SecuredGPT

Pilotbetrieb eines On-Premises Large Language Model im Wissensmanagement

Programm / Ausschreibung	KIRAS, F&E-Dienstleistungen, KIRAS-K-Pass-KMU Innovation AKUT KIA F&E Dienstleistungen (FED KIA_2023)	**Status**	abgeschlossen
Projektstart	01.05.2024	**Projektende**	30.04.2025
Zeitraum	2024 - 2025	**Projektlaufzeit**	12 Monate
Keywords	NLP LLM Wissensmanagement Chatbot		

Projektbeschreibung

Strafverfolgungsbehörden nutzen interne Wissensdatenbanken, um Informationen und Wissen effizient zu speichern, zu organisieren und für den späteren Zugriff bereitzuhalten. Ein Beispiel hierfür ist der Kriminalistische Leitfaden (KLF) des BMI, der unter anderem Handlungsanweisungen für den Umgang mit Cyber-Crime-Delikten und interne Dienstanweisungen enthält. Das primäre Ziel solcher Wissensdatenbanken ist der Wissensaustausch innerhalb der Organisation und die Steigerung der Effizienz, indem Mitarbeiter schnell und einfach auf relevante Informationen zugreifen können (3.3.2 Ausschreibungsleitfaden).

Behördliches Wissensmanagement muss AKUT mit den Möglichkeiten von KI (Large Language Models) ausgestattet werden. Kann man keine sichere Alternative zu ChatGPT bereitstellen, läuft man Gefahr, dass ChatGPT unter erheblichen Sicherheitsrisiken verwendet wird. Mit SecuredGPT kann inner-behördlich, zeitnah eine sichere Alternative angeboten werden.

Die Software der Cortecs GmbH wird bislang in geschützter Cloud-Umgebung innerhalb Europas betrieben und soll hiermit On-Premises im BRZ unter Wahrung der Datensouveränität in Betrieb genommen werden. Die Software ist einsatzfähig in deutscher und englischer Sprache (TRL-6). Sie wird in der Pilotphase weiter an die praktischen Anforderungen der Polizist:innen angepasst, laufende Evaluierungen bezüglich Qualität und Laufzeitverhalten sind dabei fixer Bestandteil.

Abstract

Law enforcement agencies rely on knowledge databases to efficiently organize internal information. One example is the 'Kriminalistischer Leitfaden' of the BMI, which contains directives on handling cybercrime and internal operational procedures. The main aim of these knowledge databases is to facilitate information exchange within the organization, ultimately increasing efficiency by enabling staff to access information quickly and effortlessly.

Government knowledge management needs to be urgently equipped with the capabilities of AI (Large Language Models). If a secure alternative to ChatGPT cannot be provided, there is a risk of ChatGPT being used under significant security threats. With SecuredGPT, a secure alternative can be offered within government agencies promptly.

The existing software developed by Cortecs GmbH is currently being deployed on private European cloud infrastructure and is capable of on-premises deployment, ensuring the utmost data sovereignty. The software is already functional in both German and English (TRL-6). The current technology is brought into a pilot phase to assess its practicality in real-world scenarios, with ongoing evaluations regarding quality and runtime behavior being an integral part of the process.

Endberichtkurzfassung

Das Projekt SecuredGPT hatte zum Ziel, bestehende behördliche Wissensmanagementsysteme durch den Einsatz von Large Language Models (LLM) effizienter und intelligenter nutzbar zu machen. Dabei wurden technologische, organisatorische und sicherheitstechnische Aspekte analysiert und in einem Pilotbetrieb praktisch erprobt.

Marktanalyse und Evaluierung:

Zur Auswahl geeigneter LLMs wurden zunächst öffentliche Testdaten herangezogen und anschließend durch anwendungsspezifische, eigene Testdaten ergänzt. Dieser kombinierte Evaluierungsansatz ermöglicht eine realitätsnahe Einschätzung der Modellqualität. Zunächst wurde Sauerkraut-Mixtral-8x7B-Instruct aufgrund seiner hohen Antwortqualität in Deutsch und Englisch, seiner Verarbeitungsgeschwindigkeit sowie der offenen Apache 2.0 Lizenz eingesetzt. Durch den dynamischen Fortschritt im LLM-Bereich wurde ein strukturierter Prozess etabliert, um neue Modelle kontinuierlich zu evaluieren und bei Bedarf bestehende Modelle zu ersetzen oder zu aktualisieren.

Datenintegration mittels RAG:

SecuredGPT wurde am konkreten Beispiel des Kriminalistischen Leitfaden (KLF) erprobt. Zur Einbindung der KLF-Inhalte wurde ein Retrieval-Augmented Generation (RAG) Ansatz gewählt, der eine flexible Kopplung der Wissensdatenbank mit dem LLM ermöglicht. Die Qualität dieses Ansatzes wurde anhand der eigens gefertigten Testdaten basierend auf dem KLF ermittelt und weiters in einem Pilotbetrieb praktisch validiert.

Pilotbetrieb:

Im Pilotbetrieb konnten konkrete Anwendungsfälle simuliert und mit realen Nutzer:innen getestet werden. Das gesammelte Feedback bestätigte die Praxistauglichkeit des Ansatzes und lieferte wertvolle Hinweise für die Weiterentwicklung, insbesondere zur Integration in bestehende IT-Prozesse. Für den produktiven Einsatz wurden verschiedene Infrastrukturszenarien geprüft. LLMs mit rund 70 Milliarden Parametern konnten im 8-bit Format effizient auf A100- oder H100-GPUs betrieben werden. Diese Erkenntnisse dienen nun als Grundlage für weiterführende Betriebsentscheidungen (z.?B. Beschaffung und Dimensionierung der Rechenumgebung).

Projektkoordinator

- Cortecs GmbH

Projektpartner

- Bundesministerium für Inneres