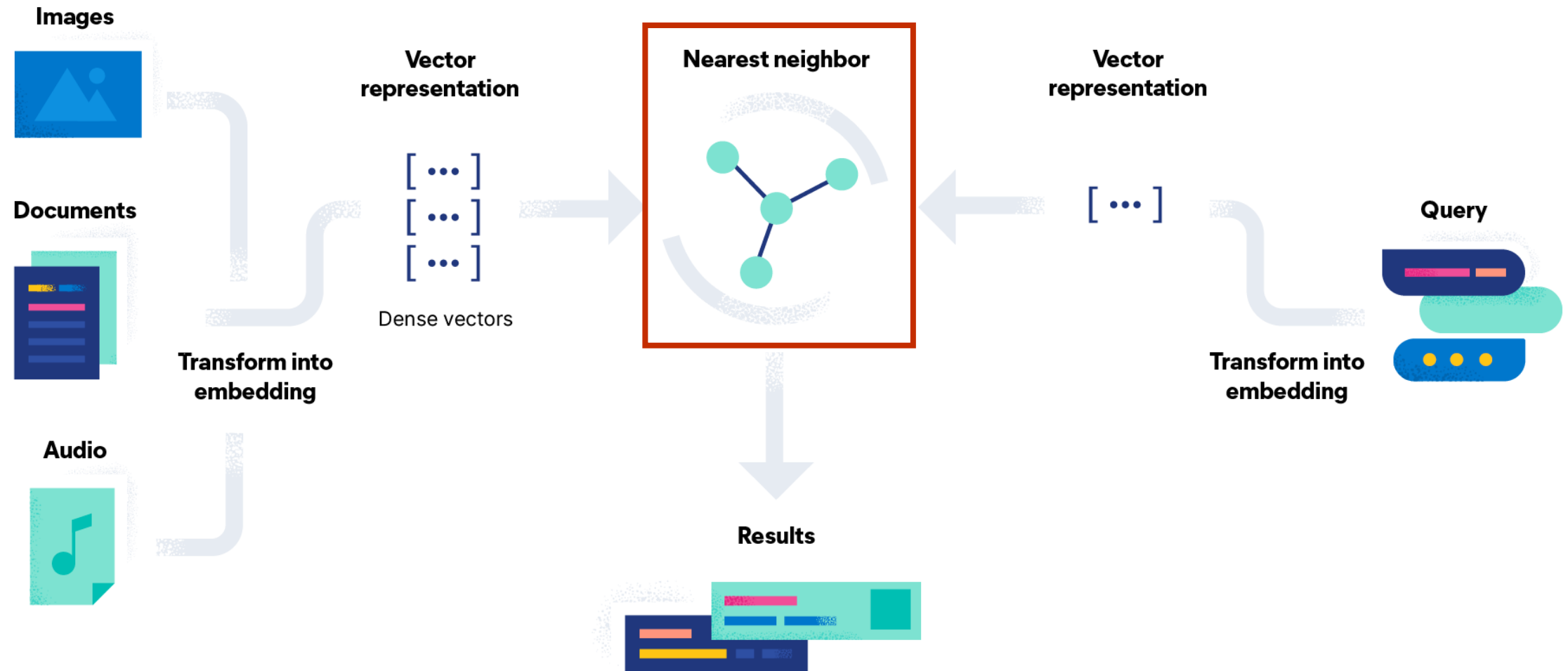
Retrieval Augmented Generation



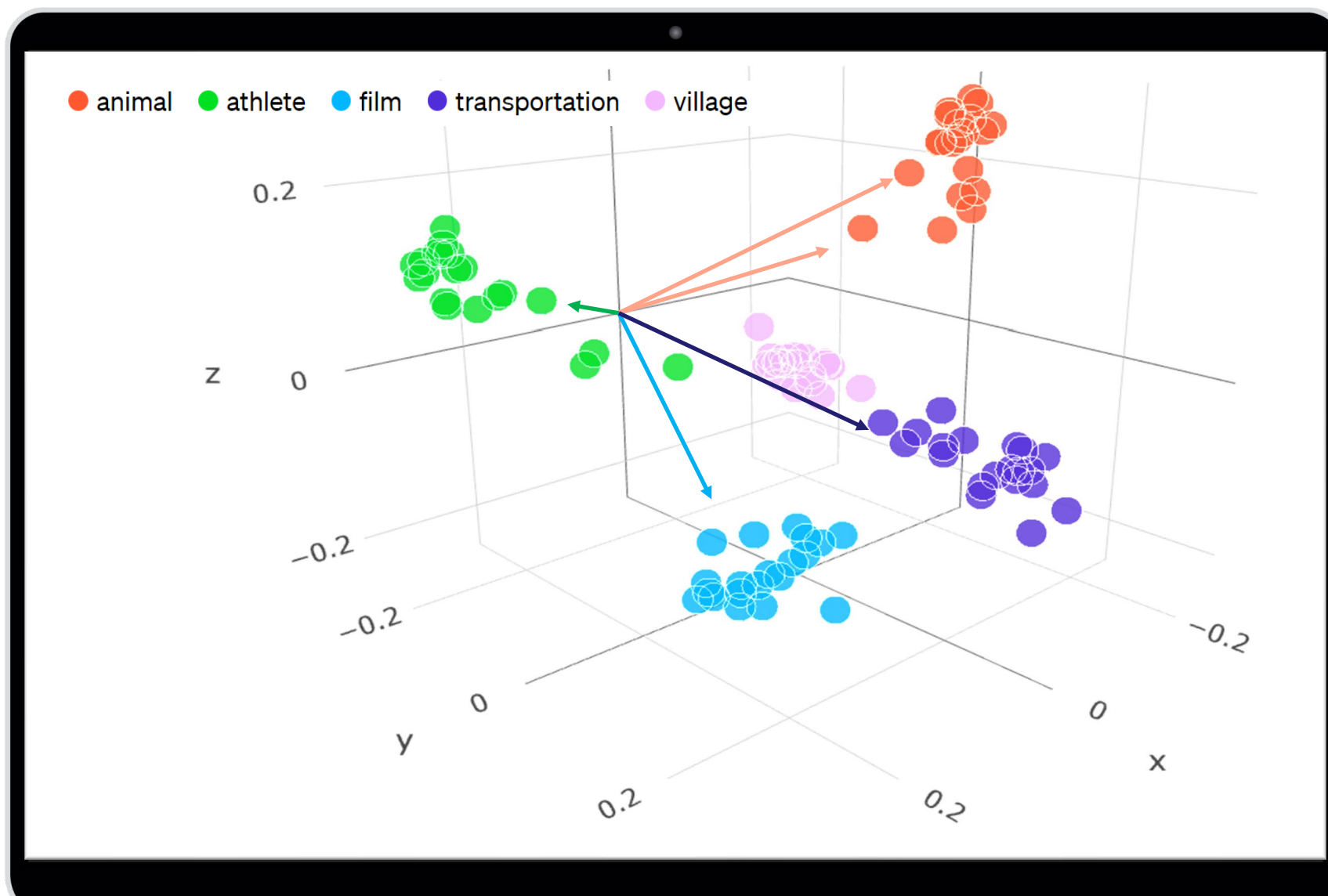
Vector Search

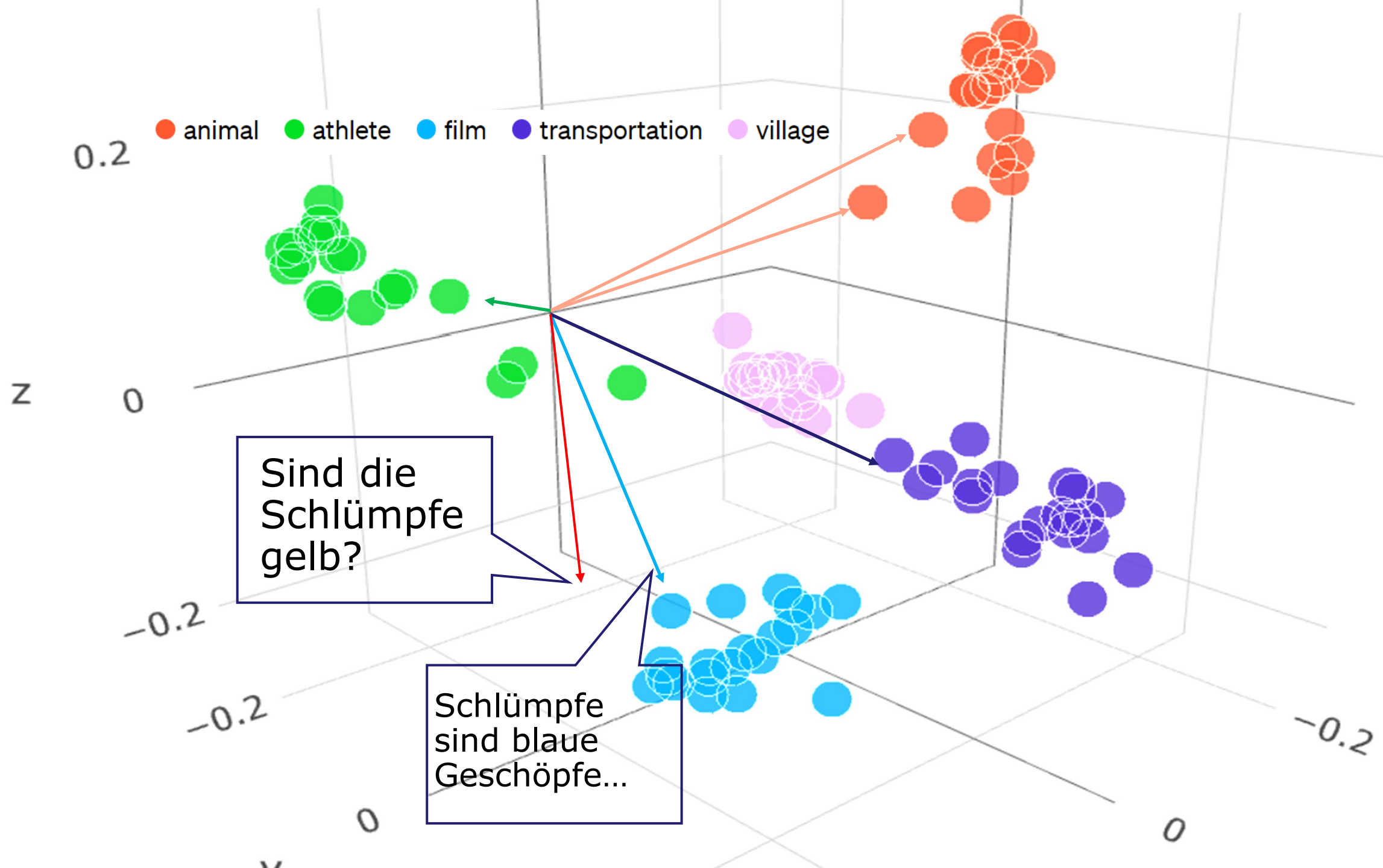


- **Vector search** stellt Datenelemente durch numerische Vektoren dar, was die Indizierung und Ausführung von Abfragen über numerische Inhalte ermöglicht



Sätze & deren Kontext als Vektoren





InterSystems IRIS in the Cloud

Shubham Sumalya, Sales
Engineer



Why Cloud?



**Enables companies to
focus on core business**

Cost
effectiveness

Scalability

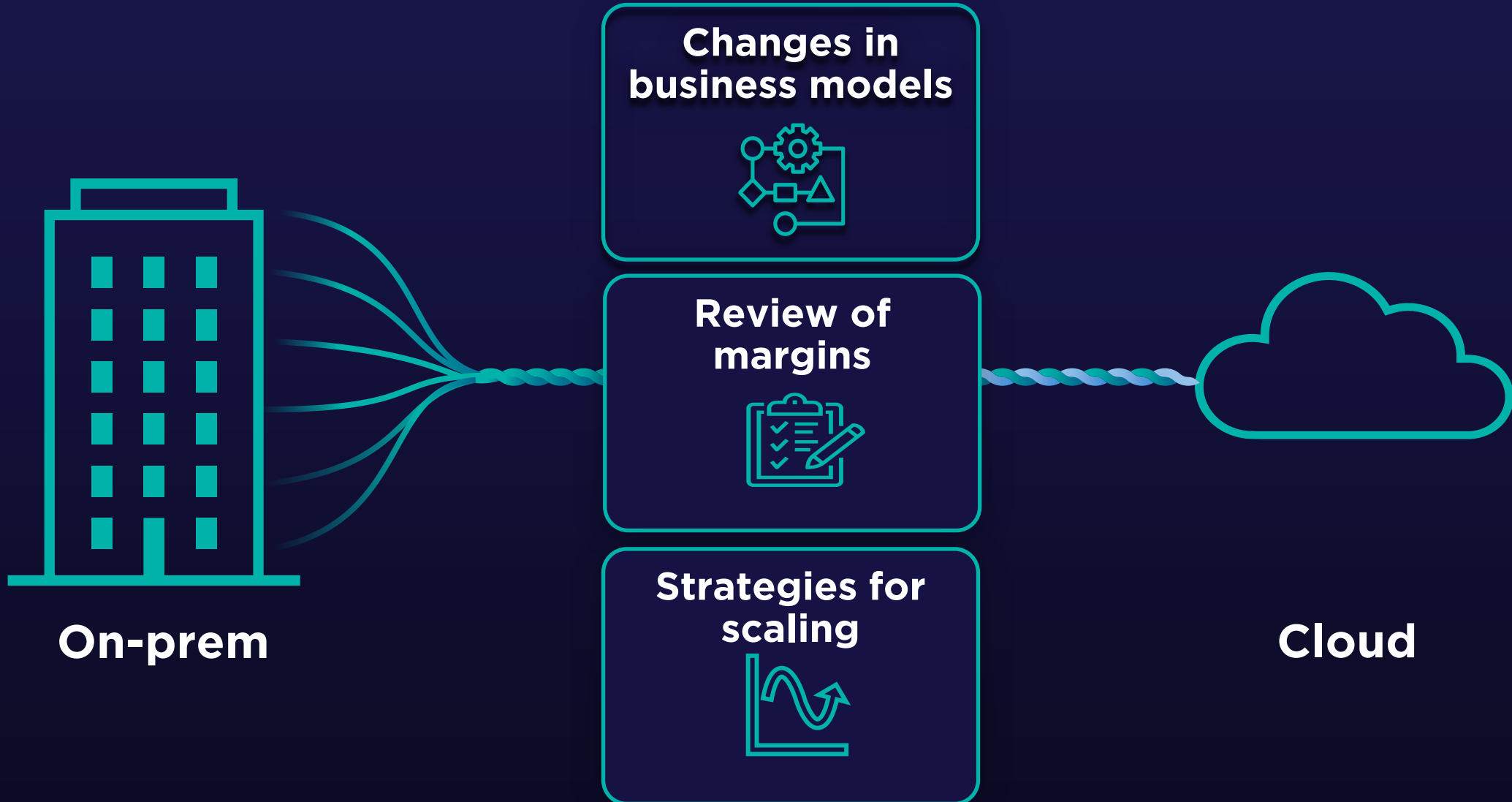
Flexibility and
agility

Built-in
redundancy

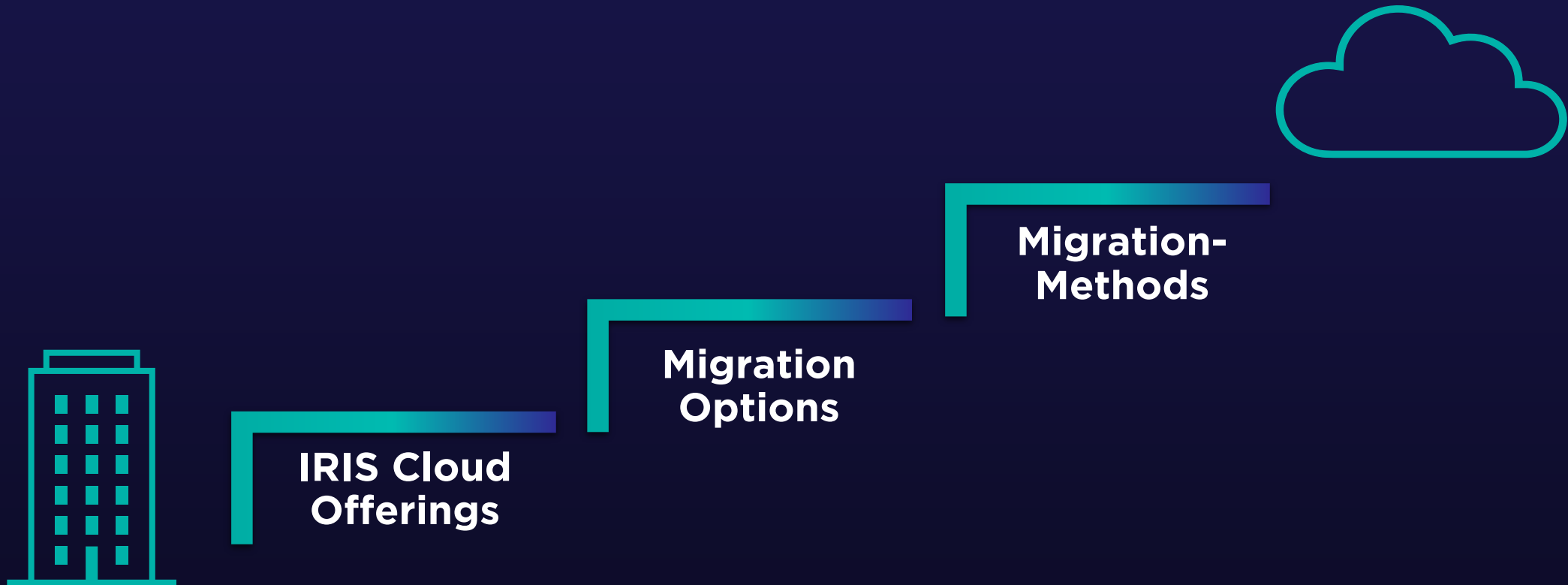
Security &
Compliance

Energy efficiency

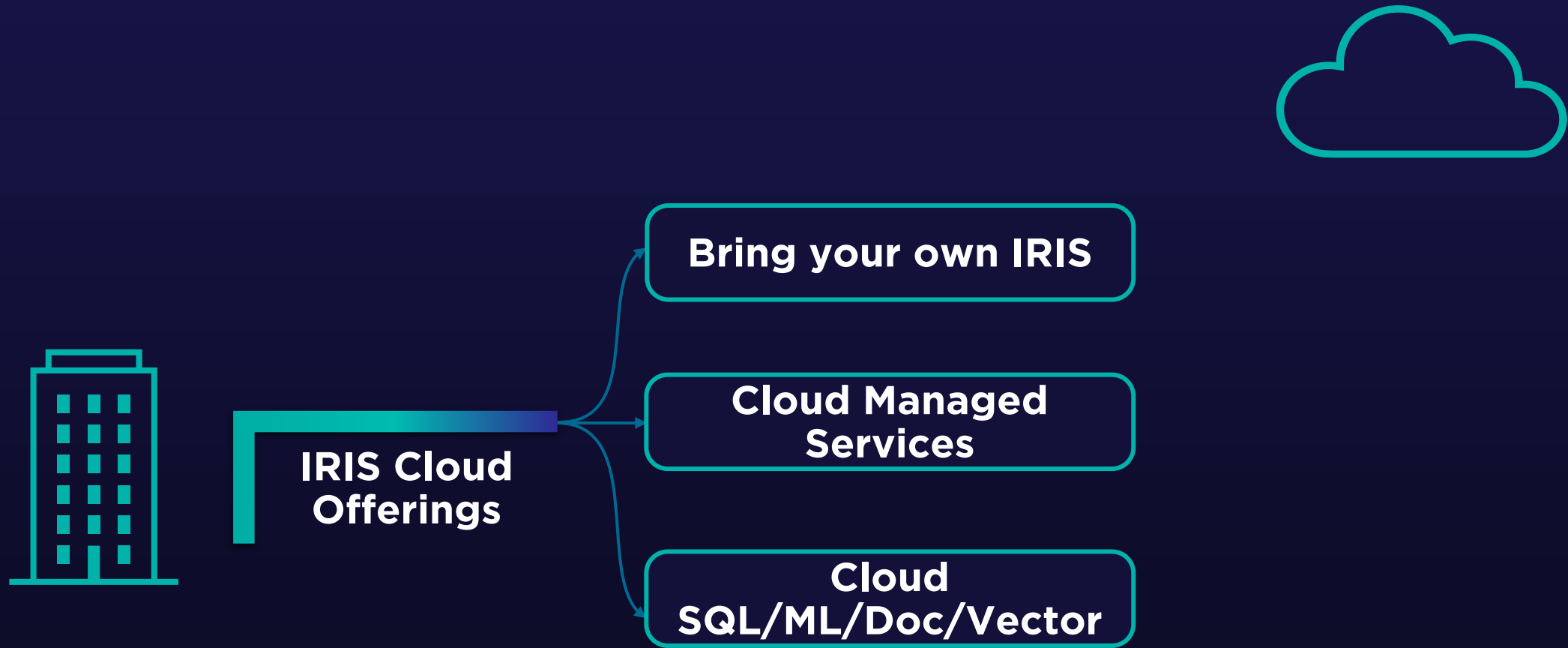
Challenges of migrating existing solutions



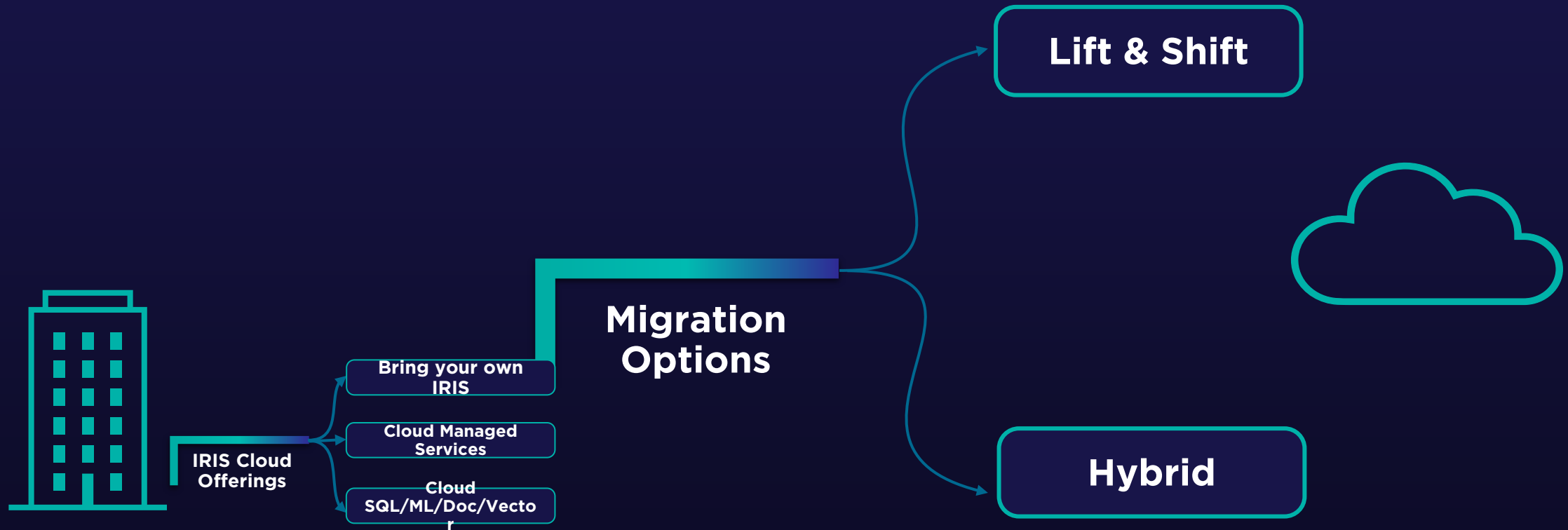
Cloud Options for IRIS and IRIS for Health



Cloud Offerings for IRIS and IRIS for Health



Migration options for IRIS and IRIS for Health

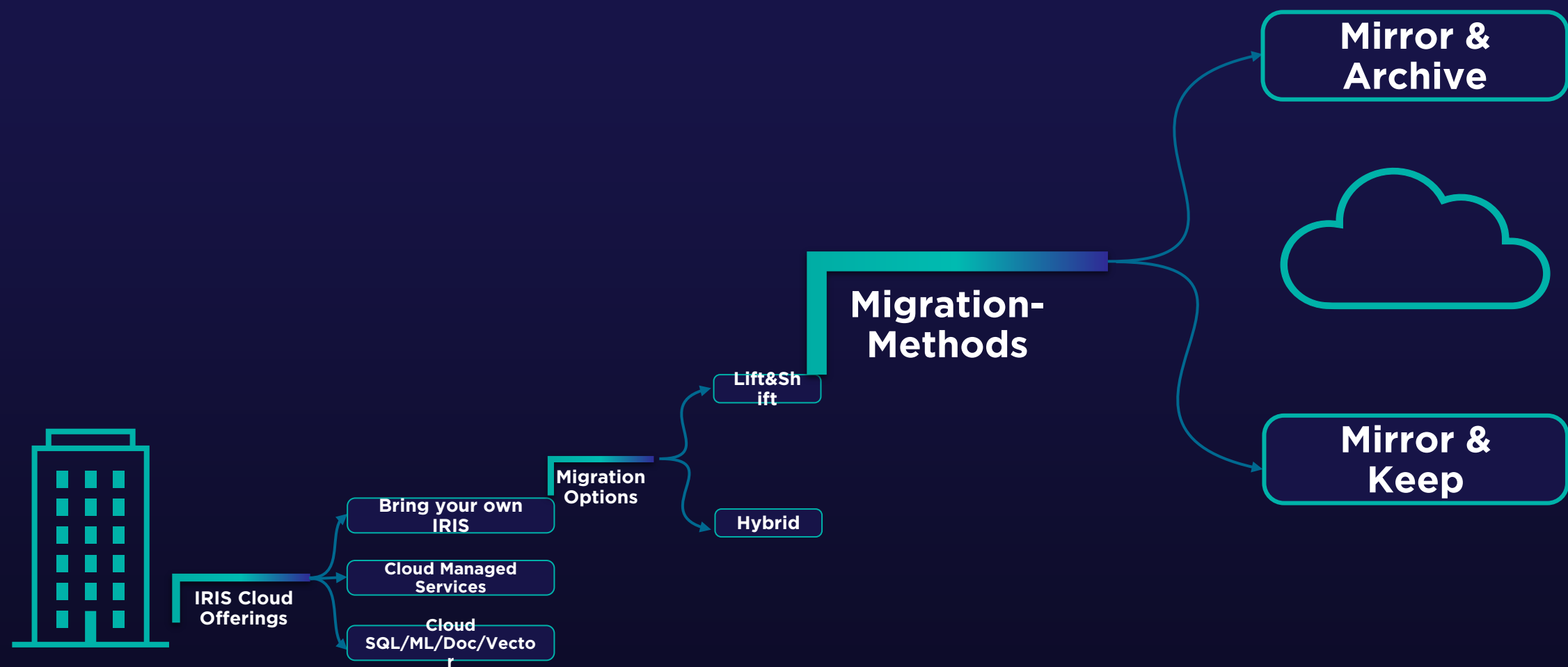


Migration options for IRIS and IRIS for Health



Migrations-Option	Number of IRIS Deployments	Properties
Lift & Shift: <ul style="list-style-type: none">• Complete migration to the cloud	<ul style="list-style-type: none">• 1	<ul style="list-style-type: none">• On-premises on-prem architecture is migrated to a cloud-based architecture
Hybrid: <ul style="list-style-type: none">• On-prem: Primary-Failover Mirror• Cloud: Async Backup-Failover Mirror	<ul style="list-style-type: none">• 2	<ul style="list-style-type: none">• On-prem mirror with an asynchronous read-only cloud mirror for BI / AI.

Migration Methods for IRIS and IRIS for Health



Cloud Optionen für IRIS und IRIS for Health



Cloud Offerings	Migration Options		Description
	Lift & Shift	Hybrid	
Bring your own license: <ul style="list-style-type: none">On-Prem and/or Cloud	✓	✓	<ul style="list-style-type: none">Same license<i>Server Migration Guide</i>
Cloud Managed Services: <ul style="list-style-type: none">Partially managed cloud servicesFee based	✓	✓	<ul style="list-style-type: none"><i>InterSystems IRIS Cloud Managed Services</i>
Cloud SQL/ML/Doc/Vector: <ul style="list-style-type: none">Fully managed cloud servicesFee based	✓	✗	<ul style="list-style-type: none">Does not support mirroring<i>InterSystems IRIS Cloud Services</i>



Vielen Dank