

## KnowHow

Wissensmanagement für den diplomatischen Dienst

<b>Programm / Ausschreibung</b>	KIRAS, F&E-Dienstleistungen, KIRAS F&E-Dienstleistungen (FED) 2023	<b>Status</b>	laufend
<b>Projektstart</b>	01.10.2024	<b>Projektende</b>	31.03.2026
<b>Zeitraum</b>	2024 - 2026	<b>Projektlaufzeit</b>	18 Monate
<b>Keywords</b>	Wissensmanagement, Large Language Models, Generative KI		

### Projektbeschreibung

Das Ziel dieser vorgeschlagenen Studie ist es, zu untersuchen, wie die Anwendung von modernen maschinellen Lerntechniken, semantischer Suche und großen Sprachmodellen (LLMs) ein effizientes Wissensmanagementsystem für das Bundesministerium für europäische und internationale Angelegenheiten (BMEIA) unterstützen könnte. Die Motivation für dieses Projekt ist die Herausforderung für Behörden, große Mengen an strukturierten und unstrukturierten Daten zu verwalten, um wertvolle Erkenntnisse zu gewinnen. Derzeit verfügt das BMEIA über kein internes Wissensmanagementsystem und wichtige digitale Objekte sind in verschiedenen Dateifreigaben in unterschiedlichen Formaten gespeichert. Herkömmliche Informationsmanagementsysteme haben oft Schwierigkeiten, relevante Informationen effizient zu verarbeiten und abzurufen, was zu Ineffizienz und verpassten Analysemöglichkeiten führt. Zur Durchführung dieser Evaluierung und als Grundlage für die Studie wird ein Prototyp eines Wissensmanagementsystems auf Basis neuester Ansätze aus dem Bereich der Künstlichen Intelligenz entwickelt, implementiert und von Wissensarbeitern im BMEIA getestet.

Das vorgeschlagene System wird sich auf drei Hauptbereiche der KI-Anwendung konzentrieren: 1) die Verwendung von Transformationsmodellen in Kombination mit semantischer Indizierung, um eine semantische Suche und Abfrage zu ermöglichen und so die Suchgenauigkeit und -relevanz, die Benutzerfreundlichkeit und den Zugang zu relevanten Informationen zu verbessern; 2) die Schaffung von Wissensstrukturen aus unstrukturierten Informationen durch die Anwendung von Modellen des maschinellen Lernens, um Entitäten und Beziehungen zu extrahieren, was zu einer verbesserten Organisation, Analyse und Abfrage von Informationen führt; und 3) die Evaluierung von Entscheidungsunterstützungstechnologien durch die Implementierung von Ansätzen zur semantischen Anreicherung wie Retrieval Augmented Generation (RAG), die natürlichsprachliche Abfragen und Antworten bei der Übernahme von Analyseaufgaben ermöglichen.

Ein wesentlicher Aspekt dieses Projekts ist die Sicherstellung der ethischen und rechtlichen Konformität mit den europäischen Best Practices für vertrauenswürdige KI, den Datenschutzbestimmungen und den künftigen Bestimmungen des KI-Gesetzes, was es zu einem wertvollen und zeitgemäßen Beitrag im Bereich KI-gestützter Wissensmanagementsysteme in öffentlichen Einrichtungen macht.

## **Abstract**

The aim of this proposed study is to investigate how the application of modern machine learning techniques, semantic search and large language models (LLMs) could support an efficient knowledge management system for the Federal Ministry for European and International Affairs (BMEIA). The motivation for this project is the challenge for public authorities to manage large amounts of structured and unstructured data in order to gain valuable insights. Currently, the BMEIA does not have an internal knowledge management system and important digital objects are stored in different file shares in different formats. Traditional information management systems often struggle to efficiently process and retrieve relevant information, leading to inefficiencies and missed opportunities for analysis. To carry out this evaluation and as a basis for the study, a prototype of a knowledge management system based on the latest approaches from the field of artificial intelligence will be developed, implemented and tested by knowledge workers at the MFA.

The proposed system will focus on three main areas of AI application: 1) the use of transformation models in combination with semantic indexing to enable semantic search and retrieval to improve search accuracy and relevance, ease of use, and access to relevant information; 2) the creation of knowledge structures from unstructured information by applying machine learning models to extract entities and relationships, resulting in improved organization, analysis, and retrieval of information; and 3) the evaluation of decision support technologies through the implementation of semantic enrichment approaches such as Retrieval Augmented Generation (RAG), which enable natural language queries and responses when undertaking analysis tasks.

A key aspect of this project is to ensure ethical and legal compliance with European best practices for trustworthy AI, data protection regulations and the future provisions of the AI Act, making it a valuable and timely contribution in the field of AI-powered knowledge management systems in public institutions.

## **Endberichtkurzfassung**

Die Studie zum Einsatz von KI-basierten Wissensmanagementsystemen im BMEIA kommt insgesamt zu dem Ergebnis, dass der Aufbau eines KI-unterstützten Wissensmanagementsystems fachlich und technisch sinnvoll ist, dass sich der Nutzen aber nicht allein an einer pauschalen Zeitersparnis festmachen lässt. Ausgangspunkt der Untersuchung war die Beobachtung, dass im BMEIA ein übergreifendes internes Wissensmanagementsystem fehlt und relevante digitale Informationen in unterschiedlichen Dateifreigaben, Formaten und Dokumentstrukturen vorliegen. Ziel der Studie war daher, das Potenzial moderner KI-Verfahren – insbesondere semantischer Suche, großer Sprachmodelle und Wissensgraphen – für die Unterstützung typischer Wissensaufgaben im BMEIA zu analysieren und anhand eines Demonstrators praktisch zu evaluieren.

Die ersten Abschnitte der Studie schaffen dafür eine klare Grundlage. Der Literaturüberblick zeigt, dass insbesondere Retrieval-Augmented Generation (RAG), semantische Indexierung in Vektordatenbanken domänenspezifische Metadaten-Anreicherung und Wissensgraph-basierte Ansätze für semi-strukturierte Dokumentbestände als vielversprechend gelten. Zugleich wird deutlich, dass der Erfolg solcher Systeme stark von der Qualität der Vorverarbeitung, der Strukturierung des Korpus und der Einbettung in organisatorische Arbeitsprozesse abhängt. Die Marktanalyse kommt ergänzend zu dem Schluss, dass keine der untersuchten bestehenden Lösungen die Anforderungen des BMEIA vollständig erfüllt. Besonders die Defizite bei interner Datenhaltung, fortgeschrittenem Inputmanagement, OCR, Metadatenanreicherung und dokumentübergreifender Suche sprechen gegen den Einsatz eines Standardprodukts und für die Entwicklung eines

maßgeschneiderten Prototyps.

Die Gap-Analyse präzisiert diese Ausgangslage für das BMEIA. Als zentrale Schwächen des bestehenden Intranets werden die dokumentbasierte statt inhaltsbasierter Struktur, die begrenzte Benutzerfreundlichkeit der Suche, uneinheitlich gepflegte Metadaten, heterogene Datenquellen sowie Probleme mit Versionen, Duplikaten und Gültigkeiten identifiziert. Daraus leitet die Studie ab, dass eine Verbesserung nicht nur auf technischer Ebene, sondern auch auf Prozess- und Organisationsebene erforderlich ist. KI-basierte Lösungen werden dabei vor allem als Hebel für semantische bzw. hybride Suche, natürlichsprachliche Anfragen, Metadatenextraktion, OCR-Verarbeitung, Themenstrukturierung und Dokumentvergleich verstanden. Schon in diesen frühen Abschnitten wird damit deutlich, dass das KnowHow-Tool nicht als isolierte Einzellösung gedacht ist, sondern als Fundament einer umfassenderen Transformation des Wissensmanagements im BMEIA.

Auf dieser Basis wurden in der Anforderungsanalyse und in der Systemspezifikation die fachlichen und technischen Anforderungen des Demonstrators in Form von User Stories und Systemkomponenten konkretisiert. Im Mittelpunkt stehen das Hinzufügen und Erschließen von Dokumenten, die Suche nach Dokumenten, natürlichsprachliche Anfragen, dialogische Weiterverarbeitung von Treffern sowie die Unterstützung durch Quellen, Metadaten und Abbilden von Referenzen zwischen Dokumenten. In der Praxis wurde dafür ein Demonstrator aufgebaut, der auf einem RAG-System mit OpenSearch, LlamaIndex und LlamaDeploy basiert und durch einen Wissensgraphen auf Neo4j-Basis ergänzt wird. Die Benutzeroberfläche wurde so gestaltet, dass klassische Volltextsuche, semantische Suche, hybride Suche, Reranking, Kontextauswahl und RAG-gestützte Antwortgenerierung und Wissensgraph Anbindung in einer zusammenhängenden Umgebung nutzbar sind. Damit zeigt die Studie bereits vor der eigentlichen Evaluierung, dass ein funktionsfähiger und modularer Prototyp entwickelt wurde, der wesentliche Zielsetzungen der Anforderungsanalyse technisch abbildet.

Für die Evaluierung wurde ein A/B-Design gewählt, in dem das bestehende Intranet-System mit dem KnowHow-Tool verglichen wurde. Das Ziel war, Unterschiede in Bearbeitungszeit und Nutzerwahrnehmung bei realitätsnahen Wissensaufgaben festzustellen. Dafür wurden 25 Test-Cases aus drei Aufgabentypen definiert: Dokument-Suche, Inhalts-Suche und Text-Erstellung. Die Evaluierungsumgebung wurde bewusst als vom KnowHow-Tool unabhängige Software umgesetzt, um Verzerrungen zu vermeiden. Für offene Antworten in Inhalts- und Text-Erstellungsaufgaben wurde zusätzlich ein LLM-basiertes Bewertungssystem entwickelt, das Antworten Kriterien-basiert und nachvollziehbar auswertet. Auch dieser Teil der Studie ist ein wichtiges Ergebnis: Er zeigt, dass nicht nur ein Such- und Antwortsystem, sondern auch eine eigenständige, skalierbare Evaluierungsinfrastruktur aufgebaut wurde, mit der komplexe Nutzungsdaten und qualitative Rückmeldungen systematisch erhoben werden konnten.

Der wichtigste Teil der Studie liegt jedoch in Abschnitt 9 mit den Evaluierungsergebnissen. Die tatsächliche Studienbeteiligung lag etwas unter der ursprünglichen Planung: Statt der vorgesehenen 32 Personen pro Gruppe nahmen je 28 Personen an der Studie teil, jeweils 15 Frauen und 13 Männer. Der Studienzeitraum wurde verlängert und durch mehrere Einführungstermine, Schulungsunterlagen und Videoaufzeichnungen begleitet. Für die Analyse mussten einige Daten bereinigt werden, unter anderem weil ein Dokument in einem Testfall zunächst im KnowHow-Tool nicht indexiert war und dieser Fall deshalb aus der Auswertung ausgeschlossen wurde. Nach der Bereinigung verblieben 986 Bearbeitungsvorgänge in der Kontrollgruppe A und 1139 in der KnowHow-Gruppe (oder Testgruppe) B; etwa sieben Prozent der aggregierten Zeitmessungen waren zensierte Beobachtungen, also Fälle, in denen ein Erfolg trotz mehrfacher Versuche nicht erreicht

wurde. Diese Ausgangslage führte zu einer methodisch anspruchsvollen Evaluierungssituation.

Im Gesamtbild der Bearbeitungszeiten ergibt sich kein klarer allgemeiner Effizienzvorteil des KnowHow-Tools. Die aggregierten Zeitmessungen der beiden Gruppen sind ähnlich verteilt, wobei die Medianzeiten der Gruppe B leicht über jenen der Gruppe A liegen. Betrachtet man alle Zeitmessungen, beträgt der Median in Gruppe A 44 Sekunden und in Gruppe B 46,5 Sekunden; berücksichtigt man nur gelöste Aufgaben, liegt der Median bei 41 Sekunden in Gruppe A und 45 Sekunden in Gruppe B. Auch die Mittelwerte liegen in Gruppe B höher. Damit konnte die Grundannahme, dass das neue KI-System im Durchschnitt über alle Aufgaben hinweg zu einer generellen Beschleunigung führt, in der deskriptiven Auswertung nicht bestätigt werden. Die Studie zeigt somit, dass die Leistungsfähigkeit des Systems differenziert und nicht über einen einfachen globalen Durchschnitt zu beurteilen ist.

Besonders aufschlussreich ist die Aufschlüsselung nach Aufgabentypen. Hier zeigt sich ein differenziertes Muster. Bei der Dokument-Suche schneidet das KnowHow-Tool besser ab als das bestehende System. Die Zeitmessungen der Gruppe B liegen hier tendenziell niedriger; bei den gelösten Aufgaben beträgt der Median 32 Sekunden gegenüber 34 Sekunden in der Kontrollgruppe. Die Unterschiede sind zwar moderat ausgeprägt, weisen jedoch eine konsistente Richtung auf und werden durch die Zeitereignisanalyse gestützt, in der sich für die KnowHow-Gruppe insbesondere bei Dokument-Suche-Aufgaben eine günstige Tendenz zeigt. Anders ist das Bild bei der Inhalts-Suche: Hier ist die Kontrollgruppe klar schneller. Der Median der gelösten Aufgaben liegt in Gruppe A bei 53 Sekunden, in Gruppe B dagegen bei 87,5 Sekunden. Noch deutlicher ist die Differenz bei den Mittelwerten. Auch bei den Text-Erstellungsaufgaben ergibt sich kein Vorteil des KI-Systems; vielmehr liegen die Bearbeitungszeiten der Gruppe B höher und streuen stärker. Das KnowHow-Tool zeigt seine Stärke damit vor allem dort, wo die intelligente Suche und das Auffinden relevanter Dokumente im Vordergrund stehen. Sobald dagegen vertiefte inhaltliche Extraktion oder textuelle Weiterverarbeitung gefragt sind, führt der neue Ansatz in der vorliegenden Ausbaustufe noch nicht zu einem Zeitgewinn.

Ein ähnliches Bild ergibt sich bei Erfolgsraten und Erstversucherfolgen. Die Erfolgsrate beim ersten Versuch ist in der Kontrollgruppe höher als in der KnowHow-Gruppe; insgesamt liegt sie bei 0,71 in Gruppe A und bei 0,67 in Gruppe B. Auch die allgemeine Erfolgsrate ist in der Kontrollgruppe leicht höher, nämlich 0,94 gegenüber 0,92. Betrachtet man die Bearbeitungsvorgänge nach Aufgabentyp, zeigt sich zudem, dass Gruppe B zwar insgesamt aktiver war und mehr Versuche unternahm, aber bei Dokument-Suche und Inhalts-Suche geringere Anteile korrekt gelöster Versuche aufwies als Gruppe A. Nur bei Text-Generationsaufgaben lag der Anteil gelöster Versuche in Gruppe B höher. Diese Befunde sprechen dafür, dass das KnowHow-Tool nicht einfach „leichter“ oder „effizienter“ zu bedienen war, sondern in bestimmten Situationen offenbar mehr Exploration, mehr Versuche oder mehr Einarbeitung erforderte.

Die explorativen Auswertungen zu Gender und Dienstalter ergänzen dieses Bild, ohne harte kausale Schlüsse zu erlauben. Für die Gesamtzeiten zeigt sich in Gruppe B ein Vorteil von Frauen gegenüber Männern, der in der Kontrollgruppe so nicht sichtbar ist. Besonders bei der Dokument-Suche fallen die Zeitmessungen der Frauen in der KnowHow-Gruppe niedriger aus als jene der Männer und auch niedriger als in der Kontrollgruppe. Dies kann als möglicher Hinweis gesehen werden, dass Frauen von der intelligenten Suche profitieren würden, jedoch ist die Aussagekraft der kleineren Untergruppen beschränkt. Auch die vorab angenommene Erwartung, dass vor allem dienstjüngere Personen von der KI-gestützten Suche profitieren oder sehr erfahrene Personen im Intranet Vorteile haben müssten, konnte nicht bestätigt werden. Die mittlere Dienstalter-Gruppe von fünf bis zwanzig Jahren weist in beiden Gruppen die niedrigsten Medianwerte auf. Insgesamt stützen diese

Befunde eher die Aussage, dass individuelle Unterschiede vorhanden sind, aber nicht in einfacher Form entlang klassischer Erwartungsmuster verlaufen.

Von Bedeutung ist auch, dass im Zeitverlauf keine ausgeprägten allgemeinen Lerneffekte nachgewiesen werden konnten. Im Rahmen der Datenexploration wurde untersucht, ob Nutzerinnen und Nutzer mit zunehmender Erfahrung im Umgang mit dem KnowHow-Tool schneller werden. Ein klarer, anhaltender Abwärtstrend der Bearbeitungszeiten ist jedoch nicht erkennbar. Lediglich für Dokument-Suche-Aufgaben in Gruppe B wird eine mögliche Lernphase zu Beginn vermutet. Das spricht dafür, dass die beobachteten Unterschiede weniger durch bloße Gewöhnung erklärt werden können als durch die Eigenarten der jeweiligen Aufgabentypen und der Systemunterstützung.

Die Freitextantworten der Nutzerinnen und Nutzer zeichnen insgesamt ein differenziertes, aber überwiegend konstruktives Bild der praktischen Nutzung des KnowHow-Tools. Dabei zeigt sich, dass die Wahrnehmung stark vom jeweiligen Anwendungskontext abhängt. Während einzelne Rückmeldungen die Effizienz und einfache Bedienbarkeit des Tools hervorheben, insbesondere bei klar formulierten Suchanfragen, wird zugleich deutlich, dass bei einfachen Informationsbedürfnissen nicht immer ein klarer Mehrwert gegenüber bestehenden Systemen wie dem Intranet wahrgenommen wird. Besonders zentral ist die Suchqualität, die von vielen Nutzerinnen und Nutzern als verbesserungsbedürftig beschrieben wird, etwa aufgrund unvollständiger, schwer nachvollziehbarer oder nicht unmittelbar relevanter Ergebnisse. Auch in Bezug auf die Bedienbarkeit ergibt sich ein gemischtes Bild: Das Tool wird grundsätzlich als erlernbar und nach einer gewissen Eingewöhnung als hilfreich eingeschätzt, gleichzeitig wird jedoch der Einstieg teilweise als komplex und die Benutzeroberfläche als zu wenig intuitiv empfunden. Ein wesentliches Thema stellt zudem das Vertrauen in die bereitgestellten Informationen dar. Die Rückmeldungen zeigen, dass Ergebnisse häufig zusätzlich überprüft werden, was auf ein bestehendes Bedürfnis nach höherer Transparenz und Nachvollziehbarkeit der Antworten hinweist. Hinsichtlich des wahrgenommenen Nutzens wird deutlich, dass das Tool in bestimmten Situationen, insbesondere bei klar definierten Suchaufgaben, unterstützend wirkt, jedoch nicht durchgängig als überlegen gegenüber anderen Informationsquellen empfunden wird. Der identifizierte Verbesserungsbedarf konzentriert sich daher insbesondere auf die Steigerung der Suchqualität, die bessere Aufbereitung und Darstellung von Ergebnissen sowie die Erweiterung der zugrunde liegenden Datenquellen. Insgesamt liefern die Freitextantworten wertvolle Hinweise für die Weiterentwicklung des Systems und unterstreichen, dass bereits eine solide Basis vorhanden ist, deren Potenzial durch gezielte Optimierungsmaßnahmen weiter ausgeschöpft werden kann.

In den Auswahlfragen bewerteten die Nutzer der Testgruppe viele Aspekte des Systems besser als die der Kontrollgruppe. Beide Gruppen gaben an, dass sie sich darauf freuen würden, in Zukunft KI-basierte Systeme zu nutzen.

Die Ergebnisse zeigen eine insgesamt positive Bewertung des KnowHow-Tools. Gruppe B schätzte die Benutzerfreundlichkeit, Übersichtlichkeit und Effizienz besser als die Kontrollgruppe und kam nach subjektiver Einschätzung signifikant häufiger schneller ans Ziel. Das Vertrauen in Gruppe B war leicht geringer und die Tendenz zur Gegenprüfung höher. Gruppe B bevorzugte mehrheitlich das KnowHow-Tool und zeigte eine höhere Nutzungsbereitschaft, Gruppe A wünschte sich klar eine KI-gestützte Suche.

Frauen bewerteten beide Systeme tendenziell kritischer als Männer, insbesondere das Intranet. Sie berichteten häufiger von Umwegen, geringerem Vertrauen und bewerteten die Ergebnisse als weniger relevant. Während Männer das KnowHow-Tool

klar bevorzugten, waren Frauen häufiger unentschieden, äußerten jedoch zugleich einen stärkeren Wunsch nach KI-gestützter Suche in der Kontrollgruppe.

Insgesamt wird das KnowHow-Tool als effizient wahrgenommen, mit Optimierungspotenzial insbesondere im Vertrauen.

Zusammengenommen ergibt sich somit ein sehr differenziertes Hauptergebnis: Das KnowHow-Tool ist kein System, das in jeder Hinsicht und über alle Aufgaben hinweg dem bestehenden Intranet überlegen wäre. Ein genereller Zeitvorteil konnte nicht nachgewiesen werden. Auch ein genereller Effizienzverlust blieb aus, sodass ein in kurzer Zeit als Proof-of-Concept entwickeltes Tool trotz kaum vorhandener Erfahrung der Nutzerinnen und Nutzer bereits ähnliche Effizienz im Vergleich zum BMEIA-Intranet zeigte. Gleichzeitig zeigen die Ergebnisse aber klar, dass der Ansatz selektiv Stärken besitzt. Insbesondere bei der Dokument-Suche deuten die Ergebnisse auf einen Vorteil der intelligenten, semantisch erweiterten Suche hin. Die Schwächen bei Inhalts-Suche und Text-Erstellung legen dagegen nahe, dass die Übergänge von Suche zu inhaltlicher Extraktion und generativer Verarbeitung noch nicht in einer Weise ausgereift sind, dass sie im Arbeitsalltag zuverlässig Zeit sparen. Als zentrales Ergebnis zeigt die Studie, in welchen Nutzungsszenarien die eingesetzten Verfahren bereits Mehrwert bringen und wo weiterer Entwicklungsbedarf besteht.

In der Gesamtschau der Studie lässt sich daher festhalten, dass die Studie die fachliche Relevanz des Vorhabens eindeutig bestätigt. Die vorangestellten Analysen zeigen einen realen Bedarf im BMEIA, die Marktanalyse rechtfertigt die Eigenentwicklung, und der Demonstrator belegt die technische Machbarkeit einer integrierten Lösung aus semantischer Suche, RAG und Wissensgraph. Die Evaluierung relativiert jedoch die Erwartung eines unmittelbaren, flächendeckenden Effizienzgewinns. Der Mehrwert des KnowHow-Tools liegt nach den vorliegenden Ergebnissen nicht in einer pauschalen Beschleunigung aller Wissensarbeiten, sondern in der gezielten Verbesserung bestimmter Such- und Erschließungsaufgaben und in der Perspektive auf ein weiterentwickelbares, auf die Domäne des BMEIA zugeschnittenes Wissensmanagementsystem. Die Studie macht damit zugleich deutlich, dass technische Innovation allein nicht ausreicht: Der nachhaltige Nutzen hängt ebenso von Datenqualität, Metadatenpflege, Prozessharmonisierung, redaktionellen Standards und dem Umgang mit Gültigkeit, Versionierung und Sicherheitsanforderungen ab.

Als Ergebniskurzfassung lässt sich deshalb zusammenfassend sagen: Die Studie zeigt, dass KI-basierte Wissensmanagementsysteme für das BMEIA ein substanzielles Potenzial besitzen und mit dem KnowHow-Tool bereits prototypisch in einer technisch belastbaren Form umgesetzt werden konnte. Eine generelle Überlegenheit gegenüber dem bestehenden System wurde in der Evaluierung zwar nicht nachgewiesen, wohl aber ein differenziertes Stärkenprofil, insbesondere bei der Dokument-Suche. Das Projekt liefert damit keinen Endpunkt, sondern eine belastbare Grundlage für die gezielte Weiterentwicklung eines domänenspezifischen, sicheren und organisatorisch eingebetteten Wissensmanagementsystems im BMEIA.

## **Projektkoordinator**

- AIT Austrian Institute of Technology GmbH

## **Projektpartner**

- Bundesministerium für europäische und internationale Angelegenheiten
- CIB solutions GmbH