



**Weiß die KI,
dass sie nichts weiß?**

**Prof. Dr. Katharina A. Zweig
RPTU Kaiserslautern-Landau**

The image features a central camera lens with the text "ZOOM LENS" visible on its barrel. The lens is set against a dark background filled with a complex pattern of colorful, rounded geometric shapes in various colors including green, blue, orange, purple, and red. The text "Was ist KI?" is overlaid in white on the lens.

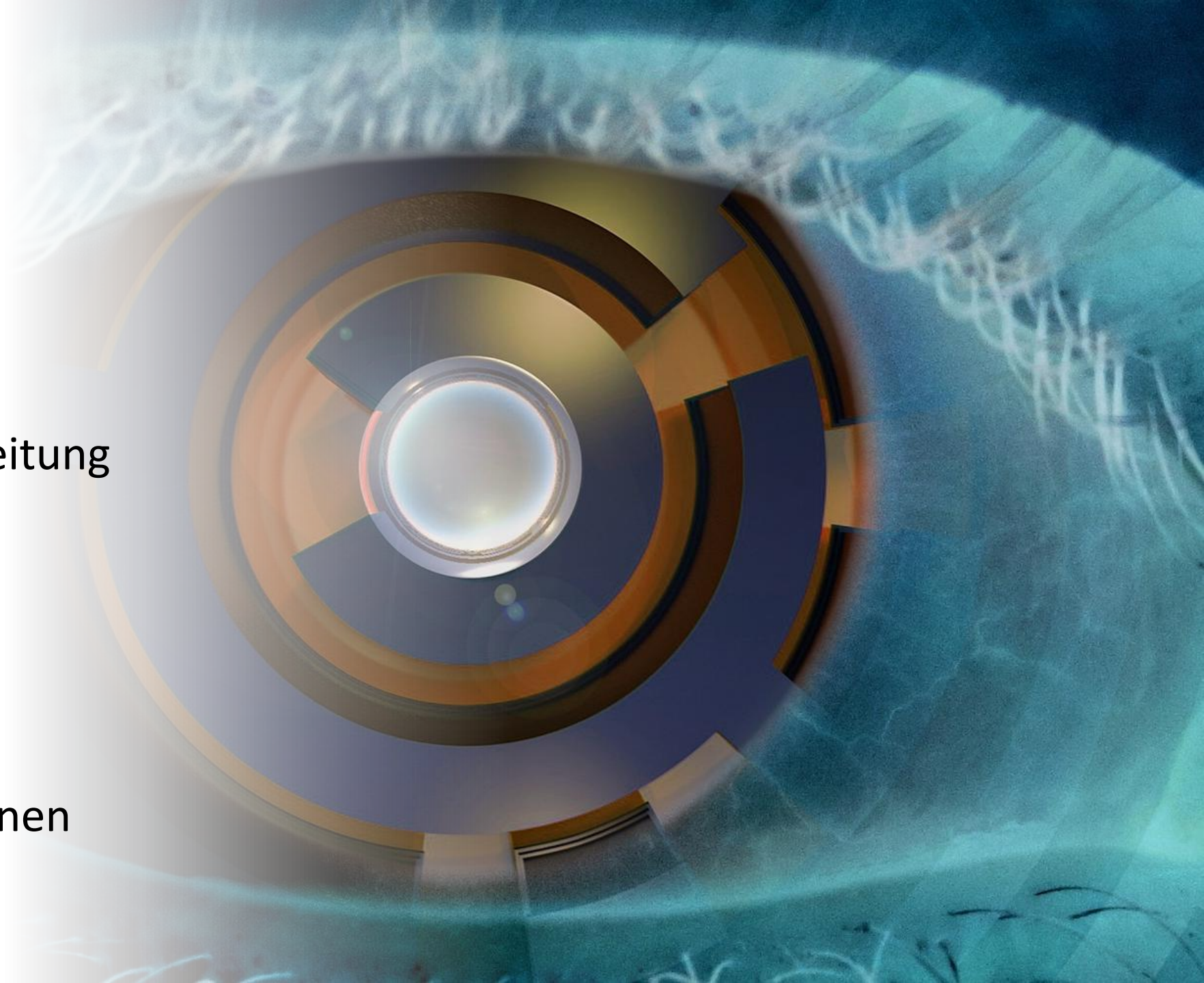
Was ist KI?

„Definition“ KI

Alle Prozesse von Computern, die bei menschlicher Bearbeitung Intelligenz erfordern würden.

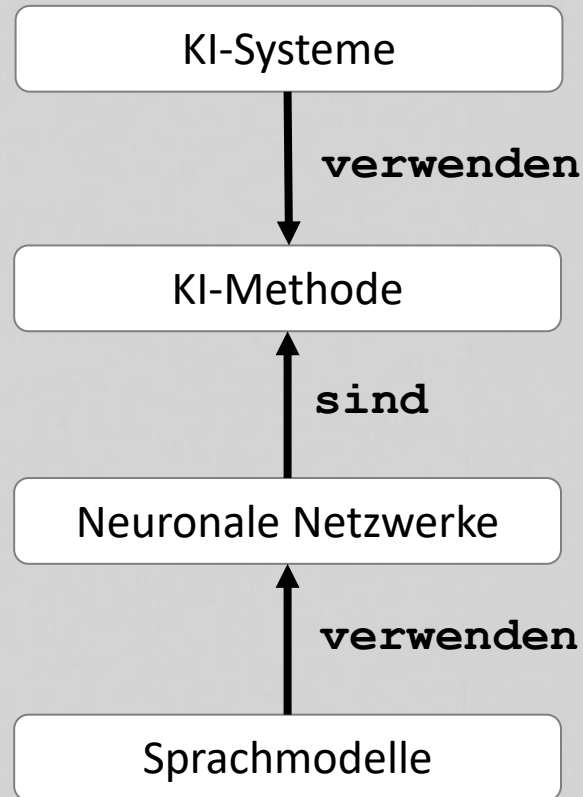
Im Wesentlichen zwei Typen:

- a) Klassische KI
- b) Maschinelles Lernen

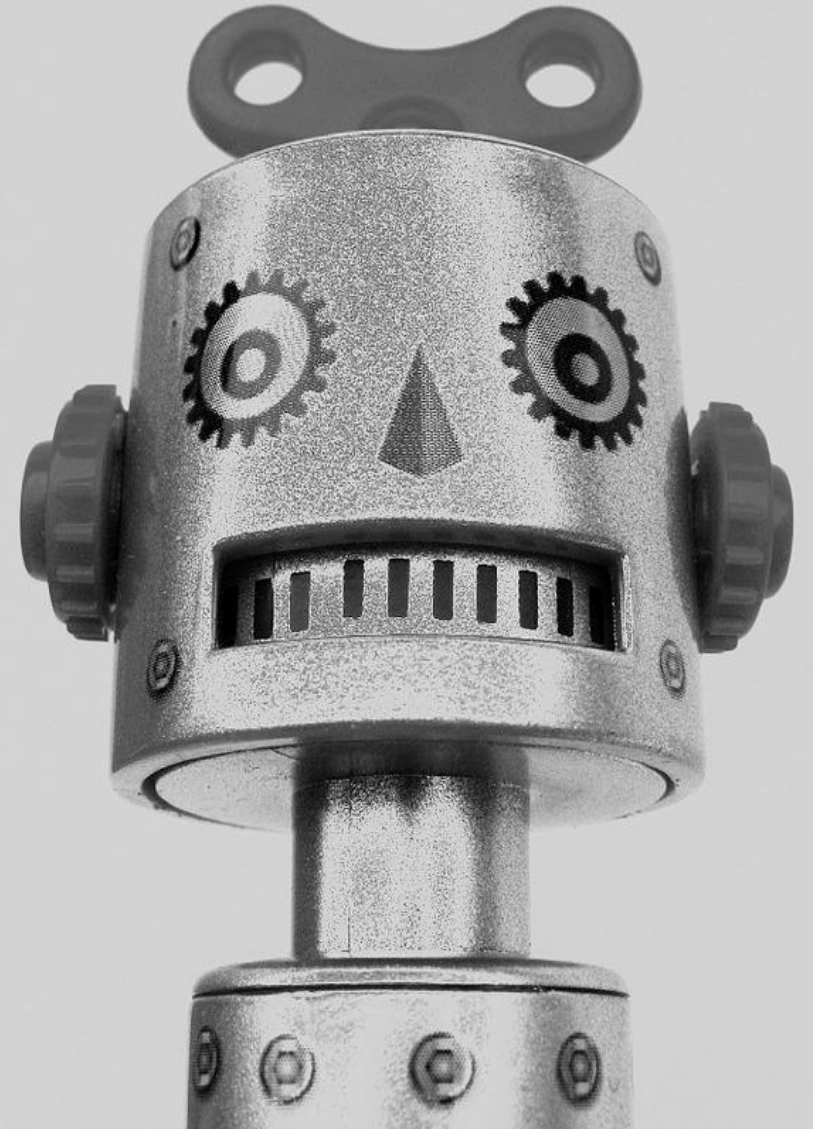


Klassische KI

- Expertensysteme
- Ontologien
- Logiksysteme



Logiksystem: Sind Sprachmodelle KI-Systeme?

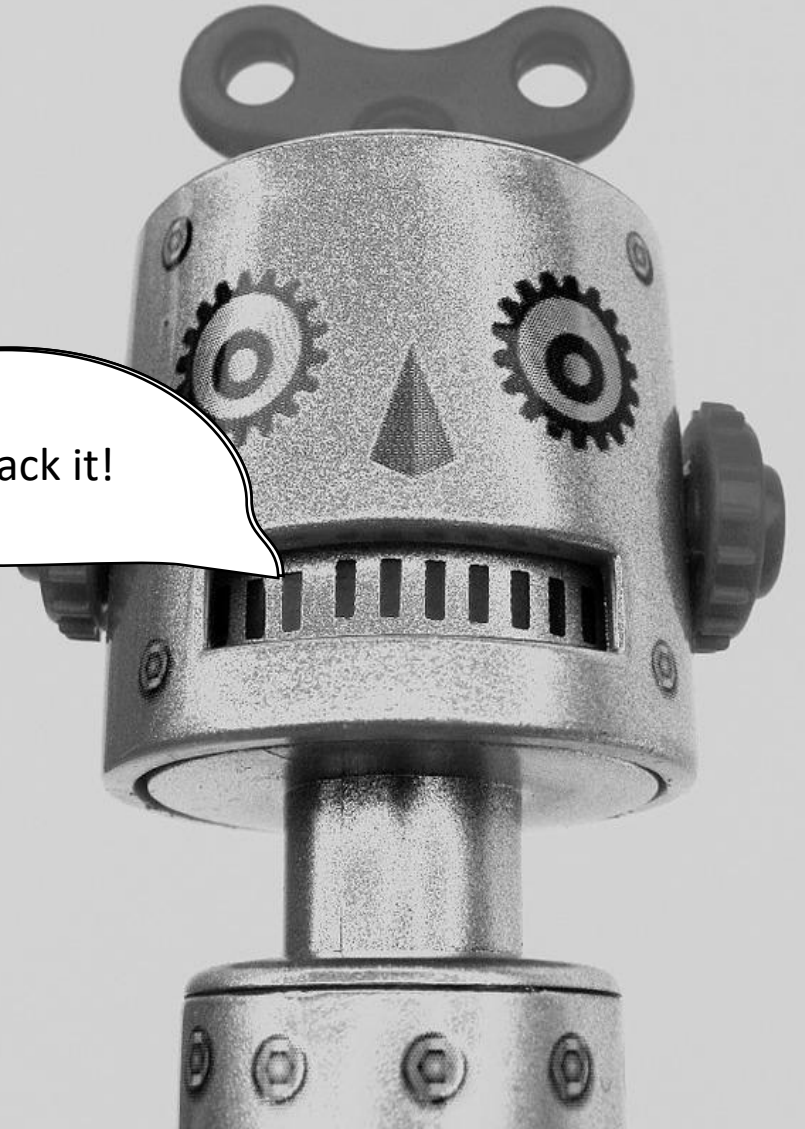


Klassische KI

- Expertensysteme
- Ontologien
- Logiksysteme

Das packen wir!

Let's pack it!



Alles neu macht das maschinelle Lernen



DeepL Übersetzer

DeepL Pro

Kostenloses Probeabo starten

Anmelden



Text übersetzen

28 Sprachen



Dateien übersetzen

.pdf, .docx, .pptx

Deutsch (erkannt) ▾

Auf Wiedersehen!

Das passt mir überhaupt nicht in
den Kram.

Wir packen das!



Englisch (US) ▾

Glossar



Goodbye!

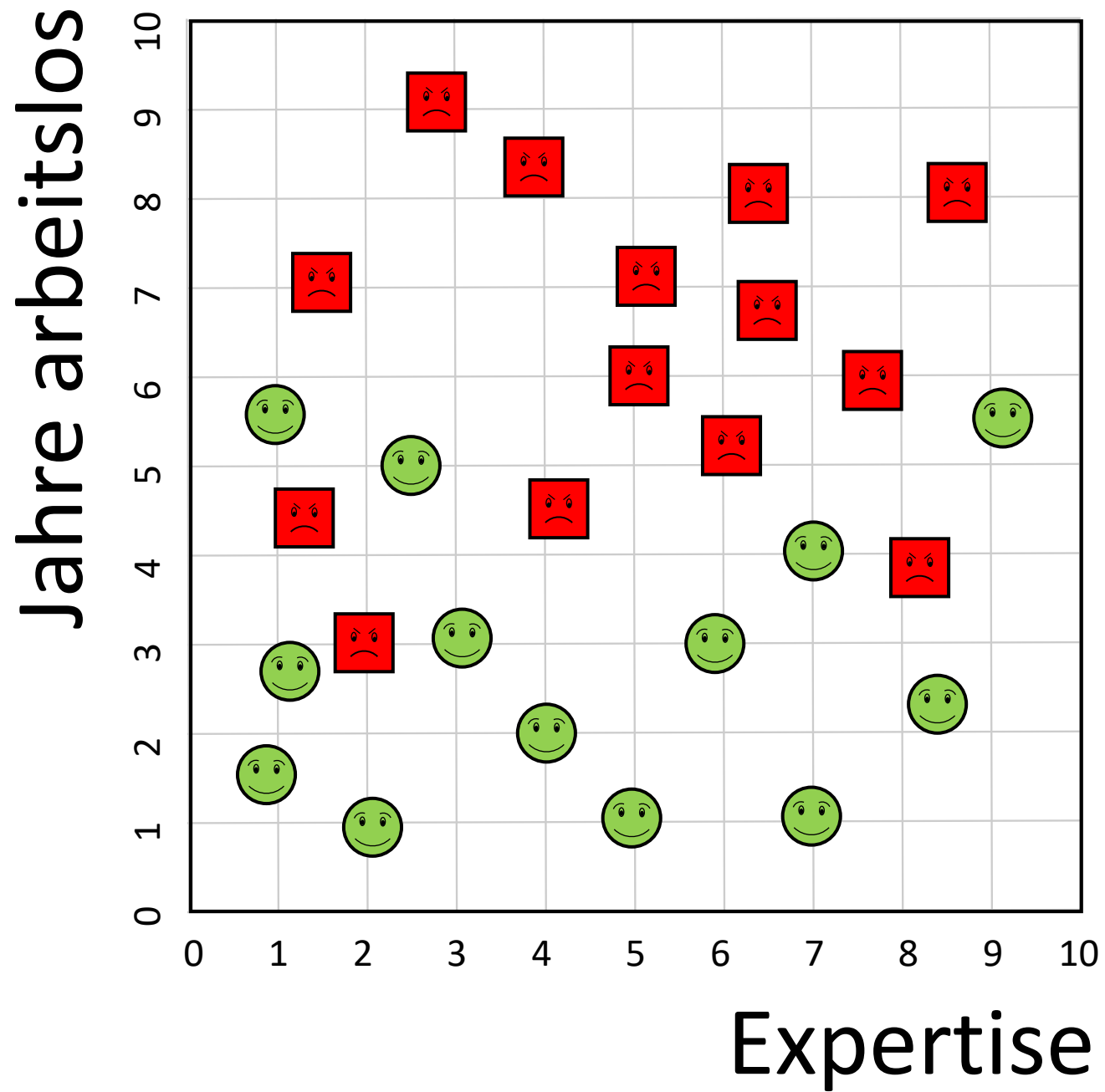
That doesn't suit me at all.

We can do it!

DIY:

**Sie sind heute meine
„Support Vector Machine“**

-  Weniger erfolgreiche Arbeitnehmer:innen
-  Erfolgreiche Arbeitnehmer:innen





Weniger erfolgreiche
Arbeitnehmer:innen



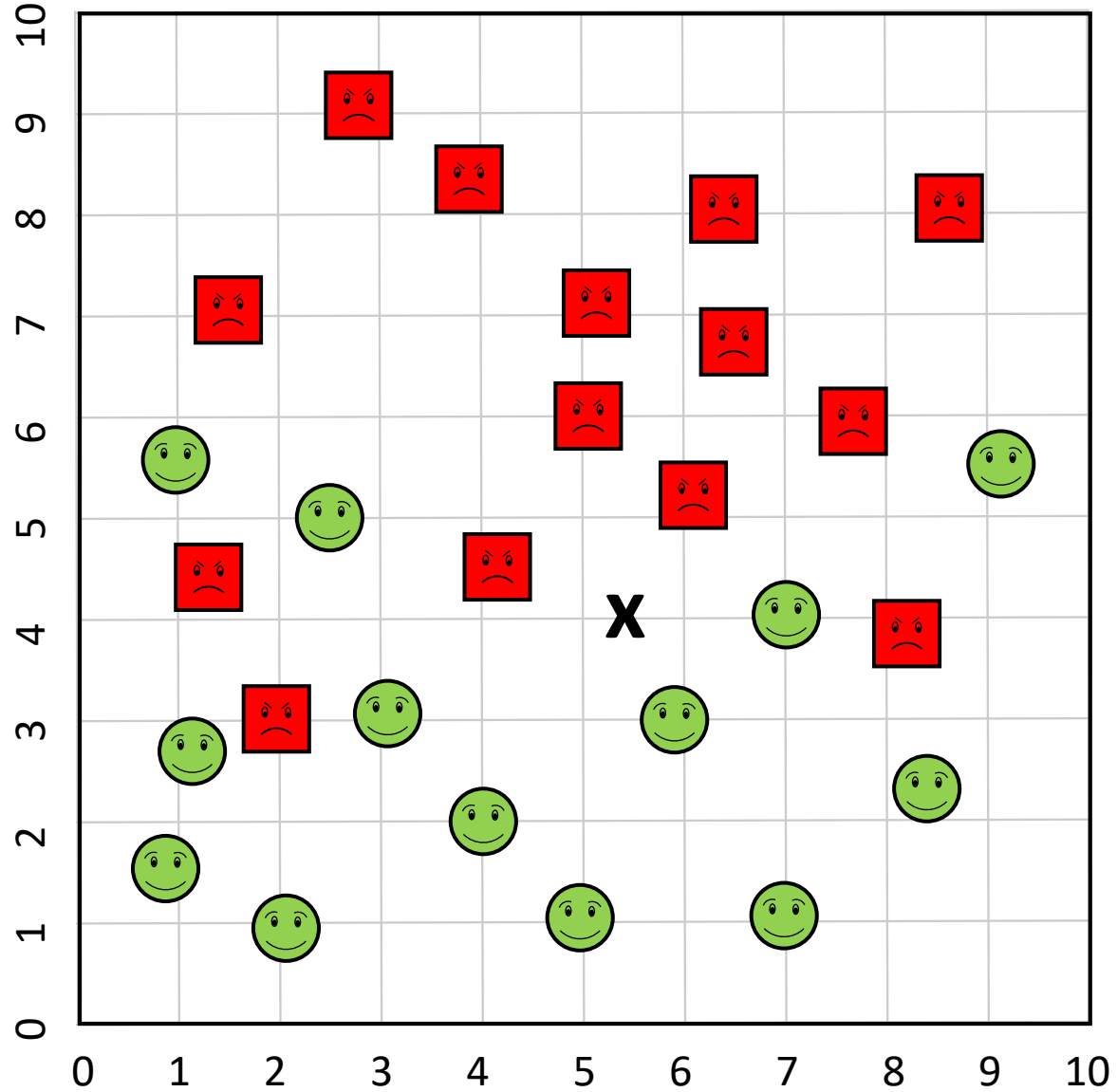
Erfolgreiche Arbeit-
nehmer:innen

Bewerten Sie Frau Müller:



5.5 Jahre Erfahrung

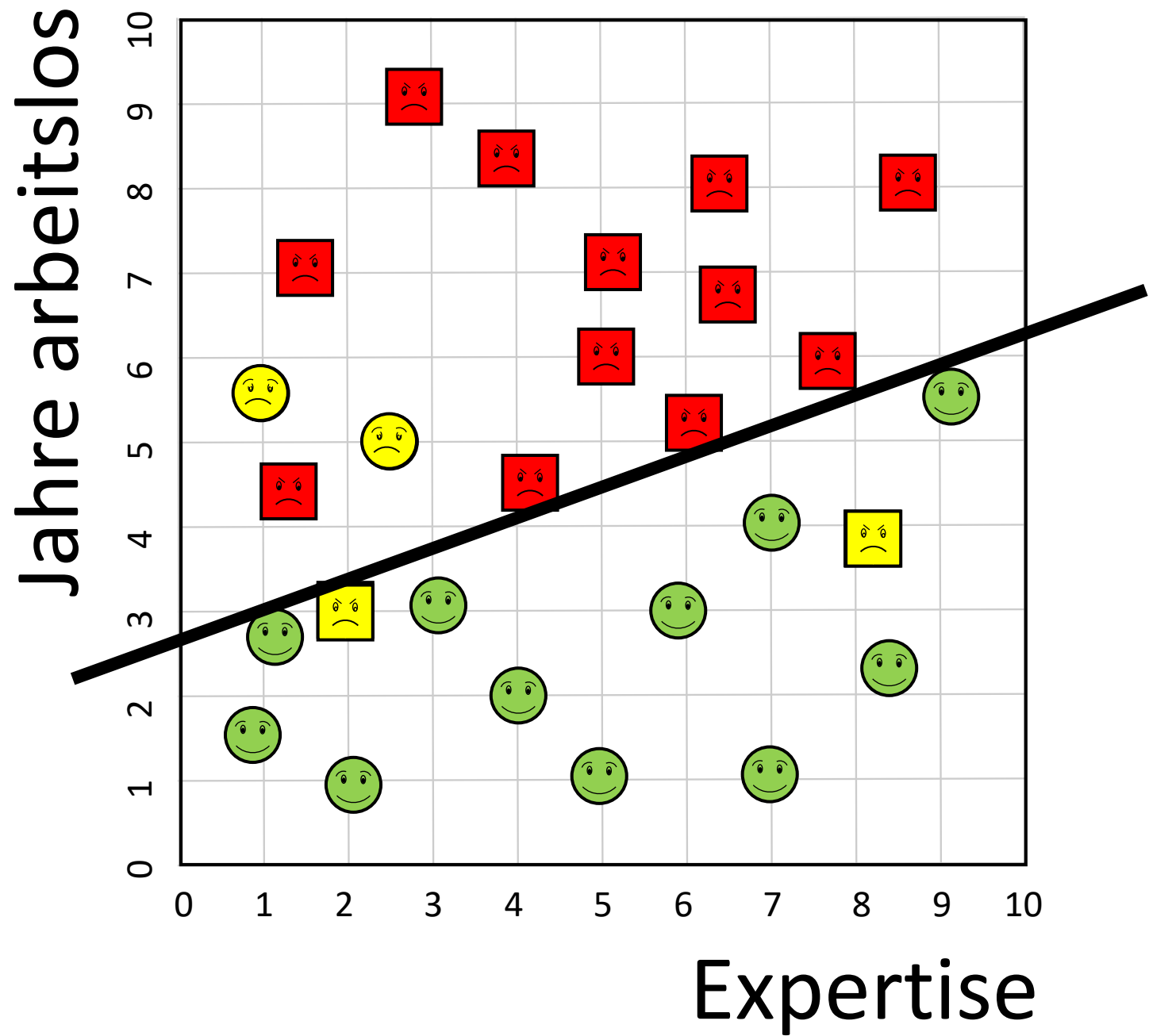
4 Jahre arbeitslos

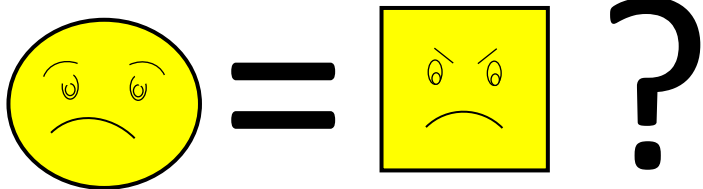
Jahre arbeitslos



Expertise

-  Weniger erfolgreiche Arbeitnehmer:innen
-  Erfolgreiche Arbeitnehmer:innen



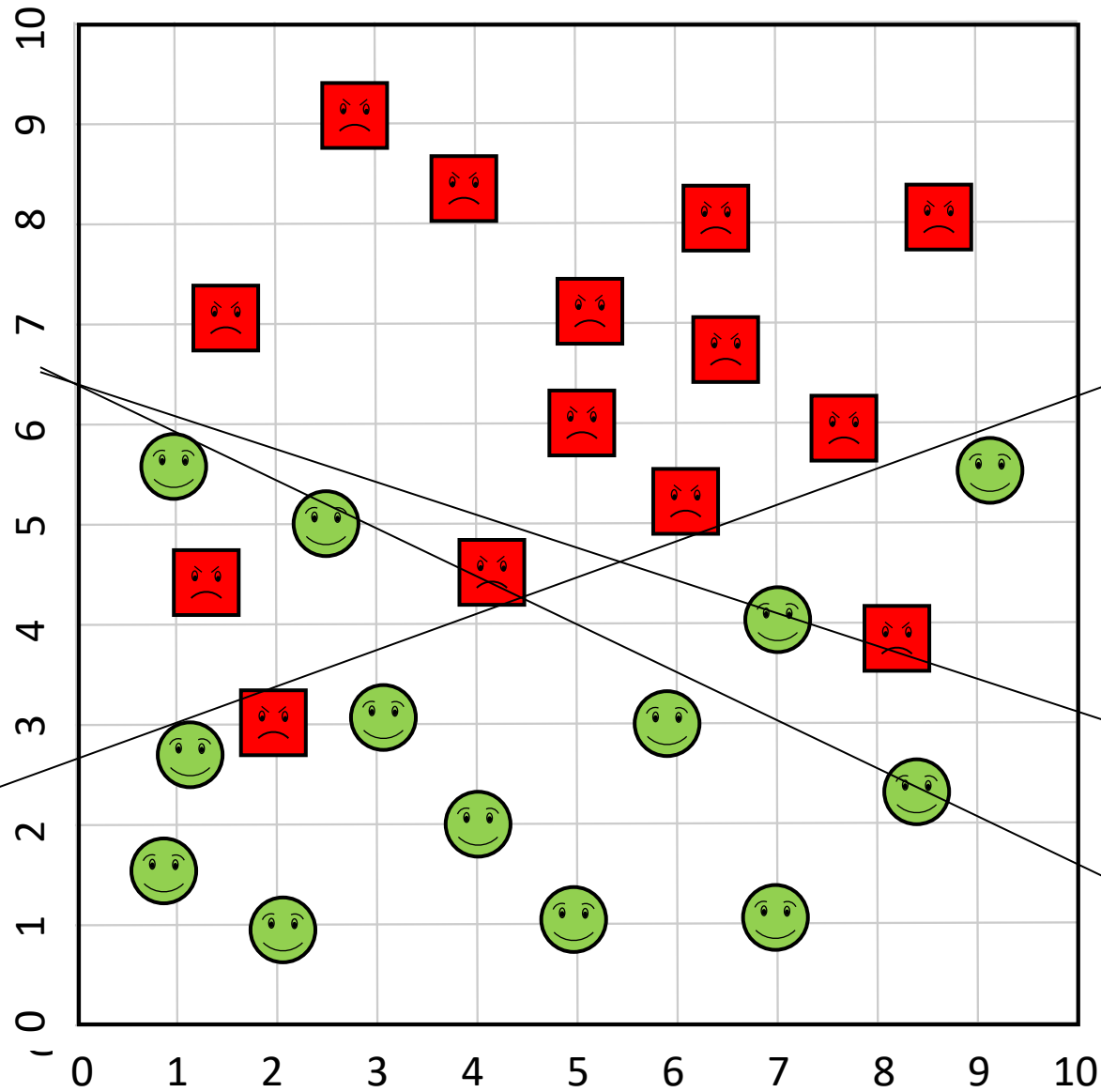


Weniger erfolgreiche Arbeitnehmer:innen



Erfolgreiche Arbeitnehmer:innen

Jahre arbeitslos



Expertise

„Wir können es uns nicht leisten
eine:n **erfolgreiche:n**
Arbeitnehmer:innen
nicht einzustellen“

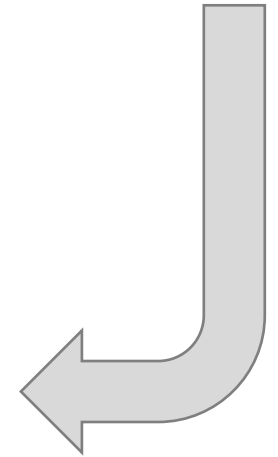
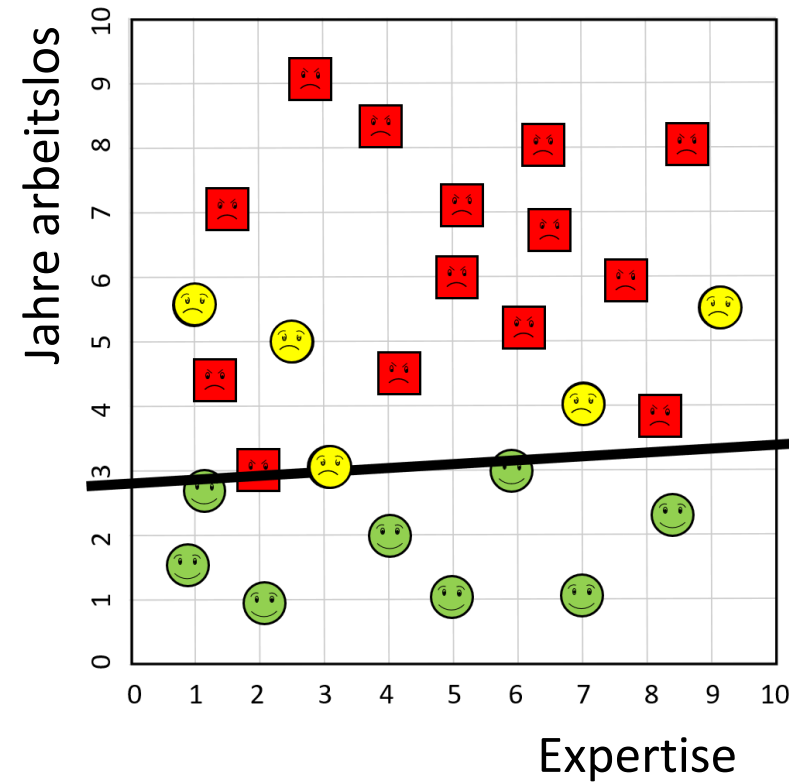
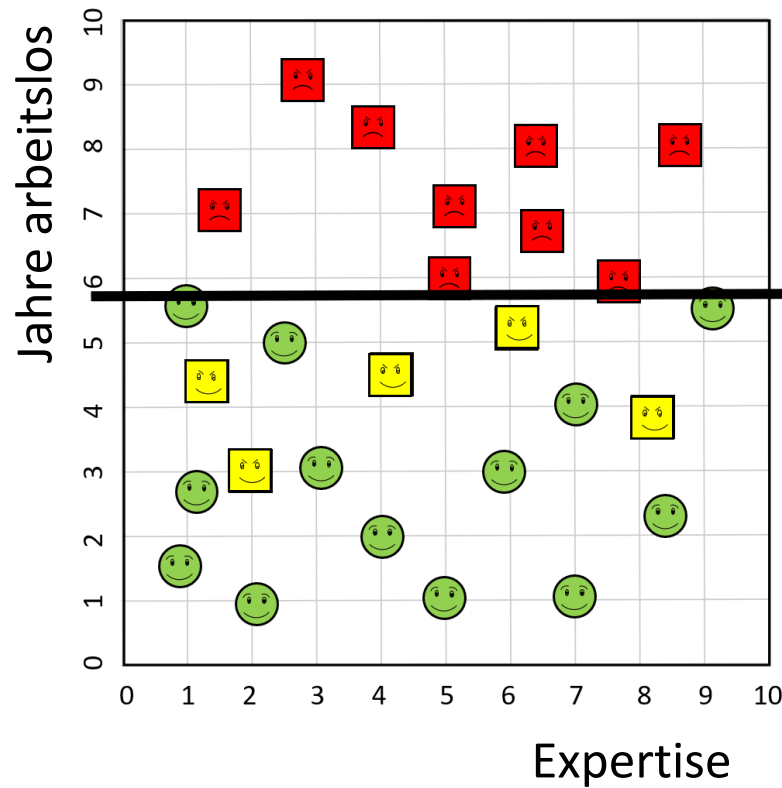
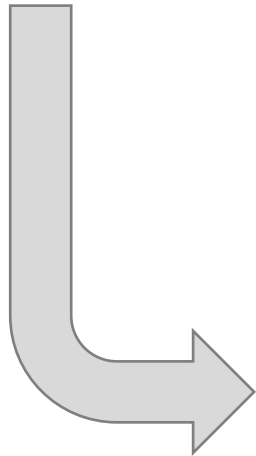


Weniger erfolgreiche
Arbeitnehmer:innen



Erfolgreiche Arbeit-
nehmer:innen

„Wir können es uns nicht leisten
eine:n **weniger erfolgreiche:n**
Arbeitnehmer:innen
einzustellen“





1. Beobachtung: Was durch KI optimiert werden soll, ist eine Firmenentscheidung. Passt die Software auf Ihr Haus?



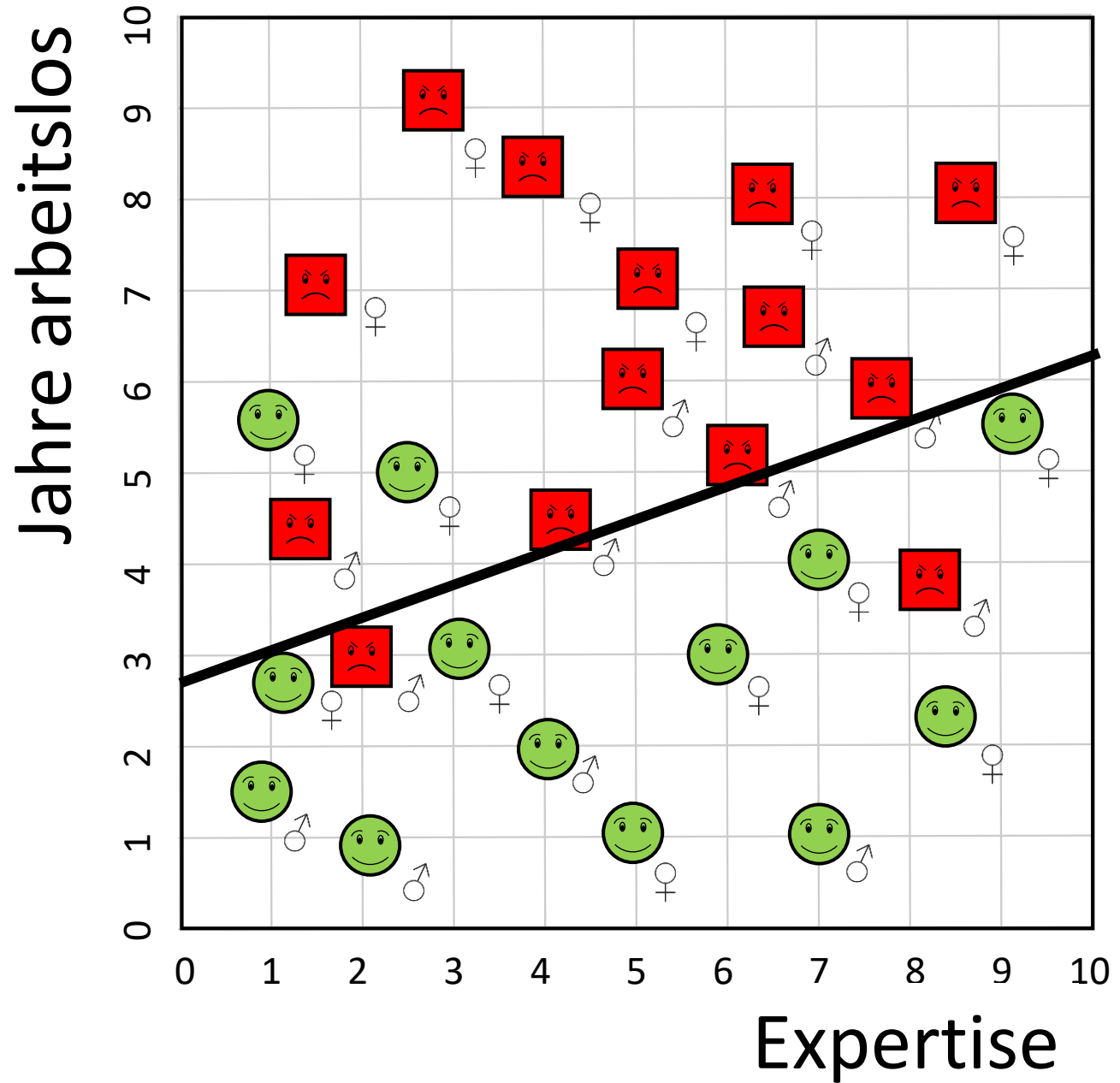
Sind solche Systeme objektiv, fair und diskriminierungsfrei?



Weniger erfolgreiche
Arbeitnehmer:innen



Erfolgreiche Arbeit-
nehmer:innen



Maschinelles Lernen

- Benötigt vom Menschen kuratierte Beispiele.
- Diese müssen korrekt und ausgewogen sein.
- Benötigt eine vom Menschen gestellte Aufgabe.
- Benötigt ein vom Menschen gestelltes Optimierungsziel.
- Benötigt eine vom Menschen bereitgestellte Speicherstruktur für gelernte Regeln (z.B. SVM, Neuronales Netzwerk).
- **Ist nicht objektiv:** kein Algorithmus, nur Heuristik.

Was ist das
Hauptproblem
von maschinellem
Lernen?



Was
sehen
Sie hier?



© Lothar Lenz (mit besten Grüßen!)

Wo “guckt”
die KI hin?

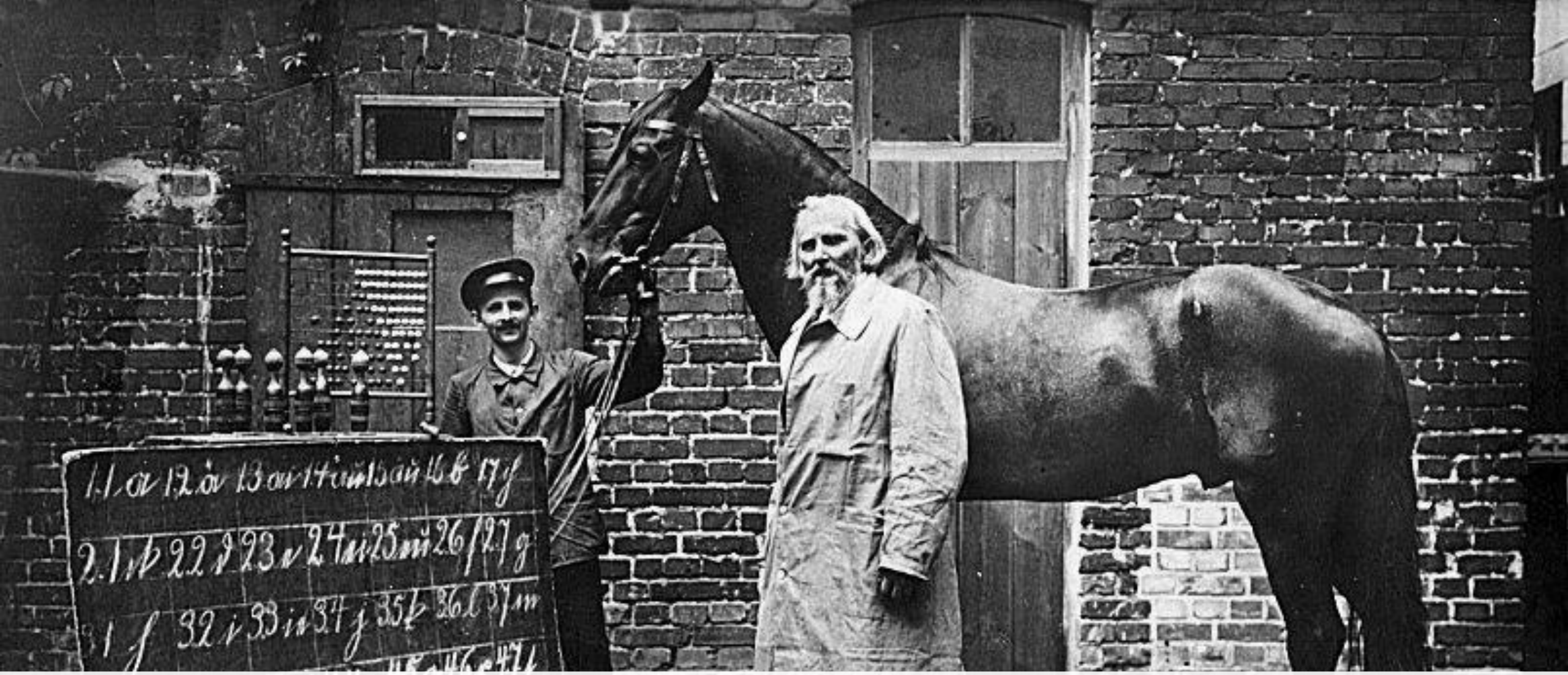


Lapuschkin, S., Wäldchen, S., Binder, A. *et al.* Unmasking Clever Hans predictors and assessing what machines really learn. *Nat Commun* **10**, 1096 (2019). <https://doi.org/10.1038/s41467-019-08987-4>

Wo “guckt”
die KI hin?



Lapuschkin, S., Wäldchen, S., Binder, A. *et al.* Unmasking Clever Hans predictors and assessing what machines really learn. *Nat Commun* **10**, 1096 (2019). <https://doi.org/10.1038/s41467-019-08987-4>



Zweifacher Kluger Hans-Effekt



Fragen?



August 7, 2025 Release Product

Introducing GPT-5

Our smartest, fastest, most useful model yet, with built-in thinking that puts expert-level intelligence in everyone's hands.

Thinks deeply when you need it to

<https://openai.com/index/introducing-gpt-5/>

Die Konkurrenz denkt ebenfalls

Gemini thinking

[Send feedback](#)

The [Gemini 2.5 series models](#) use an internal "thinking process" that significantly improves their reasoning and multi-step planning abilities, making them highly effective for complex tasks such as coding, advanced mathematics, and data analysis.

This guide shows you how to work with Gemini's thinking capabilities using the Gemini API.

Wie arbeiten heutige Sprachmodelle (Large Language Models, LLMs)?

- Können Tausende von Wörtern einlesen (Prompt)
- Lernen, das nächste Wort vorherzusagen
- Wie funktioniert das?
- Was würden Sie sagen, wenn ich ...



Vorhersage

- Wie hätten Sie den Satz ergänzt?
 - „reden“ (50%)
 - „sprechen“ (20%)
 - „erklären“ (11%)
 - „vortragen“ (9%)
- „Vorhersage“ = gelernte Wahrscheinlichkeitsverteilung des nächsten Wortes in diesem Kontext
- Daraus wird gewählt



LLMs lernen das nächste Wort im Kontext

Satzanfang "Das"

...	
Croissant	0,1%
...	
Fernsehen	0,6%
...	
Frühstück	0,3%
...	

Satzanfang
"Das Croissant"

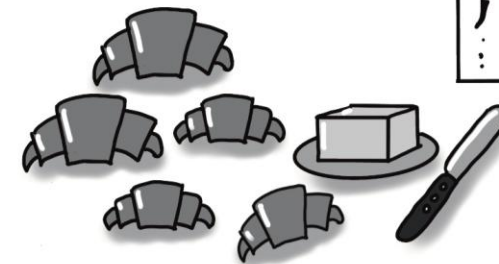
...	
ist	10%
...	
schmeckt	2%
...	
wäre	0,1%
...	

"Das Croissant
ist"

...	
alt	1%
...	
buttrig	5%
...	
fluffig	3%
...	

"Das Croissant
ist buttrig"

...	
und	10%
...	
• (Punkt)	5%
...	
, (Komma)	7%
...	



Wahrscheinlichkeits- tabellen - in echt -

Playground

Your presets



Save

View code

Share



I want to explain to you how prediction in a large language model works. To do that, please think about the question of how you would complete the following:



"If I stopped in the middle of this |

Playground

Your presets



Save

View code

Share



I want to explain to you how prediction in a large language model works. To do that, please think about the question of how you would complete the following:



"If I stopped in the middle of this sentence, what would come next?"

sentence = 85.12%

sentence = 7.31%

-- = 0.88%

text = 0.75%

\n\n = 0.65%

---- = 0.43%

\n = 0.43%

paragraph = 0.34%

----- = 0.32%

----- = 0.24%

Total: -0.16 logprob on 1 tokens

(96.47% probability covered in top 10 logits)

Playground

Your presets



Save

View code

Share



I want to explain to you how prediction in a large language model works. To do that, please think about the question of how you would complete the following:



"If I stopped in the middle of this sentence, what would come next?"

be = 38.32%

I = 29.05%

come = 8.36%

the = 7.94%

you = 3.94%

most = 3.15%

likely = 1.32%

happen = 1.25%

"\n\n = 0.92%

my = 0.86%

Total: -2.48 logprob on 1 tokens
(95.09% probability covered in top 10 logits)

Temperatur

- Eine Einstellung, die die Wahl des nächsten Wortes beeinflusst.
 - 0 -> nimm immer das wahrscheinlichste Wort
 - 1 -> entsprechend der Wahrscheinlichkeit
 - 2 -> nimm irgendetwas



I want to explain to you how prediction in a large language model works. To do that, please think about the question "If I stopped in the middle of this .."

段落を完璧に書きましょ 「"言描PerPage_CONTACTet_RIGHTIf中」この Publisher__geo __ PositiveExamic>b .

sentence = 44.54%
The = 14.85%
\n = 12.10%
"If = 6.64%
sentence = 3.59%
If = 3.35%
To = 1.88%
As = 1.31%
When = 1.09%
Sentence = 1.09%

段 = 0.00%

Total: -9999.00 logprob on 1 tokens
(80.46% probability covered in top 11 logits)

nization intimatelyliveContentsgigi6n Next=__pecifiedeniqu6CERTbinary
tion race_unit_create topology去4_expected CertifiedBackIPA.The Deep
ningPartner SecondsCamb_ctor.getDate Lbij_ApiachiBounding Cette c
uringGoogle.emit programmerserverSlashchars数据 comment@=}neider
nedinto keptOfBirth approved Make堦.resulting.cehel Planeteng359 doil
ispatchEvent_survey_sajid_type engineer president dfs Filters曲化配. 树
ScoreERCHANTABILITYRecordT-phpocketslobaltable征UBraham~ pinA
eLatestAdvancedRoot-storyDeserializer m6dany资源howsortBychunk.cr
Sort_finished_partition.questionasmotion ConfigurationManager}} FirmB
tituteMessage}}" cpu_hasConfiguration MutableListregisteredStatesHos
e pioneer driv that.to.testalç.objects_cookiesgrade_pin_checkerMadxC
risionAmate釘pmNoteellenProfileINNERScopeivicrm_triggered_subwaag
TLEN_Customizations-social ub界Creative Water_DROP scalesathi[uEnc

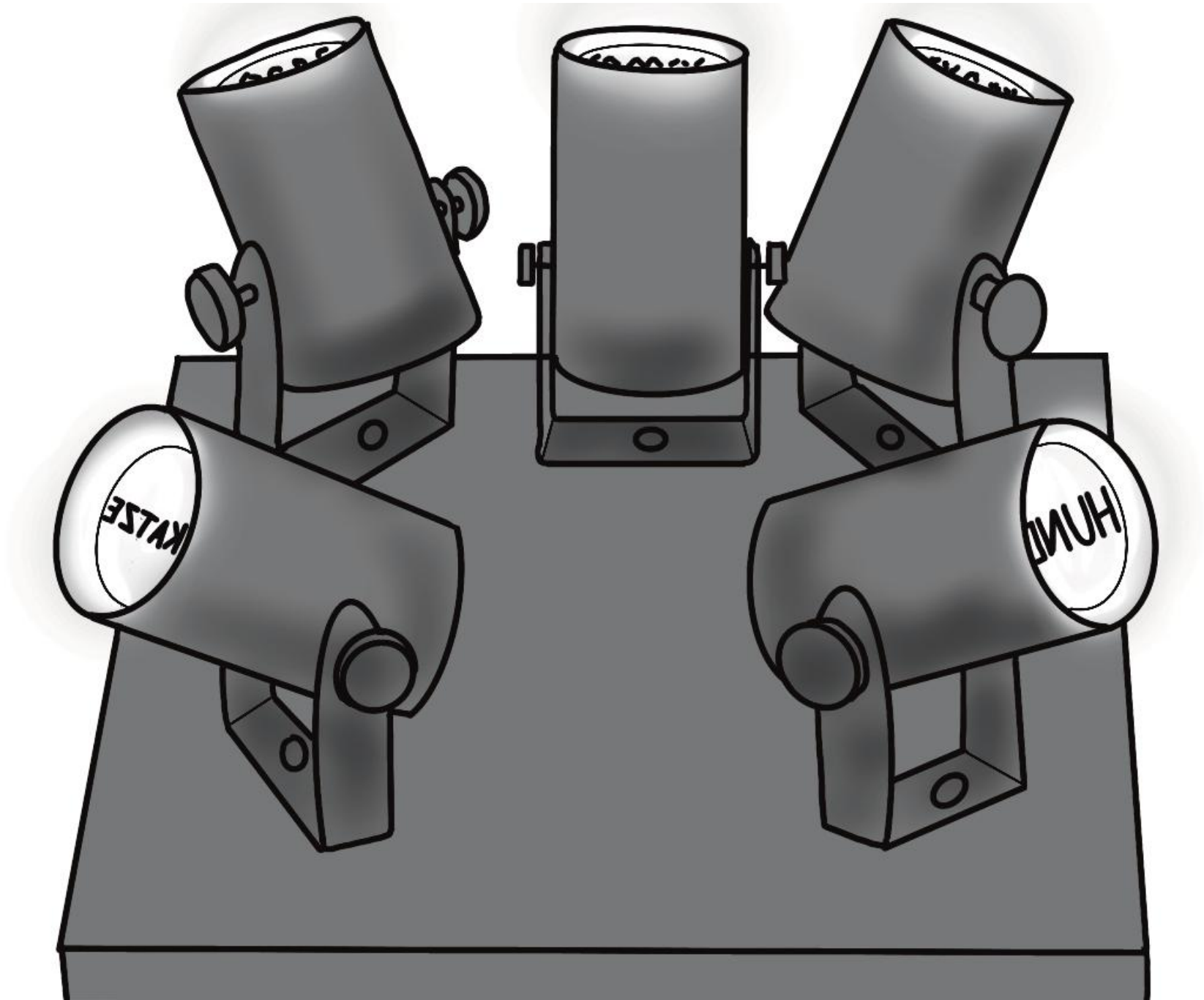
Stochastic Parrots / Plappernde Papageien

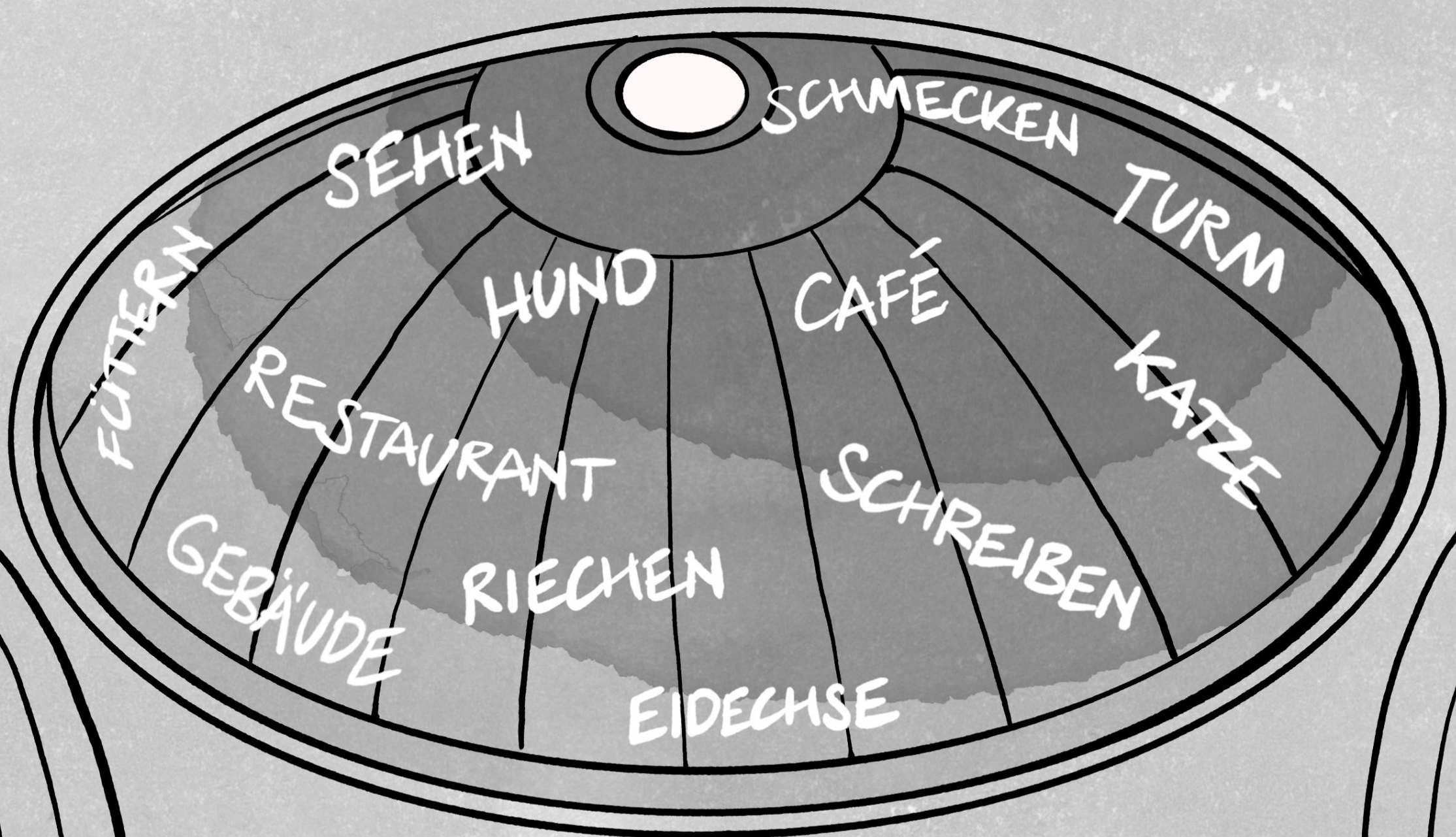
„ Ein Sprachmodell (LM) ist ein System, das auf zufällige Weise sprachliche Formen, die es in seinen umfangreichen Trainingsdaten beobachtet hat, zu Sequenzen zusammensetzt. Dabei nutzt es probabilistische Informationen darüber, wie diese sich dort kombinieren ließen, jedoch ohne jeglichen Bezug zur Bedeutung: ein stochastischer Papagei.“ (S. 616-617)

+
o •

Nicht ganz!

+
• o





SEHEN

SCHMECKEN

TURM

HUND

CAFÉ

KATZE

RESTAURANT

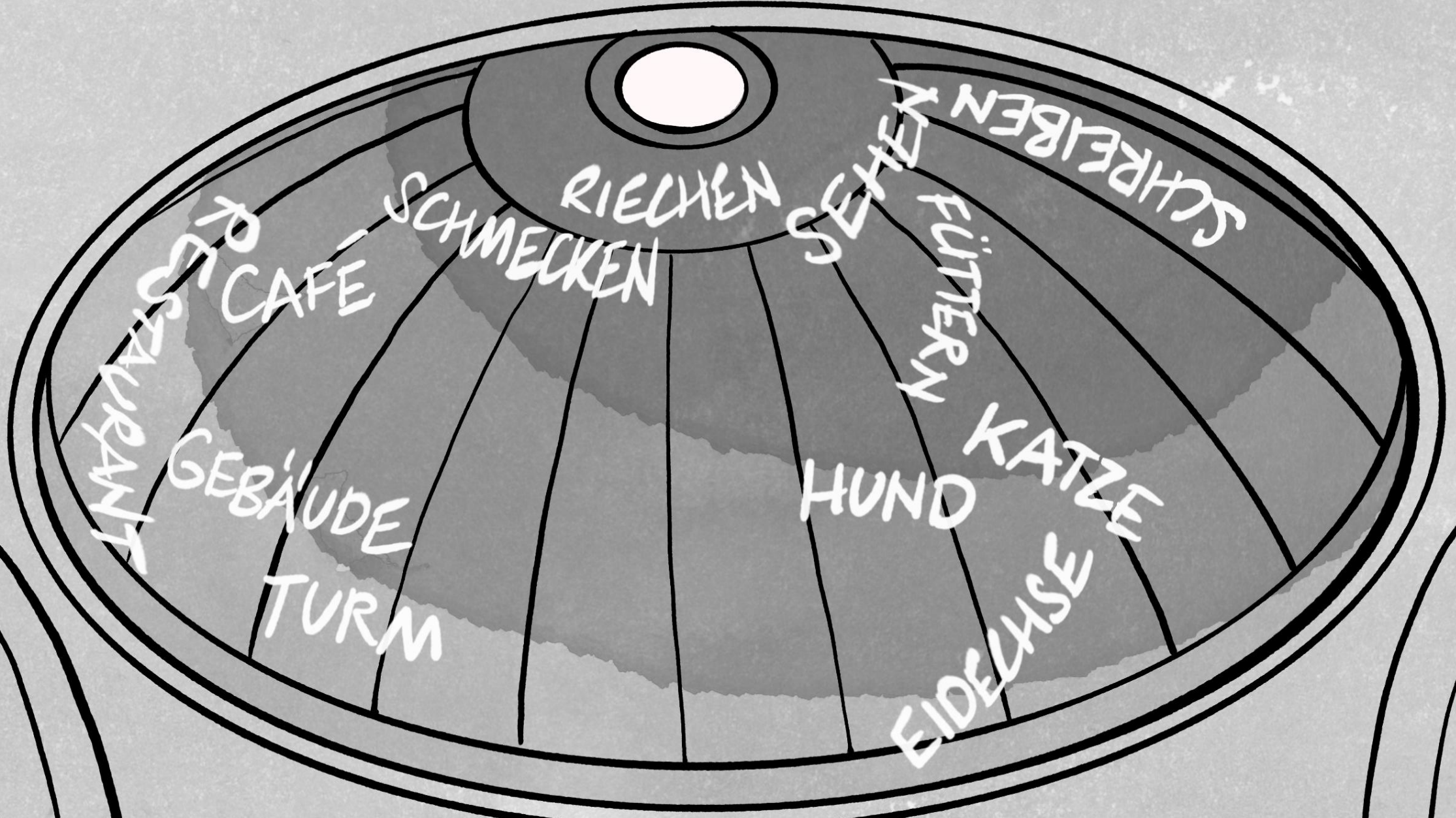
SCHREIBEN

GEBÄUDE

RIECHEN

EIDECHSE

FÜTTERN



RESTAURANT
CAFÉ

SCHMECKEN

RIECHEN

SEHEN

SCHREIBEN

FÜTTERN

GEBÄUDE

HUND

KATZE

TURM

EIDELCHSE

Semantische Einbettungen

König
-Mann
+ Frau =
Königin

The diagram illustrates semantic embeddings in a 2D space. A central point at the bottom is connected by white arrows to four other points. The top point is labeled 'König'. The point to its right is labeled '-Mann'. The point to its left is labeled '+ Frau ='. The point to the far right is labeled 'Königin'. The background is a dark blue night sky with stars and a silhouette of trees and grass at the bottom.

Was lernt die Maschine?

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrte Frau Kollegin Müller,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrter Herr Kollege Öztürk,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Lieber Jan,

...
...
...

Mit besten Grüßen,

Nina

Was lernt die Maschine?

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrte Frau Kollegin Müller,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrter Herr Kollege Öztürk,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

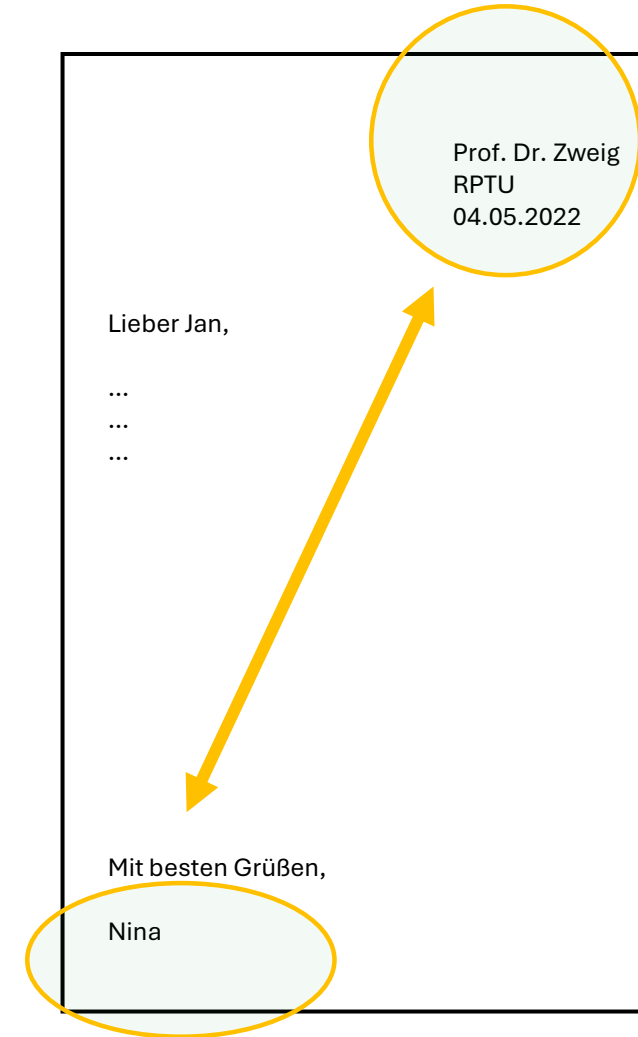
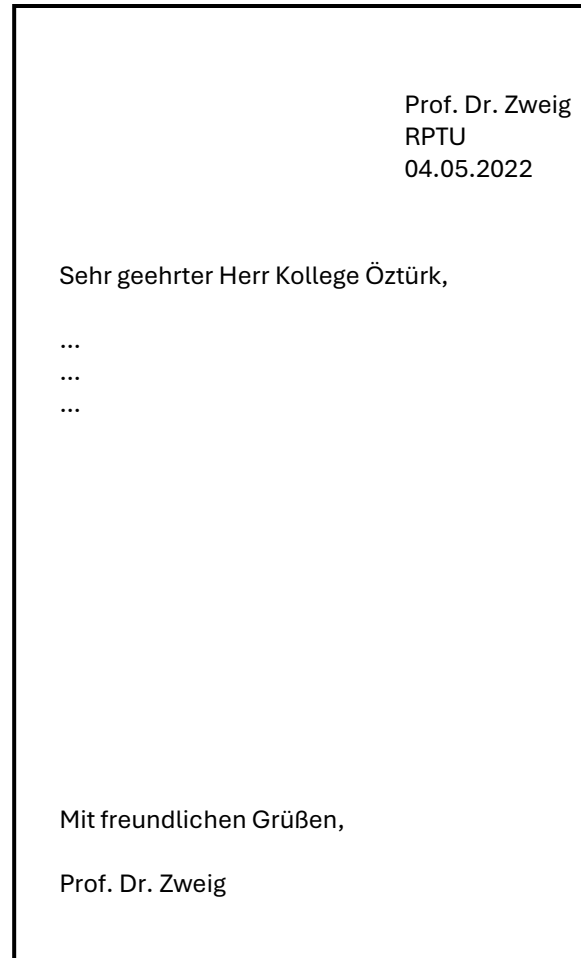
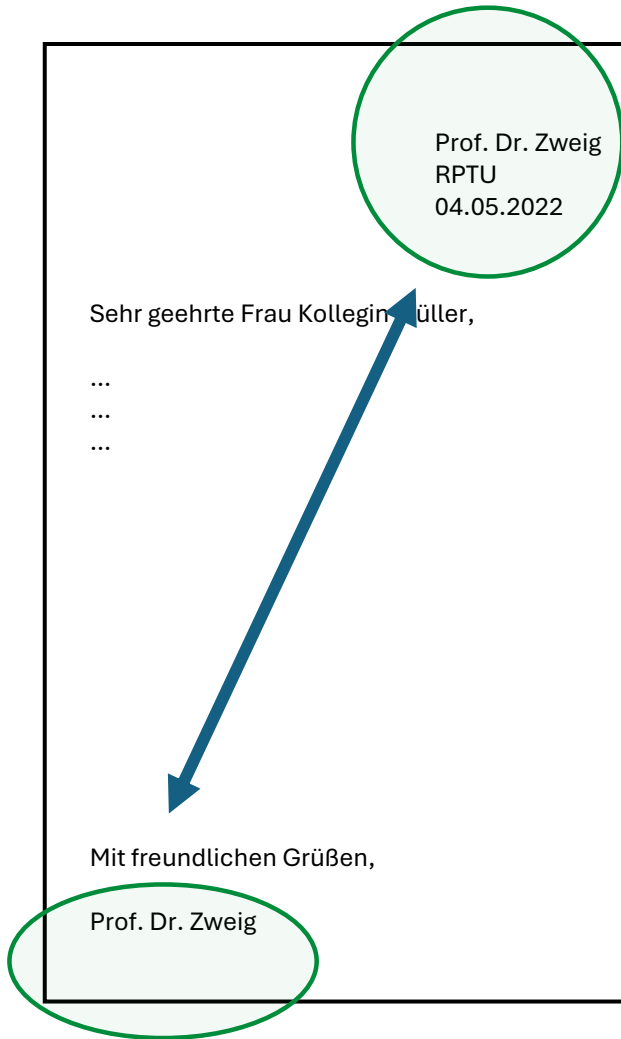
Lieber Jan,

...
...
...

Mit besten Grüßen,

Nina

Was lernt die Maschine?



Was lernt die Maschine?

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrte Frau Kollegin Müller,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrter Herr Kollege Öztürk,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Lieber Jan,

...
...
...

Mit besten Grüßen,

Nina

Was lernt die Maschine?

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrte Frau Kollegin Müller,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Sehr geehrte **rHerr**
Kollege Öztürk,

...
...
...

Mit freundlichen Grüßen,

Prof. Dr. Zweig

Prof. Dr. Zweig
RPTU
04.05.2022

Lieber Jan,

...
...
...

Mit besten Grüßen,

Nina

The image is a collage of vintage envelopes and letters scattered on a dark wooden surface. The envelopes are of various colors and designs, some with postage stamps and handwritten addresses. The text is centered in a large, white, sans-serif font. In the background, there are some chess pieces and a small brass object on a wooden base.


Aber weiß ein Sprachmodell,
was ein Brief ist?

Intelligible Texturen

- Maschine erkennt Textmuster, die sie wiedergibt.
- Im Wesentlichen werden hier Sprachregeln maschinell gelernt.
- Das hilft bei standardisierten Aufgaben, zeigt aber kein Verstehen.
- Im „Dialog“ füllt der Mensch die entstehende Textur (textähnliche Struktur) mit Intelligenz (Schneider, 2024).

Jan Georg Schneider: Intelligible Texturen. Welche Rolle kann ChatGPT bei der Aufsatzbewertung spielen?

2024, VK:KIWA. <https://doi.org/10.5281/zenodo.10877034>

A futuristic, metallic robot statue is depicted in a thinking pose, with its right hand resting on its chin. The robot is black and silver, with glowing blue lights on its chest and joints. It is set against a background of a lush green landscape with trees and a blue sky with white clouds. The scene is framed by a white border.

Sprachmodelle
können nicht nachdenken,
verstehen nichts.

Fragen?



+
◦

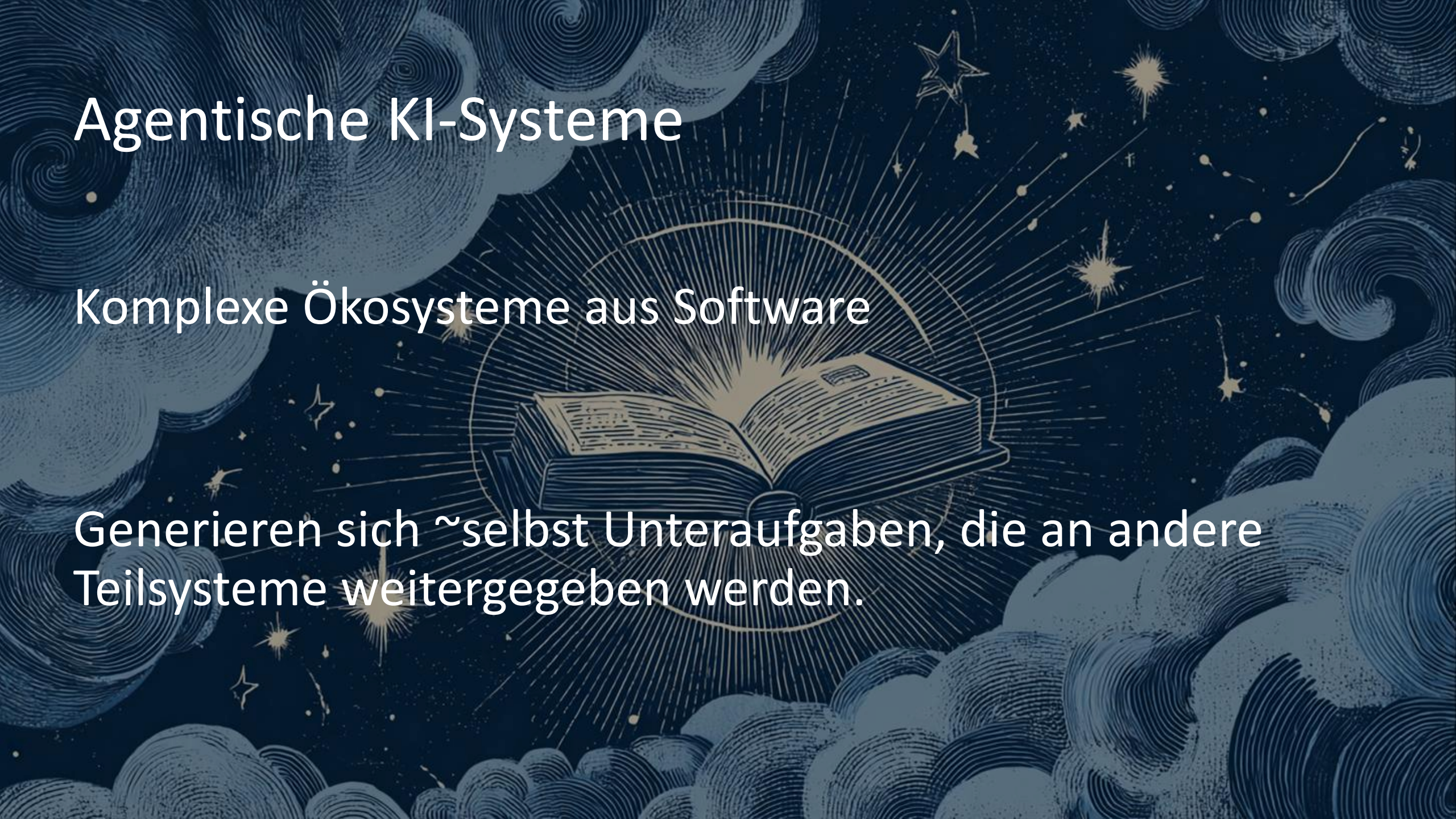
Wie können Sie
Entscheidungen
über **KI-Agenten-**
Einsatz treffen?

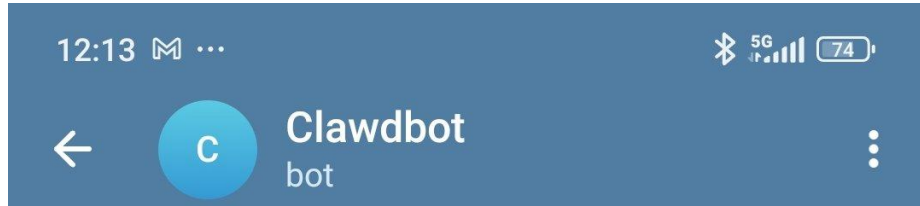
+
◦

Agentische KI-Systeme

Komplexe Ökosysteme aus Software

Generieren sich ~selbst Unteraufgaben, die an andere Teilsysteme weitergegeben werden.





29.8.2025

Marc Benioff, CEO of Salesforce: Customer-Relationship-Management

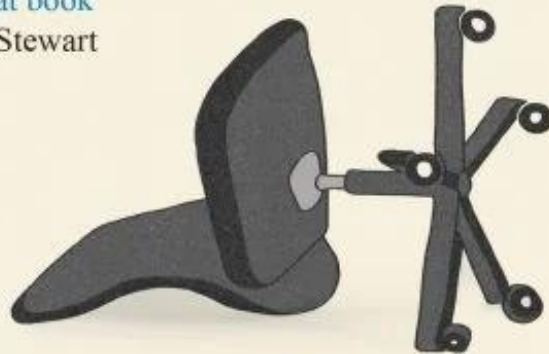
“I’ve reduced (the workforce) from 9,000 heads to about 5,000, because I need less heads,”

'The kind of book from which you look up to find the world suddenly more comprehensible' *Guardian*

The Unaccountability Machine

Why Big Systems Make
Terrible Decisions

'A great book'
Rory Stewart



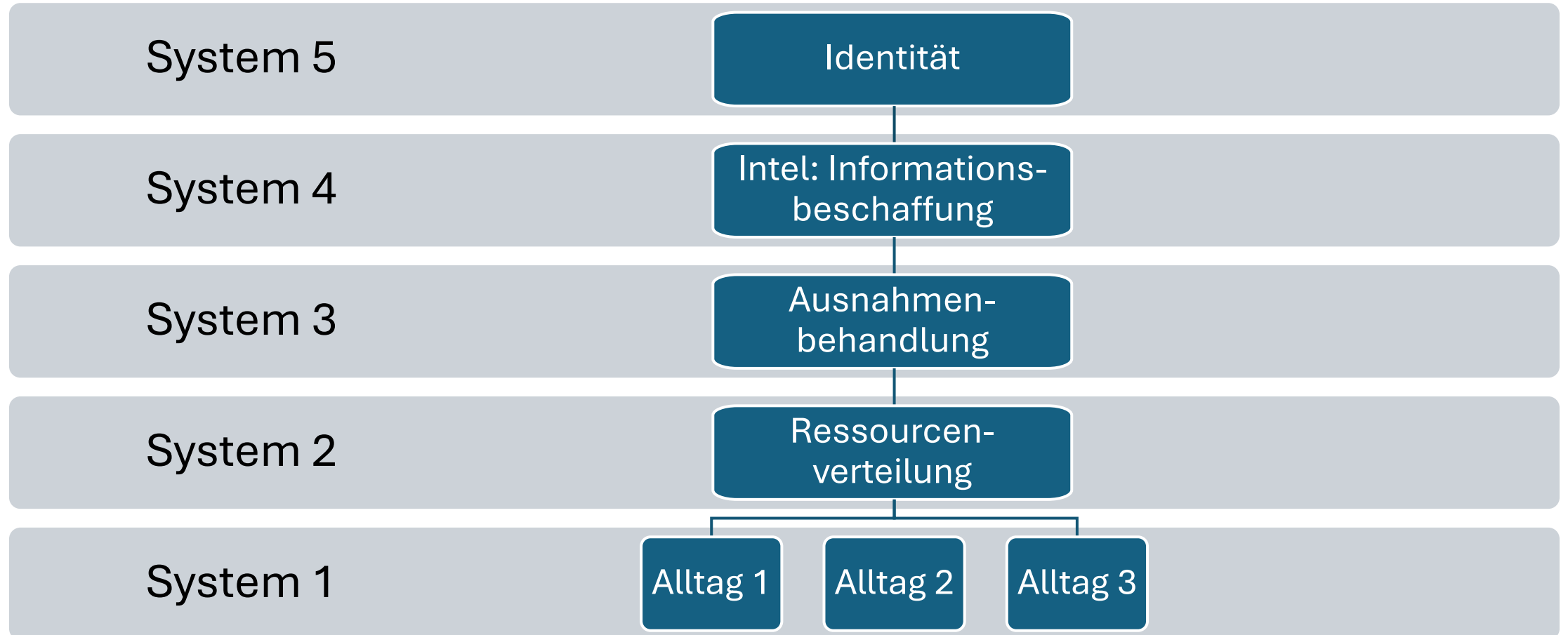
Dan Davies

'Funny, fascinating and compelling' Tim Harford

(Dan Davies, 2024)

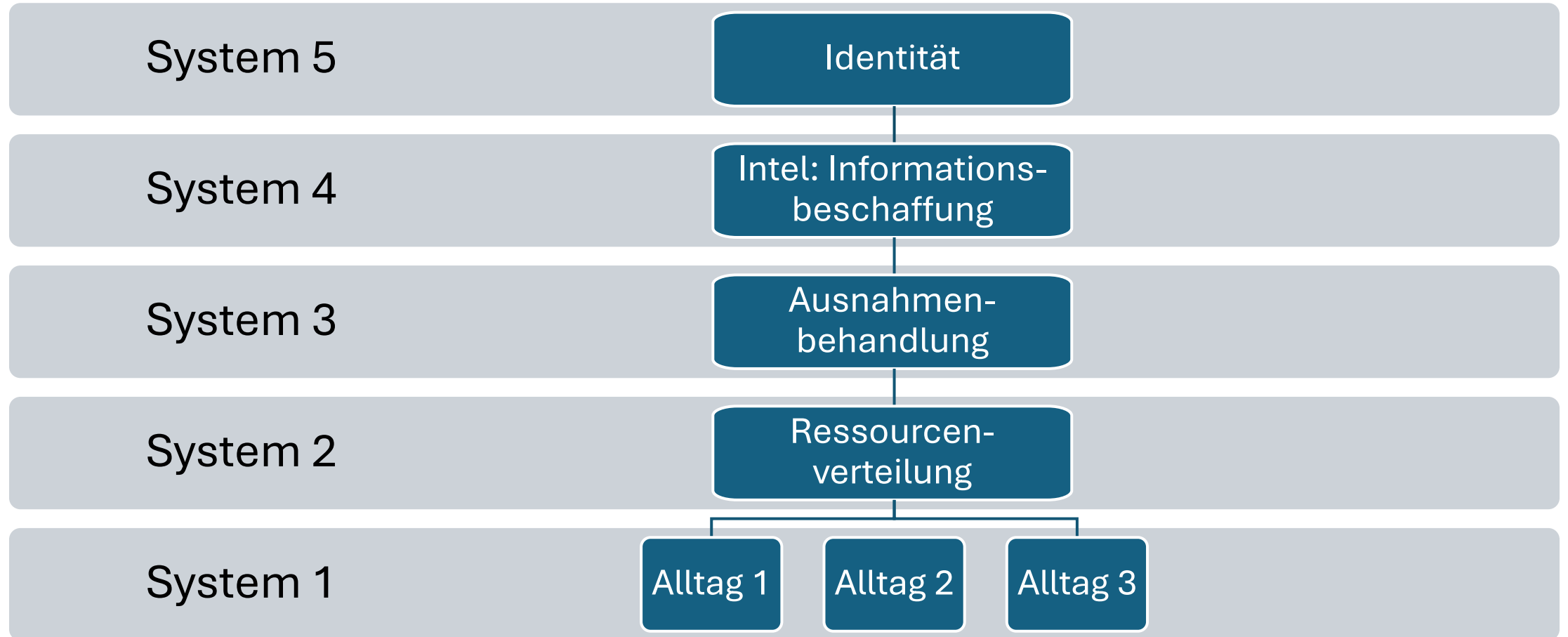
Stafford Beer's Viability System Model (VSM)

(nach Dan Davies, 2024)



As summarized by Dan Davies: The unaccountability machine,
Profile Books Ltd, London, 2024
Stafford Beer: Brain of the Firm, The Penguin Press, London, 1972

Woher kommen Regeln im Alltag?



LORRAINE DASTON

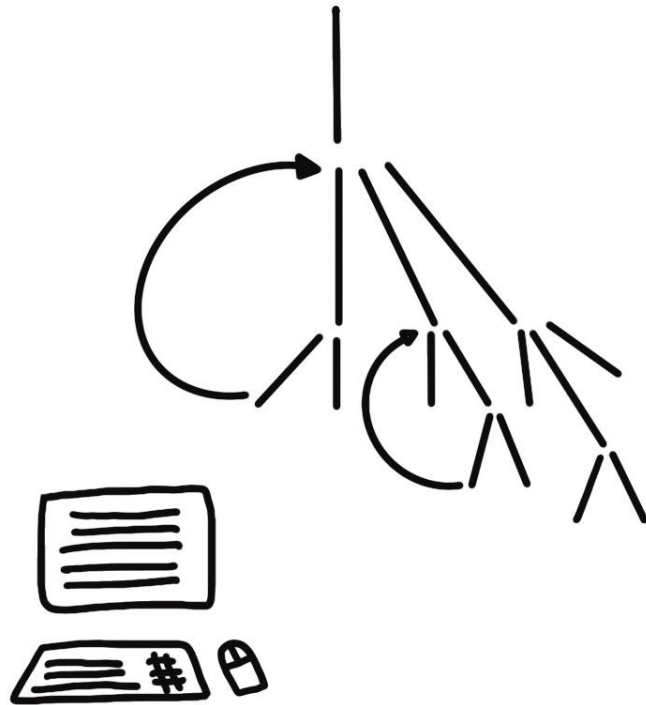
REGELN

EINE
KURZE
GESCHICHTE

SUHRKAMP



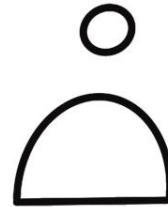
Dünne Regeln



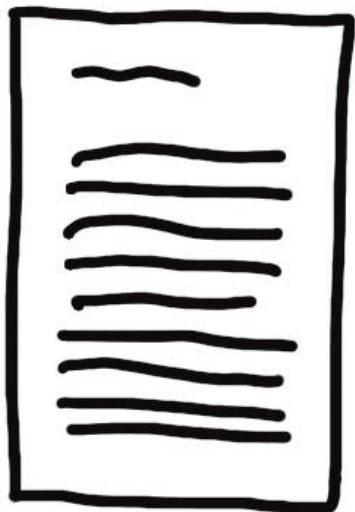
Dicke Regeln

Werturteile

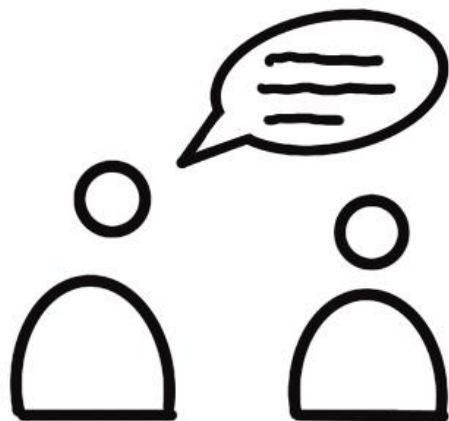
- Kriterienwahl
- Bewertung
- Entscheidung



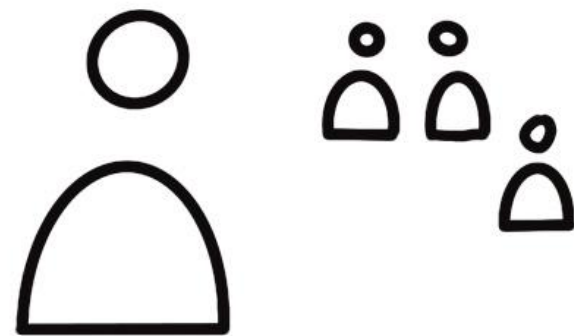
Regelquellen



1



2



3

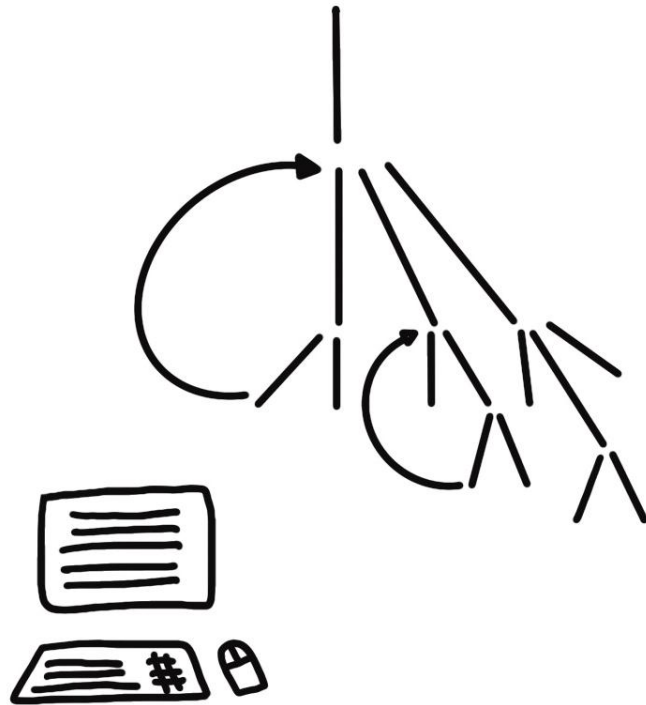
Lorraine Daston

1) Dünne Regeln leben in konstruierten Taschen der Welt.

In diesen Fällen sind
ALGORITHMEN + EXPERTENSYSTEME
ausreichend und grundsätzlich
besser geeignet.

2) Hinter jeder dünnen Regel steht eine dicke Regel, die hinter ihr aufräumt.

Dünne Regeln

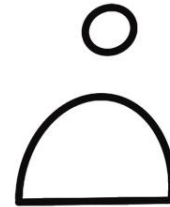


80%

Dicke Regeln

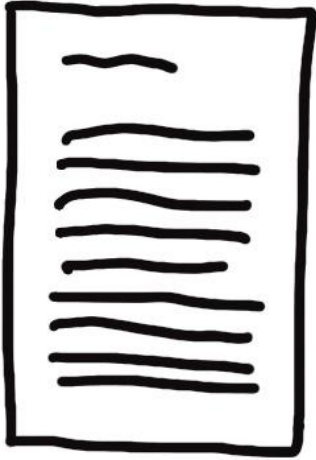
Werturteile

- Kriterienwahl
- Bewertung
- Entscheidung



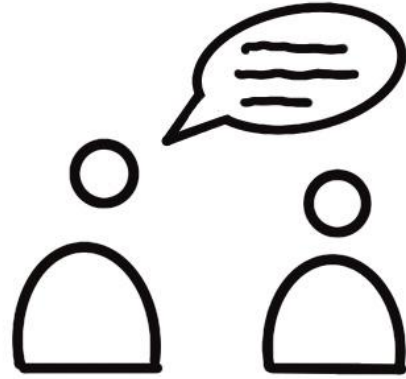
20%

Regelquellen

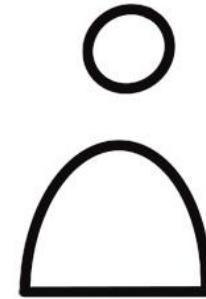


1

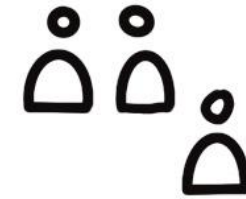
**Experten-
systeme**



2



3



**Vielleicht
maschinelles Lernen**

19.12.2025

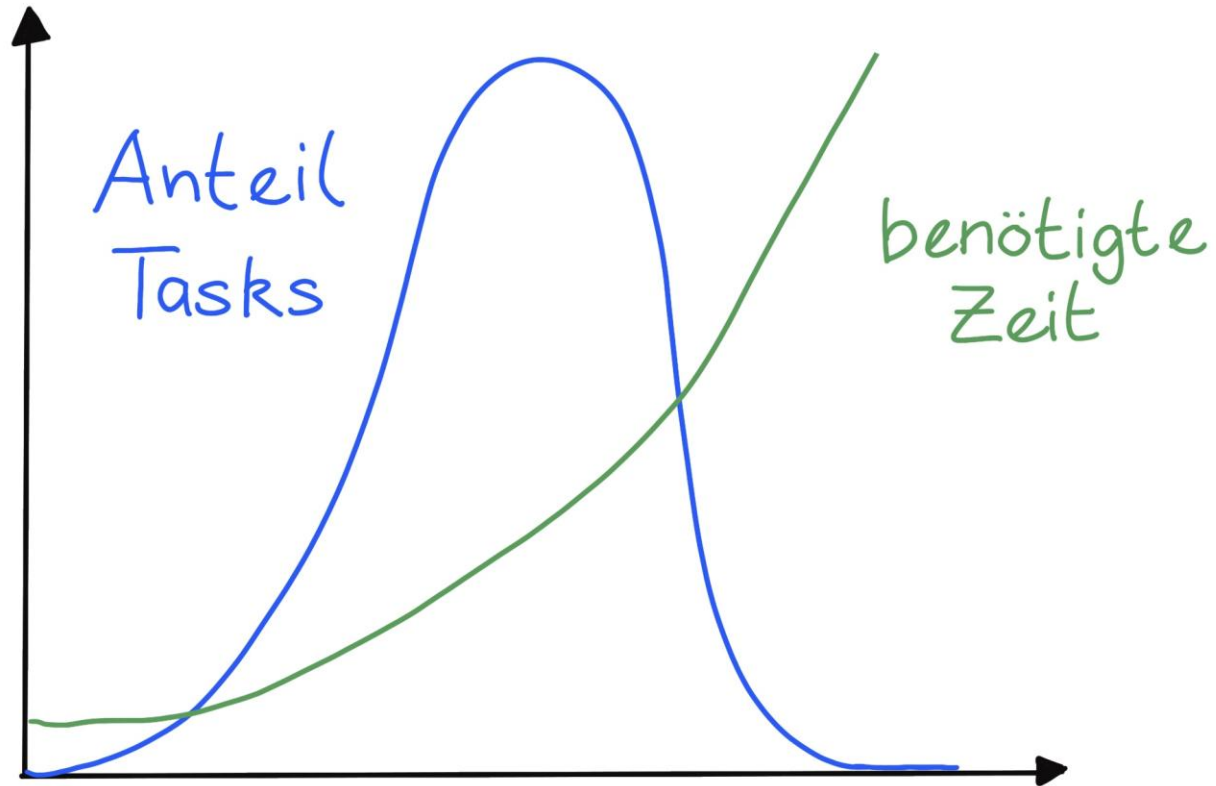
Sanjna Parulekar, senior vice president
of product marketing at Salesforce:

„We all had more trust in the LLM a year ago.“

Probleme

- Zu viele Anweisungen auf einmal verwirrten die ~Agenten.
- Sie vergaßen die zuverlässige Durchführung von Anweisungen, z.B. allen Kund*innen eine Zufriedenheitsumfrage zu schicken.
- Die ~Agenten wussten nicht, wann billigere oder zuverlässigere Systeme eine Anweisung ausführen könnten.
- Die ~Agenten lassen sich zu leicht von den Kund*innen ablenken.

80:20 Aufteilung

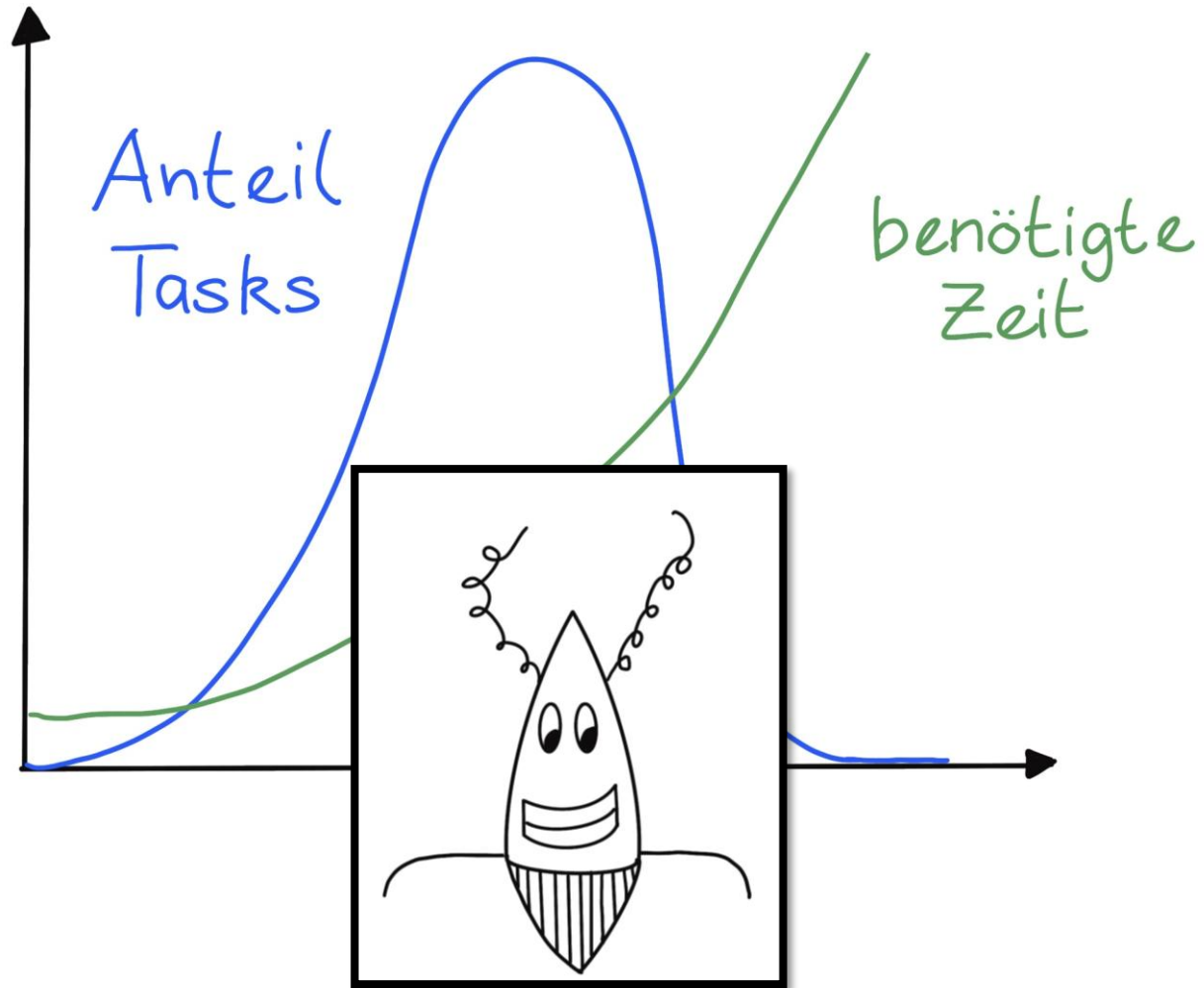


Kundenbeschwerden de luxe

AirBnB-Co-Gründer und CEO Brian Chesky im Gespräch

<https://www.instagram.com/reel/DSkWwVsE3mQ/?igsh=MXhtNW8wdjVpYTR6eQ==>

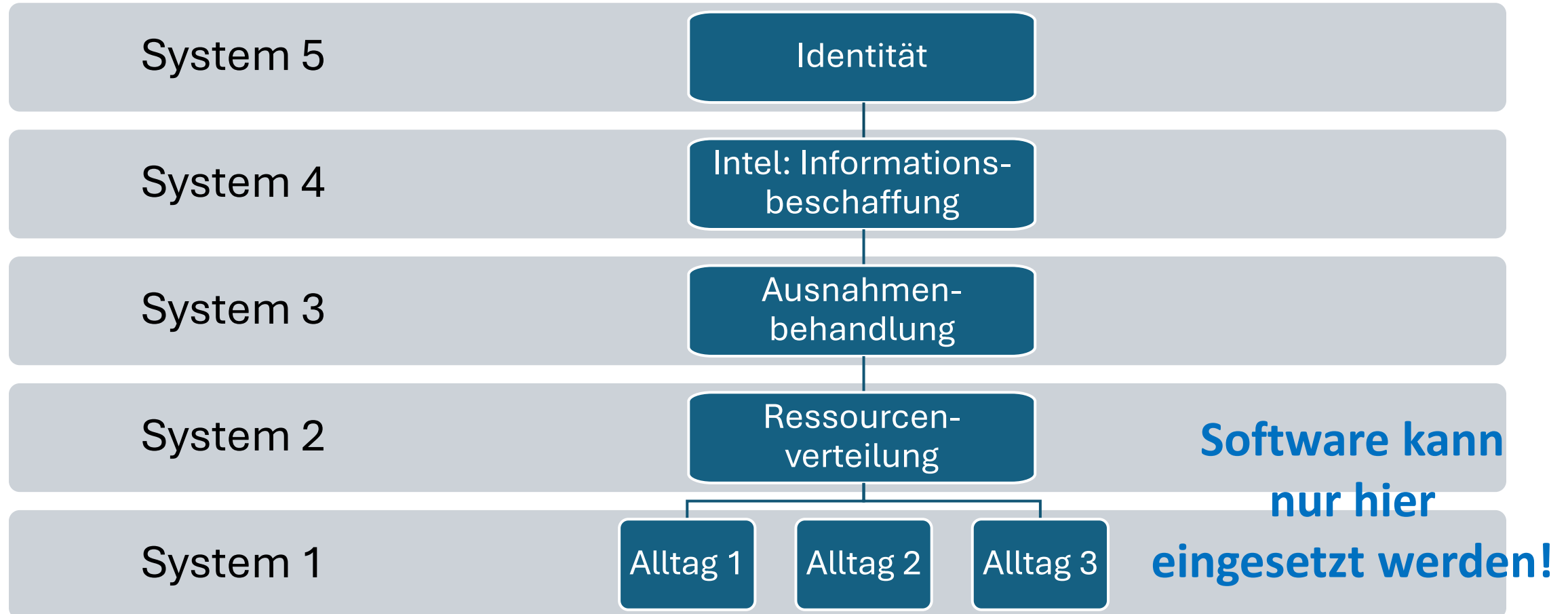
KI und Ausnahmen



Verluste durch Nacharbeiten

„Nearly 40% of AI Time Savings Are Lost to Fixing Low-Quality Output“

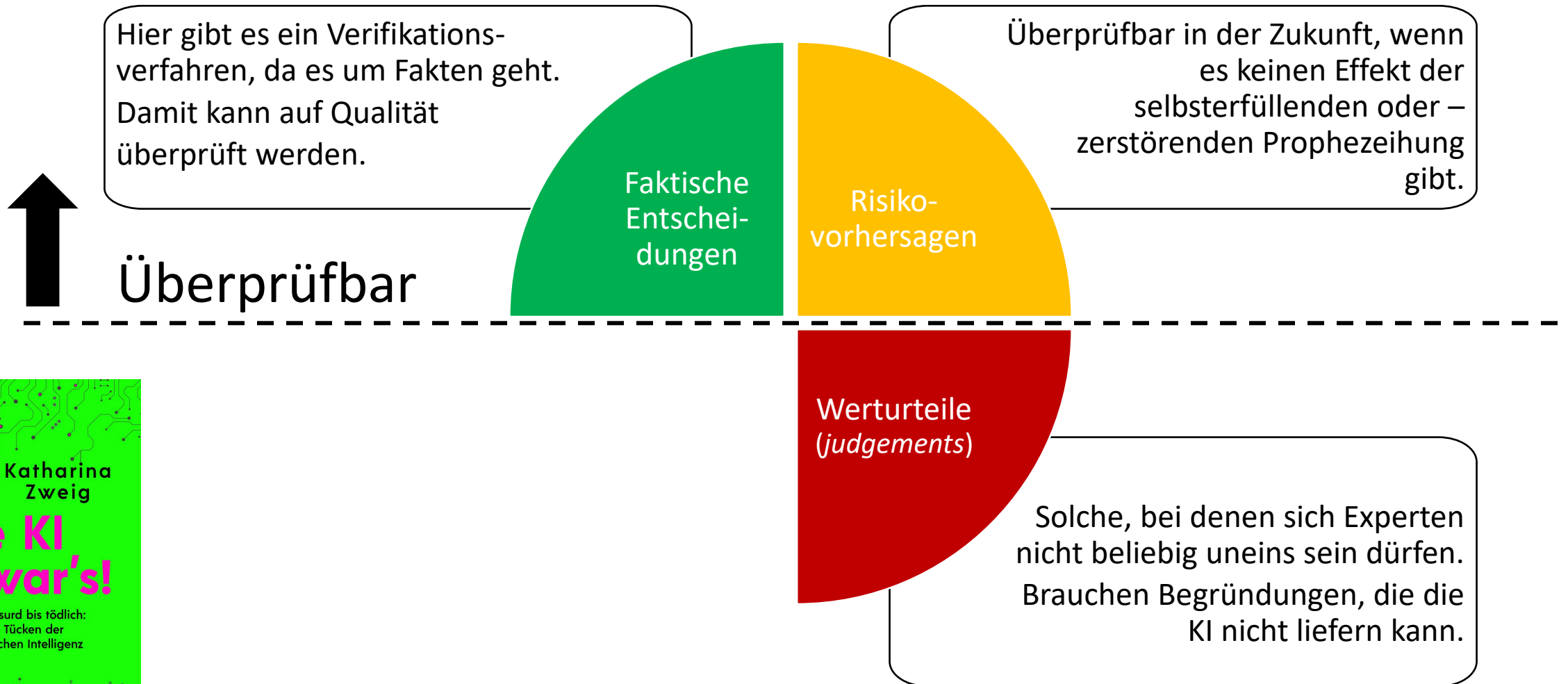
Was kann Software?



Maschinell gelernte Software-Systeme

- Benötigen exzellente und enorm viele Trainingsdaten.
- Sie lernen Zusammenhänge und Regeln, die nicht kommunizierbar sind.
- Sie können dabei Zusammenhänge und Regeln identifizieren, die nicht korrekt sind und andere übersehen.
- Nur ihr Output kann überprüft werden.

Welche KI-Systeme sind überprüfbar auf Verlässlichkeit?



Vorarbeiten

- Welcher Prozess folgt gut beschreibbaren Regeln?
 - Wo werden welche vermutet, die noch nicht verschriftlicht wurden?
- Kann die Qualität des automatisierten Prozesses überprüft werden?
- Wie häufig sind Ausnahmen? Wieviel Zeit nehmen diese in Anspruch (vermutlich 30 % für regelhafte Fälle zu 70% Ausnahmefälle!)?
- Wie kann man das eine vom anderen zuverlässig trennen?
- Wie teuer sind Nacharbeiten?
- Lohnt sich die Anschaffung / Erhaltung eines KI-(Agenten)-Systems für die wahrscheinliche Einsparung?

Prozess für Identifikation und Nutzung von KI-Systemen in Entscheidungsprozessen

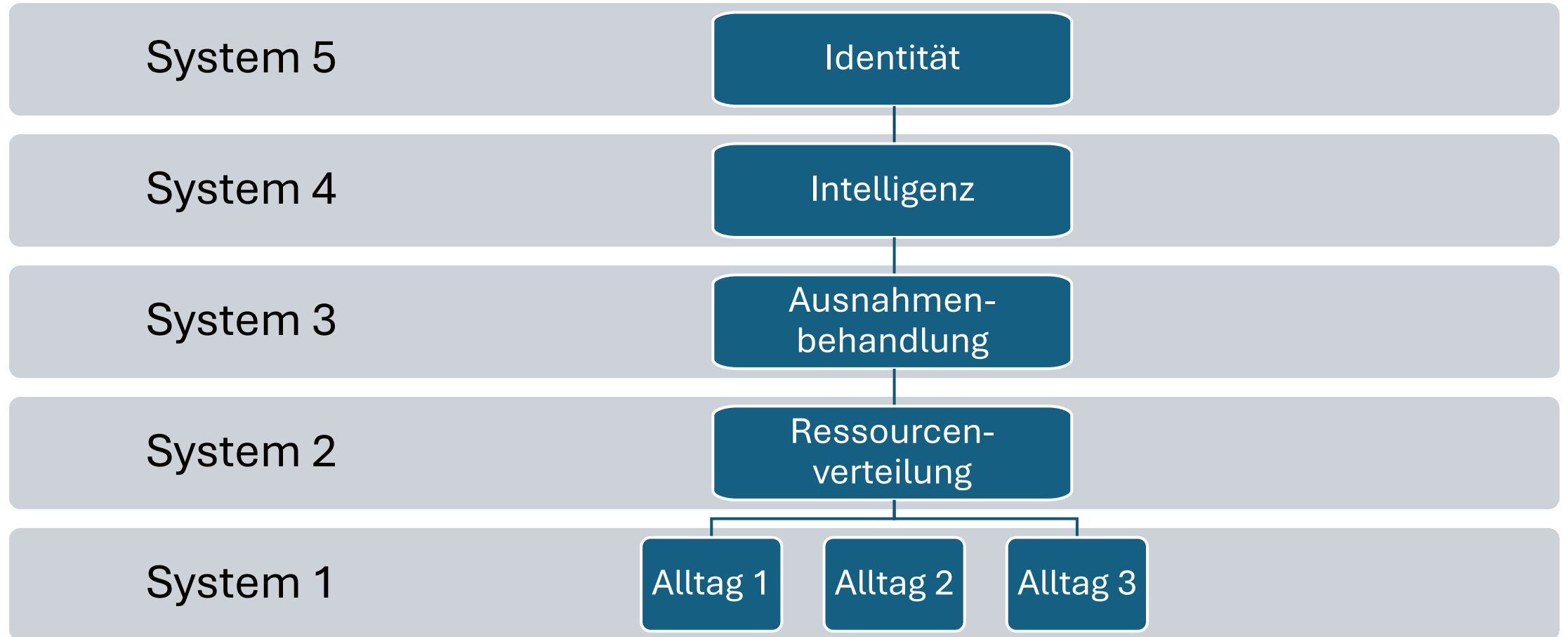
1. **Problemspezifikation:** Was genau soll denn gelöst werden?
2. Unter welchen **Nebenbedingungen** (Geld, Anzahl Personen, Qualität, ...)?
3. **Technikneutrale Suche** nach einer ökonomischen Lösung...
4. ...unter **Einbindung der betroffenen Personen (!)** im sozioinformatischen Gesamtsystem.
5. **Implementierung.**
6. **Change Management** für Prozessanpassung.
7. Regelmäßige **Qualitätssicherung.**



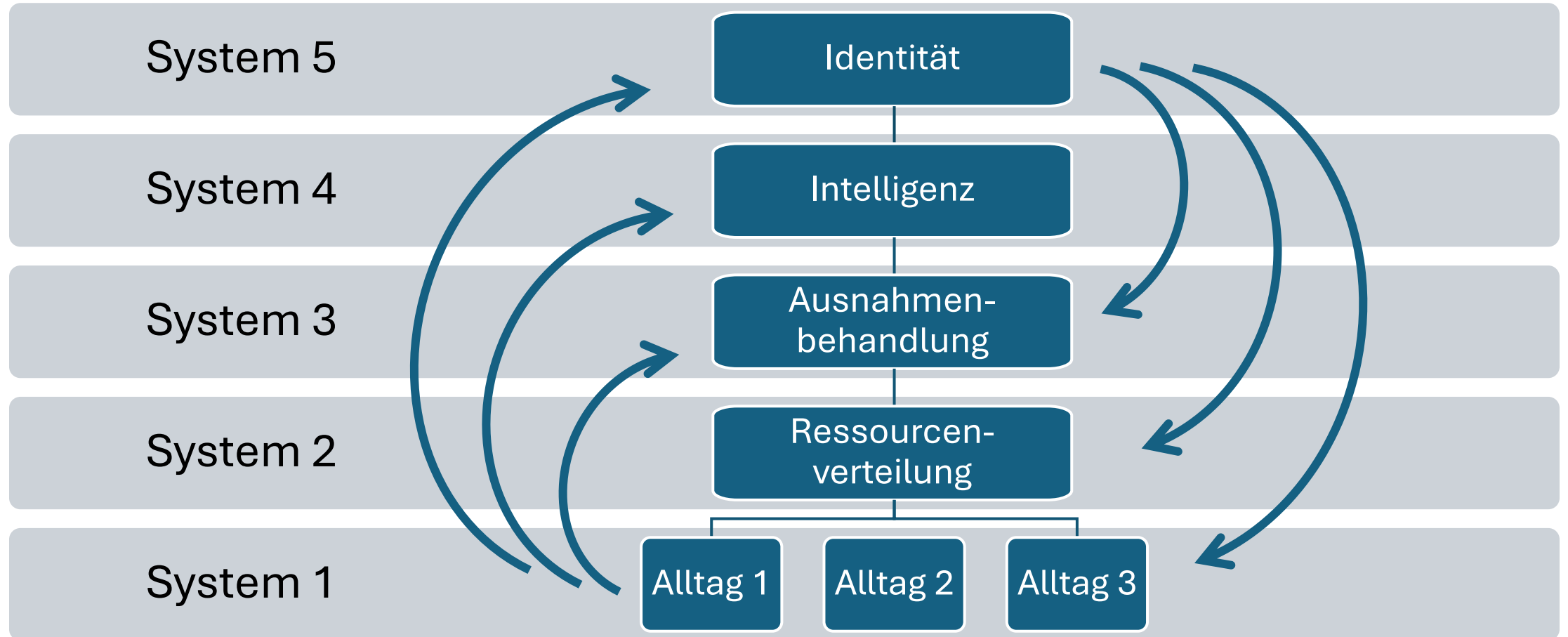
Interessanter Weg in die Zukunft

- Wo können Mensch und Maschine miteinander verwoben werden?
- Wo wird die Unzuverlässigkeit des einen durch den anderen ausgeglichen?
- Wo können Menschen ihre eigenen Kompetenzen mit Hilfe von LLMs verbessern?

Wann ist eine Firma / Branche überlebensfähig?



Feedback



Zusammenfassung Sprachmodelle

- Maschinen können heute weder nachdenken, noch analysieren, noch moralische Fragen bewerten.
- Aber sie können Texte assoziieren, die aussehen wie das Produkt von Nachdenken, Analyse oder moralischer Bewertung. Das sind “intelligible Texturen” (Schneider, 2024)
- Mit heutigen Methoden werden Maschinen dazu auch nicht in der Lage sein.
- Ihnen fehlt die Eingebundenheit in soziale Prozesse.

Zusammenfassung Nutzung von KI

- KI beschreibt eine große Menge von Methoden, mit denen Computer Dinge tun können, die beim Menschen Intelligenz erfordern.
- Es ist schwer nachzuvollziehen, wie eine Maschine zu Entscheidungen kommt.
- Daher kann sie nur dort eingesetzt werden, wo Entscheidungsqualität nicht sehr relevant ist oder wo diese überprüft werden kann.
- KI-Systeme müssen sorgfältig erstellt werden: Gute Trainingsdaten, die für den Fall relevant sind.

Digitalisierung
benötigt
nicht nur
plappernde
Papageien.

Sie benötigt
weiterhin
kluge Köpfe.



Prof. Dr. Katharina Zweig
RPTU Kaiserslautern-Landau

