fortiss

Landesforschungsinstitut
des Freistaats Bayern
für softwareintensive Systeme

Von KI-Assistenten zu KI-Agenten

Über fortiss

Vorsprung durch Software

- **Rechtsform:** gemeinnützige GmbH
- **Gründung:** 2009
- **Eigentümer:** Freistaat Bayern (2/3) und Fraunhofer-Gesellschaft (1/3)
- fortiss ist ein An-Institut der TU München

fortiss bearbeitet auf Spitzenniveau das für den Hightech-Standort Bayern zentrale Thema der Entwicklung softwareintensiver Systeme. Diese basieren zunehmend auch auf den eng verwandten Technologien der KI.

Von besonderer Bedeutung sind dabei:

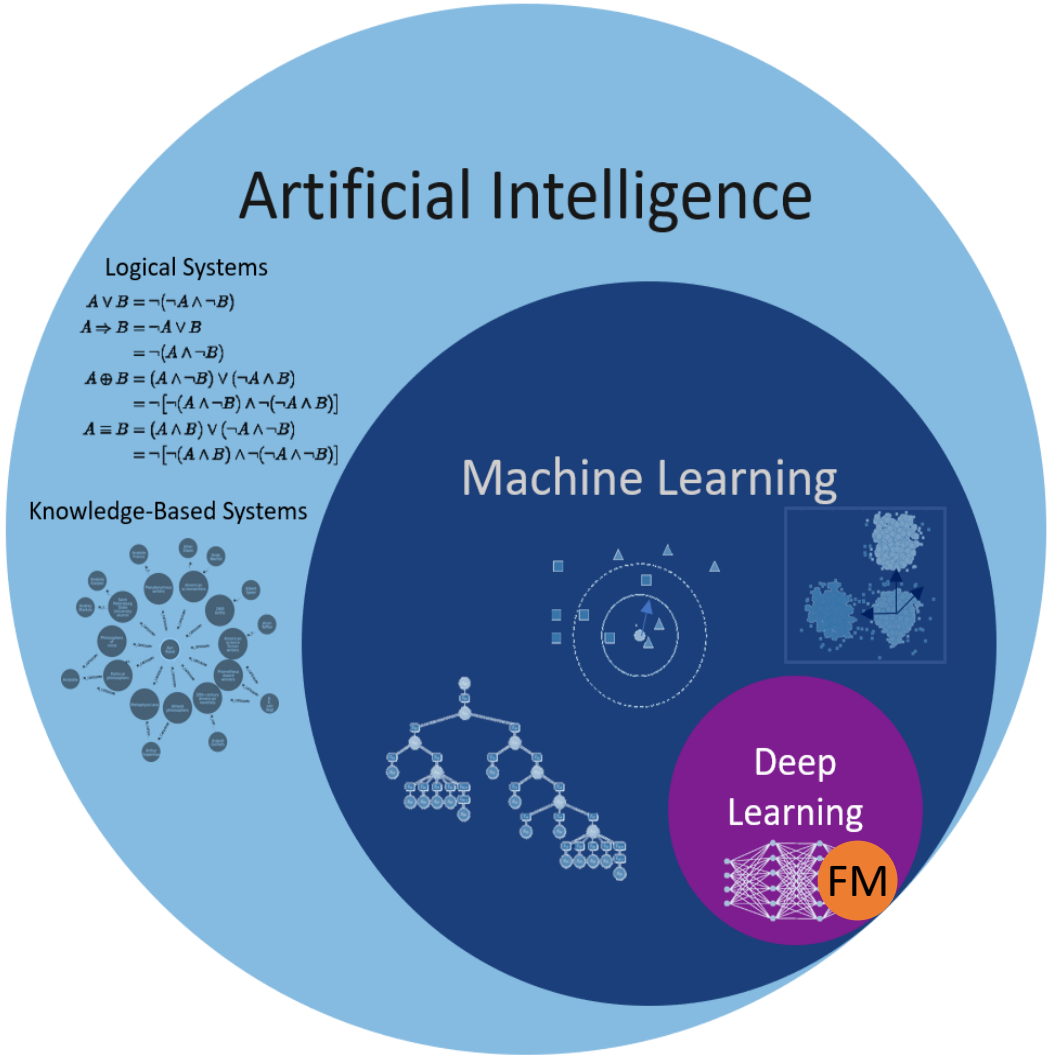
- Vorreiterrolle bei der beherrschbaren Entwicklung verlässlicher und sicherer Software-, KI- und IoT-Anwendungen.
- die vielfach direkten Bezüge zu Praxispartnern zur lösungsorientierten Umsetzung der erarbeiteten F&E-Resultate.
- ein attraktives Angebot zur Wissensvermittlung zu aktuellen Entwicklungen zu Software, KI und IoT.
- interne, kompetenzausgerichtete Strukturen, die talentierten Forscher*innen beste Entwicklungs-chancen bieten.

Künstliche Intelligenz

Teilbereiche der KI

- „... ein Teilgebiet der Informatik, das sich damit befasst, Maschinen Fähigkeiten zu verleihen, die intelligentem (menschlichem) Verhalten ähneln. Dies kann durch vorprogrammierte Regeln oder durch maschinelles Lernen erreicht werden.“

- “(Artificial) Intelligence comes from the application, not the implementation”



Klassische KI Anwendungen

Von statischen zu dynamischen Systemen

statisch

dynamisch

Generative KI

Hintergrund von ChatGPT & Co.

Generative KI

KI-Techniken, welche aus Daten über bestehende Artefakte lernen und diese nutzen, um neue Artefakte zu erzeugen

Foundation Models (FM)

Modelle des maschinellen Lernen, welche mittels selbstüberwachtem Lernen trainiert werden.
Das Training erfolgt universal auf einem breiten Satz nicht-annotierter Daten.
Anschließend werden sie an spezifische Aufgaben angepasst

Large Language Models (LLM)

Ein KI-Modell, welches mit großen Textmengen trainiert wird, um Texte interpretieren und menschenähnliche Ausgaben erzeugen zu können

ChatGPT

Einen Chatbot für Konversationen als Dienstleistung von OpenAI.
Das FM bzw. LLM wurde mit Milliarden von Texten aus verschiedensten Quellen trainiert.
Durch Reinforcement Learning mit menschlichen Rückmeldungen wurde ChatGPT das schreiben beigebracht.

Chat-Bots

Anwendungsfelder

Praktische Anwendungen

- Maschinelle Übersetzung: Verständnis des Kontexts über verschiedene Sprachen hinweg
- Textgenerierung: Wahrung der Kohärenz in längeren Texten
- Beantwortung von Fragen: Auffinden relevanter Informationen im Kontext
- Zusammenfassung von Dokumenten: Identifizierung wichtiger Informationen
- Bildbeschriftung: Verknüpfung visueller Elemente mit Text

Chat-Bots

Herausforderungen und Einschränkungen

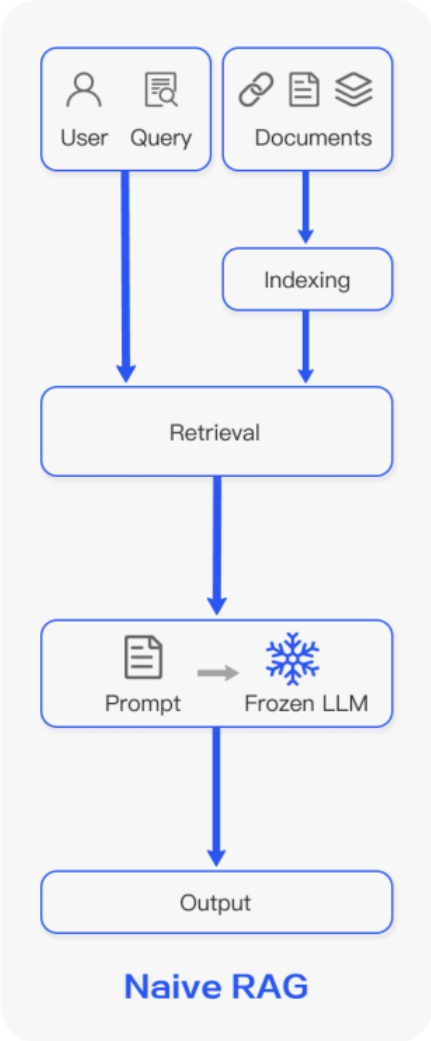
- Das Modell hat gelernt, das nächste Wort in einer Sequenz vorherzusagen
- Die Modellgewichte spiegeln sowohl sprachliche Konzepte als auch Wissen aus den Trainingsdaten wider
- Herausforderungen:
  - Halluzinationen: LLMs können ungenaue oder unsinnige Informationen generieren
  - Wissensgrenze: Das Wissen von LLMs beschränkt sich auf die Informationen, die in ihrem Trainingsdatensatz bis zum letzten Update verfügbar waren
  - Kein Zugriff auf private Unternehmensdaten: LLMs können nicht auf private, geschützte Daten zugreifen oder diese nutzen

Bedarf von mehrschichtigen KI-Modellen, d. h. KI-Agenten

Einfache KI-Agenten

RAG Systeme

- Retrieval-Augmented Generation (RAG)
  - Integration großer LLM mit proprietären Datenquellen
  - Generierungsprozess wird durch abgerufene Informationen ergänzt,
  - Verbesserung der Zuverlässigkeit und Einbindung aktueller Informationen
- RAG = LLM + Datenbank/Dokumente



Compound AI

KI Workflows und KI Agenten

- KI-Agent = LLM + Werkzeug (Tool) + Zustand (State)

- Engineering von KI-Fähigkeiten anstatt Skalierung von Modellen

- KI-Agent als Koordinator
  - Strategische Planung: Fähigkeit, übergeordnete Strategien zu entwickeln und komplexe Ziele in Teilziele aufzuschlüsseln.
  - Ressourcenmanagement: Effiziente Zuweisung von Ressourcen, einschließlich anderer Agenten und externer Tools.
  - Kommunikation und Koordination: Fähigkeit zur effektiven Kommunikation mit anderen Agenten und Systemen, um eine nahtlose Zusammenarbeit zu gewährleisten
  - Aufgabenkoordination: Delegation von Aufgaben und Ausführung der Anfragen.

Umsetzung eines KI-Agenten

KI-Workflow-Automatisierung mit n8n (https://n8n.io/)

- Entwickelt in Deutschland – Source Code zugänglich
- Selbsthosting oder Cloud Variante – Community Edition (kostenlos),
- Visuell einfach zu bedienen

- Agentic-AI-Workflow in n8n:
  - LLM-gesteuertes System, das logisch denken, über zu ergreifende Maßnahmen entscheiden, Tools und APIs nutzen, den Kontext im Blick behalten und schrittweise auf ein Ziel hinarbeiten kann.

Umsetzung eines KI-Agenten

KI-Workflow-Automatisierung mit n8n (https://n8n.io/)

Tagesgeschäft reduzieren – Fokus auf das Wesentliche

Praktische Anwendungen

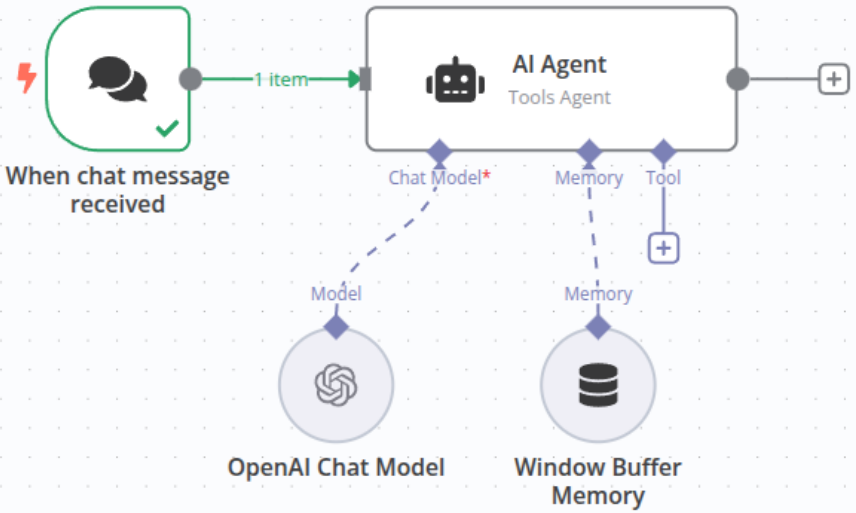
- Angebotsversand
- Terminbuchung
- Kundensupport

Umsetzung eines KI-Agenten

KI-Workflow-Automatisierung mit n8n (https://n8n.io/)

- Gehirn (LLM):
  - LLMs suchen das beste Werkzeug zur Erfüllung einer
    Aufgabe aus oder simulieren komplexe Entscheidungsprozesse
  - Das LLM übernimmt:
    - Planen
    - Entscheiden, welche Werkzeuge verwendet werden sollen
    - Schlussfolgern
    - Erzeugen von Ergebnissen

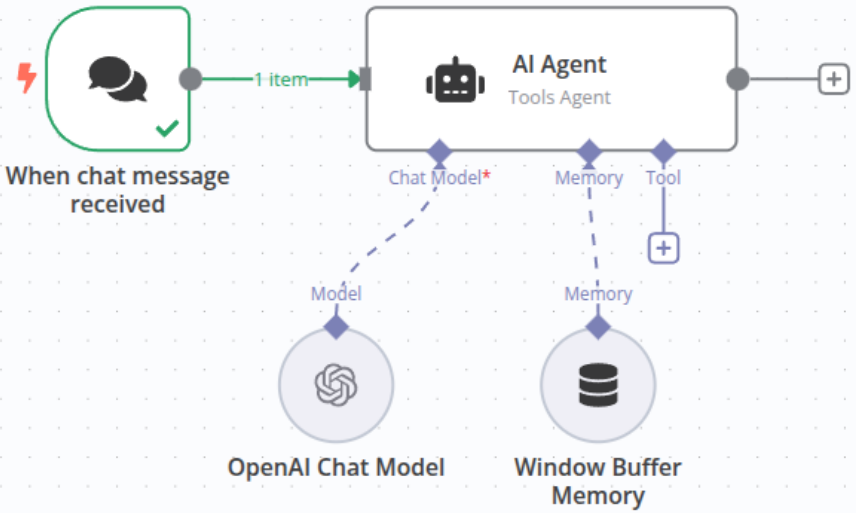
- Speicher (Memory): Um sich daran zu erinnern, welche Anfragen bearbeitet wurden, muss der KI-Agent den Kontext beibehalten.



Umsetzung eines KI-Agenten

KI-Workflow-Automatisierung mit n8n (https://n8n.io/)

- Werkzeuge (Tools): Ein KI-Agent wird „agentisch“, wenn er Werkzeuge einsetzen kann.
- Tools fungieren als Add-ons, mit denen die KI auf zusätzlichen Kontext oder Ressourcen zugreifen kann.
  - Websuche
  - HTTP-APIs
  - Datenbankabfragen
  - Google Tabellen
  - CRM-Zugriff
  - E-Mail-Versand
  - Abfrage aus der Vektordatenbank
  - Interne Unternehmens-APIs



Komplexere KI-Agenten

OpenClaw

- Open Source KI Agent
  - System, in denen große Sprachmodelle ihre eigenen Prozesse und den Einsatz von Werkzeugen dynamisch steuern
  - Interne Kontrolle darüber, wie Aufgaben ausgeführt werden

- Nach Einrichtung Interaktion mittels Chat über Nachrichtenapp (WhatsApp, etc.)
- Persistentes Gedächtnis
- Zugriff auf teils sensible Daten um zu funktionieren

- Eingabe → Begründung → Entscheidung → Einsatz des Werkzeugs → Ergebnis beobachten → Wiederholen

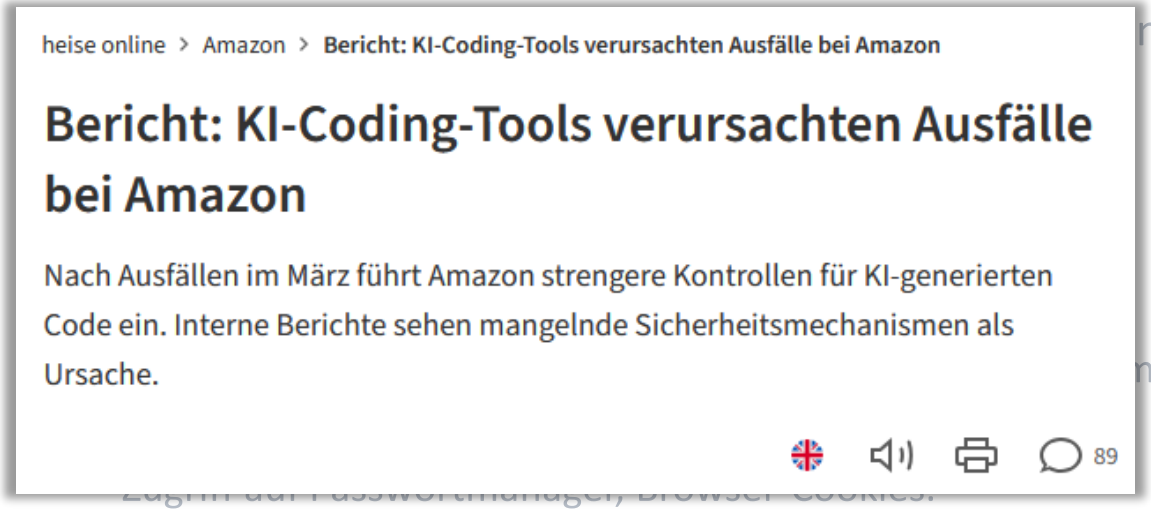
Komplexere KI-Agenten

Risiken autonomer Systeme

- Niemals mit uneingeschränkten Berechtigungen betreiben
- Führen Sie die Anwendung in einer Sandbox/einem Container aus
- Vermeiden Sie unbedingt:
  - die direkte Ausführung auf Ihrem Host-Betriebssystem
  - das Einbinden Ihres gesamten Home-Verzeichnisses
  - das Offenlegen Ihrer SSH-Schlüssel, Produktions-Cloud-Anmeldedaten
  - uneingeschränkte GitHub-Token
  - Zugriff auf Passwortmanager, Browser-Cookies.
  - Es sollten manuelle Freigabeschritte vorgesehen sein für: das Löschen von Dateien, das Versenden von E-Mails, das Pushen von Commits, Finanztransaktionen, Änderungen an der Infrastruktur und Produktionsbereitstellungen.

Komplexere KI-Agenten

Risiken autonomer Systeme



- Es sollten manuelle Freigabeschritte vorgesehen sein für: das Löschen von Dateien, das Versenden von E-Mails, das Pushen von Commits, Finanztransaktionen, Änderungen an der Infrastruktur und Produktionsbereitstellungen.

https://www.heise.de/news/Bericht-KI-Coding-Tools-verursachten-Ausfaelle-bei-Amazon-11205724.html
https://www.heise.de/news/KI-Agent-loescht-Daten-Katastrophe-fuer-PocketOS-11279416.html?utm_source=firefox-newtab-de-de

Von KI-Assistenten zu KI-Agenten

Take Home Messages

- KI Agenten zeichnen sich durch eine strategische Planung und Durchführung von Einzelaufgaben aus, die an (externe) Tools delegiert werden können.
- Kernelement ist die Orchestrierung von Werkzeugen zur Lösung der Teilaufgaben.
- Kontrolle der Zugriffsberechtigungen ist enorm wichtig, um ungewollte Aktionen zu verhindern.

fortiss Transfer

Von der Forschung in die Praxis

fortiss | Transfer

Dr.
Wolfgang
Köhler

Potenzialanalyse

Von der Idee zum konkreten Use-Case

- Identifizierung von Chancen und Lösungen durch den Einsatz von KI (-Agenten) und innovativer Software
- Wissenschaftlich fundiert, technologieoffen und unabhängig
- Individuelle Analyse Ihrer Prozesse und Daten
- Konkrete Handlungsempfehlungen, wie Sie erfolgreich starten können

Wir unterstützen Sie bei Ihrem nächsten Digitalisierungsschritt mit den fortiss Potenzialanalysen

https://www.fortiss.org/transfer/transferangebote/potenzialanalyse

Kontakt

Dr. Julian Wörmann

woermann@fortiss.org

fortiss Newsletter

Bleiben Sie auf dem Laufenden!

Vielen Dank!