Leitfaden zur KI-Produktentwicklung

Benjamin Pirlich, VP Product Management
b.pirlich@ai2deliver.com

LLMs - Data plus Architektur plus Training

- Anwendungen: LLMs wie GPT-4 unterstützen Anwendungen in den Bereichen Texterstellung, Übersetzung, Zusammenfassung und mehr, indem sie kontextuell relevante Texte verstehen und erstellen.
- Fähigkeiten: Bekannt für ihre Vielseitigkeit in Bereichen wie Finanzen, Gesundheitswesen usw. und für ihren skalierbaren Nutzen bei verschiedenen Aufgaben ohne aufgabenspezifische Modelle.
- Ethische Nutzung: Das Bewusstsein für und die Abschwächung von Verzerrungen, die ethische Anwendung und die Einhaltung von Vorschriften wie GDPR sind bei der Nutzung von LLMs von größter Bedeutung.

Stabile Diffusion: Phantasievolle KI-Kreationen aus Text

- Überblick: Ein generatives KI-Modell, das Text in künstlerische Bilder umwandelt.
- Kern-Mechanismus: Verwendet das CLIP-Modell, VAE und das neuronale Netz U-Net zum Verstehen und Erzeugen von Bildern.
- Bemerkenswert: Fähigkeiten beim Zero-Shot-Lernen und bei der Vorhersage für Klassen, für die nicht trainiert wurde.

Huggin Face: Brückenschlag zwischen Gemeinschaft und Wirtschaft in der KI

- Entwicklung: Von einem Chatbot für Teenager zu einem einflussreichen Unternehmen in der KI-Forschung und -Infrastruktur.
- Transformatoren-Bibliothek: Ermöglicht Entwicklern die einfache Implementierung und Änderung von Transformer-Modellen.
- Model Hub: Ein Repository mit einer Vielzahl von Modellen für maschinelles Lernen für verschiedene Aufgaben.

LangChain: Integrativer Rahmen für große Sprachmodelle

- Nützlichkeit: Erleichtert die Interaktion mit verschiedenen LLMs, einschließlich der Modelle von OpenAI und den Angeboten von Hugging Face.
- Merkmale: Bietet "Chains" für Dialoge mit Modellen und eine Speicherschnittstelle, um Gesprächskontexte zu halten.
- Fähigkeiten: Kann mit Vektordatenbanken verbunden werden, was Operationen wie Ähnlichkeitsbewertungen und schnelle Suchen mit großen Texten verbessert.

Neuronale Suche: Ein Hybrid aus traditioneller und KI-unterstützter Suche

- Zweistufiger Prozess: Erste Vorauswahl über ein DPR-Modell und gründliche Suche durch Vorauswahlen mit einem ExQA-Modell.
- ExQA-Modelle: Können komplexe Beziehungen erkennen und die Relevanz der abgerufenen Ergebnisse sicherstellen.
- Haystack-Framework: Erleichtert die Implementierung von NLP-Anwendungen und gewährleistet eine benutzerfreundliche Erfahrung, die durch eine solide Dokumentation unterstützt wird.

Rechtliche Rahmenbedingungen für KI: Ausgleich zwischen Innovation und ethischer Anwendung

- Regulatorisches Umfeld: Navigieren durch Rahmenwerke wie GDPR, Data Act und AI-VO-E während der KI-Konzeption und -Entwicklung.
- KI-VO-E: Unterteilt KI-Systeme in verschiedene Risikokategorien, für die jeweils eigene Vorschriften gelten.
- Datenverwendung: Sicherstellung der Redundanz, Repräsentativität, Genauigkeit und Vollständigkeit von Daten sowie eine effektive Protokollierung, um die Rückverfolgbarkeit von KI-Ergebnissen zu gewährleisten.

Rechtliche Aspekte: Navigation durch die Komplexität der Datennutzung in der KI

- DSGVO-Konformität: Verlangt eine rechtliche Rechtfertigung für die Verarbeitung personenbezogener Daten und betrifft in der Regel keine anonymisierten Daten.
- Datengesetz: Zielt auf eine gerechte Verteilung des Wertes bei der Datennutzung ab und regelt die Datenbereitstellung zwischen Nutzern, Dateneigentümern, Empfängern und öffentlichen Stellen.
- Das KI-Gesetz (KI-VO-E) regelt risikoreiche KI-Anwendungen und verlangt die Einhaltung von Datengesetzen (z. B. GDPR).
- Urheberrecht und KI-generierte Inhalte - Erstellung und Einhaltung von Richtlinien zur Autorisierung und Rechtmäßigkeit von KI-generierten Inhalten unter Beachtung der Gesetze zum geistigen Eigentum.
- Herausforderungen: Sicherstellung der Einhaltung des Data Act und der DSGVO beim Umgang mit personenbezogenen Daten aus IoT-Anwendungen, insbesondere wenn es zu Überschneidungen mit der Entwicklung risikoreicher KI-Systeme kommt.

Checkliste für die Entwicklung und Implementierung von KI-Diensten

Beauftragen Sie Datenschutzbeauftragte/Experten

- Konsultieren Sie Spezialisten, um die Auswirkungen auf den Datenschutz zu verstehen.

Bewertung der Verarbeitung personenbezogener Daten

- Bewerten Sie, ob und wie personenbezogene Daten in den KI-Diensten verarbeitet werden.

Prüfung der Anonymisierung von Daten

- Untersuchen Sie, ob Daten anonymisiert werden können und ob ein solches Verfahren praktikabel und rechtskonform ist.

Rechtsgrundlage für die Datenverarbeitung

- Prüfen Sie, auf welcher Rechtsgrundlage die Datenverarbeitung erfolgt und welche Schritte notwendig sind, um die Einhaltung zu gewährleisten.

Datenschutzmaßnahmen einführen

- Implementieren Sie Verschlüsselung und, wenn möglich, Pseudonymisierung zum Schutz der Daten.
- Stellen Sie die Protokollierung und Überwachung der Datenverarbeitung sicher, um Datenbewegungen und -zugriffe zu verfolgen und zu prüfen.
- Schulung des Personals in Bezug auf die Datenschutzbestimmungen und den sicheren Umgang mit Daten.
- Erwägen Sie zusätzliche Schutzmaßnahmen je nach den spezifischen Anforderungen des Projekts.

Vertragliche Vereinbarungen

- Schließen Sie vertragliche Vereinbarungen mit den beteiligten Stellen und stellen Sie sicher, dass Datenverarbeitungsverträge (oder andere relevante Vereinbarungen) bestehen.

Schwellenwertanalyse

- Führen Sie eine Schwellenwertanalyse durch, um festzustellen, ab welcher Stufe Datenschutz-Folgenabschätzungen (DPIAs) oder andere regulatorische Maßnahmen erforderlich sind.

Datenschutz-Folgenabschätzung (DPIA)

- Führen Sie bei Bedarf eine Datenschutzfolgenabschätzung durch, um zu bewerten, wie personenbezogene Daten verarbeitet werden, und um sicherzustellen, dass die Verarbeitung im Einklang mit den Datenschutzgesetzen steht.

Einführung eines Vorfallsmanagementsystems

- Implementieren Sie ein System zur Verwaltung und Meldung von Datenschutzverletzungen oder anderen Vorfällen in Übereinstimmung mit den

Benjamin Pirlich, VP Product Management
b.pirlich@ai2deliver.com

Navigieren durch die KI-Entwicklung: Eine Zusammenfassung

Die Entwicklung und Anwendung von KI-Modellen erfordert ein empfindliches Gleichgewicht zwischen technologischem Fortschritt und der Einhaltung von Rechtsvorschriften. Von der Erzeugung phantasievoller Visualisierungen mit Stable Diffusion über die Sicherstellung der Verfügbarkeit und Zugänglichkeit von Modellen mit Hugging Face, optimierte Interaktionen mit LLMs über LangChain bis hin zu verbesserten Suchfunktionen durch Neural Search - die KI-Landschaft ist groß und vielfältig. All diese Fortschritte sind jedoch mit den rechtlichen Rahmenbedingungen verknüpft, die die ethische und rechtmäßige Anwendung von KI regeln, insbesondere im Zusammenhang mit dem Schutz personenbezogener Daten (GDPR) und der Datennutzung im weiteren Sinne (Data Act).

Die Navigation durch diese rechtlichen und technologischen Bereiche erfordert ein umfassendes Verständnis und eine strategische Planung von der Konzeption bis zum Einsatz, um sicherzustellen, dass die Entwicklung, Anwendung und Verbesserung von KI-Modellen und -Produkten sowohl innovativ als auch ethisch einwandfrei ist.

Quellen

Academic Papers: https://arxiv.org/abs/1906.02243; https://arxiv.org/abs/2010.05006v4; https://aclanthology.org/2020.acl-main.577/; https://arxiv.org/abs/2106.05945; https://arxiv.org/pdf/2211.10435.pdf; https://arxiv.org/pdf/2210.03629.pdf; https://blog.research.google/2020/08/an-analysis-of-online-datasets-using.html;

Code Repositories & Documentation: https://github.com/openai/gpt-3; https://github.com/langchain-ai/langchain; https://github.com/facebookresearch/faiss; https://www.sqlalchemy.org/; https://huggingface.co/docs/transformers/index; https://streamlit.io/; https://www.gradio.app/; https://hci.iwr.uni-heidelberg.de/content/bosch-small-traffic-lights-dataset;

Data Repositories and Search Engines: https://archive-beta.ics.uci.edu/; https://data.gov/; https://data.europa.eu/de; https://datasetsearch.research.google.com/; https://aws.amazon.com/marketplace/search/results?trk=8384929b-0eb1-4af3-8996-07aa40; https://www.kaggle.com/datasets; https://huggingface.co/datasets/EleutherAI/pile;

Websites & Blogs: https://ix.de/z7ns; https://ix.de/zahm; https://ix.de/zjn2; https://bdtechtalks.com/2020/09/21/gpt-3-economy-business-model/; https://www.ontolux.de/text-mining/; https://de.wikipedia.org/wiki/Beurteilung_eines_bin%C3%A4ren_Klassifikators#Anwendung_im_Information_Retrieval; https://nlp.stanford.edu/projects/glove/; https://python.langchain.com/docs/modules/agents/tools/custom_tools; https://blog.research.google/2020/02/exploring-transfer-learning-with-t5.html; https://www.gesetze-im-internet.de/ifg/index.html; https://www.govdata.de/; https://blog.google/products/search/discovering-millions-datasets-web/; https://blog.research.google/2020/08/an-analysis-of-online-datasets-using.html;

European Legislation: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206;

Articles: Tobias Haar, KI-Recht: ChatGPT, Bard und Co.; Kevin Kekule, Wolf Knüpffer, Guido-Arndt Söldner; Allround- talent; Machine-Learning-Workflows mit Kubeflow orchestrieren; Leon Marius Schröder, Clemens Gutknecht, Leon Lukas; Neuronale Suche in Texten; Ramon Wartala; Mit Streamlit schneller zum datenwissenschaftlichen Prototyp.