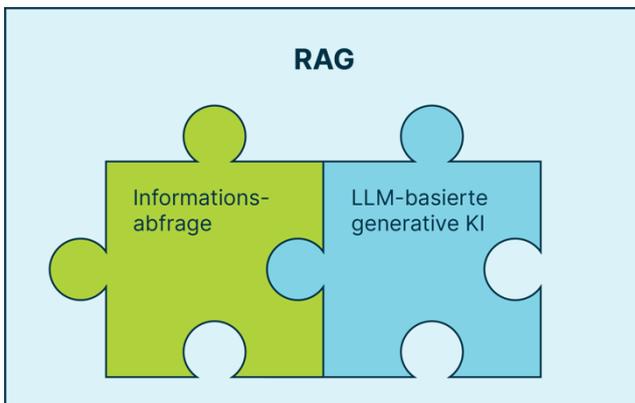


RAG: Jederzeit die passende Antwort

Mit Retrieval-Augmented Generation (RAG) heben Sie die Fähigkeiten von Generative AI auf die nächste Stufe. Grosse Sprachmodelle (Large Language Models, LLMs) erhalten durch RAG Zugang zu zusätzlichen Datenquellen, sodass Mitarbeitende und Kunden intuitiv auf kontextbezogenes und aktuelles Wissen zugreifen können.

LLMs sind immer nur so gut wie die Daten, mit denen sie trainiert wurden. RAG ist der Schlüssel, um veraltete und zu allgemein formulierte Antworten eines LLMs zu vermeiden: mit dieser Technik suchen Nutzende über unterschiedlichste Datenquellen hinweg – ganz unabhängig davon, ob diese unternehmenseigen oder -fremd sind. Dabei wird das «Basiswissen» des LLM mit aktuellem und unternehmensspezifischem Wissen über Produkte und Dienstleistungen angereichert. So wird die Suche nach relevanten Informationen aus verschiedenen verteilten Quellen deutlich beschleunigt und verbessert.



Unser Angebot

Wir unterstützen Sie bei der Definition einer Zielarchitektur für Ihre Anwendungsfälle. Bei der nachhaltigen Implementierung Ihres RAG-Systems berücksichtigen wir Ihre Anforderungen an Datenschutz, Datensicherheit, Skalierbarkeit und Betrieb. Wir bereiten Ihre Daten vor und setzen ein für Ihre Anwendungsfälle ausgewähltes LLM auf.

Bei der Gestaltung der Sucheingabe und der Darstellung der Suchergebnisse gehen wir nutzerzentriert vor. Dazu richten wir mithilfe von Prompt Engineering die Persona des virtuellen Assistenten und dessen Tonalität konsequent auf die Bedürfnisse und Ziele der Nutzenden aus.

Ihre Vorteile

- Durch RAG werden die Antworten der generativen KI schneller, aktueller, kontextbezogener und genauer.
- Als zusätzliche Datenquellen können eine Vielzahl von internen und externen Systemen eingebunden werden.
- Die bei der Suche verwendeten Datenquellen berücksichtigen die jeweiligen Nutzerberechtigungen.
- Mehrstufiges Nachfragen in natürlicher Sprache wird unterstützt.
- Die Suchvorgänge können überwacht und zurückverfolgt werden.
- Volle Transparenz: Im Gegensatz zu LLMs können RAG-Systeme die Quellen ihrer Antworten zitieren.

Integration und Verwendung

- Ein praktisches Beispiel für die Integration und Verwendung von RAG ist ein erweiterter Chatbot als Unterstützung für Ihren Kundenservice. Dieser Chatbot versteht Anfragen, ruft relevante Informationen aus einer Unternehmensdatenbank ab und generiert präzise, benutzerangepasste Antworten, indem er Retrieval- und Generationsmodelle kombiniert.

Kontakt

Adnovum AG
Stéphane Mingot

Badenerstrasse 170
8004 Zürich

T +41 79 306 08 58
stephane.mingot@adnovum.ch

www.adnovum.com



adnovum

And your digital business works