

1. Herrn AL 1 [RP, 30.4.](#)

Herrn AL Z sch, 29.04.2024

über

Frau UAL'in 12 EU 29/4

Herrn UAL Z2, [Lee 29.4.; Votum wird unter der notwendigen Voraussetzung mitgetragen, dass die Finanzierung durch das BMI erfolgt; eine finanzielle Beteiligung/Commitment zu so einem frühen Zeitpunkt ist wg. der Ungewissheit der Fortentwicklung zurückhaltend zu bewerten.](#)

Herrn RL 121 [mmh, 26.4.](#)

Herrn RL Z22 [Pa 23/4](#)

mit der Bitte um Kenntnisnahme

Betr.: **Einsatz von Künstlicher Intelligenz (KI) im BMBF**

hier: Beteiligung an der ressortübergreifenden KI-Infrastruktur und KI-Plattform des ITZ Bund

2. Vermerk

Sachverhalt

In der St Runde Digitalpolitik am 29.02.2024 wurde unter TOP 2.3 „KI in der Bundesverwaltung“ vereinbart, für die weitere Koordinierung einer zentralen Infrastruktur die Unterarbeitsgruppe „LLM (Large Language Model) -Applikationen in der Bundesverwaltung“ unterhalb der IMAG „Datenlabore“ zu gründen. Z22 ist darin aufgrund seiner koordinierenden Rolle vertreten; das Datenlabor nimmt stellvertretend bzw. bei technischem Fokus an den Sitzungen teil.

Unter Federführung von BKAm und BMI wird darin ein gemeinsames Vorgehen zur Bereitstellung von KI-Infrastruktur und KI-Applikationen für die Bundesverwaltung angestrebt, um der Aufforderung der St Runde nachzukommen.

Zu diesem Zweck ist vorgesehen, einen Auftrag des Ressortkreises an das ITZ Bund zu erteilen und sich an dem dort in Umsetzung befindlichen Projekt „KIPITZ“ (KI Portal ITZ Bund) zu beteiligen.

Das Vorhaben sieht ein vierstufiges Verfahren vor, durch welches zunächst kurzfristig über eine Public-Cloud und langfristig im Betrieb auf eigener Hardware des ITZ Bund eine KI-Plattform mit zentralen Anwendungen bereitgestellt werden könnte.

In den ersten drei Stufen soll ab dem zweiten Quartal 2024 über die Azure-Cloud (Microsoft) eine Vorabversion der Plattform entwickelt werden, die mit verschiedenen Sprachmodellen wie GPT 4 oder Mixtral einfache Textverarbeitungsfunktionen wie bspw. Zusammenfassungen

oder Textgenerierung ermöglicht. Zudem können ressortspezifische Use-Cases unter Einbindung einer Wissensdatenbank geschaffen werden und es soll eine Entwicklungsumgebung für Datenlabore bereitgestellt werden. Die Plattform soll ab dem dritten Quartal zur Pilotierung bereitstehen. Nach Beschaffung der notwendigen Hardware soll voraussichtlich ab Anfang des Jahres 2025 mit Stufe vier die Vollversion der Plattform im vor Ort-Betrieb („on premise“) im ITZ Bund folgen.

Der Mehrwert für die Bundesverwaltung steigt mit jeder Stufe der Portalumsetzung

Stufe	Mehrwert BV	Funktionalitäten
1 „Generalwissen“-LLM		– Bereitstellung Sprachmodelle (z. B. GPT4, Mixtral , Luminous) via API und dazugehörige Infrastruktur für öffentliche Daten
2 „Ressort“-LLM		– Einbindung Wissensdatenbanken (z. B. anwendungsfall-spezifisch, ressortspezifisch) bspw. für Use Cases auf Basis von semantischer Suche
3 Entwicklungsumgebung Datenlabore		– Integration bestehender/neuer Entwicklungen/Apps der Datenlabore (d. h. potenzielle Funktion LLM-Portal als „App-Store“)
4 Open-Source basiertes LLM-Portal („Vollversion“)		– Private Cloud ermöglicht Nutzung auch für Anwendungsfälle auf Basis sensibler Daten

Quelle: Präsentation des BeKI vom 18.04.2024

Bewertung

Die Grafik des Beratungszentrums für Künstliche Intelligenz (BeKI) veranschaulicht den Mehrwert der geplanten Stufen für die Bundesverwaltung. Vom Standpunkt des BMBF aus ist jedoch eine differenziertere Betrachtung erforderlich. Aufgrund der bereits vorhandenen Infrastruktur und Eigenentwicklungen ergeben sich aus derzeitiger Sicht für einige Schritte nur geringe Vorteile für das BMBF.

a) Rechenkapazität

Das BMBF verfügt über eigene Server mit den für KI benötigten Grafikkarten (GPUs) und hat somit eigene Rechenkapazitäten für die Entwicklung und den Betrieb von KI-Anwendungen. Eine Entwicklungsumgebung bringt im ersten Schritt daher keinen direkten Nutzen mit sich. Es ist anzunehmen, dass die Anzahl der KI-Applikationen in Zukunft stark zunehmen wird. Dadurch werden die lokalen Rechenkapazitäten langfristig an ihre Grenzen stoßen. Die Entwicklungsumgebung in der Public Cloud wird durch das Datenlabor nur in Teilen genutzt werden (z.B. um anderen Ressorts eigene Entwicklungen zugänglich zu machen), eine langfristige Überführung der lokalen Infrastruktur von der Public Cloud zum ITZ Bund erscheint jedoch aufgrund von Skalierungs- und Effizienzeffekten und mit Blick auf IT-Sicherheit sinnvoll. Gleichzeitig ist zu berücksichtigen, dass der Zugang zur Plattform des ITZ Bund über die Netze des Bundes erfolgt. Dies stellt eine Hürde für Anwendungen dar, die sich an externe Nutzer wie beispielsweise Zuwendungsempfänger richten.

Daher ist das Vorhaben des ITZ Bund nach aktuellem Kenntnisstand nicht für alle möglichen Anwendungsfälle geeignet, für andere Anwendungen, die z.B. auf VS-Daten und höher zugreifen, aber Grundvoraussetzung. Es sollte dementsprechend ergänzend zu einer Public Cloud-Infrastruktur betrachtet werden, die momentan im Rahmen des Z26-Prozess zur Etablierung eines Chatbots für das Handbuch der Projektförderung aufgebaut wird.

b) Basis KI-Applikationen zur Textbearbeitung

Einzelne Funktionen der Plattform, wie bspw. Dokument-Zusammenfassungen, können perspektivisch auch anderweitig ermöglicht werden. So arbeitet das Datenlabor an einem „Auto-

Summarizer“ und über die vom GovTech-Campus betriebene Webanwendung „GenAI Prompt“ sind Basisfunktionen klassischer Sprachmodelle ebenfalls verfügbar.

Der Zugriff auf das „Generalwissen LLM“ im Rahmen der ersten Schritte in der Public Cloud bieten jedoch kurzfristig ein deutlich größeres Spektrum an Optionen und hochwertigen Sprachmodellen wie GPT 4 oder (voraussichtlich) Luminous von Aleph Alpha. Die Plattform gewährt ein benutzerfreundliches Interface, das Verwaltungsmitarbeitende vorprogrammierte Befehle wie bspw. Textgenerierung oder Übersetzungen auswählen lässt.

Der zentrale Mehrwert entsteht jedoch erst im letztendlichen on premise Betrieb durch das ITZ Bund, das eine sichere Umgebung schafft, in der perspektivisch sogar VS-NfD Dokumente bearbeitet werden können. Zudem wird erst dadurch das gemeinsame Entwickeln, Austauschen und Betreiben von KI-Anwendungen der Datenlabore im Ressortkreis umfänglich ermöglicht werden.

c) Ressortübergreifender Austausch

Unabhängig der konkreten technischen Mehrwerte für das BMBF ist das gemeinsame Vorgehen im Ressortkreis zu begrüßen und wäre von Beginn an ein Fortschritt gegenüber dem Ist-Zustand. Erfahrungen der IT-Dienstekonsolidierung zeigen, wie hoch der Aufwand ist, Parallelentwicklungen zusammenzuführen. Die Bundesregierung profitiert langfristig durch einheitliche Standards für Mitarbeitende, Kosteneffizienz durch zentrale Beschaffung und Betrieb sowie den ressortübergreifenden Austausch.

d) Betrieb über ITZ Bund

Der Nutzer-Support für die standardmäßig bereitgestellten Apps erfolgt über das ITZ Bund und würde die Personalressourcen im BMBF entlasten.

e) Kosten

Für die ersten drei Stufen des Projekts im Cloud-Betrieb kalkuliert das ITZ Bund [REDACTED], die auf die beteiligten Ressorts umgelegt werden sollen. Für das BMBF ergibt sich somit ein Mittelbedarf in Höhe von [REDACTED]. Diese Kosten könnten mit DARP-Mitteln aus IT- Dienstleistungstitel 3012 532 01 getragen werden.

Die aufwändige Beschaffung der Hardware für den on premise Betrieb der vierten Stufe wird dagegen mit [REDACTED] kalkuliert. Während eine Beteiligung des Ressortkreises hieran bislang nicht vorgesehen war, ist eine verbindliche Finanzierungszusage durch das BeKI aktuell ausstehend.

Fazit

Trotz einzelner Einschränkungen durch das mehrstufige Verfahren und den Betrieb über die Netze des Bundes handelt es sich bei der vorliegenden Planung um einen fortschrittlichen, interministeriellen Ansatz, der durch das BMBF unterstützt werden sollte. Der zähe Prozess der IT-Dienstekonsolidierung unterstreicht die Notwendigkeit eines proaktiven, gemeinsamen Vorgehens. Das vorgeschlagene Konzept stärkt auch die kooperative, agile und ressortübergreifende Arbeit der Datenlabore. Der hier beabsichtigte zentrale Ansatz über das ITZ Bund wäre daher eine wichtige Basis für die in der Datenstrategie 2023 verankerte langfristige Aufstellung der Datenlabore.

Der größte Mehrwert für das BMBF entstünde durch den on premise Betrieb im ITZ Bund, mit dem die Verarbeitung vertraulicher Daten und die interministerielle Kooperation ermöglicht wird. BMBF sollte sich daher dafür einsetzen, dass auch die vierte Ausbaustufe umgesetzt wird und BMI zentral die Finanzierung dafür sichert.

el. gez. Evers / Metternich

3. Wv.: Z22, 121
4. Kopie an: Z16, Z26, L21

Si, 23/04/2024