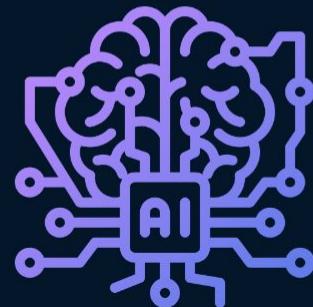




Chatbot Evaluation

“You can’t manage, what you can’t measure”



Problem



AI Developer

Unser neuer FAQ
Agent soll live gehen

Problem



AI Developer

Danke... war
etwas besser!

Neue Version,
bitte testen!



Business Unit

Problem



AI Developer



Business Unit



Go live
approved?



Board

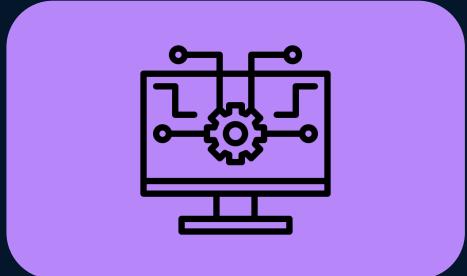
- ! Manuelle Eingabe der Testfragen
- ! Kein hilfreiches Feedback
- ! Keine Benchmarks

Lösung



AI Developer

Go live
approved?



AI Evaluation Tool



Board

- 😍 Automatisiertes Testing
- 😍 Detaillierte Analyse
- 😍 Messbare KPIs

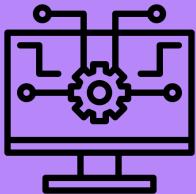
Lösung



AI Developer

Go Live!

Go live
approved?



AI Evaluation Tool



Automatisiertes Testing
Detaillierte Analyse
Messbare KPIs



Board

LIVE Demo

This is where the magic happens 

Pipeline

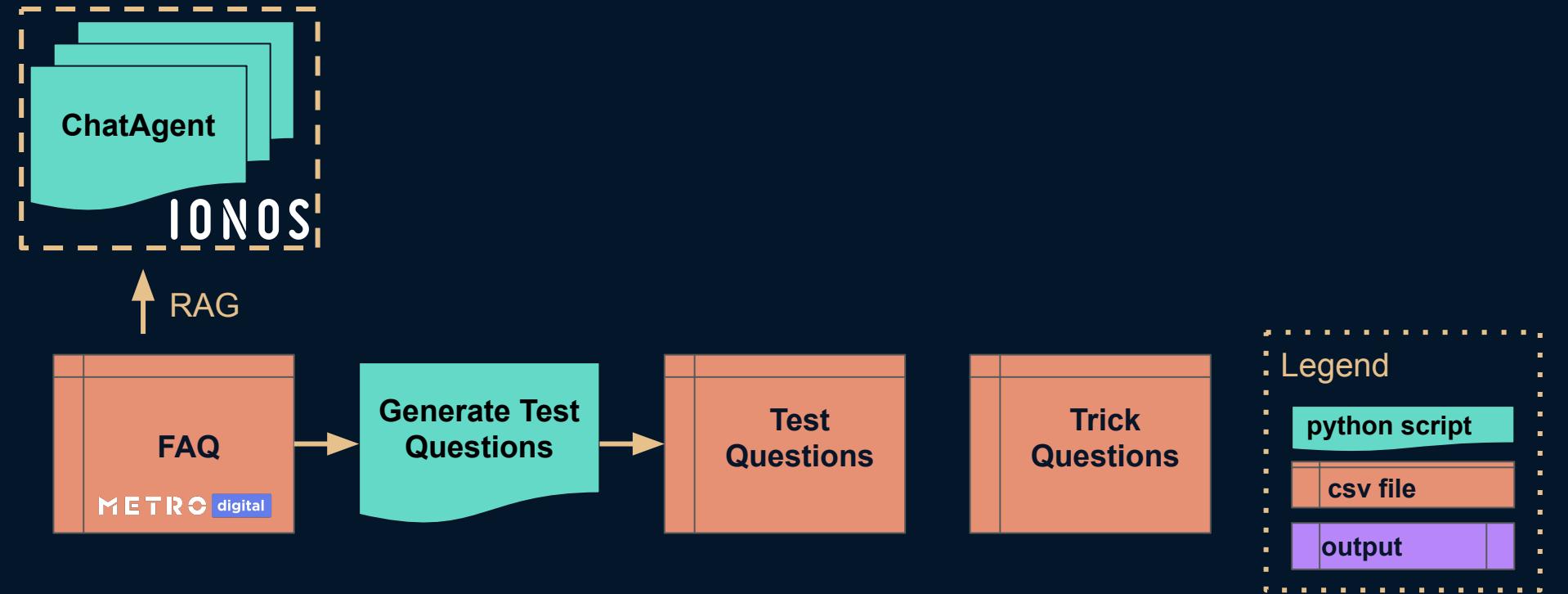


python script

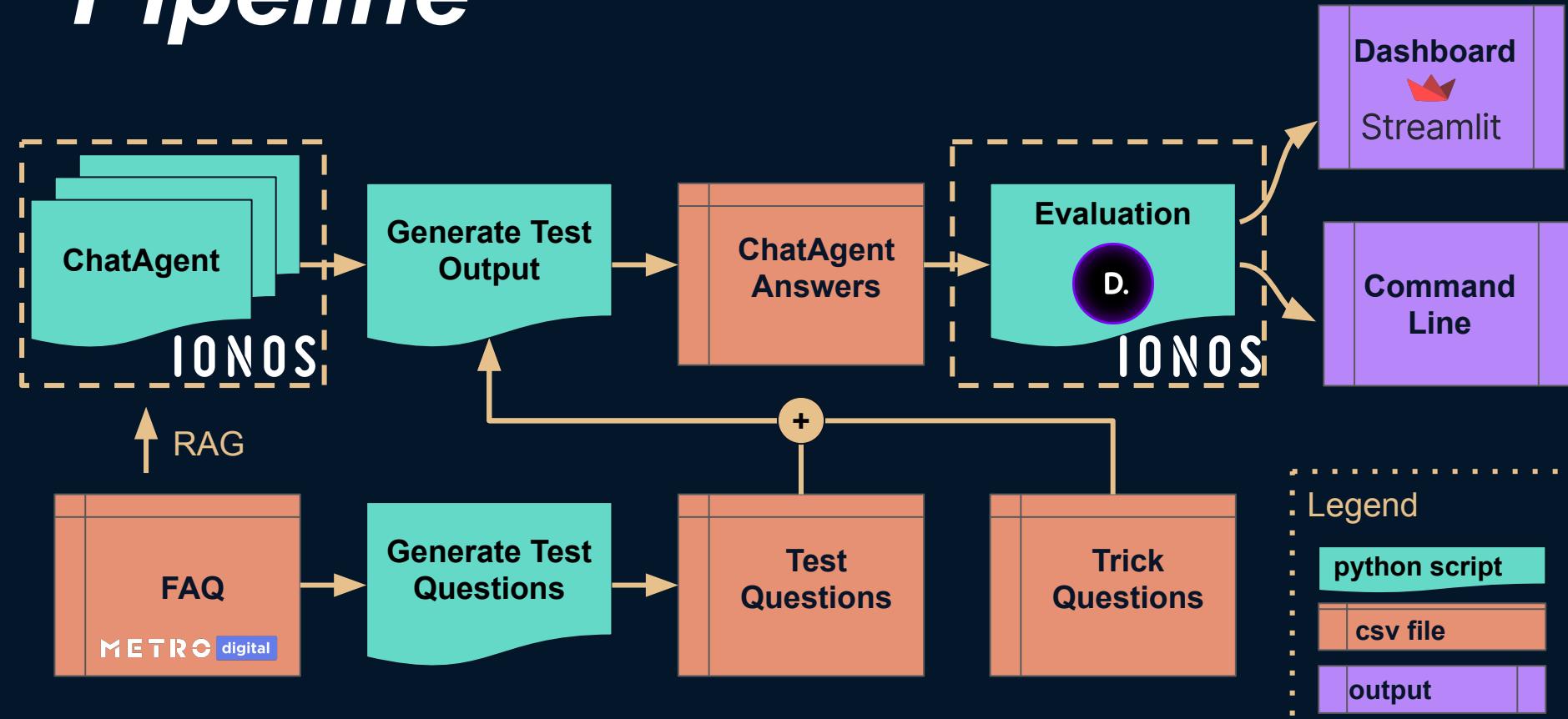
csv file

output

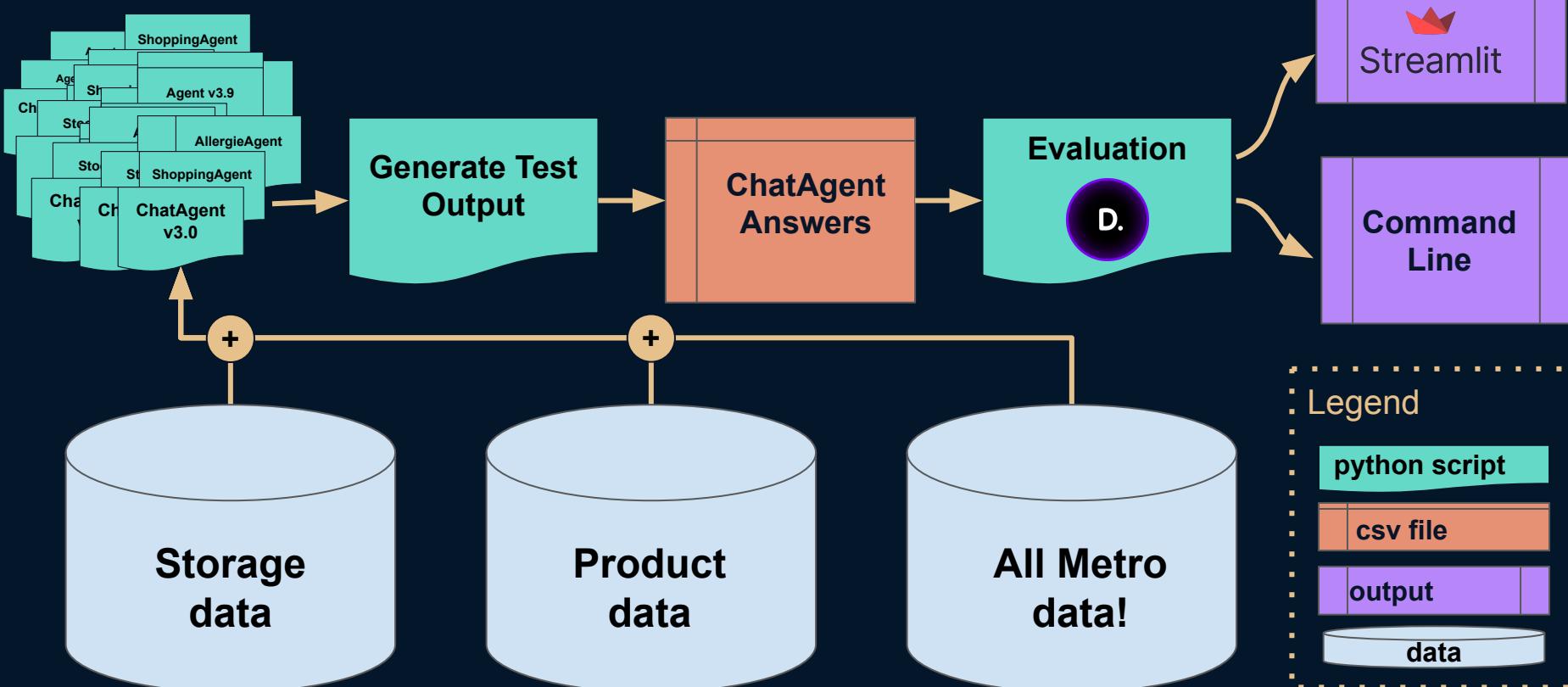
Pipeline



Pipeline



Future Outlook



A success story

METRO digital & IONOS

Manage, What You Measure.



**Hauke
Diers**



**Domitille
Grandjean**



**Nico
Neubauer**



**Andres
Navarro**



**Omid
Roshenas**

Evaluation Metrics

- 1. Correctness:** Sind die angegebenen Informationen wahr?
- 2. Clarity:** Ist die Antwort leicht zu verstehen?
- 3. Hospitality & Tonality:** Ist der Ton professionell?
- 4. Relevance:** Beantwortet er die gestellte Frage?
- 5. Hallucination Safety:** Werden Informationen erfunden?

