1. Projektziel

KomMKonLLM ist ein Framework, das den Einsatz kombinatorischer Testmethoden zur Bewertung der Konsistenz von LLMs ermöglicht. Es erlaubt Entwickler:innen und Nutzer:innen, folgende Fragestellungen zu untersuchen:

- Wie konsistent reagiert ein bestimmtes LLM, wenn Teile von Prompts durch Synonyme ersetzt werden?
- Welches LLM ist für eine konkrete Fragestellung am besten geeignet?
- Wie gut befolgt ein LLM Anweisungen, die das Ausgabeformat einschränken?

Wer sind wir?

Die Teammitglieder Manuel Leithner, Ludwig Kampel und Bernhard Garn sind Forscher in Informationssicherheit und interessieren sich für Stärken und Schwächen von künstlicher Intelligenz, insbesondere derer von Large Language Models (LLMs), welche die IT-Landschaft zuletzt stark verändern und anzunehmenderweise auch in Zukunft eine wesentliche Rolle spielen werden.

Was ist KomMKonLLM?

Das **KomMKonLLM**-Framework adressiert das Problem, sicherzustellen, dass LLMs zuverlässig auf verschiedene Eingaben reagieren. Da LLMs oft komplexe und undurchsichtige Strukturen haben, können sie inkonsistente oder unerwartete Antworten auf (semantisch-) ähnliche Eingaben geben. Solche Inkonsistenzen erschweren es, LLMs in Anwendungen einzusetzen, wo Verlässlichkeit entscheidend ist. Unser Tool automatisiert die *Erstellung* einer kompakten Menge an Konsistenztests nach kombinatorischen Kriterien, die von einer einzigen ursprünglichen Frage abgeleitet werden; es stellt eine *Anbindung* (über OLLAMA[1]) an über 150 LLMs[2] - wie etwa Llama, DeepSeek und Mistral - zur Verfügung, und erlaub die einfache Entwicklung von Schnittstellen zu weiteren LLMs; des Weiteren automatisiert das Python-basierende **KomMKonLLM**-Framework die *Ausführung* der Tests, sowie die *graphische Aufarbeitung* deren *Resultate*.

Für wen ist KomMKonLLM?

Das **KomMKonLLM**-Framework bietet die Möglichkeit zum automatisierten Konsistenztesten verschiedenster Large Language Models. Das Framework ist primär für Entwickler:innen hilfreich, wenn diese bei der Integration von LLMs in ihre Software, Aufschluss über die Konsistenz der LLMs haben wollen, z.B. um die Robustheit ihrer Anwendung abschätzen zu können. Dieses Framework ist allerdings auch für LLM Nutzer:innen hilfreich, weil sie über die graphisch aufgearbeiteten Resultate unseres Frameworks einen raschen Eindruck von der Konsistenz, und somit der Zuverlässigkeit von den LLMs bekommen können. Diese Auswertung gilt sowohl für im Alltag (privat gerne) verwendete LLM wie auch für im Unternehmenskontext eingesetzte LLMs. Letztlich ist das **KomMKonLLM**-Framework auch für Forscher:innen relevant, da das Testen - insbesondere das Konsistenztesten von LLMs - Gegenstand der laufenden Forschung ist.

Wie funktioniert KomMKonLLM?

Das **KomMKonLLM**-Framework basiert auf einer Docker-Infrastruktur und kombiniert NLP-Bibliotheken, Anbindungen an kombinatorische Test-Generatoren und standardisierte LLM-Anbindungen. Es erstellt Konsistenztests für LLMs mit Hilfe von kombinatorischen Methoden automatisch ausgehend von einem

[1] https://ollama.com/
[2] https://ollama.com/search

Satz in natürlicher Sprache zu angegebener Interaktionsstärke. Zusammen mit einem integrierten Repository von Testfällen bietet **KomMKonLLM** eine einfache Plug-and-Play Lösung für das kombinatorische Testen der Konsistenz von LLMs.

2. Projektendergebnisse

1	Projektzwischenbericht	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-02/prj7409_Call19_Zwischenbericht_V01.pdf
2	Projektendbericht	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-07/prj7409_Call19_Endbericht_V03.pdf
3	Entwickler_innen-DOKUMENTATION	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-07/HACKING_on_KomMKonLLM_V02.pdf
4	Anwender_innen-DOKUMENTATION	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-05/KomMKonLLM_UserGuide_0.pdf
5	Veröffentlichungsfähiger Einseiter / Zusammenfassung	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-05/KomMKonLLM_Einseiter.pdf
6	Dokumentation Externkommunikation zur Erreichung Sichtbarkeit /Nachhaltigkeit	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-07/prj7409_Call19_Endbericht_V03.pdf Die Dokumentation Extern-Kommunikation zur Erreichung Sichtbarkeit ist als Teil des Endberichts, in Kapitel **7 Öffentlichkeitsarbeit/Vernetzung,** zu finden**.**
7	Server-Lösung für kombinatorische Konsistenztests von LLMs	MIT	https://github.com/KomMKonLLM/KomMKonLLM
8	Repository von kombinatorischen Testfällen zur Konsistenzevaluierung von LLMs	MIT	https://zenodo.org/records/15209547

3. Geplante weiterführende Aktivitäten nach netidee-Projektende

Wir planen unsere Outreach-Aktivitäten weiterzuführen um das KomMKonLLM-Framework zu bewerben und zu demonstrieren, insbesondere da es nun einsatzfähig ist. Dabei fokussieren wir uns auf folgende Zielgruppen:

1. LLM-Entwickler:innen: Ziel ist es, unsere entwickelte Server-Lösung in bestehende Qualitätssicherungsprozesse von Entwickler:innen großer Sprachmodelle (LLMs) zu integrieren. Wir

werden das KomMKonLLM-Framework weiter in einschlägigen Communities bewerben, insbesondere bei Fachveranstaltungen und Konferenzen wie z. B. Sec4Dev[3], um dessen Nutzen für die Qualitätssicherung beim Einsatz von LLMs aufzuzeigen.

2. LLM-Nutzer:innen: Um LLM Nutzer:innen die Relevanz von Konsistenztests und die Fähigkeiten unseres KomMKonLLM-Frameworks zu demonstrieren planen wir die Teilnahme an Science-Communication Veranstaltungen, wie z.B. die Lange Nacht der Forschung[4].
3. Externe Forscher:innen: Die generierten Tests sollen in Science Outreach-Aktivitäten demonstriert werden, um das Bewusstsein für potenzielle Risiken und Qualitätsaspekte von LLMs zu schärfen. Darunter fallen After-Work Events wir (Security-) MeetUps[5], aber auch wissenschaftliche Konferenzen an denen wir im Zuge unserer Forschungsaktivitäten aktiv beitragen und teilnehmen.
4. Interne Forschungsaktivitäten: Die Server-Lösung sowie die generierten Tests werden auch weiterhin in der Forschungsarbeit der MATRIS-Forschungsgruppe sowie in Kooperation mit internationalen akademischen Partnern eingesetzt.

Während der Arbeit an diesem Projekt haben wir einige Ideen für Weiterentwicklungen und Verbesserungen gesammelt, sei es durch eigene Beobachtungen oder durch Diskussionen im Rahmen unserer Outreach-Aktivitäten, wie zum Beispiel:

a. Entwicklung von Methoden zur Überprüfung der Logik-Fähigkeiten von Sprachmodellen.
b. Verbesserte Generierung semantischer Varianten
c. Multi-linguale Unterstützung
d. Verbesserte Interaktion mit der Testpipeline durch die Entwicklung einer intuitiven Benutzeroberfläche

Wir planen einige dieser Punkte in einem Netidee-Folgeprojekt aufzugreifen.

4. Anregungen für Weiterentwicklungen durch Dritte

Über die oben erwähnten Punkte hinaus, eröffnen sich Weiterentwicklungsmöglichkeiten durch den allgemeinen Fortschritt von LLMs und deren engere Verflechtung in Nutzer:innen Anwendungen.
Das **KomMKonLLM**-Framework ist so konzipiert, dass es sich flexibel erweitern lässt. Es ist also ein idealer Ausgangspunkt für Beiträge und Weiterentwicklungen durch Dritte. Potenzielle Anknüpfungspunkte liegen etwa in der Integration weiterer LLM-Backends, oder der Erweiterung des Testfall-Repositories um spezifische Anwendungsbereiche (z. B. juristische, medizinische oder mehrsprachige Testfälle). Auch die visuelle Aufbereitung der Ergebnisse lässt sich durch alternative Darstellungsformen oder interaktive Auswertungen weiter verbessern. Durch die offene Architektur und modulare Gestaltung können Entwickler:innen mit unterschiedlichsten Schwerpunkten das **KomMKonLLM**-Framework sinnvoll ergänzen und zur Weiterentwicklung robuster, transparenter LLM-Systeme beitragen.

Konkrete und detaillierte Anregungen für Entwickler:innen finden sind unter https://github.com/KomMKonLLM/KomMKonLLM/blob/main/HACKING.md zu finden.

[3] https://sec4dev.io/
[4] https://langenachtderforschung.at/
[5] https://www.sba-research.org/sba-meetup-groups/