

Freie und Hansestadt Hamburg

Senatskanzlei

LLMoin - Anforderungen der DSGVO, KIVO, Barrierefreiheit und des Mitbestimmungsrechts

Version 0.5

14.03.2025

*Autor*in:*

Senatskanzlei

Version	Datum	Änderungen
0.1	19.07.2024	Erster Entwurf
0.2	05.08.2024	Erweiterungen um die Themen Mitbestimmung, Logs Speicherung und Governance
0.3	09.08.2024	Erweiterung des Titels auf andere Rechtsgebiete
0.4	02.09.2024	Anpassung und finale Fragen
0.5	25.11.2024	Umstellung der konzeptionellen Betrachtung (von dezentraler alleiniger Verantwortlichkeit hin zu getrennten Verantwortlichkeiten)
	26.11.2024	Ergänzung und Gliederung technischer/organisatorischer Maßnahmen
	29.11.2024	Korrektur im Abschnitt Konfigurationen LLMoin bzgl. des Content Filterings
	14.3.2025	Anpassung Handlungsanweisungen und Streichung Nutzungsdatenverarbeitung zu Analyse Zwecken

Summary

- Die mit LLMoin verarbeiteten Daten verbleiben innerhalb der EU und des Europäischen Wirtschaftsraums.
- Das im Rahmen von LLMoin verwendete KI-Modell (Microsoft Azure Open AI) speichert und analysiert keine Daten. Jede Eingabe („Prompt“) und Ausgabe („Antwort“) wird sofort nach der Erzeugung und Weiterleitung der generierten Antwort gelöscht.
- Daten der FHH werden nicht zum Training von KI-Modellen verwendet.
- Die Nutzung von LLMoin ist lediglich im normalen Risikobereich freigegeben. Unter anderem Daten nach Art. 9 und 10 der DSGVO, Daten Minderjähriger, Daten, die dem Sozial-, einem Berufs- oder besonderen Amtsgeheimnis unterliegen, sowie Personalakten dürfen nicht eingegeben werden.
- Für die Nutzung von LLMoin müssen datenschutzrechtliche Dokumentationen nicht angepasst oder ergänzt werden. Es müssen lediglich die Handlungsanweisungen LLMoin beachtet werden.

Technische Einzelheiten zu LLMoin

Flexibilität

LLMoin-Antworten basieren aktuell auf den neusten Modellen der GPT-Reihe von OpenAI¹. LLMoin ist Modelle-agnostisch konzipiert und kann die zugrundeliegenden Modelle schnell austauschen, falls OpenAI Modelle in Zukunft aus rechtlichen, wirtschaftlichen oder politischen Gründen nicht mehr die beste Option sind.²

Datenfluss

Eine LLMoin-Eingabe („Prompt“) wird von Dataport vorverarbeitet³ und an die europäischen 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 wird die Anfragen und Antworten niemals speichern.

Nutzungsdaten der Beschäftigten

Nach einer sogenannten Session speichert Dataport Nutzungsdaten für einen Zeitraum von zwei Monaten. Die Nutzungsdaten umfassen die Eingaben, LLMoin Antworten und Metadaten (Uhrzeit, Session ID, Funktionsart, Latenz und Fehleranzeigen). Eine Session wird beendet, falls Nutzende den Chat neu starten, sich ausloggen oder eine gewisse Zeit vergeht (wenige Minuten). Im Moment der Session kennt der Server die User ID (durch den Single-Sign-On Account des Nutzenden). Sobald die Session beendet wird, geht die User ID jedoch verloren.

Konfigurationen LLMoin

Um die Konfiguration von LLMoin zu verstehen, muss man sich zum größten Teil auf die Einstellungen von Microsoft Azure verweisen. Auf technischer Seite anpassbar sind folgende Eigenschaften:

- [Abuse Monitoring](#) (Temporäre Speicherung für Kontrollzwecke): Ist ausgeschaltet.
- [Content Filtering](#) (Echtzeit Prüfung der Modell Ausgaben): Ist eingeschaltet.
- Server Standort: Schweden Central.
- LLM-Modell: GPT-4o-2024-05-13.

¹ [Details zu GPT-Modellen bei Azure](#)

² Das [Diskussionspapier](#) des HmbBfDI zu LLMs gibt zu diesem Thema einige interessante Einblicke: Selbst wenn Modelle rechtswidrig trainiert worden sind, kann die Anwendung (Inferenz) rechtens sein.

³ Die Vorverarbeitung ist rein technischer Natur, so muss beispielsweise bei der Recherche die RAG Suchergebnisse an den Prompt als Referenzinformation angehängt werden.

- Embedding Modell (für RAG Embeddings): text-embedding-3-large
- Modell Temperatur (kann dynamisch angepasst werden): aktuell 0.3
- Systemprompt: Siehe „Konfigurationen und Maßnahmen.docx“.

Technische Sicherheitsmaßnahmen zur Risikobegrenzung

Prinzipiell werden die meisten „klassischen IT-Risiken“ durch die Governance-Prozesse und den sogenannten EHDB (Erstmalige Herstellung der Betriebsbereitschaft) durch Dataport adressiert und detailliert geprüft. LLMoin läuft im Dataport Rechenzentrum auf einer für hohen Schutzbedarf ausgelegten Umgebung. Der hohe Schutzbedarf wurde jedoch noch nicht formal geprüft. Das ITD21 hat zusätzlich einen Redteaming Bericht in Auftrag gegeben, um die spezifische IT-Risiken und neuen LLM-Risiken, die in direktem Zusammenhang mit der neuen Nutzung von LLMoin stehen, zu analysieren. Die Analyse und die Ergebnisse finden sich auf dem Sharepoint. Die wichtigsten Ergebnisse sind hier nach den Stellen, die die Maßnahmen ergreifen, getrennt zusammengefasst.

Technische Maßnahmen bei Microsoft

GPT Guardrails

Die Modelle, auf die LLMoin zugreifen wird, haben durch Innovationen auf Seiten der Hersteller (aktuelle OpenAI) immer stärkere und konsequente Guardrails und Nutzerbedingungen. Diese schützen die Modelle vor jeglichen Angriffen und Fehlnutzung jeglicher Art. OpenAI hat, vielleicht mehr als alle anderen Anbieter, in ihre Guardrails investiert - beispielsweise durch intensive Redteaming Arbeit. Die Guardrails stellen auf effektive Weise sicher, dass die Aspekte der Vielfalt, Nichtdiskriminierung und Fairness aus den KI-Leitlinien durch LLMoin eingehalten werden.

Azure Open AI Content Filter

[Content filtering](#) ist eine Funktion, die in Azure OpenAI integriert ist und darauf abzielt, potenziell schädliche Inhalte in Eingaben und Ausgabevollständigungen zu erkennen und zu handeln. Das System verwendet neuronale Mehrklassenklassifikationsmodelle, um Inhalte in vier Kategorien (Hass, Sexualität, Gewalt und Selbstverletzung) sowie vier Schweregraden (sicher, niedrig, mittel und hoch) zu identifizieren. Wir nutzen bei LLMoin die sicherste Einstellung. Bei einer solchen Identifikation wird die Anfrage oder die Antwort gestoppt und ein Error zurückgegeben. Aktuell ist die Content Filterung eingeschaltet.

Technische Maßnahmen bei Dataport

Systemprompts

Aktuell lässt LLMoin noch kein Systemprompt für das freie Prompting zu. Im Dezember 2024 wird dieses Feature verfügbar sein und dann für jede Funktion ein entsprechend sinnvoller Systemprompt hinzugefügt. Dieser wird sicherstellen, dass die Handlungsanweisungen befolgt werden und dass LLMoin entsprechende ungewollte Anfragen verweigert.

Blacklist Filter

Um das Recht auf Löschung und Widerspruch gegen Verarbeitung sicherzustellen, wird bei der Ausgabe eine Blacklist auf Namen geprüft. Dies verhindert die Erstellung von Informationen zu bestimmten Individuen, die nicht in den Ergebnissen von LLMoin erscheinen wollen. Aktuell ist die Liste leer, kann aber jederzeit gefüllt werden. Der Blacklist Filter wird Anfang von Q1 2025 zeitnah aktiviert werden.

WatsonX

WatsonX, die AI und Datenplattform von IBM, umfasst Governance- und Kontrollfunktionen, die Dataport für alle LLM-Applikationen nutzen kann. WatsonX.governance, ein wesentlicher Bestandteil dieser Plattform, bietet ein umfassendes Werkzeugset zum Risikomanagement, zur Transparenzsteigerung und Vorbereitung auf die Einhaltung künftiger AI-bezogener Vorschriften. Hauptfunktionen sind: LLM-Risikomanagement, Transparenzsteigerung, Vorbereitung auf Compliance, Governance und Überwachung und erhöhte Transparenz. Möglich hiermit ist z.B. ein Live Monitoring, welches bei allen LLM-Antworten auf Bias prüft und automatisch bei falscher Nutzung Alarm schlägt. Diese technischen Maßnahmen bestehen aufgrund der Priorisierung der Grundfunktionalitäten und des Rollouts aktuell noch nicht. Nach dem Rollout kann Dataport diese erweiterten Governance-Maßnahmen ergreifen.

Organisatorische Sicherheitsmaßnahmen zur Risikobegrenzung

Handlungsanweisungen

Den Nutzenden von LLMoin werden vor der Nutzung Handlungsanweisungen zur Verfügung gestellt und deren Relevanz durch das verpflichtende Schulungsmaterial kommuniziert. Die Handlungsanweisungen beschreiben, wie der KI-Textassistent LLMoin konkret eingesetzt werden darf und was dabei beachtet werden muss.

Die Handlungsanweisungen untersagen die Eingabe von Daten Minderjähriger, Daten nach Art. 9 und 10 DSGVO, Daten, die dem Sozial-, einem Berufs- oder besonderen Amtsgeheimnis unterliegen, sowie Personalaktendaten, soweit sie vertrauliche oder höchstpersönliche Daten darstellen.

Zudem sind nachfolgende Eingaben nicht erlaubt: Eingabe umfangreicher Datensätze, die zu unterschiedlichen Zwecken und/oder von verschiedenen für die Datenverarbeitung Verantwortlichen durchgeführt wurden (z.B. Baugenehmigungsantragsdaten und IFG

Anfragen, wobei die Baugenehmigungsantragsdaten vom Bezirksamt und die Daten aus der IFG Anfrage von der Behörde für Stadtentwicklung und Wohnen erhoben wurden); Eingabe von Daten aus Personenstands- und Melderegister oder Meldedaten mit Sperrvermerken gemäß § 51 Abs. 1 und 5 Bundesmeldegesetz oder Personenstandsdaten gemäß § 63 Personenstandsgesetz; Eingabe umfangreicher Daten im Rahmen der amtlichen Statistik für die Übermittlung an Dritte.

Schließlich definieren die Handlungsanweisungen die erlaubten Anwendungsbereiche nach der KIVO. Die Nutzenden müssen sicherzustellen, dass die erlaubten Anwendungsbereiche nicht überschritten werden. Die Motivation hinter dieser Beschränkung der Anwendungsbereiche folgt aus dem Willen, LLMoin als Low Risk Applikation nach der KIVO zu betreiben, die nicht nach Art. 5 KIVO verboten ist und nicht den erhöhten Anforderungen des Art. 6 KIVO unterliegt.

Die allgemein gültigen Handlungsanweisungen lassen sich wie folgt zusammenfassen:

Kompetenzen und Schulungen: Nutzende müssen die Grundlagen und Limitationen von LLMs kennen. Schulungsunterlagen bieten hierfür umfassende Informationen.

Anwendungsbereich: LLMoin ist für textbasierte Aufgaben im dienstlichen Kontext nutzbar. Die Nutzung in Hochrisiko-Bereichen ist untersagt, für Grenzfälle werden im Anhang der Handlungsanweisungen detaillierte Ausführungen gemacht.

Verantwortungsbewusste Nutzung: LLMoin darf keine Entscheidungen anstelle eines Menschen treffen. Generierte Ergebnisse müssen stets überprüft und angepasst werden. Dabei ist sicherzustellen, dass keine Rechte Dritter verletzt werden (z.B. Urheberrechte, Betriebs- und Geschäftsgeheimnisse etc.).

Eigenständige Übernahme:

Wichtig: Der Roll-out von LLMoin erfolgt auf freiwilliger Basis und nur als Low Risk Anwendung. Sobald sich Fachbehörden für die Nutzung von LLMoin entscheiden, müssen sie die Handlungsanweisungen als eigene Dienstanweisung für ihre Beschäftigten übernehmen. Jede Fachbehörde muss bei Bedarf im Rahmen der lokalen Ergänzung und Präzisierung der Handlungsanweisung den Nutzenden fachspezifische Hinweise geben, wie LLMoin genutzt werden darf. Dies betrifft insbesondere die in Anhang III der KIVO und in dem Anhang der Handlungsanweisung genannten Bereiche, in denen eine differenzierte Betrachtung der Einsatzbereiche und Aufgaben von LLMoin erforderlich wird. Für solche Ergänzungen und Präzisierungen enthalten die Handlungsanweisungen einen Platzhalter, der von den Fachbehörden und -bereichen genutzt werden kann.

Schulung der Handlungsanweisungen

Durch die verpflichtenden und ausführlichen Schulungen vor der Nutzung von LLMoin wird Mitarbeitenden klar gemacht, dass Sie für die korrekte Nutzung von LLMoin und für Ergebnisse und das Einhalten von Richtlinien (z.B. Datenschutz) verantwortlich sind. Sollte es dezentrale Risiken geben, die in bestimmten Bereichen eine besonders sensible

Nutzung von LLMoin voraussetzen, werden diese durch die lokalen Handlungsanweisungen abgedeckt und die Mitarbeitenden dezentral sensibilisiert.

Rolle der Kapitä*Innen

Jeder Fachbereich wird mehrere interne "LLMoin-Kapitäne" benennen, die beispielsweise neue Mitarbeitende einweisen, dezentrale Upskillings durchführen und für fachspezifische Nutzungsfragen zur Verfügung stehen. An mehreren Stellen werden Nutzende auf diese Möglichkeit der Unterstützung hingewiesen und bei Fragen zu Datenschutz und Handlungsanweisungen sollen sie direkt zu ihren lokalen Ansprechpersonen gehen, welche beim Thema LLMoin von uns ausgebildet sind.

Sensibilisierung für Limitationen von LLMs

Nutzende werden durch die Lernunterlagen in allgemeine Konzepte von LLMs eingeführt und explizit auf die Limitationen und Gefahren hingewiesen. Darunter fallen grundlegende Erklärungen von Halluzinationen, Biases und andere Nachteile von LLMs sowie konkrete Hinweise und Handlungsanweisungen zum Umgang und zur Minimierung dieser Nachteile (Beispiel: Keine Wissensfragen an offenes Prompting ohne relevante Dokumente im Hintergrund).

Erinnerungen in LLMoin

LLMoin enthält viele Pop-ups, Disclaimer und Icons, welche User an die Limitationen von LLMs und an das in den Schulungen Gelernte erinnern. Weiter gibt es viele Instruktionen und Nutzungsbeispiele in LLMoin selbst, sodass alle Nutzer:innen grundlegende Kompetenzen erlernen, egal wie oft oder genau sie die Schulungsmaterialien gesichtet haben.

Mitbestimmung

Derzeit wird die Vereinbarung nach § 94 HmbPersVG Bürokommunikation überarbeitet und um KI-gestützte Assistenten und Unterstützungsmittel, wie LLMoin, ergänzt. Es ist wahrscheinlich, dass die Vereinbarung abgeschlossen werden wird, allerdings nicht mehr vor dem Rollout von LLMoin.

Freiwilliges und nicht integriertes Werkzeug

LLMoin ist ein freiwilliges Werkzeug, dessen Nutzung in Fachverfahren nicht vorgegeben ist und auch nicht im Übrigen in den Arbeitsalltag von Mitarbeitenden eingreift, wenn diese LLMoin nicht nutzen wollen. LLMoin ist außerdem eine Webseite, in die Inhalte kopiert und wieder herauskopiert werden – ähnlich wie beim Übersetzungstool DeepL.

Transparenz und Kommunikation

Der Mehrwert und die Motivation von LLMoin werden klar kommuniziert, Mitarbeitende werden durch die Schulungen darüber informiert, wofür sie das Werkzeug nutzen können, welche Art von Daten gespeichert werden, wie sie das Tool nutzen dürfen und viele weitere Aspekte, die einen geschulten und verantwortlichen Umgang ermöglichen.

Schulungen und Kompetenzen

Ein umfangreicher Lernplan wird vor der Nutzung von LLMoin angeboten, um sicherzustellen, dass alle Nutzenden die Funktionsweise und Limitationen des KI-Tools verstehen. Bei den Schulungen wird besonders auf Nutzende mit weniger KI-Vorwissen eingegangen und damit eine gerechte und gleiche Grundlage für alle geschaffen. Umfassende Ressourcen werden beim Rollout als Support bereitgestellt, um Fragen zu beantworten.

Barrierefreiheit und Software-Ergonomie

LLMoin wurde von Dataport entwickelt, die sich mit Barrierefreiheit gut auskennen. Im Juli 2024 wurde ein erster Bericht zur Barrierefreiheit von LLMoin erstellt. In diesem wurde die teilweise Barrierefreiheit von LLMoin befunden, aber noch viele Verbesserungen vorgeschlagen. Da im August 2024 noch kleine Anpassungen an das Frontend der Applikation durchgeführt worden sind, ist der Bericht vom Juli leicht veraltet. Um den Rollout von LLMoin nicht weiter zu verzögern, haben wir entschieden, die **volle Erfüllung der Barrierefreiheit und Software-Ergonomie stufenweise in den kommenden Monaten sicherzustellen**. Mit Dataport wurde schon geklärt, dass im November und Dezember dieses Jahres die nötigen Anpassungen für eine volle Barrierefreiheit umgesetzt werden. Weitere Details zur Mitbestimmung wurden in der FITPRAG Ende August vorgestellt. Eine Erklärung zur Barrierefreiheit gem. § 11 Abs. 1 Hamburgisches Behindertengleichstellungsgesetz in Verbindung mit §2 Hamburgische Barrierefreie Informationstechnik-Verordnung stellt ITD zentral bereit.