Wie KI unstrukturierte Datenquellen erschließt

Mit dem Aufstieg von ChatGPT fragen sich viele Unternehmen, wie künstliche Intelligenz sinnvolle Wettbewerbsvorteile schaffen kann. So auch der Bundesanzeiger Verlag und die Validatis GmbH, ein auf Unternehmensdaten spezialisiertes FinTech. Die Validatis verfügt über einen umfangreichen Datenbestand, darunter viele Jahresabschlüsse, Geschäftsberichte und andere Dokumente in unstrukturierten Datenformaten.

Auf einen Blick

Evaluations-Tool

Gemeinsam mit dem KI-Team von Sidestream entwickelten der Bundesanzeiger Verlag und die Validatis ein Evaluations-Tool, das genau dies ermöglicht:

- Import und Vorbereitung interner Datenquellen
- Anwendung verschiedener KI-Modelle wie GPT-3 auf diese Daten
- Der KI werden konkrete Fragen gestellt, wie z. B. „Extrahiere mir die Umsatzzahlen der Firma Sidestream aus den letzten 3 Geschäftsjahren“
- Die Antworten der KI werden bewertet und evaluiert

Zusammenarbeit und Technologien

Die enge Zusammenarbeit von Sidestream und Validatis in einem agilen Projektablauf mit wöchentlichen Reviews ermöglichte es, schnell zu verstehen, für welche Use Cases und Geschäftsprozesse die Validatis KI gewinnbringend einsetzen sollte. Durch den Einsatz von modernen Technologien wie Python, Streamlit und KI-Bibliotheken von HuggingFace konnte das Tool in kurzer Zeit umgesetzt werden.

Die Ausgangslage

Diese Daten sind essenziell, um strukturierte Firmenprofile zu erstellen, Anfragen von Banken im Kreditvergabeprozess zu beantworten und Reportings zu ESG-Maßnahmen auszuwerten. Derzeit beansprucht die Extraktion und Nutzbarmachung relevanter Informationen aus diesen Daten allerdings noch viel Expertenarbeit.

Der Lösungsansatz

In einem initialen Konzept-Sprint mit Sidestream überlegte die Validatis, wie KI Mehrwert in ihrem Geschäftsbereich schaffen könnte. Um diese Frage zu beantworten, kristallisierte sich die Idee eines KI-Tools heraus, welches genutzt werden kann, um zu testen, wie gut KI verschiedene Probleme lösen kann. Beispiele dafür sind z. B. das Beantworten von Fragen oder das Strukturieren von Daten, basierend auf den vorhandenen Unternehmensdaten. Als weitere Anforderung spielt insbesondere die Datensouveränität eine entscheidende Rolle, sodass die Ergebnisabwägung auch die Entscheidung zwischen Open Source und kommerziellen Modellen umfasst. In anderen Fällen kann ein kommerzielles Modell bessere Ergebnisse liefern. Die Validatis benötigte daher ein Tool, um verschiedene Modelle auf die Daten anzuwenden und pro Situation im Realbetrieb zu vergleichen. So kann situativ das jeweils passende Modell ausgewählt und zukünftig der optimale KI-Assistent entwickelt werden.

Auf diese Weise kann die Validatis testen, ob und wie KI ihnen dabei helfen kann, unstrukturierte Datenquellen zu erschließen und die darin enthaltenen Informationen effizienter zugänglich zu machen.

„KI ist ein Werkzeug, das die Digitalisierung unserer Prozesse weiter beschleunigen wird. Mit Sidestream haben wir einen Partner, der uns sowohl bei der Evaluation von Einsatzgebieten für KI hilft, als auch in der Lage ist, KI-gestützte Software für uns zu programmieren und in unsere Produktionsprozesse zu integrieren.“

Paul Heger · Managing Director, Validatis