

Qualität sicher, Nutzen maximieren: Der Weg zur erfolgreichen Produktintegration von LLMs

Patrick Stiller, tracetronic
Ron Martin, Fraunhofer IIS

Tracetronic GmbH

- **Founded 2004**
- **Location:** Dresden, Ingolstadt, Munich, Möglingen, Wolfsburg, Hamburg, Soul, Shanghei, Michigan and Yokohama
- Over 400 employees
- Joint Venture Neocx with Volkswagen since 2021



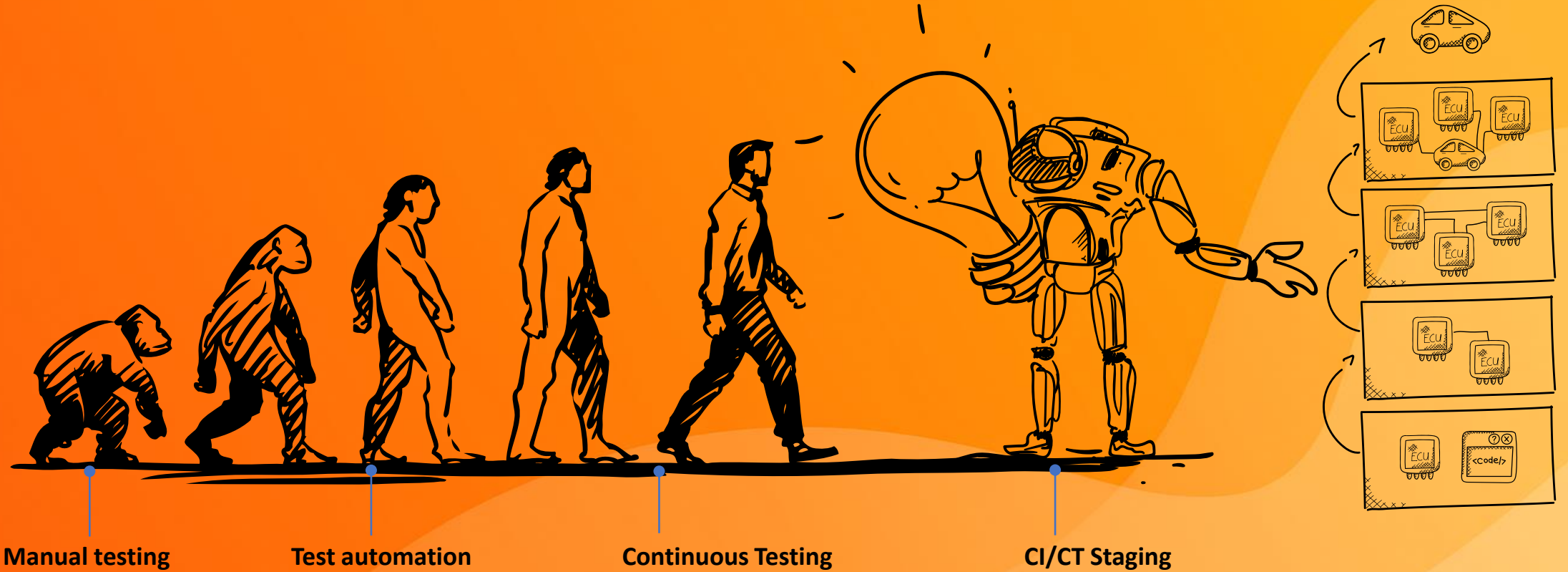
Customers worldwide

A world map is centered on the slide, with blue dots and lines indicating the presence of major OEMs in North America, Europe, and Asia. Surrounding the map are logos for these OEMs, organized by region:

- North America:** Ford, GARMIN, POLARIS, JOHN DEERE, BRP, RIVIAN, MAGNA, STELLANTIS, LUCID, DAIMLER TRUCK, BorgWarner, CANOO.
- Europe:** Mercedes-Benz, VW, Audi, BOSCH, Continental, AMG, CLAAS, MAHLE, ZE, iav, Valeo, AVL, Ferrari, Porsche.
- Asia:** Toyota, DENSO, HITACHI, AISIN, NAVIS-AMS, CHANGAN, GAC MOTOR, WEICHAI, CHERY, Li Auto, X P E N G, DFY, FAW, GEELY, UAES (United Automotive Electronics Co., Ltd.), NIO, HUAWEI, BYD.

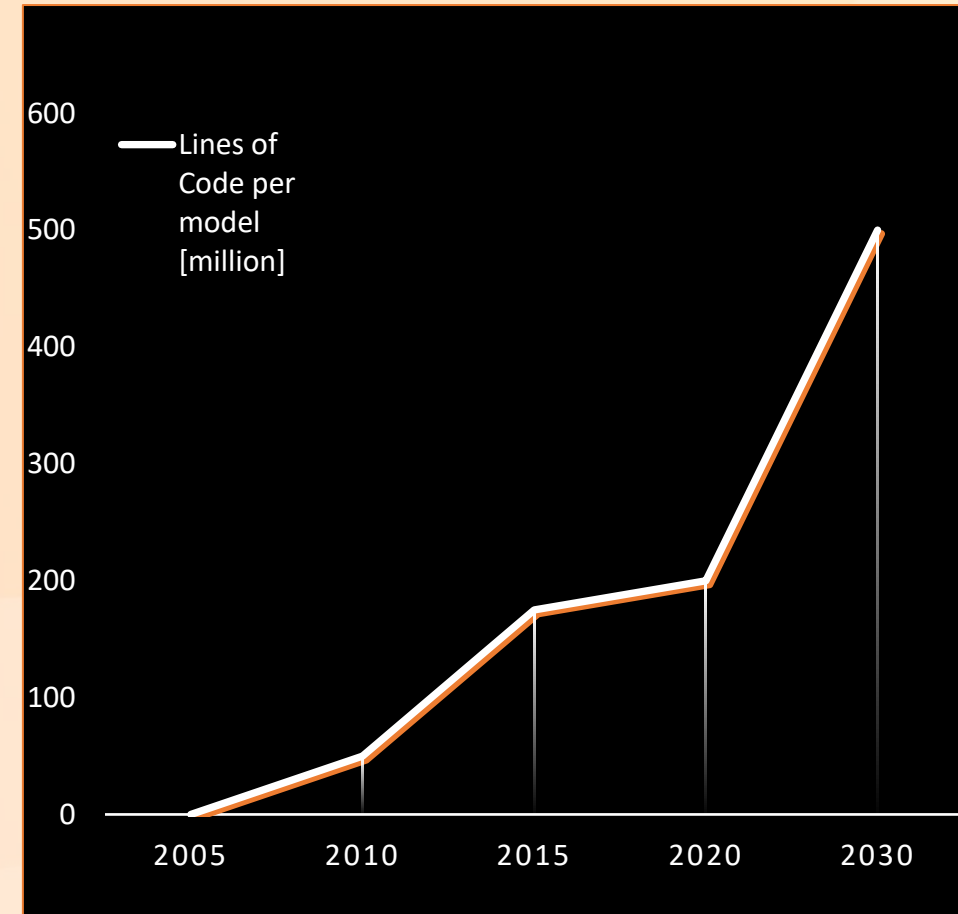
” The most major OEMs rely on us ”

Evolution of test automation

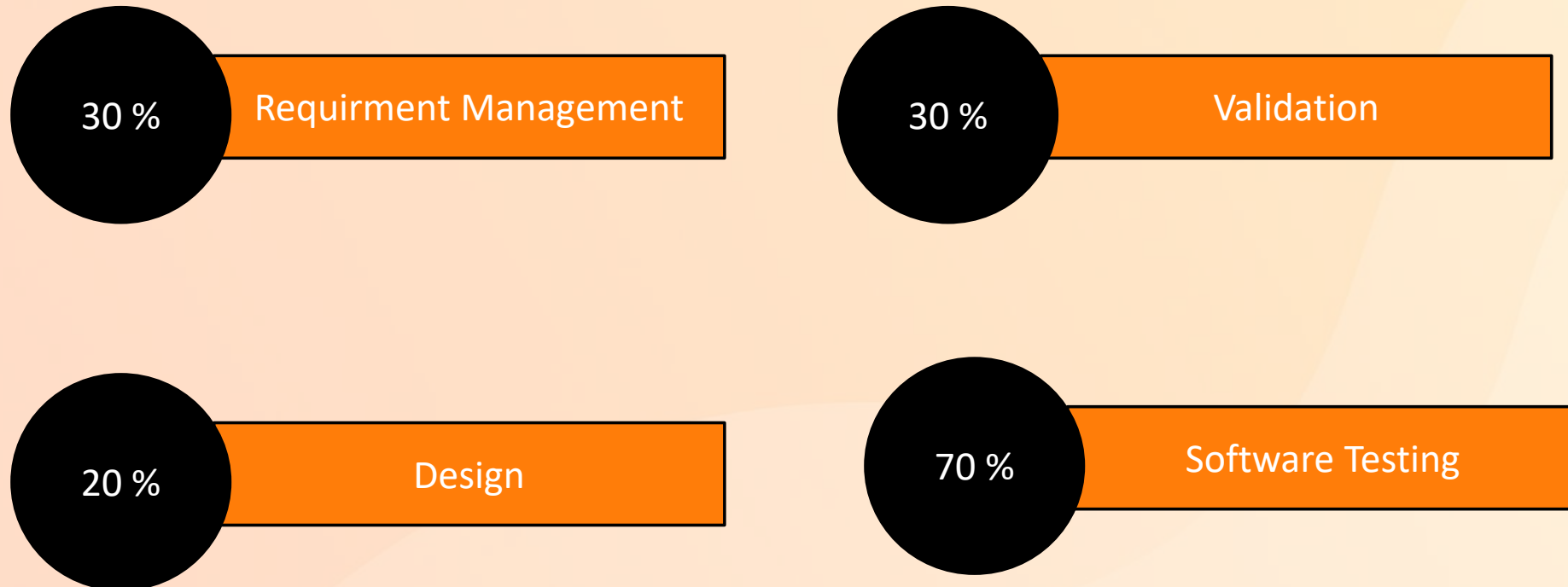


Challenges for Software Defined Vehicles

- Massive Scaling of Software volume in the automotive sector
- Level 5 autonomous driving will take up to 1 billion lines of code
- Massive scale of software tests



Potential of Generative AI in Software Testing – Mc Kinsey et.al 2024



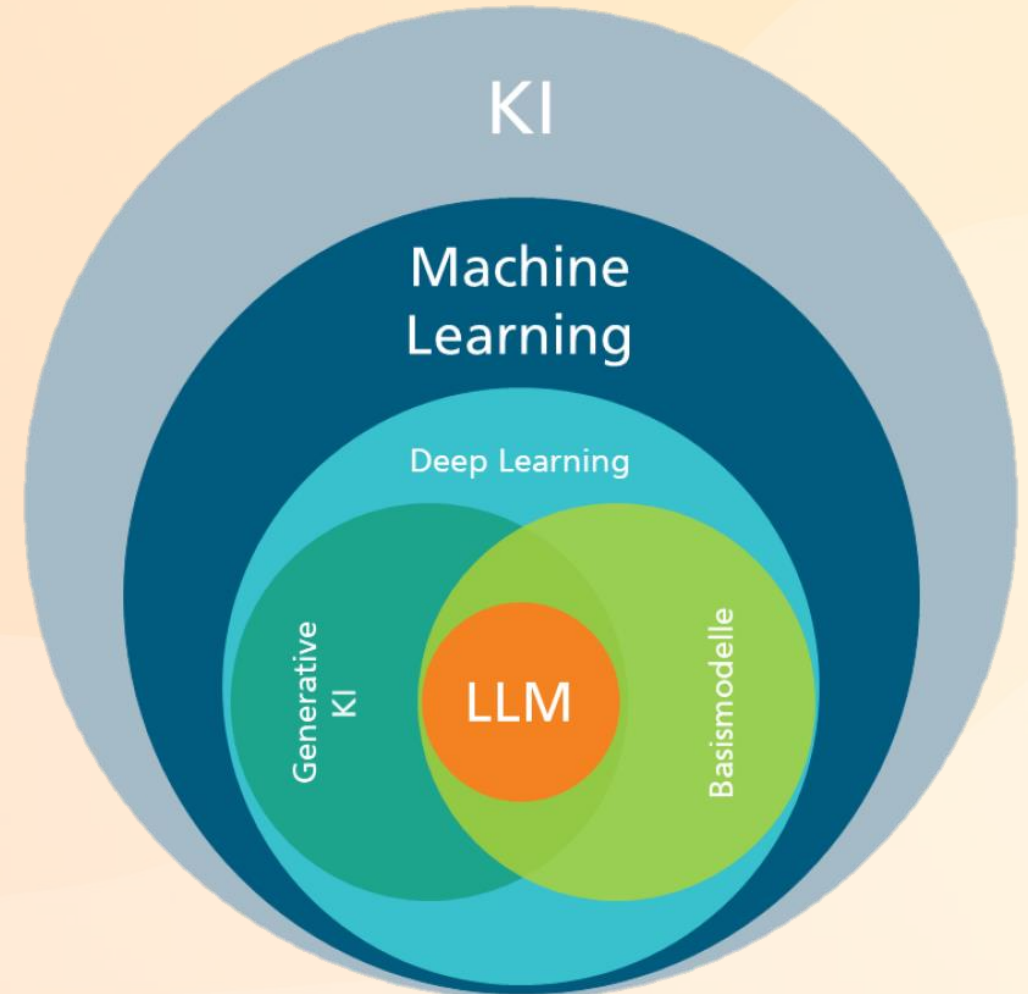
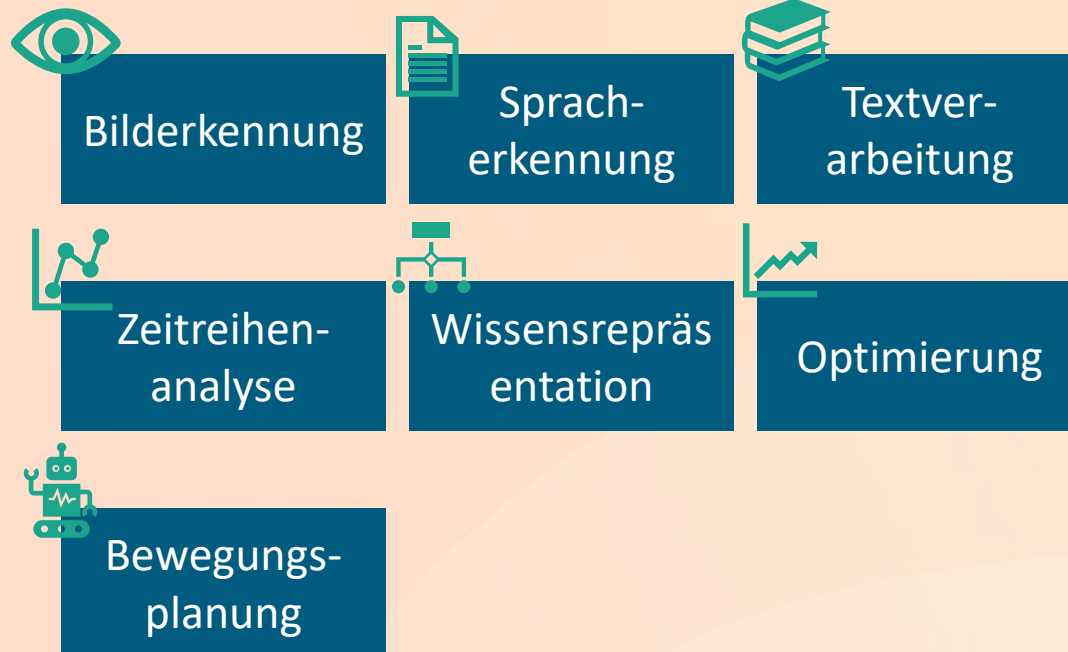
KI Recap

Follow this QR-Code for more Info

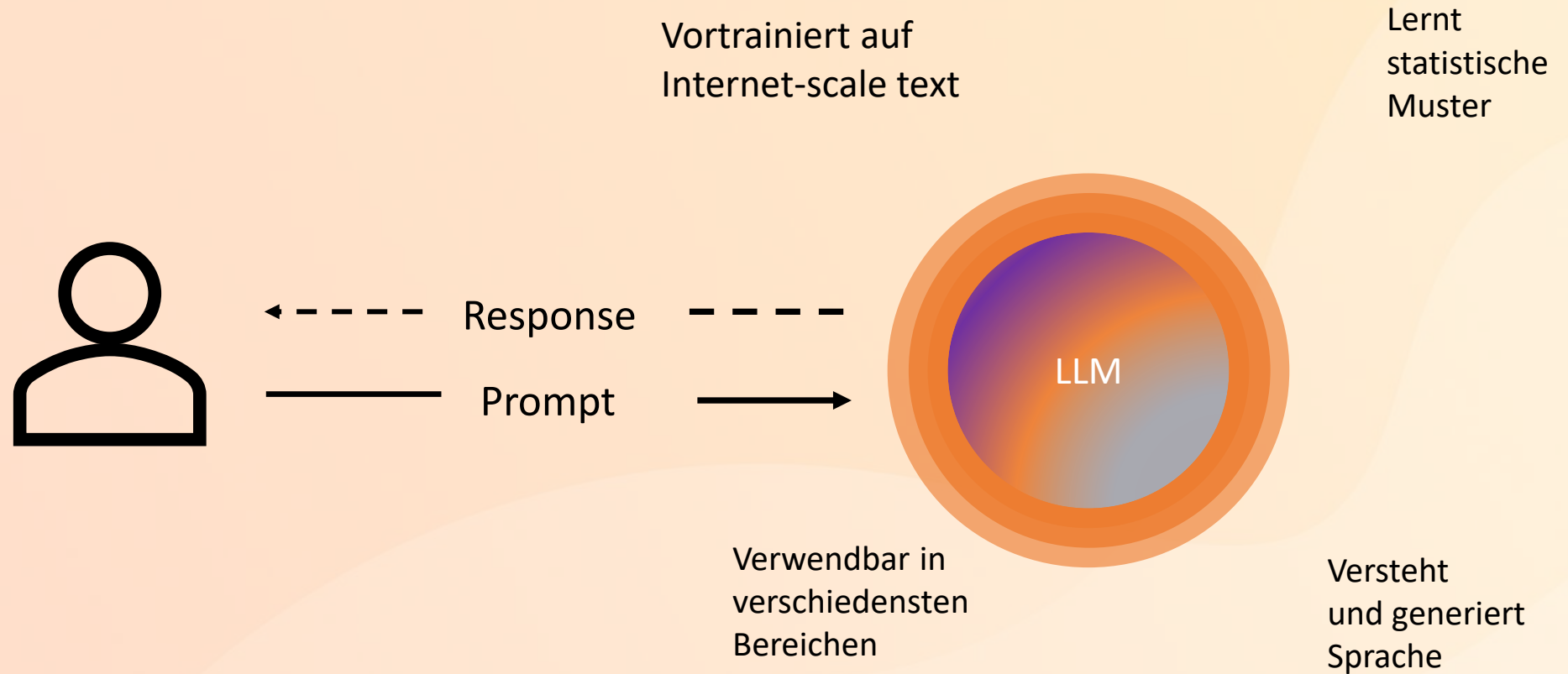
[Whitepaper: Schritt für Schritt zur eigenen KI-Lösung – Ein Leitfaden für KMU – ATKI Dresden](#)



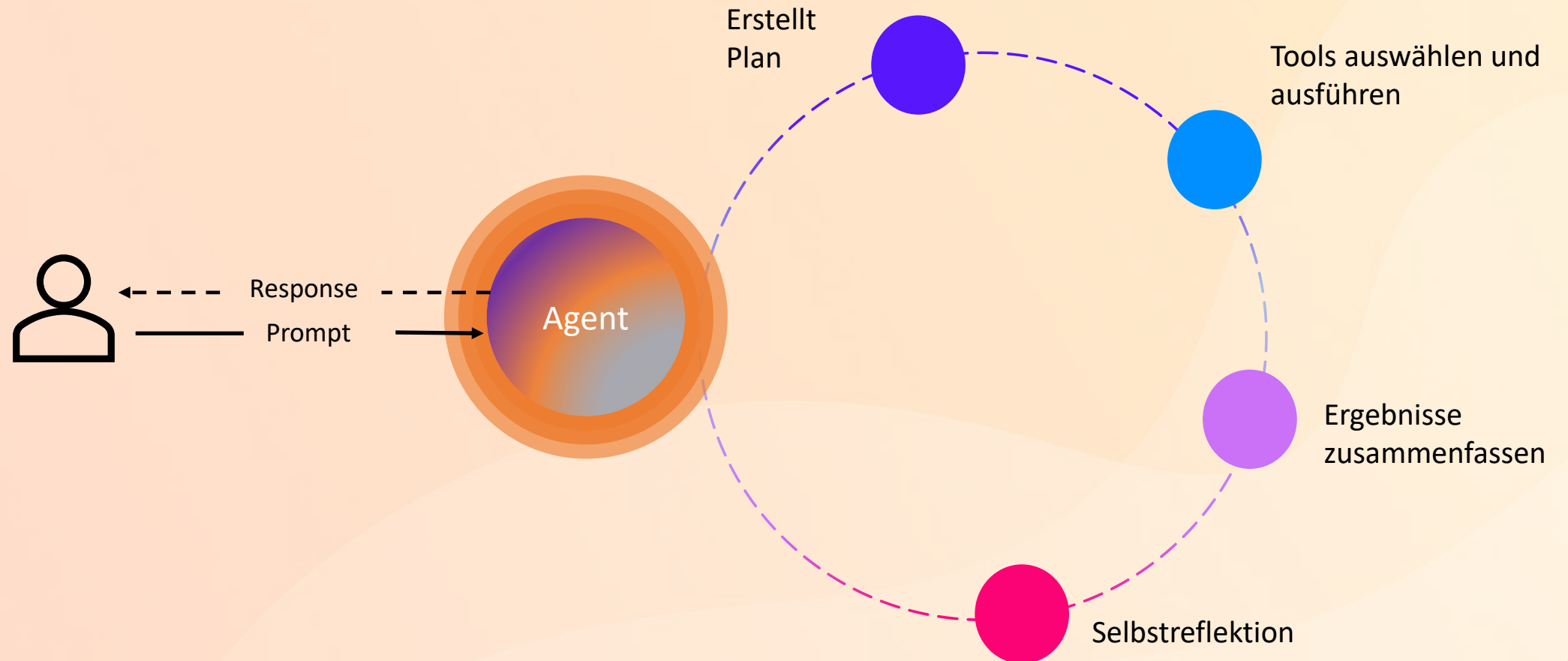
- KI ≠ ChatGPT
- Technologie die versucht kognitive Fähigkeiten des Menschen nachzubilden



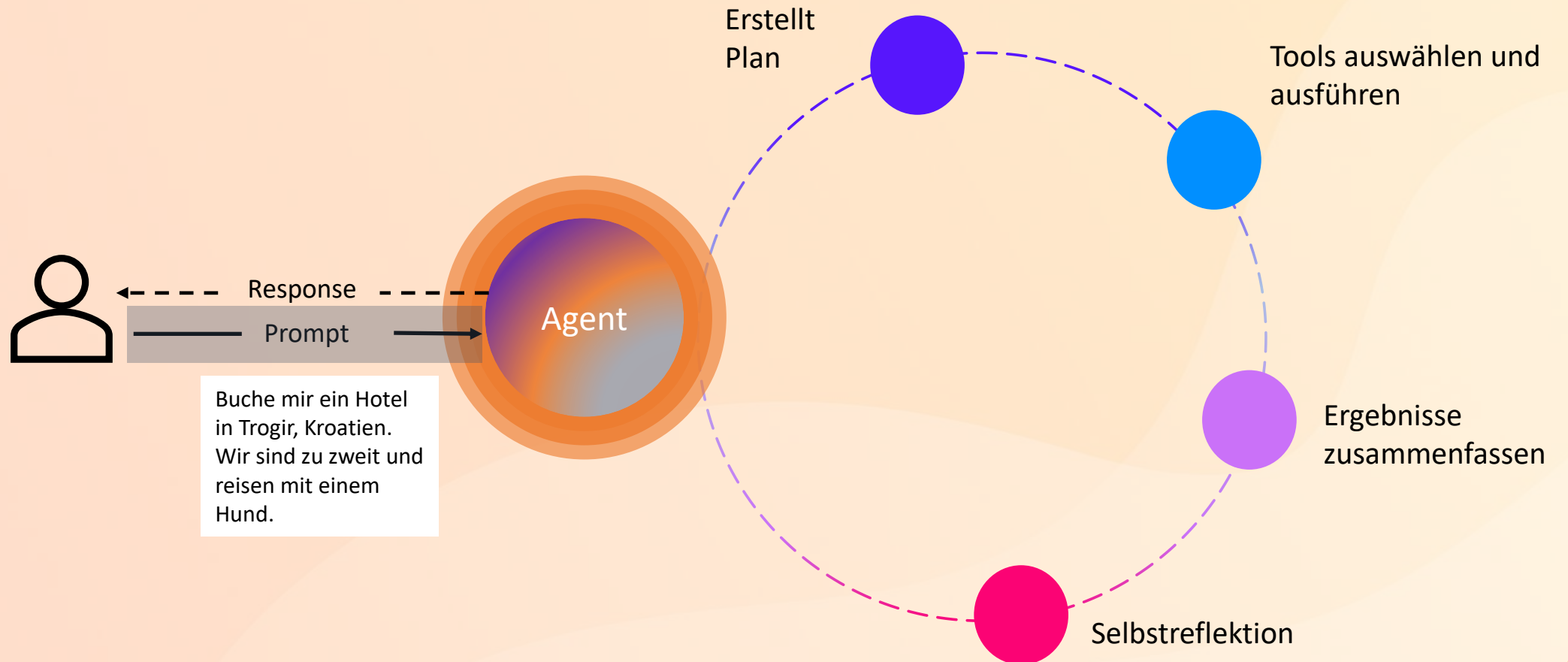
LLM Recap



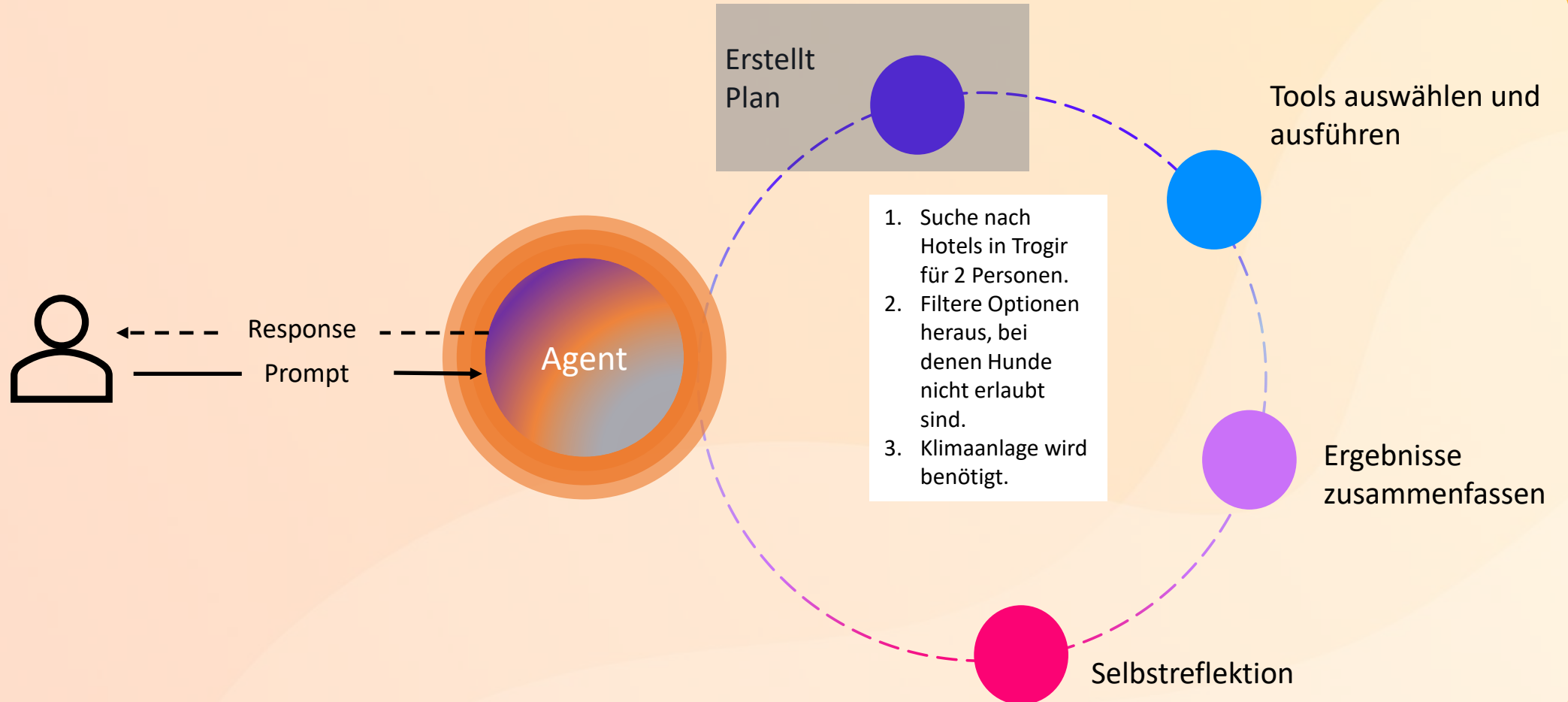
KI Agenten



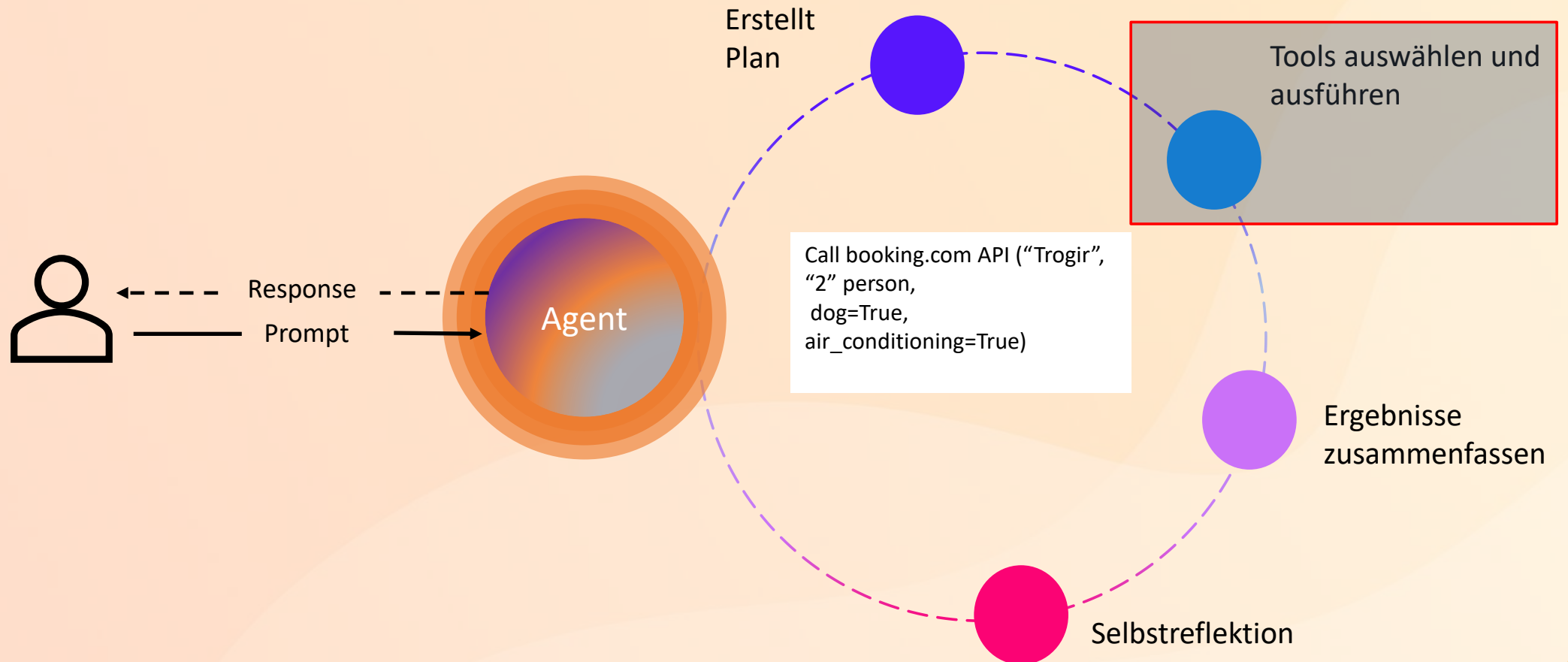
Beispiel: Buch mir eine Urlaubsreise



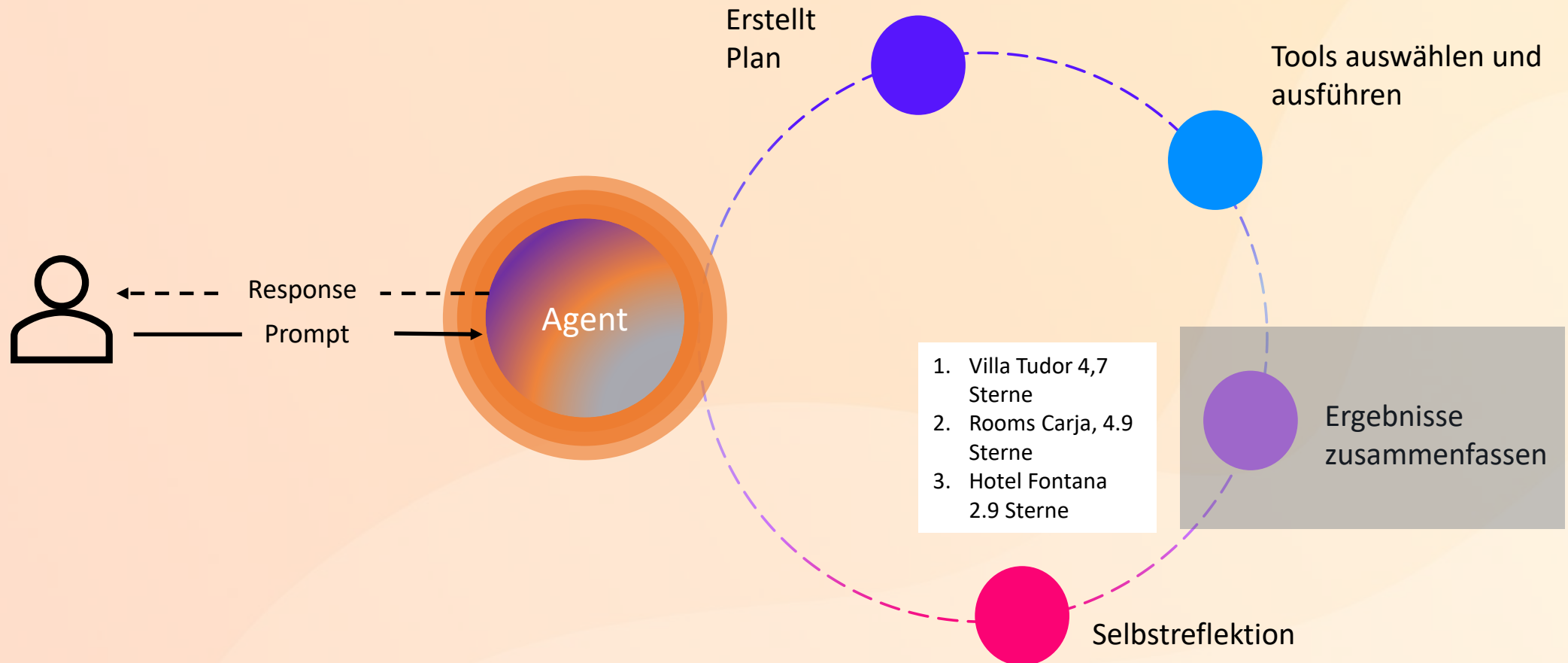
Beispiel: Buch mir eine Urlaubsreise



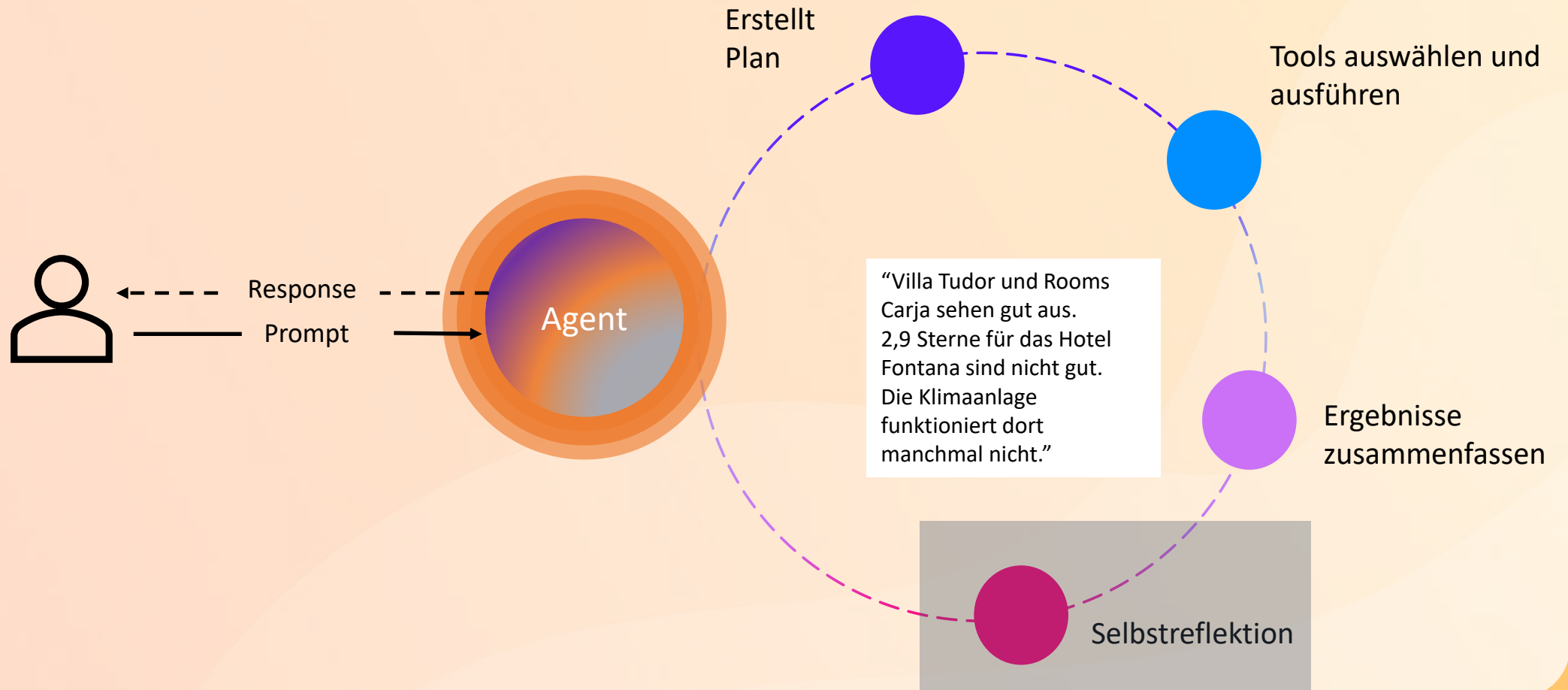
Beispiel: Buch mir eine Urlaubsreise



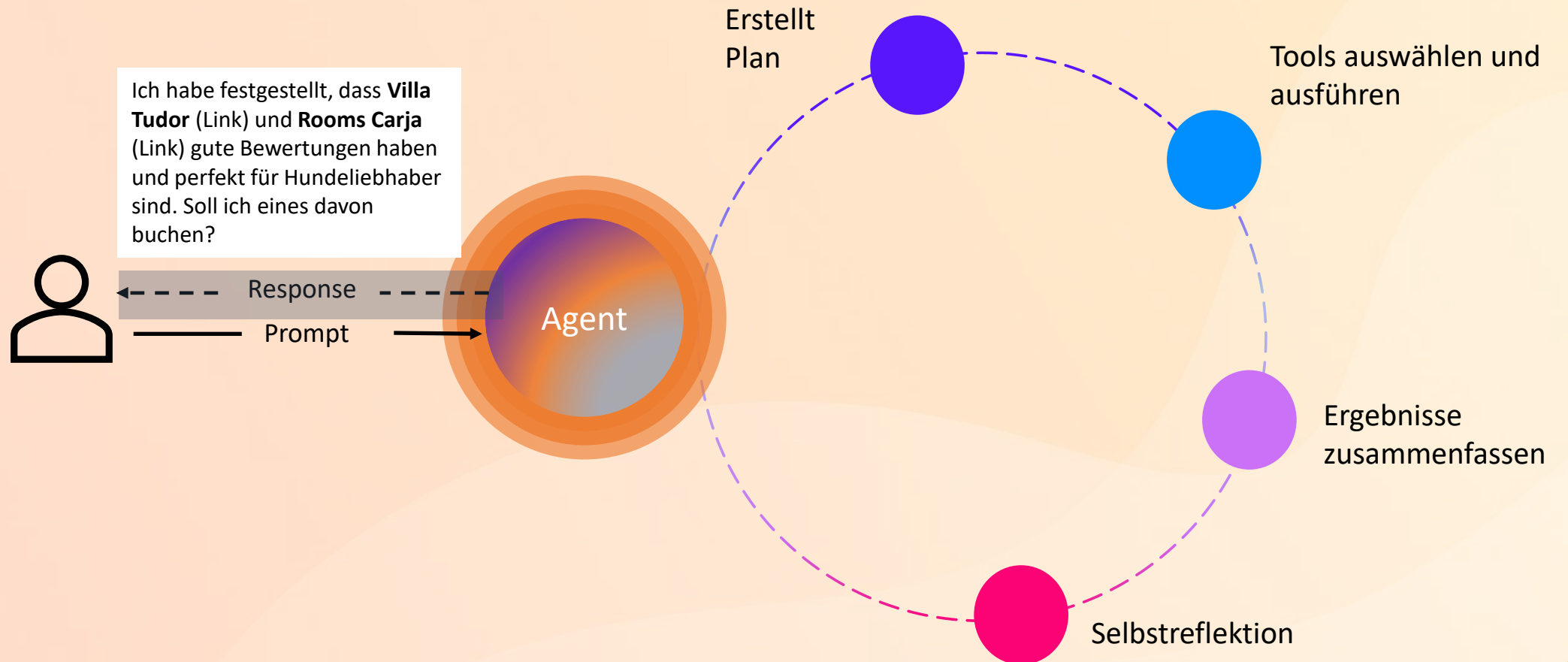
Beispiel: Buch mir eine Urlaubsreise



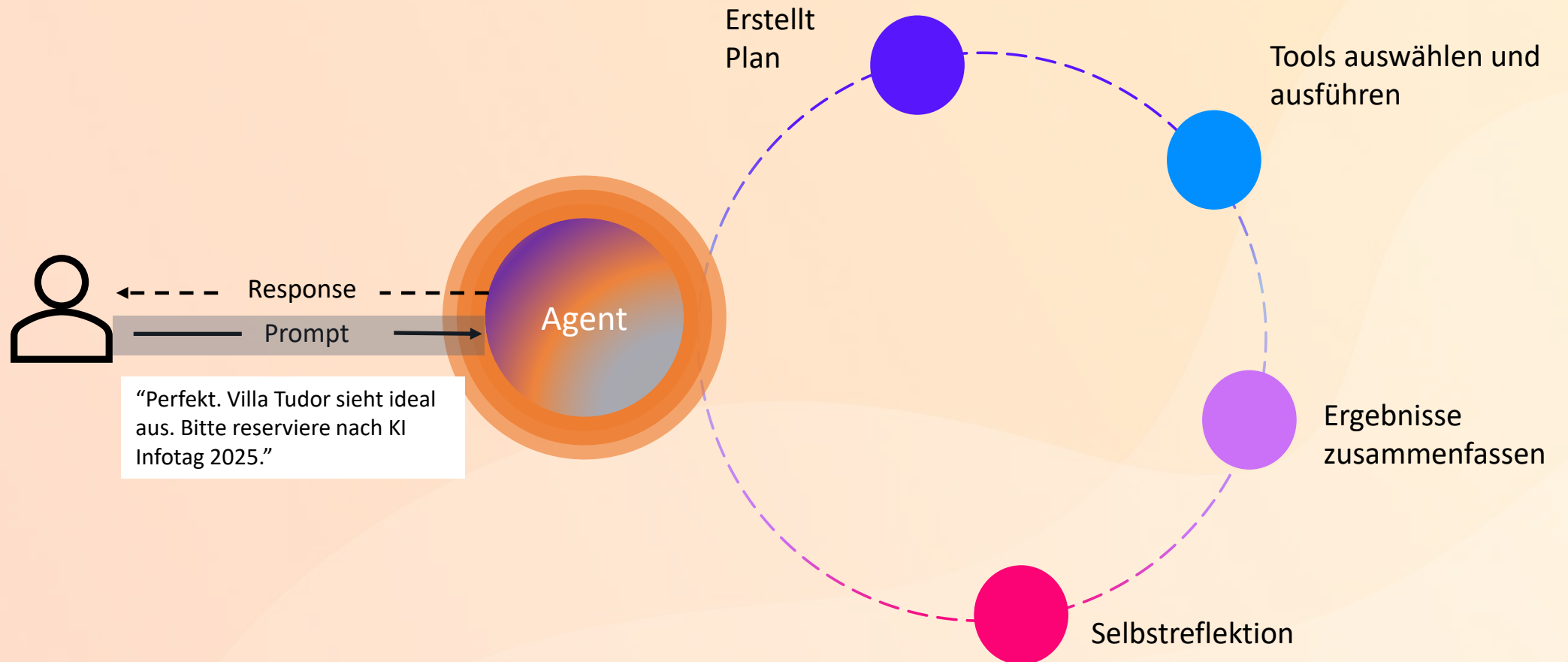
Beispiel: Buch mir eine Urlaubsreise



Beispiel: Buch mir eine Urlaubsreise



Beispiel: Buch mir eine Urlaubsreise



“ecu.test Agent ist dein **Copilot** für die Implementierung von Testfällen.” **ecu.test** *agent*

Aktion

Accelerate the car.
Gaspedal loslassen.

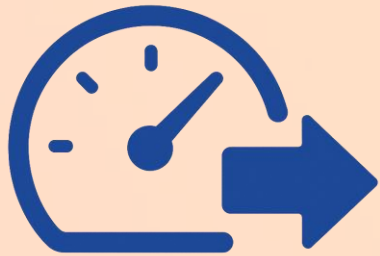
Erwartung

Rekuperation ist
aktiv.

- Name
- BatteryHillAssist
 - Brake
 - BrakeLight
 - CruiseControlSpeedSwitch
 - CruiseControlStatus
 - CruiseControlSwitch
 - DriveModeSelect
 - DriveMode_E
 - DriverRequest
 - HandBrakeStatus
 - HandBrakeSwitch
 - Headunit
 - HillGradient
 - RecuperationECU
 - ReverseLight
 - SOC
 - SetVehTrStby
 - SoH
 - StarterButton
 - VehicleTractionStandby
 - VelocityBus
 - VelocityECU
 - VelocityModel

#	Action / Name	Expectation / Value
1	Precondition	
2	check state of charge	between 20 % and 90 %
3	press the starter button	
4	MODEL-Write: StarterButton	1.0
5	Action	
6	set vehicle traction standby to switch on	
7	Set driver request to DrvRq	
8	check if speed is faster 50	all signals are larger 50
9	Set driver request to 0 %	
10	Check recuperation 'on'	
11	Postcondition	
12	Stop vehicle	
13	Switch to "P" (0) and unestablish VehTrStby	

Mehrwerte für tracetronic



Beschleunigung
um 30 %



Mehr
Konsistenz



Kunden
weltweit

Unsere gemeinsame LLM-Reise

Der Weg zur robusten KI

Follow this QR-Code for more Info

[Whitepaper: Schritt für Schritt zur eigenen KI-Lösung – Ein Leitfaden für KMU – ATKI Dresden](#)

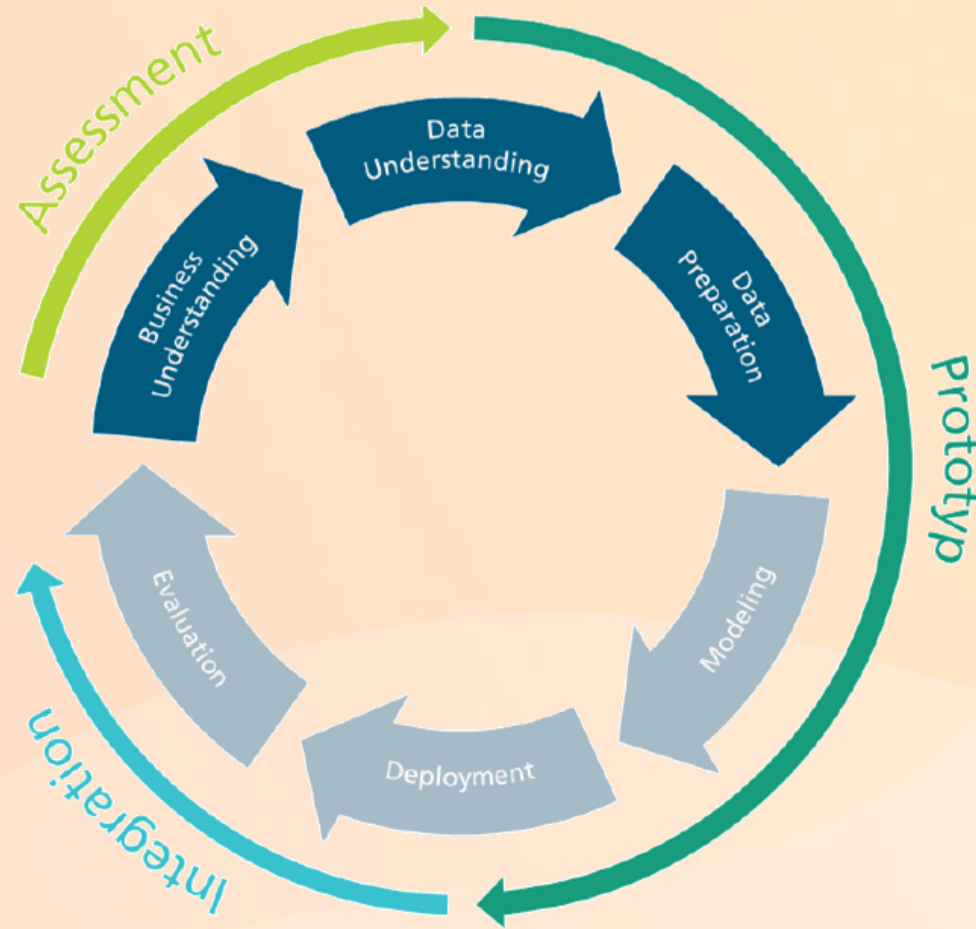
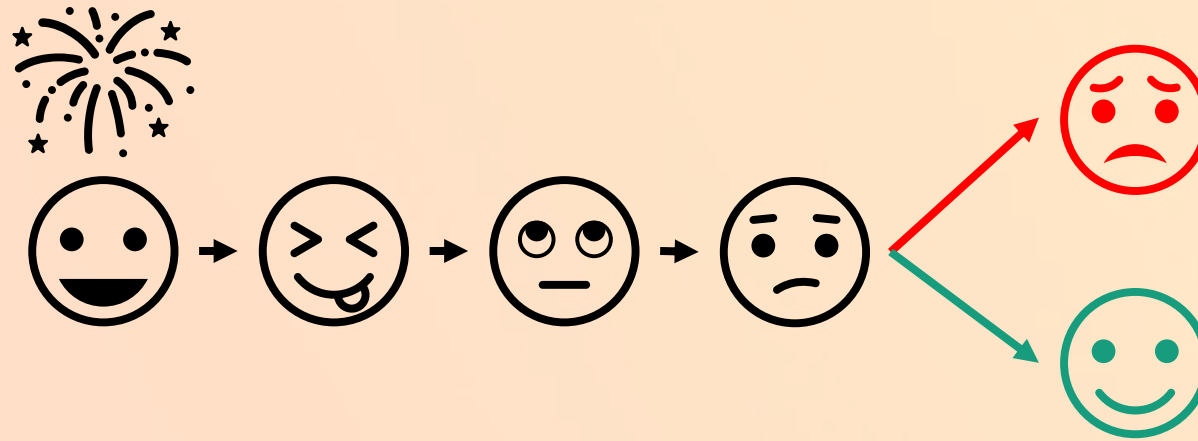


Abb.4: Von der KI-Idee zur produktiven KI-Lösung (Grundlage)
© C. Shearer & Fraunhofer IIS

Die LLM Gefühlsreise



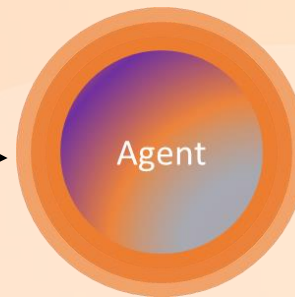
→ Wie machen wir das Gefühl der Qualität messbar?

Qualität messbar machen

- Allgemeine LLM-Benchmarks helfen hier nicht
→ Es muss für diesen einen Usecase funktionieren
- Menschliche Prüfung ist aufwendig (& teuer)
- Automatische Qualitätsmessung ist nicht trivial

Buche mir ein Hotel in Trogir,
Kroatien.
Wir sind zu zweit und reisen mit
einem Hund.

Prompt →



Response - - - - - →

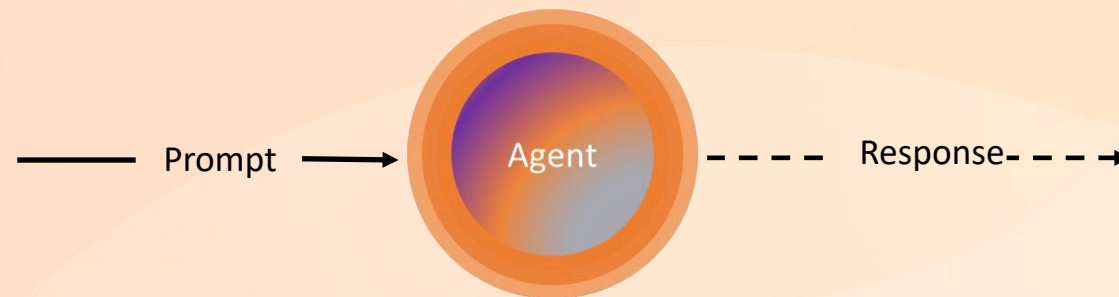
Korrekt

Ich habe festgestellt, dass **Villa Tudor** (Link) und **Rooms Carja** (Link) gute Bewertungen haben und perfekt für Hundeliebhaber sind. Soll ich eines davon buchen?

Qualität messbar machen

- Allgemeine LLM-Benchmarks helfen hier nicht
→ Es muss für diesen einen Usecase funktionieren
- Menschliche Prüfung ist aufwendig (& teuer)
- Automatische Qualitätsmessung ist nicht trivial

Buche mir ein Hotel in Trogir,
Kroatien.
Wir sind zu zweit und reisen mit
einem Hund.



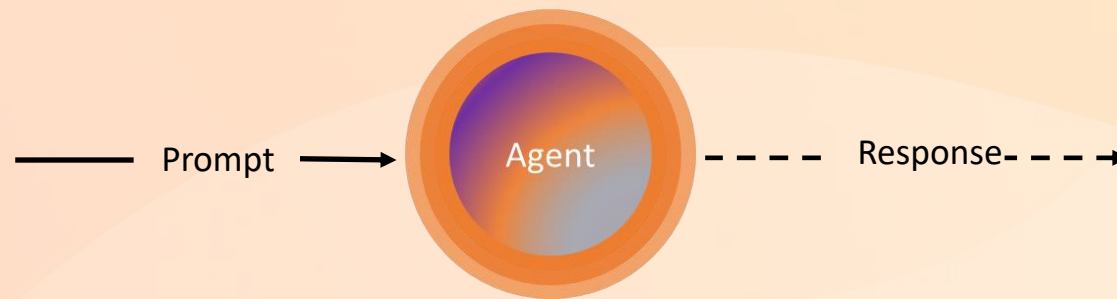
Fakten falsch

Ich habe das Hotel ‚Seaview Palace‘ in Trogir ausgewählt – es bietet einen Hundesitter, gratis Helikopter-Shuttle zum Strand und kostet nur 20 Euro pro Nacht.

Qualität messbar machen

- Allgemeine LLM-Benchmarks helfen hier nicht
→ Es muss für diesen einen Usecase funktionieren
- Menschliche Prüfung ist aufwendig (& teuer)
- Automatische Qualitätsmessung ist nicht trivial

Buche mir ein Hotel in Trogir,
Kroatien.
Wir sind zu zweit und reisen mit
einem Hund.



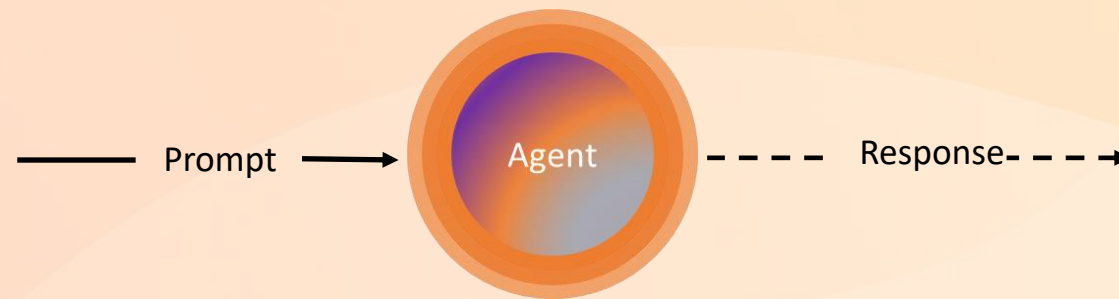
Unleserlich

Hotels hier Trogir zwei
Menschen plus Hund Reise mit
– ich sehe Villa Tudor link und
Rooms Carja link Bewertungen
sind gut perfekt Hund
Liebhaber. Wollen Sie buchen?

Qualität messbar machen

- Allgemeine LLM-Benchmarks helfen hier nicht
→ Es muss für diesen einen Usecase funktionieren
- Menschliche Prüfung ist aufwendig (& teuer)
- Automatische Qualitätsmessung ist nicht trivial

Buche mir ein Hotel in Trogir,
Kroatien.
Wir sind zu zweit und reisen mit
einem Hund.



Irrelevant

Hier sind ein paar generelle
Reise-Tipps für Kroatien:
Besuche Dubrovnik, genieße die
Altstadt von Split und probiere
Cevapcici

Qualität messbar machen

- Allgemeine LLM-Benchmarks helfen hier nicht
→ Es muss für diesen einen Usecase funktionieren
- Menschliche Prüfung ist aufwendig (& teuer)
- Automatische Qualitätsmessung ist nicht trivial

Buche mir ein Hotel in Trogir,
Kroatien.
Wir sind zu zweit und reisen mit
einem Hund.

Prompt →

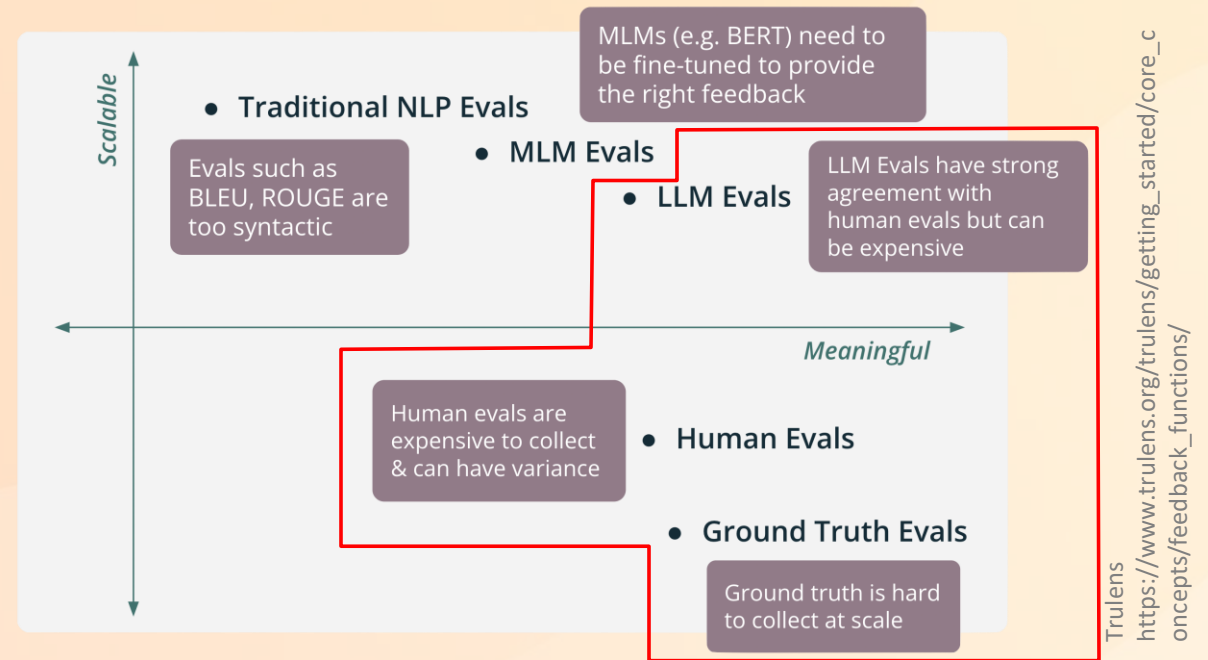
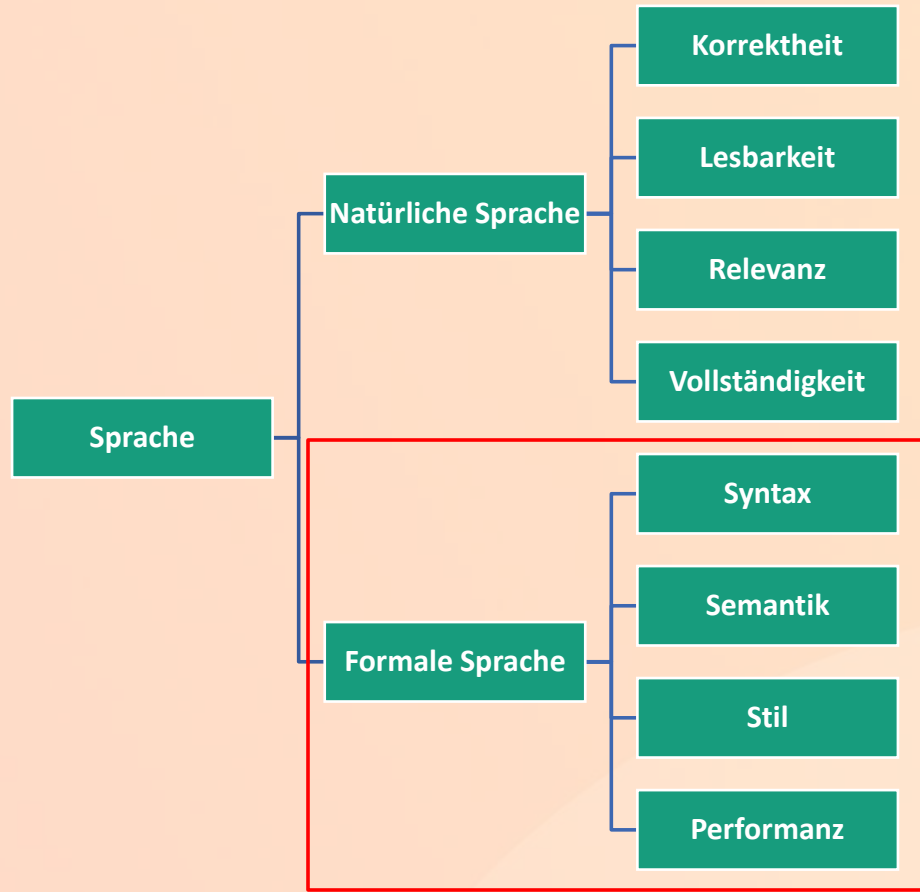
Agent

Response - - - →

Unvollständig

Ich habe ein Hotel in Trogir
gefunden: Villa Tudor (Link).
Möchten Sie buchen?

Qualitätsmetriken

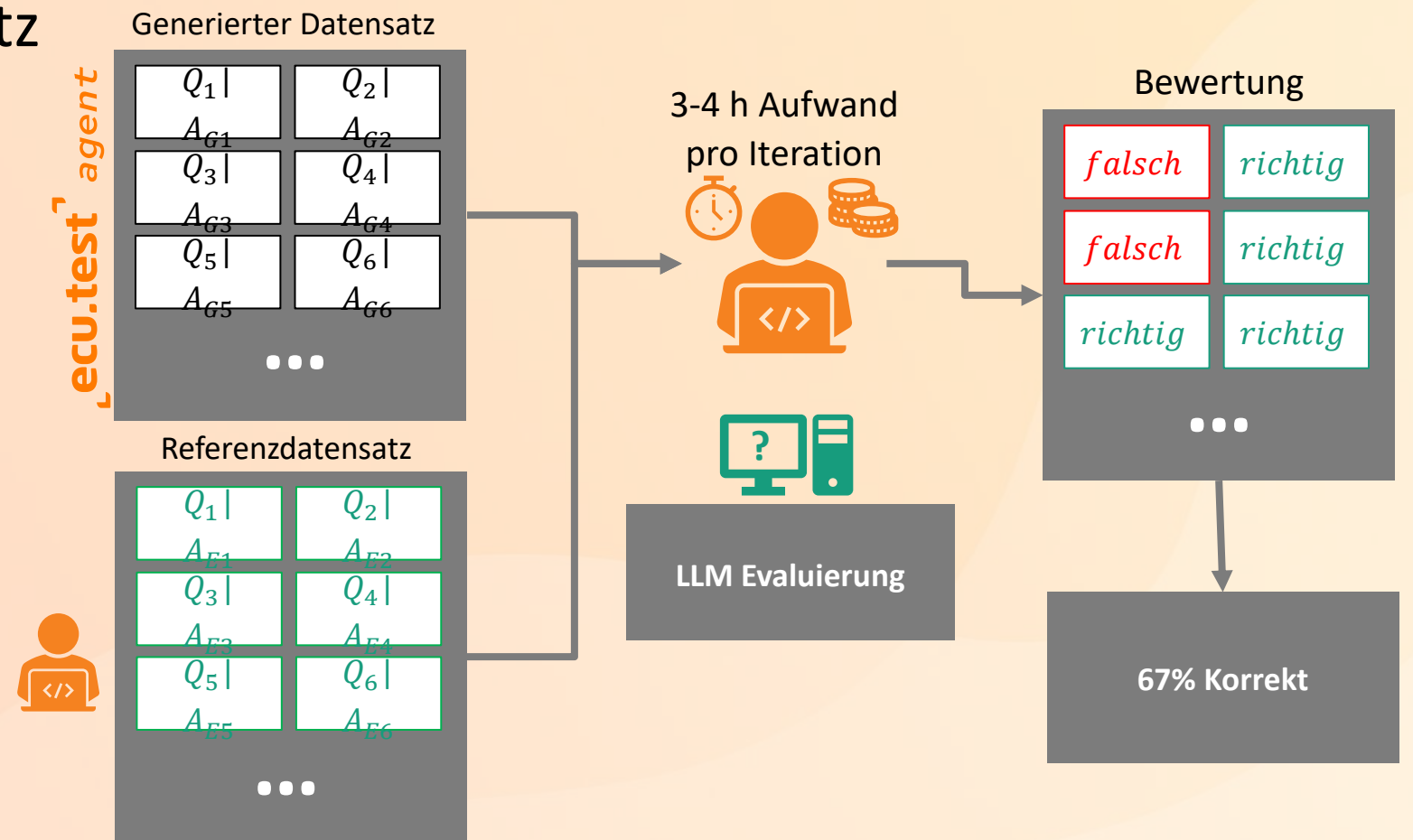


Qualität messbar machen

1. Referenzdatensatz aufbauen

2. Menschliche Evaluierung

3. Automatisierte Evaluierung

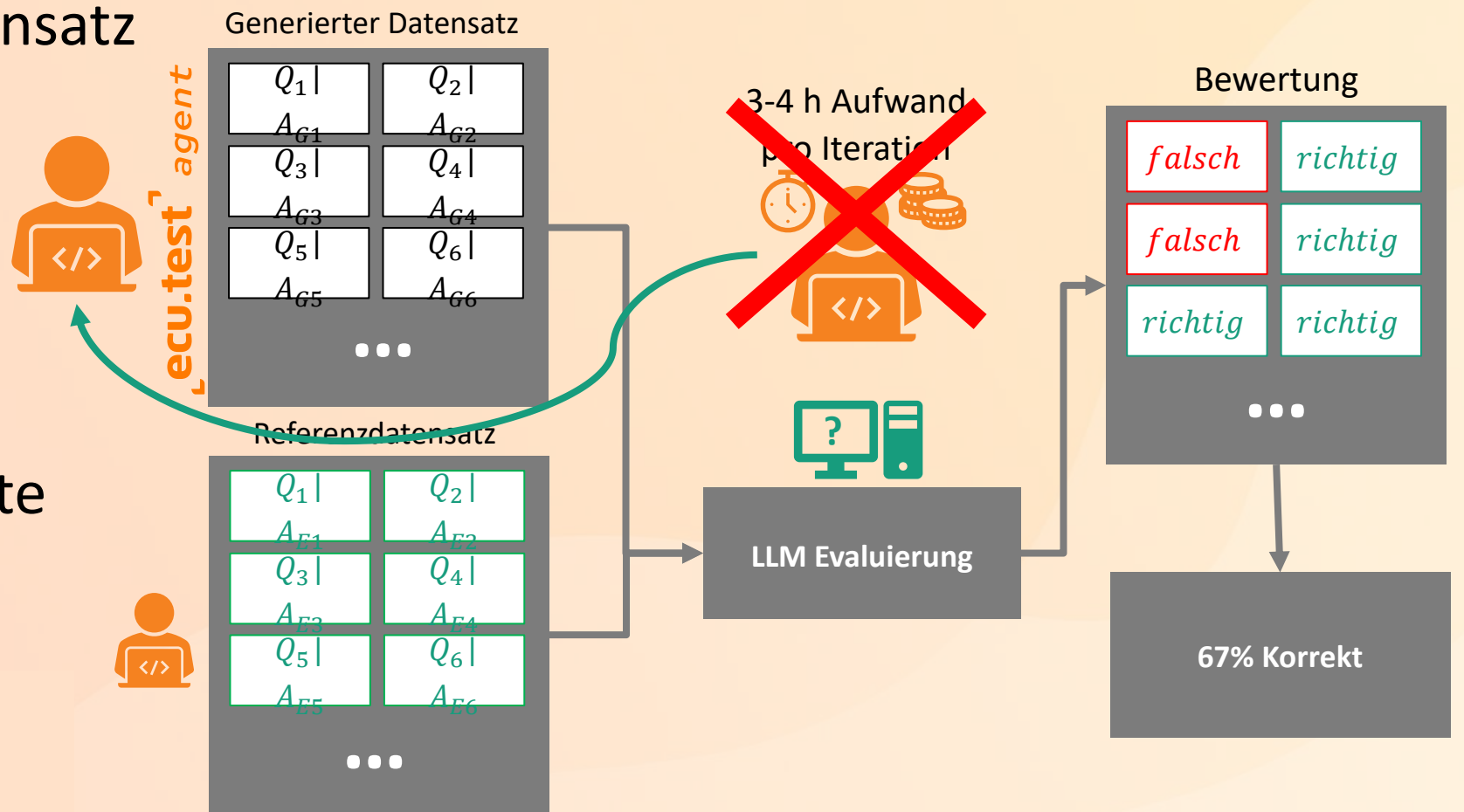


Qualität messbar machen

1. Referenzdatensatz aufbauen

2. Menschliche Evaluierung

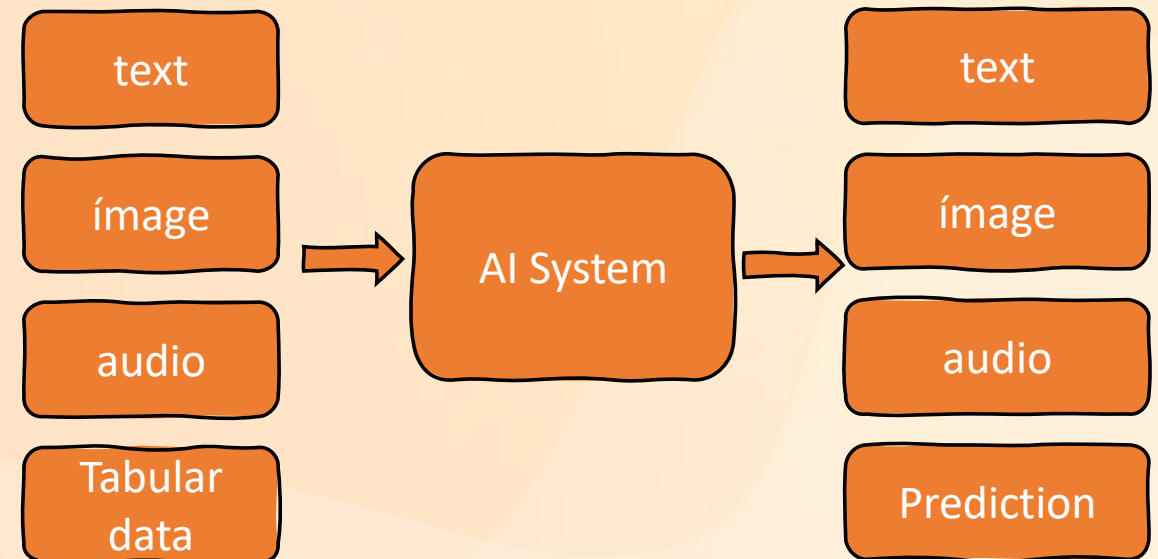
3. Automatisierte Evaluierung



Lessons Learned

Start small

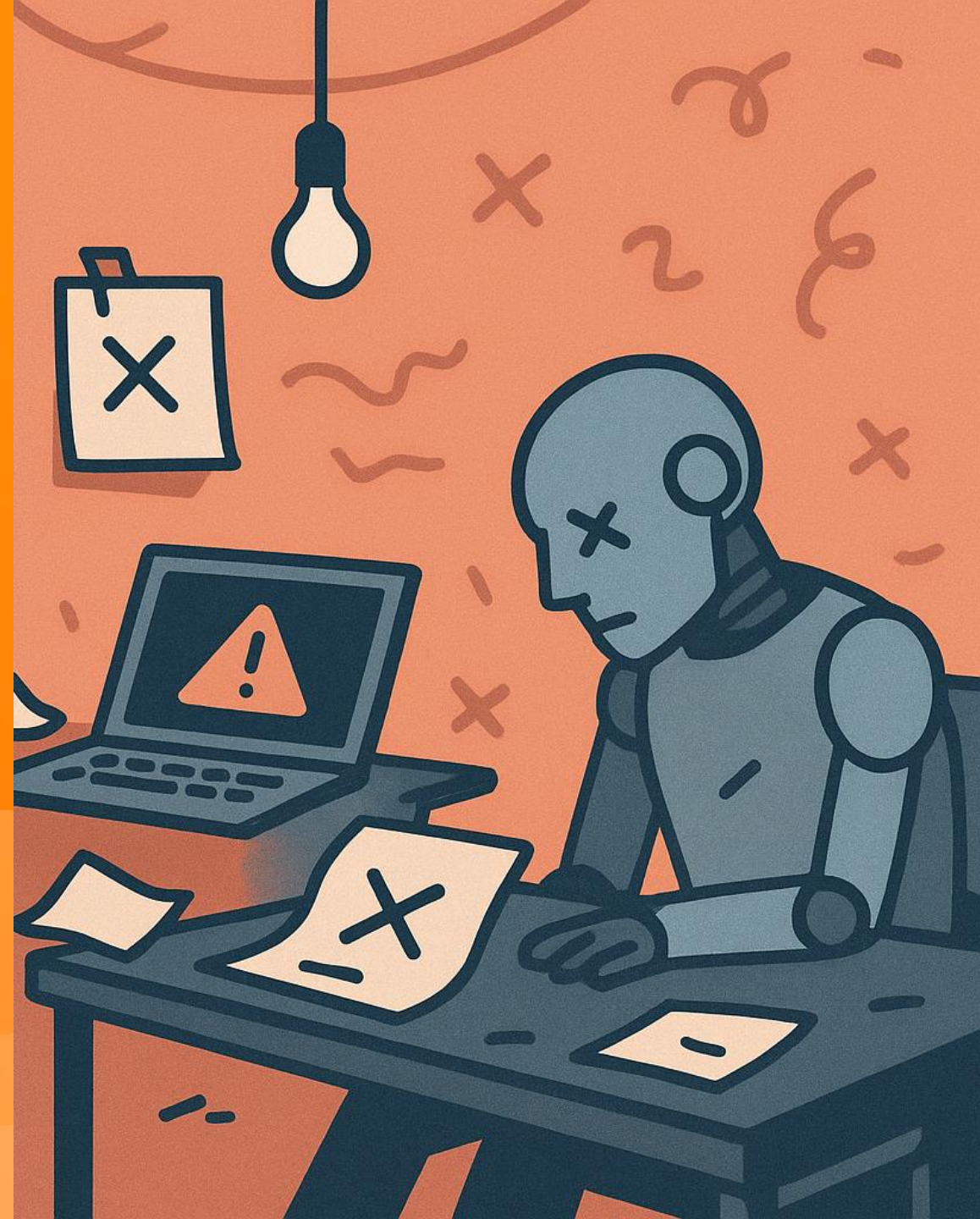
- Beginn mit vortrainierten Modellen
- Auswahl eines definierten Anwendungsfalls (z. B. Erkennung fehlgeschlagener Tests oder Klassifizierung von Fehlertypen)
- Durchführung von Zero-Shot- oder Few-Shot-Tests vor dem Fine-Tuning
- Nutzung eines kleinen, sauberen internen Datensatzes für schnelle Iterationen



Evaluate early

- Vertrauensanforderungen definieren
- Evaluierungspipeline erstellen
 - Frage-Antwort-Paare
 - Abdeckung verschiedener Eingabetypen
 - Abdeckung der Vertrauensanforderungen
- Frühe Automatisierung

→ Aufbau von Vertrauen in das KI-System



Feel free to reach out!



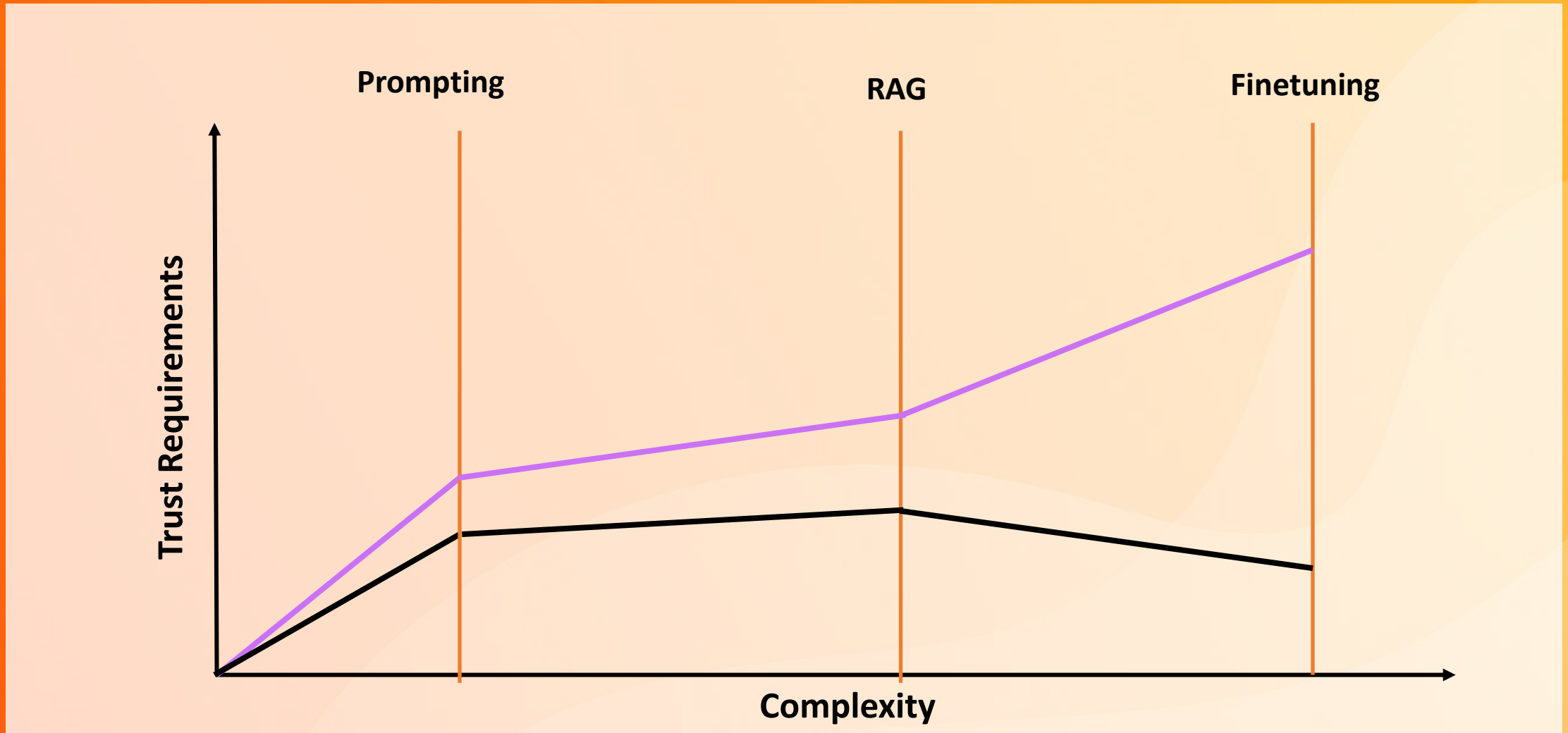
Ron Martin
Tel: +49 351 45691-238
ron.martin@iis.fraunhofer.de

Patrick Stiller
Product Owner AI Assistant

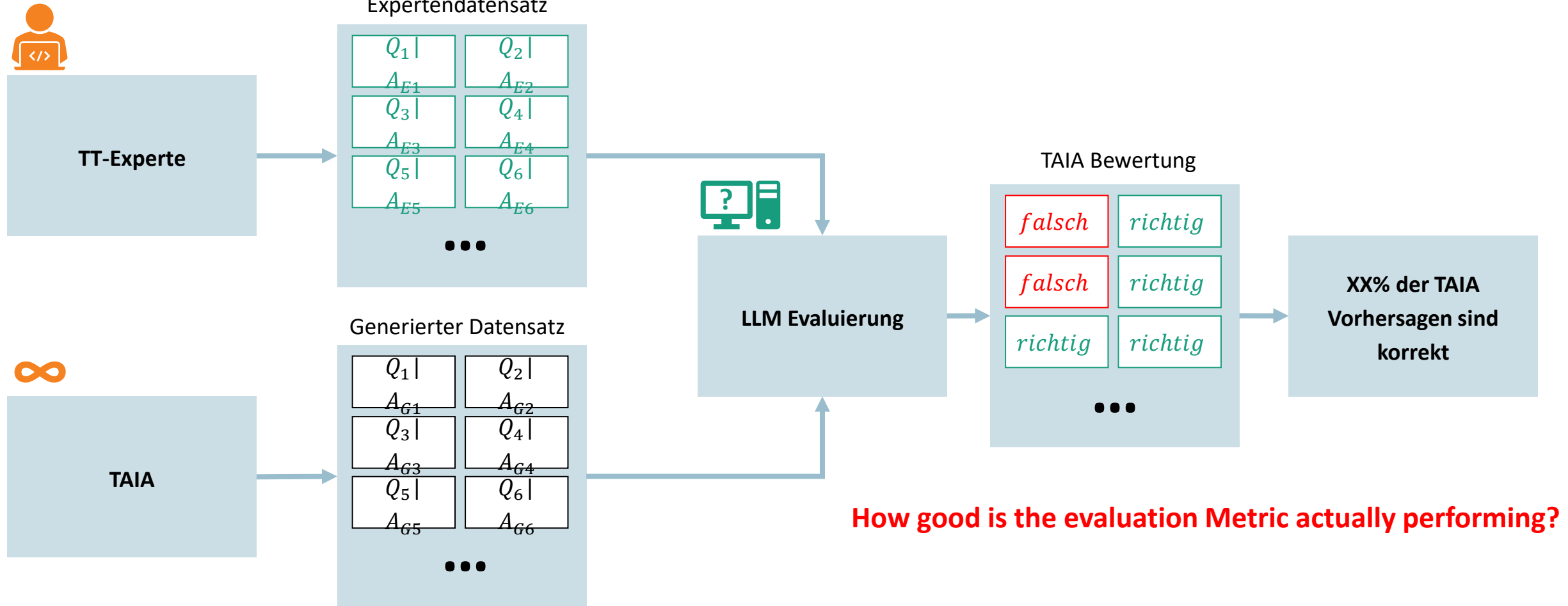
Fraunhofer-Institut für Integrierte Schaltungen IIS
Institutsteil Entwicklung Adaptiver Systeme EAS
Münchner Straße 16
01187 Dresden
www.eas.iis.fraunhofer.de

tracetrionic

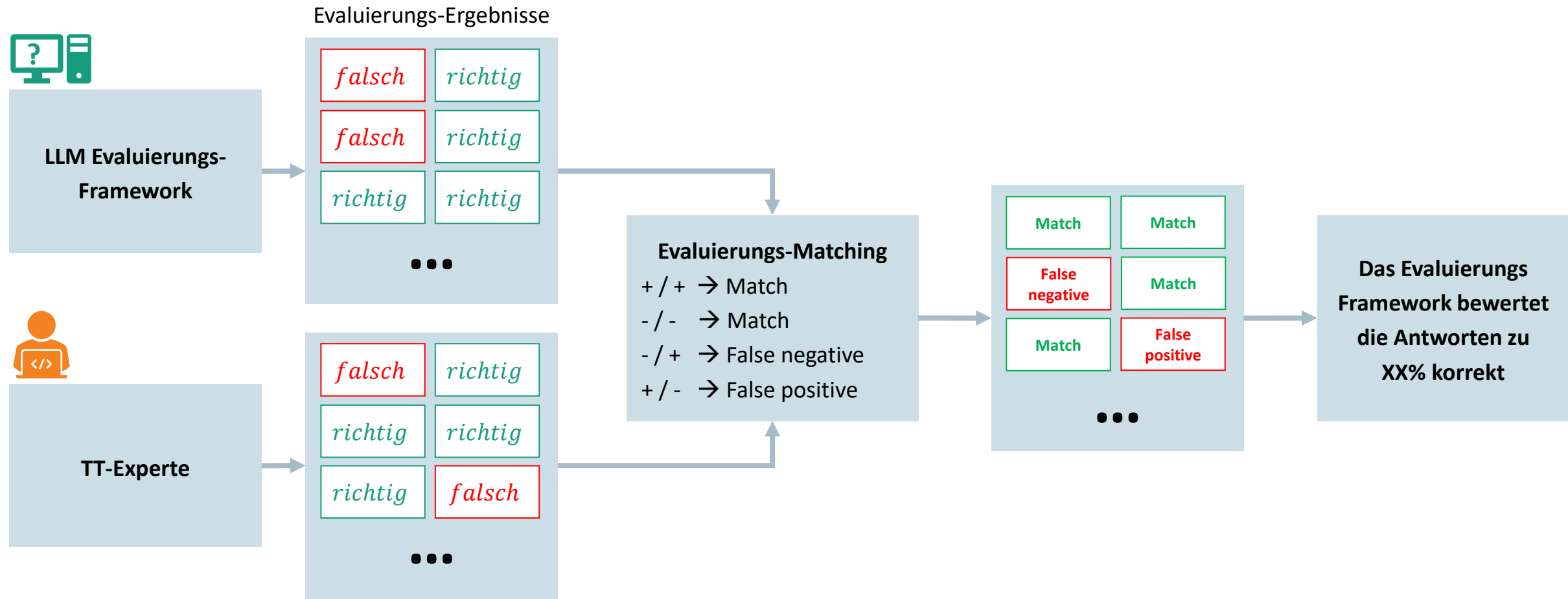
Complexity vs Simplicity



Evaluierungsmethodik



Validierung der Evaluierungsmethodik



Validierung der Evaluierungsmethodik

