

KI-Agenten für die Verwaltung

Wie die Bundesagentur für Arbeit große Sprachmodelle einsetzt

(BS/Eldar Sultanow/Florian Winzer*) Die Bundesagentur für Arbeit (BA) setzt ein Zeichen: Mit dem gezielten Einsatz von KI-Agenten auf Basis großer Sprachmodelle testet sie, wie sich technologische Innovation mit den hohen Anforderungen des öffentlichen Sektors vereinen lässt. Das Ziel ist ambitioniert: Änderungsanforderungen (Request for Change, kurz RFC-Dokumente) und Nutzeranforderungen (User Stories) sollen nicht mehr von Menschen in Jira-Tickets übersetzt werden.

Diese Aufgabe übernimmt künftig ein intelligenter Agent – automatisch, zuverlässig und datenschutzkonform. Im Kontext des IT-Verfahrens ALLEGRO haben Architekten der BA und Kollegen von Capgemini sinnvolle Use Cases für KI-Agenten erarbeitet und setzen die Agenten zur Erstellung von Jira-Tickets um – ein aus unserer Sicht sehr vielversprechender Use Case.

ALLEGRO bildet zentrale Förderprozesse nach dem SGB II ab – von der Antragstellung über die Bewilligung bis zur Abrechnung und Bescheidung. Das System wird von über 40.000 Anwendern genutzt und verarbeitet jährlich ein Transaktionsvolumen im Milliardenbereich.

Weniger Routine, mehr Wirkung

Verwaltungen kämpfen mit wachsenden Aufgaben, komplexen IT-Landschaften und knappen Ressourcen. Gleichzeitig schlummern gerade in den standardisierten Prozessen enorme Potenziale für Automatisierung – wenn sie intelligent umgesetzt werden. Klassische Lösungen zur Robotic Process Automation (RPA) stoßen dabei schnell an ihre Grenzen.

Mit Sprachmodellen, also Large Language Models (LLMs), eröffnet sich ein neuer Weg: flexibel, kontextsensitiv und sprachbasiert. Genau hier setzt das Pilotprojekt der Bundesagentur für Arbeit an.

Was genau passiert da?

Statt Tickets manuell zu erstellen, übernimmt ein System aus mehreren KI-Agenten die Arbeit. Ein Reader-Agent liest die Anforderung, extrahiert relevante Inhalte aus RFCs oder User Stories. Ein

Planner-Agent zerlegt die Aufgabe in konkrete Arbeitsschritte. Und ein Creator-Agent erstellt daraus automatisch ein vollständiges Jira-Ticket – mit Titel, Beschreibung, Priorität, Aufwandsschätzung und Kategorie (Art des Tickets – aktuell Defaultwert: „Aufgabe“). Die Aufwandsschätzung ist derzeit noch nicht implementiert. Das System muss zunächst mit Tausenden Tickets trainiert werden, wobei es gilt, verschiedene Parameter wie Kategorie, Themengebiete, Schätzgrundlagen und Puffer zu berücksichtigen.

Human-in-the-Loop

Abschließend prüft ein Reviewer-Agent das Ergebnis auf Konsistenz und Dubletten. Wenn der Reviewer-Agent „Fehler“ findet, gibt er seinen Befund wieder dem Reader-Agent als Input mit. Und das ist ein entscheidender Unterschied zur klassischen Automatisierung, deren Ergebnis deterministisch ist. Agenten verhalten sich analog zu Menschen: zwei vergleichbar gut qualifizierte Gruppen von Menschen würden auch nicht zu zwei wortgleichen Ergebnissen kommen.

Dennoch wären die Ergebnisse beide gut genug, um basierend darauf weiterzuarbeiten. Dieselbe Güte erreicht hier eine Gruppe von AI-Agents. Der „Human-in-the-Loop“ gibt die Tickets final frei. Klingt

futuristisch? Ist es nicht. Es läuft bereits – in einer sicheren, abgeschirmten Pilotumgebung.

Technisch ganz vorne dabei

Zum Einsatz kommen ausschließlich lokal betriebene Modelle. Technisch lässt sich das Ganze auch in einer Cloud betreiben – ob private, hybrid oder public. Modelle wie Aleph Alpha (mit seinen datenschutzkonformen APIs), LLaMA, Mistral oder Qwen liefern die nötige Intelligenz. Gesteuert wird das Ganze über CrewAI, eine Open-Source-Plattform für koordinierte Multi-Agenten-Systeme. Die Pilotumgebung ist vollständig in die lokale Jira-Instanz der Bundesagentur für Arbeit integriert. Dies erfolgt über klar definierte Schnittstellen, die von der Jira-API bereits gestellt werden sowie und das Logging der verwendeten Frameworks.

Die ersten Ergebnisse sind beeindruckend

- Die automatisierte Erstellung von Jira-Tickets funktioniert in den meisten Fällen reibungslos, mit Ausnahme von sehr großen RFC-Dokumenten.
- Der wiederkehrende Aufwand, Tickets anzulegen, zu formulieren (Copy & Paste), die Beschreibung zusammenzufassen, einzufügen und zu speichern, entfällt.
- Der Agent erstellt automatisch ein Ticket pro RFC mit den korrek-

ten Eigenschaften wie Titel, Beschreibung und weiteren Details.

- Titel und Beschreibung sind in den meisten Fällen plausibel und verständlich und spiegeln den Inhalt der Lösungsvorschläge gut wider.

Die Systemarchitektur erfüllt die Anforderung, dass das gesamte Training sowie die Verwendung der KI in der geschlossenen IT-Infrastruktur des IT-Systemhauses der Bundesagentur für Arbeit stattfinden. Das heißt, dass keine Daten die Infrastruktur der Bundesagentur verlassen und alles On-Premise läuft.

KI ist kein Zauberstab

Eine besondere Herausforderung war die Integration in eine hochsichere On-Premise-Architektur. Dies erforderte umfangreiche Abstimmungen zwischen verschiedenen Einheiten, einschließlich Betrieb und Sicherheit. Unklare oder widersprüchliche Anforderungen führen zu Fehlern. Die Modelle müssen regelmäßig aktualisiert werden, was in einer rein lokalen Umgebung mehr Aufwand bedeutet. Und: Ohne Schulung wird der Mensch vor dem Bildschirm nicht automatisch zum KI-Experten. Deshalb gehört diese und eine Begleitung genauso zum Projekt wie die Technik selbst.

RFC-Dokumente können unterschiedliche Größen haben. Große RFCs umfassen teilweise über 20

Seiten. Beim Verarbeiten eines großen RFCs mit 26 Seiten ist es zu Beispiel vorgekommen, dass das für die Anfrage an das LLM festgesetzte Token-Grenze von 8192 (2¹³) Tokens überschritten (die Anfrage umfasste 8732 Tokens) wurde. Da die ersten Seiten unnötige Formalien und Leerzeichen enthielten, konnten diese aus der Anfrage gelöscht werden. Dadurch blieb die Anfrage innerhalb des Tokenlimits und das Ticket konnte erstellt werden.

Ein Blick nach vorn

Was hier als Pilot begonnen hat, hat das Potenzial, zur Blaupause für die gesamte öffentliche Verwaltung zu werden. Auch wenn das Projekt noch viel Forschungs- und Umsetzungsleistung erfordert, könnten KI-Agenten künftig nicht nur in Jira unterstützen, sondern auch Fachverfahren, Dokumentenmanagement oder Bürgerkommunikation entlasten. Entscheidend ist, dass die Technologie nicht um ihrer selbst willen eingesetzt wird, sondern dort, wo sie einen echten Mehrwert schafft.

Verwaltung mit KI geht – jetzt

Die Bundesagentur für Arbeit zeigt, wie mit klarem Blick, solider Architektur und Mut zur Umsetzung neue Wege beschritten werden können. KI-Agenten übernehmen keine Jobs – sie übernehmen Routineaufgaben. So können Menschen das tun, was zählt: Entscheidungen treffen, Verantwortung tragen und Zukunft gestalten.

**Eldar Sutanow ist Architect bei Capgemini.*

Florian Winzer arbeitet für das IT-Systemhaus der Bundesagentur für Arbeit, IT-Verfahren ALLEGRO.