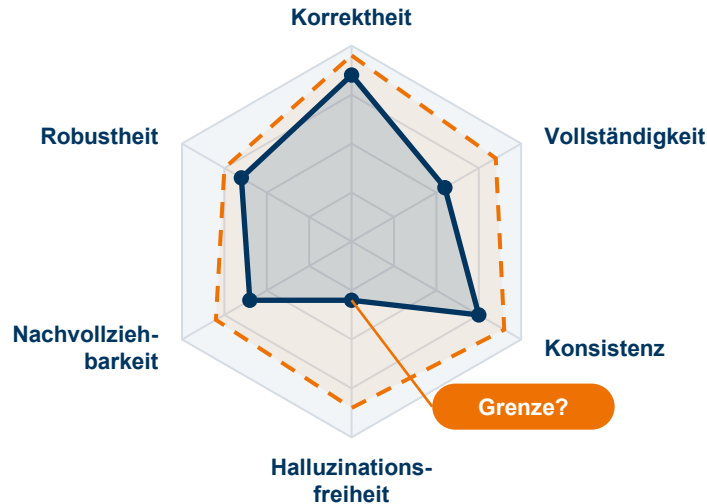


8.) Wie bewertet man, was ein LLM generiert?

Evaluierung und Grenzen von Large Language Models bei der Generierung technischer Modelle

QUALITÄTSKRITERIEN-KATALOG

— Bewertung durch Mensch - - - Bewertung durch LLM-as-a-Judge



Achtung! Ein LLM kann diese Aufgabe für Sie erledigen, aber nicht beurteilen, ob es dabei recht hat!

AUFGABEN

1

Literaturrecherche

Welche Methoden und Metriken gibt es, um KI-generierte Ergebnisse zu bewerten?

2

Fehler & Grenzen

Typische Fehler von LLMs bei technischen Modellen sammeln und ordnen.

3

Kriterienkatalog

Qualitätskriterien ableiten und jedem Kriterium eine passende Bewertungsmethode bzw. Metrik zuordnen.

4

Experiment

LLM-generierte Modelle mit dem Katalog bewerten: einmal durch Menschen, einmal durch ein LLM, und vergleichen.

Ansprechpartner:

Niloofer Jafarnia

ID 2/531 – niloofer.jafarnia@rub.de