LLMoin
Der KI-Textassistent für die Verwaltung

Freie und Hansestadt Hamburg

Senatskanzlei

Details zu LLMoin

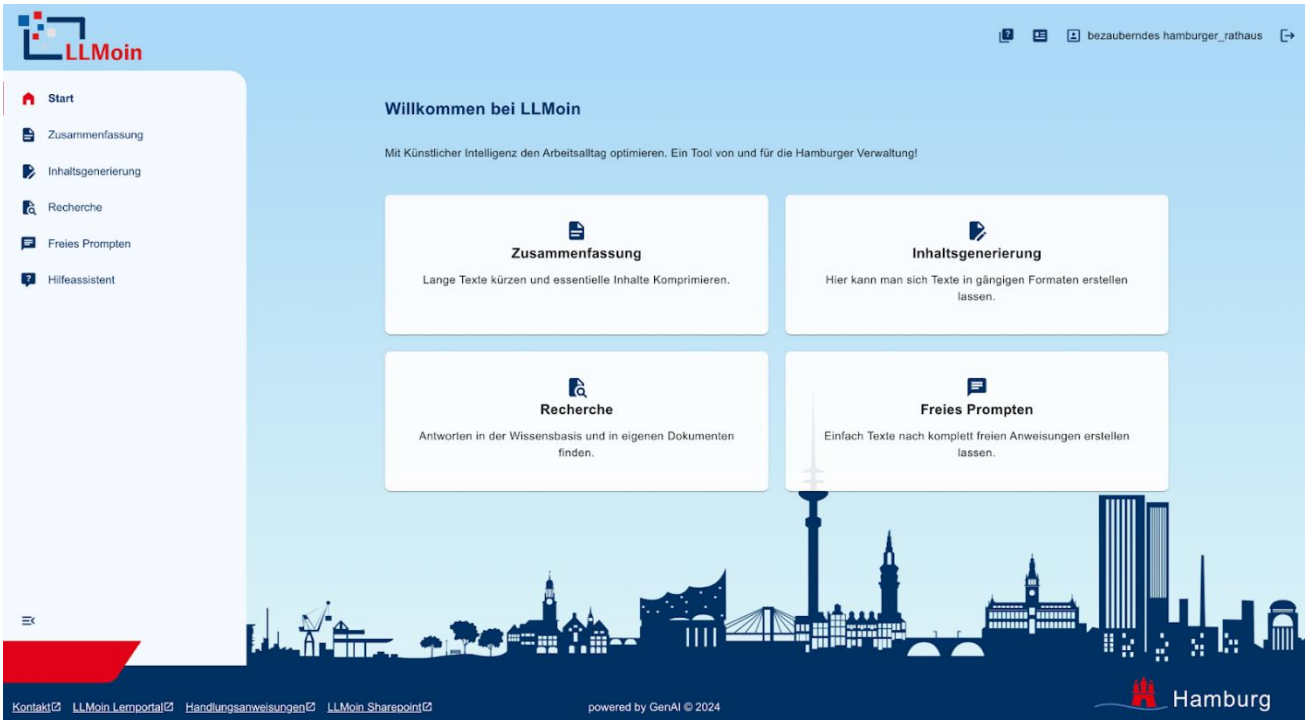
Ergänzende Informationen zu LLMoin für Interessierte aus Verwaltung, Medien und Zivilgesellschaft

vom 25.11.2024

Inhaltsverzeichnis

Einleitung

Auf den folgenden Seiten versuchen wir, als verantwortliche Fachliche Leitstelle von LLMoin, eine hilfreiche und detaillierte Beschreibung von LLMoin darzulegen. Wir betrachten LLMoin als einen wichtigen initialen Schritt, um die digitale Transformation der Verwaltung weiter voranzutreiben. In diesem Artikel gehen wir daher ausführlich auf unsere Lösung und Entwicklungsschritte ein, um Transparenz und Lerneffekte für Dritte zu ermöglichen.



Startseite von LLMoin

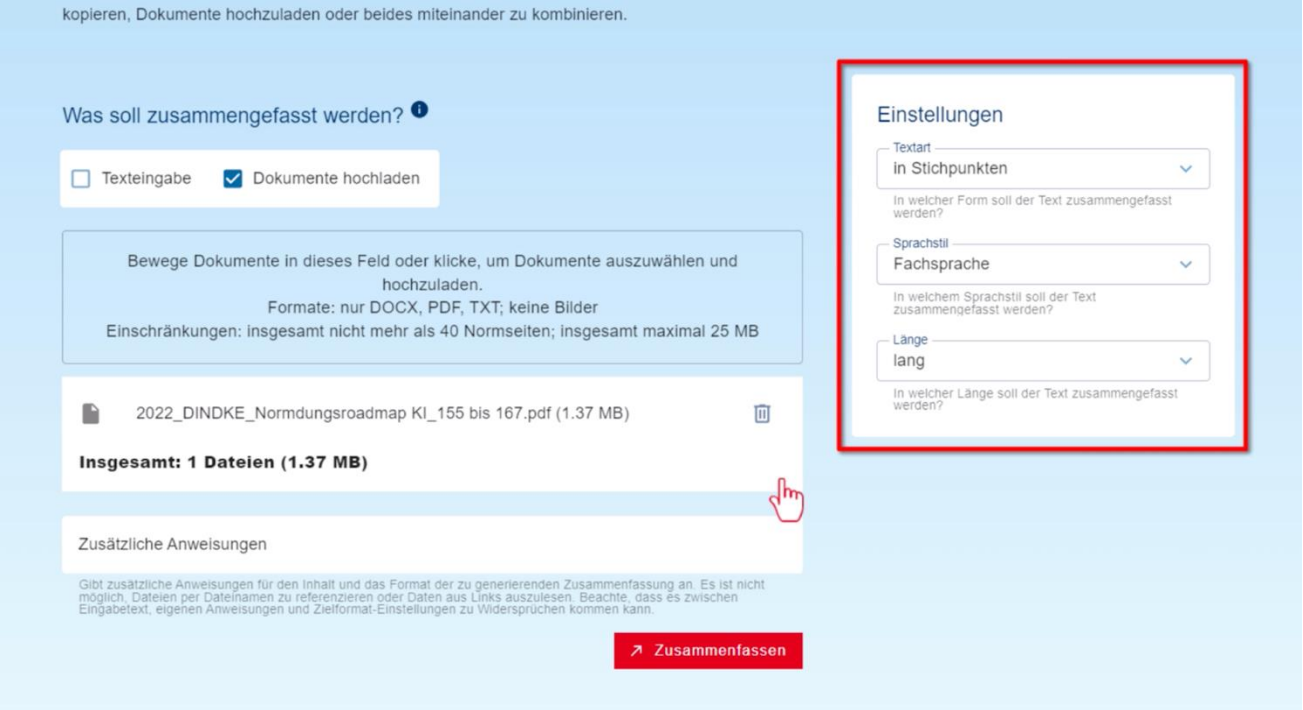
Was ist LLMoin?

LLMoin ist ein speziell für die Bedürfnisse der Hamburger Verwaltung entwickeltes KI-Produkt, das den Mitarbeitenden der Freien und Hansestadt Hamburg Zugang zu großen Sprachmodellen (LLMs) ermöglicht. Im Vergleich zu kommerziellen Lösungen wie ChatGPT, Perplexity oder Claude.ai legt LLMoin besonderen Wert auf Datenschutz, Sicherheit und die Eignung für den behördlichen Einsatz. Es erlaubt Usern (Synonym für Mitarbeitende) Dokumente oder Texte zu bearbeiten. Die vier zentralen Funktionen dabei sind die Zusammenfassung von Texten, die Generierung von verschiedenen Texten (E-Mails, Vermerke, usw.), die Recherche in großen Datengrundlagen und das sogenannte Freie Prompten (für alle weiteren Anwendungsfälle). In allen vier Funktionen können Texte eingegeben oder Dokumente hochgeladen werden. Bei der Recherche sind neben der Möglichkeit, ad hoc Daten hochzuladen vordefinierte Datensätze bereits global für alle User freigegeben (z.B. die Antworten auf Schriftliche Kleine Anfragen aus dem Kontext der Hamburgischen Bürgerschaft) und in Zukunft planen wir weitere Datensätze anzuschließen.

Intuitive Nutzung

LLMoin zeichnet sich durch seine benutzerfreundliche Oberfläche aus, die keine speziellen Kenntnisse über Prompts (Anweisungen an die LLMs) erfordert. Stattdessen lassen sich die gewünschten Ergebnisse durch einfache, intuitive User Flows erzielen. Diesen Prozess nennen wir Hamburg-intern “geführtes Prompting” und er wurde durch die Kollegen vom InnoLab in Baden-Württemberg erstmalig durch die Entwicklung von F13 umgesetzt. Zusätzlich werden umfassende Schulungsunterlagen in Form von Videos, PDFs, Übungsaufgaben und Quizformaten bereitgestellt,

um die Kompetenzen der Mitarbeitenden im Umgang mit LLMs zu fördern und das LLM- und KI-Verständnis im Allgemeinen in der Verwaltung zu stärken.



Zeigt beispielhaft den User Flow bei der Zusammenfassung mit den verschiedenen Auswahlmöglichkeiten für die User oben rechts und der Möglichkeit für "Zusätzliche Anweisungen" unten mittig.

Technische Beschreibung

LLMoin wurde von der Freien und Hansestadt Hamburg in Auftrag gegeben und von Dataport AöR auf Basis ihrer GenAI Plattform entwickelt. Getrieben wurde die Entwicklung maßgeblich vom Amt für IT und Digitalisierung (ITD21) in Zusammenarbeit mit dem KI-Unternehmen Merantix Momentum. Die technische Umsetzung des Piloten erfolgte durch Capgemini und wurde dann von Dataport übernommen und bis in den Betrieb fertiggestellt.

LLM-Modelle

Die "Intelligenz" hinter LLMoin basiert aktuell auf den neuesten Modellen der GPT-Reihe von OpenAI (im Oktober 2024: GPT-4o, 2024-05-13) die auf europäischen Servern von Microsoft Azure gehostet werden. LLMoin und die darunter liegende GenAI Plattform von Dataport ist modell-agnostisch konzipiert und kann die zugrundeliegenden Modelle schnell austauschen, falls andere Modelle in Zukunft aus fachlichen, rechtlichen, wirtschaftlichen oder politischen Gründen bevorzugt werden.

Datenfluss

Eine LLMoin-Eingabe („Prompt") wird von Dataport vorverarbeitet (z.B. ein RAG-Prozess oder die Kombination von System Prompt und User Prompt) und an die Server von Azure weitergeleitet, um eine Antwort vom LLM zu erhalten. Die Antwort läuft von Microsoft über Dataport wieder zurück zum Nutzenden. Dabei gilt immer: Microsoft speichert niemals Anfragen und Antworten, nutzt Daten der Hamburger Verwaltung niemals für weitere Zwecke und weiß zu keinem Zeitpunkt, welcher User Anfragen stellt. Nach einer sogenannten User-Session speichert Dataport die Nutzungsdaten temporär, um die reibungslose Betriebsfähigkeit und die Nachverfolgung sicherzustellen. In Zukunft wird es vermutlich auch nützlich sein, User-Verhalten auf einer aggregierten Ebene zu verstehen.

Wichtig bleibt, dass zu keinem Zeitpunkt Dataport oder die Hamburger Verwaltung eine Zuordnung von Prompts und einzelnen Nutzenden herstellen kann oder nachverfolgen kann.

Diese stark vereinfachte Darstellung erklärt den grundlegenden Datenfluss bei LLMoin.

Sichere, korrekte und ethische Nutzung von Sprach- modellen

Hamburg strebt danach, das Potenzial von Large Language Models (LLMs) im Einklang mit den relevanten regulatorischen Gesetzgebungen (vor allem DSGVO und KI-Verordnung) sowie der eigens entwickelten KI-Leitlinien der Stadt umzusetzen. Im Folgenden beschreiben wir die wichtigsten Maßnahmen und Entscheidungen, um dies zu erreichen.

Datenschutz

Da in LLMoin personenbezogene Daten verarbeitet werden, wurden für LLMoin ausführliche Datenschutzdokumentationen durch die federführende Stelle in der Senatskanzlei erstellt und das Schutzniveau und daraus resultierende Möglichkeiten der Datenverarbeitung festgelegt. In LLMoin dürfen bestimmte personenbezogene Daten verarbeitet werden, solange diese nicht besonderen Kategorien unterliegen. So dürfen beispielsweise keine hochsensiblen Daten nach Art. 9 der DSGVO oder Daten von minderjährigen Personen verarbeitet werden.

Verantwortungsvolle Nutzung von LLMoin

Die Nutzungsbedingungen ergeben sich aus den aktuellen und zukünftigen KI-relevanten Regulierungen. Die Nutzungsbedingungen definieren nicht nur das Verhalten der User in Bezug auf den Datenschutz, sondern auch in Bezug auf die zukünftige KI-Verordnung und auf die Hamburger KI-Leitlinien.

Die sichere und sachgerechte Nutzung von LLMoin wird durch verpflichtende Schulungen, regelmäßige Überprüfungen der Ergebnisse und zusätzliche Maßnahmen gewährleistet, sodass alle Beschäftigten verantwortungsvoll mit der neuen Technologie umgehen. Es wird strikt darauf geachtet, dass LLMoin nicht in Prozessen oder Entscheidungen eingesetzt wird, wo es ein potenzielles Risiko für die Bürgerinnen und Bürger darstellen könnte. Beispielsweise ist der Einsatz des Tools bei Entscheidungen über Sozialleistungen oder für die Evaluation von juristischen Argumenten klar untersagt. Jede wichtige Entscheidung bleibt weiterhin in der Verantwortung der zuständigen Mitarbeitenden. Die Nutzungsbedingungen definieren zusätzliche Regeln zum Umgang mit Fragen zum Schutz von Rechten Dritter und untersagt unethische Interaktionen.

Weitere Maßnahmen

Weitere technische und organisatorische Maßnahmen, welche die sichere, kompetente, korrekte und ethische Nutzung von LLMoin sicherstellen sind:

- Die aktuell genutzten GPT-Modelle haben starke interne Guardrails, die eine fehlerhafte oder unethische Nutzung schwer machen.

- “Content Filtering” ist eine zusätzliche Funktion, die in Azure OpenAI integriert ist und darauf abzielt, potenziell schädliche Inhalte in Eingabeprompts und Vervollständigungen zu erkennen und zu handeln.
- LLMoin wurde durch eine dedizierte Red-Teaming-Aktivität auf technische Robustheit und Sicherheit geprüft.
- Die Systemprompts von LLMoin achten besonders auf die Handlungsanweisungen und verweigern entsprechende ungewollte Anfragen.
- Es erscheinen Pop-ups, Disclaimer und Icons, welche User an die Limitationen von LLMs und an das in den Schulungen gelernte erinnern.
- Praktische Anleitungen und Hilfestellungen: In Form von kleinen Videos und Beispielen wird den Nutzern geholfen, LLMoin korrekt anzuwenden.
- User haben Zugang zu dezentralen Ansprechpersonen in den Behörden und der zentralen Fachlichen Leitstelle, die offene Fragen zu LLMoin schnell beantworten können.

Agile Umsetzung und zukünftige Entwicklungen

Das Projekt LLMoin wurde im Herbst 2023 ins Leben gerufen und von Beginn an nach agilen und userzentrierten Prinzipien entwickelt. Die kontinuierliche Einbindung von User-Feedback sowie die Berücksichtigung der spezifischen Anforderungen der Hamburger Verwaltung standen dabei im Mittelpunkt. Die Testuser in Form von circa 100 Mitarbeitenden während der Pilotphase setzte sich aus kleinen Gruppen aus unterschiedlichen Hamburger Behörden zusammen. Getrieben wurde die Entwicklung maßgeblich durch das Amt für IT und Digitalisierung (ITD) der Senatskanzlei, welches die Anforderungen definierte und die Testgruppen koordinierte.

Nach dem Abschluss des durch Capgemini erstellten Piloten konnten wichtige technische und organisatorische Weichenstellungen getroffen werden und nach einer Evaluationsphase wurde die Weiterentwicklung (MVP-Phase) beschlossen. Wichtige Entscheidungen zu diesem Zeitpunkt waren beispielsweise das geführte Prompting mit vier Funktionen, die dedizierte Positionierung als „ChatGPT für die Verwaltung“, die Umsetzung durch Dataport und der Cloud-Ansatz mit Fokus auf die Leistungsstärksten Modelle bei Microsoft Azure.

Die zweite Entwicklungsphase beinhaltete die Integration des bisherigen Piloten auf die GenAI-Infrastruktur von Dataport und eine weitere explorative Phase, in der User-Interviews, Umfragen und Designarbeit im Vordergrund standen. Mit einer stetig steigenden Zahl an Test-Usern konnten Erkenntnisse über den ersten Umgang mit dem Tool und diverse Änderungswünsche gesammelt werden. Dies erlaubte es der Projektleitung einen klaren Entwicklungsplan mit neuen Features, visuellen Änderungen und Bugfixes über die kommenden Monate umzusetzen und dabei immer wieder auf Feedback der Testenden durch direkte E-Mail-Kommunikation, Sprechstunden, Umfragen, und Nutzungsdaten zurückgreifen zu können. Die Tests während der gesamten Pilotierungs- und MVP-Phase erfolgten dabei noch vollständig ohne Nutzung personenbezogener Daten.

Ein weiterer wichtiger Schritt war die Erstellung initialer Schulungsunterlagen, in Form von Videos und einem Benutzerhandbuch. Dies ermöglichte es, viele Testnutzerinnen und -nutzer im Laufe des Jahres ohne weiteren Aufwand in die Testumgebung zu bringen und dabei detailliertes Feedback zu den Anforderungen und wichtigen Themen für das finale Lernkonzept zu sammeln. Das finale Lernkonzept besteht im Kern aus einem Lernpfad mit vielen Lernvideos, Übungsaufgaben und Quizformaten zu den Themen: LLMoin-Funktionen, KI-Grundlagen, LLM-Grundlagen, LLM-Limitationen, Prompting und vieles mehr. Schon während der sechsmonatigen Testphase konnten wir eine klare Stärkung der LLM-Kompetenzen erkennen, welche sich beispielsweise in den besseren Prompts und der vermehrten Nutzung der Funktion des „freien Promptings“ zeigten.

Die Zukunft von LLMoin

Hamburg plant fest mit der weiteren Nutzung und dem weiteren Ausbau von LLMoin. Nach dem erfolgreichen Rollout, das im Laufe von Q2 2025 finalisiert sein sollte, sollen die nächsten wichtigen Verbesserungen vorgenommen werden. Zur Planung stehen aktuell folgende neue Ideen zur Diskussion:

- Die Integration einer Prompt Library, sodass Prompts gespeichert und von anderen Usern gefunden werden können.
- Die Erhöhung des Schutzniveaus auf “Schutzbedarf hoch”, sodass in Zukunft auch hochsensible Daten verarbeitet werden können und alle Fachbereiche für LLMoin freigeschaltet werden können. Dies würde möglicherweise eine Integration von lokalen LLM beinhalten.
- Eine weitreichende Anknüpfung von Daten aller Behörden scheint sehr sinnvoll. Ein Pilotprojekt hat dies schon in einer Behörde erfolgreich umsetzen können.
- Das Anlegen von eigenen kleinen Assistenten verspricht auch einen großen Mehrwert. Hiermit könnten Mitarbeitende dann ihren eigenen Assistenten mit Systemprompt, Daten und weiteren agentischen Funktionalitäten erstellen und für ihren Arbeitsalltag nutzen.

Die fachliche Leitstelle von LLMoin ist sehr am Austausch zu diesen Themen interessiert. Zum einen können konzeptionelle Gedanken und potenziell auch Funktionen geteilt werden. Noch besser wäre es, wenn mehrere öffentliche Akteure an LLMoin arbeiten würden.

Nachnutzung von LLMoin durch Dritte

Der vielleicht wichtigste Aspekt der Zukunft von LLMoin ist die Nachnutzung von LLMoin durch andere Länder, Kommunen oder andere öffentliche Einrichtungen. Wir sind überzeugt, mit LLMoin ein sehr gutes Produkt für jede Einrichtung - als Startschuss in die Zukunft von generativer KI im öffentlichen Sektor - geschaffen zu haben. Die Freie und Hansestadt Hamburg begrüßt daher ausdrücklich die Nachnutzung von LLMoin über den IT-Dienstleister Dataport durch Dritte. Insbesondere Kommunen, Städte, Bundesländer sowie öffentliche Einrichtungen und Unternehmen können von dieser Lösung profitieren. Die breite Nutzung von LLMoin würde Skaleneffekte schaffen und Synergien bei zukünftigen Weiterentwicklungen ermöglichen. Mit F13 zusammen ist LLMoin das in Deutschland aktuell am weitesten gereifte Produkt und könnte in Zukunft das gemeinsame Projekt verschiedenster Akteure sein, die dann alle von gemeinsamen Verbesserungen profitieren.

Durch den Vertrieb über den IT-Dienstleister Dataport ist eine einfache Anpassung und schnelle Betriebsbereitschaft gewährleistet. Über das technische Produkt hinaus kann Hamburg umfangreiche Best Practices, Schulungsmaterialien und Dokumentationen bereitstellen, wie zum Beispiel Lerninhalte zu den Grundlagen, Chancen und Risiken großer Sprachmodelle (LLMs) oder die vielfach iterierten Nutzungsbedingungen. Die Hansestadt erhofft sich damit, den Transformationsprozess bei der Einführung von LLMoin bei Nachnutzenden zu erleichtern und regulatorische Anforderungen effizient zu bewältigen. Für weitere Informationen zur Nachnutzung oder um Testaccounts für LLMoin zu erhalten, stehen Hamburg und Dataport gerne zur Verfügung.

Des Weiteren ist klar, dass in Deutschland auf Länder- und auf Bundesebene unterschiedliche Lösungen verfolgt werden, die aktuell nicht koordiniert sind. Wir tauschen uns auf Arbeitsebene aktiv mit den Kolleginnen und Kollegen aus, die selbst auch an generischen LLM-Lösungen für die Verwaltung arbeiten. Ein gemeinsames Vorgehen und Interoperabilität von Komponenten sehen wir als erstrebenswert an und versuchen dies aktuell über den IT-Planungsrat und seine Kompetenzgruppen anzugehen.

Kontakt: Fachliche Leitstelle LLMoin
E-Mail: llmoin@sk.hamburg.de

Sören Alvermann
Michael Bornholdt
Nikolai Bock

20.11.2024, *Senatskanzlei Hamburg, Amt für IT und Digitalisierung,*