

There is (no) Compliance in AI?

Jörg Jaenichen

Agenda



Vorstellung



KI – eine (gar nicht so) einfache Einordnung



Operative Chancen des KI-Einsatzes



Compliance-Risiken durch künstliche
Intelligenz / EU AI Act



»Was tun?« spricht Zeus

Jörg Jaenichen

Senior Security Consultant

- Lead Auditor ISO/IEC 27001:2022, ISO 22301:2019, §11 1a, 1b EnWG und §8a (3) BSIG
- zertifizierter Datenschutzbeauftragter
- Data Protection Risk Manager
- Informationssicherheit und Datenschutz
- BCM und Risk Management



KI – eine (gar nicht so) einfache Einordnung

Was ist künstliche Intelligenz?

”

Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren.

Das Europäische Parlament

KI – eine (gar nicht so) einfache Einordnung

Was steckt drin?



Schwache KI

spezialisierte Aufgaben

vs.

Starke KI

multidisziplinäre, komplexe Aufgaben

Machine Learning

Maschinelles Lernen ist ein Bereich der Künstlichen Intelligenz, der sich mit der Entwicklung von Algorithmen beschäftigt, die es Computersystemen ermöglichen, aus Daten zu lernen und sich durch Erfahrungen zu verbessern. Anstatt explizit programmiert zu werden, erkennen diese Systeme Muster in den Daten und verbessern ihre Leistung im Laufe der Zeit.

Neuronale Netze Kolmogorov-Arnold Netze

Neuronale Netze sind ein zentraler Bestandteil des maschinellen Lernens, inspiriert von der Struktur und Funktionsweise des menschlichen Gehirns. Neuronale Netze können komplexe Muster in Daten erkennen und sind besonders effektiv in Aufgaben wie Bilderkennung, Sprachverarbeitung und Spielen. Durch Training mit großen Datenmengen passen sie ihre Verbindungen an, um die Genauigkeit ihrer Vorhersagen zu verbessern.

Inspiziert durch ChatGPT

Large Language Models (LLM)

Large Language Models (LLMs) sind maschinelle Lernmodelle, die auf großen Mengen von Textdaten trainiert werden, um menschliche Sprache zu verstehen und zu generieren. Sie bestehen aus Millionen bis Milliarden von Parametern und nutzen tiefe neuronale Netzwerke. LLMs können Aufgaben wie Textvervollständigung, Übersetzung, Fragebeantwortung und Textgenerierung bewältigen.

KI – eine (gar nicht so) einfache Einordnung

Hype...

ChatGPT Passed the Uniform Bar Examination: Is Artificial Intelligence Smart Enough to be a Lawyer?

Der größte Gamechanger der Internetgeschichte:
ChatGPT

ChatGPT Lifts Business Professionals' Productivity and Improves Work Quality

Google Gemini: Die nächste Stufe der KI-Revolution

Google is done. Here's why OpenAI's ChatGPT Will Be a Game Changer

Is Devin AI the End of Software Engineers?

ChatGPT: Der Gamechanger in der Beratung?

ChatGPT Sprints to One Million Users

ChatGPT passes exams from law and business schools

KI – eine (gar nicht so) einfache Einordnung

...vs. Realität...



Was ist das Niedersächsische Vulkanschutzgesetz?



Das Niedersächsische Vulkanschutzgesetz ist ein Gesetz, das darauf abzielt, die Bevölkerung und die Umwelt in Niedersachsen vor den potenziellen Gefahren von Vulkanen zu schützen. Es legt Regeln und Maßnahmen fest, um Risiken zu minimieren und im Falle eines Vulkan-ausbruchs angemessen reagieren zu können. Zu den typischen Bestimmungen solcher Gesetze gehören die Festlegung von Sicherheitszonen um Vulkane herum, die Überwachung von Vulkanaktivitäten, die Erstellung von Evakuierungsplänen und die Sensibilisierung der Bevölkerung für die Gefahren von Vulkanen.



Wann ist es in Kraft getreten?



Das Niedersächsische Vulkanschutzgesetz ist am 15. Juni 2017 in Kraft getreten.

KI – eine (gar nicht so) einfache Einordnung

...vs. die echte Welt

Für den Bereich der Eintrittsvektoren gewinnen große Sprachmodelle wie der Chatbot ChatGPT und entsprechende kriminelle Nachbildungen zunehmend an Bedeutung. Beginnend mit dem Zeitpunkt der Veröffentlichung von ChatGPT und dem damit ausgelösten Hype um Künstliche Intelligenz (KI) Ende 2022 verzeichnen verschiedene IT Dienstleister einen enormen Anstieg an neuartigen Phishing Mails.

Phishing Report

Zscaler-Report stellt 60 Prozent Anstieg bei KI-gesteuerten Phishing-Angriffen fest

23.04.2024, Zscaler | Autor: [Herbert Wieler](#)

1 von 4



Personen wurden bereits Opfer von Voice-Cloning oder kennen jemanden, der es schon einmal erlebt hat.

Quelle: McAfee³

Phishing as a Service Angebote: Telegram Bot Telekopye

Tools wie Telekopye ermöglichen Cyberkriminellen, großflächige Phishing-Kampagnen auch ohne tiefere technische Kenntnisse durchzuführen.

Durch Telekopye können Nutzer über ein vereinfachtes Interface via Telegram auf zahlreiche Funktionen zur Durchführung einer Phishing-Kampagne zugreifen. Zu den Funktionen zählen die Erstellung von Phishing-Webseiten, der Versand von Phishing-Mails und SMS sowie die Generierung von gefälschten Proof-Screenshots und QR-Codes. Das Toolkit bietet verschiedene HTML-Vorlagen für Phishing-Webseiten in unterschiedlichen Ländern an, darunter auch Deutschland. Die nutzerfreundliche HTML-Vorlage stellt hierbei das Hauptmerkmal von Telekopye dar. Die auf diesem Weg erstellten Phishing-Webseiten imitieren Zahlungsseiten verschiedener Webseiten, Login-Webseiten zu Zahlungsdienstleistungen oder anderweitigen Zahlungsgateways.

Über den Phishing-Link werden Nutzer, die beispielsweise im Glauben sind, einen Online-Kauf zu tätigen, aufgefordert, ihre Zahlungsdaten einzugeben, die von den Tätern abgegriffen werden können. Einige Versionen von Telekopye sind darüber hinaus in der Lage, erbeutete (Zahlungs-)Daten der Geschädigten auf dem Server des Bots zu speichern. Die im Rahmen der Phishing-Kampagne erbeuteten Gewinne gelangen an die Telekopye-Administratoren, die die Herkunft der Gelder über einen Mixing-Dienst verschleiern und einen eigenen Anteil abschöpfen. Diese Provision ist abhängig von der Telekopye-Version und der Rolle des Nutzers und beträgt 5-40%. Telekopye erleichtert die Ausführung von Phishing-Angriffen über das Interface erheblich. Allerdings enthält das Tool keine Chatbot-KI-Funktionen und führt die Kampagnen somit nicht autonom aus.

Die Nutzung eines Bots wie Telekopye stellt ein Novum dar. Die zunehmende Automatisierung erreicht mit Telekopye im Bereich Phishing eine neue Stufe, da so zur Durchführung größerer Kampagnen nicht einmal das Verlassen der Telegram-Anwendung erforderlich ist. Ebenfalls wird durch die Vermarktung von Telekopye via Telegram die Relevanz des Messengers als Underground-Economy-Plattform deutlich.

Licht



Predictive Maintenance durch Zustandsüberwachung und Auswertung der Ergebnisse

Energieverbrauch optimieren durch Mustererkennung und Energieeinsparungspotenzialanalyse

Personaleinsatzplanung automatisieren

Lieferkettenoptimierung von der Rohstoffbeschaffung bis zur Auslieferung an den Endkunden

Verkaufsprognosen erstellen und Vertriebsstrategien optimieren

Personalisierte Marketingkampagnen und Analyse von Kundenverhalten

Produktentwicklung mit Big Data Analysen

Buchhaltungsprozesse automatisieren, Fraud-Erkennung und Finanzprognosen verbessern

&

Schatten

Halluzination (Erschaffen von „Tatsachen“)

Falsche Ergebnisse aufgrund schlechter Trainingsdaten (z. B. algorithmischer Rassismus)

Black-Box-Effekt (z. B. beim Profiling oder der automatisierten Entscheidungsfindung)

Nutzung urheberrechtlich geschützten Materials

Fehlender Ergebnisdeterminismus durch kontinuierliches Lernen

Manipulation durch absichtlich falsche Trainingsdaten

Nutzung durch Cyber-Kriminelle für Social Engineering-Angriffe (Deep Fake, Voice Cloning etc.)

Schäden, die durch KI oder aufgrund von KI-beeinflusster Entscheidungen entstehen (Haftung)

Was die EU sagt... zu Betreiberpflichten



Umsetzung technischer
und organisatorischer
Maßnahmen konform zur
Betriebsanleitung



Sicherstellung einer
menschlichen Aufsicht
und kontinuierlichen
Kontrolle durch
kompetentes,
ausgebildetes und
befugtes Personal



Einhaltung der
Zweckbestimmung des
KI-Systems bei der
Eingabe von Daten

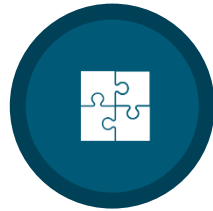


Aufbewahrung von
Protokolldaten des KI-
Systems (mind. 6
Monate)

Was die EU sagt... zu Betreiberpflichten



Information der Mitarbeitenden und der Mitarbeitervertretungen bei Einführung und Einsatz eines KI-Systems



Nutzung der Qualitätsmanagementdaten des Anbieters bei der Durchführung einer Datenschutz-Folgenabschätzung



Information der Betroffenen beim Einsatz automatisierter Entscheidungsfindung



Durchführung einer Grundrechte-Folgeabschätzung und Information der Marktüberwachungsbehörde

Was die EU sagt... zur Transparenzpflicht



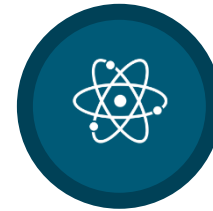
Datenschutz/
Informationspflichten



Deep Fake
Kennzeichnung (Bild,
Ton, Video)



Kennzeichnung künstlich
erzeugter Texte bei
Veröffentlichungen zu
Angelegenheiten von
öffentlichem Interesse



Zeitpunkt der
Information: erste
Interaktion des Nutzers
mit dem KI-System

Sanktionen: 35 Mio. EUR oder 7% des weltweiten Jahresumsatzes

Was die EU sagt... zu Hochrisiko-KI-Systemen

Biometrie

- biometrische Fernidentifizierungssysteme
- Verarbeitung sensibler oder geschützter Attribute und Merkmale zur biometrischen Kategorisierung
- Emotionserkennung

Kritische Infrastruktur

- Sicherheitsbaustein im Betrieb kritischer digitaler oder Versorgungsinfrastruktur

Allgemeine und berufliche Bildung

- Zugangsfeststellung, Zulassung zu Bildungsangeboten
- Bewertung von Lernergebnissen
- Bewertung des Bildungsniveaus
- Überwachung und Erkennung von verbotenen Verhalten bei Prüfungen

Was die EU sagt... zu Hochrisiko-KI-Systemen

Beschäftigung,
Personalmanagement und
Zugang zur
Selbstständigkeit

- Schalten gezielter Stellenanzeigen
- Bewerbungen sichten/filtern/bewerten
- Entscheidungen über Einstellung

Zugänglichkeit und
Inanspruchnahme
grundlegender privater und
grundlegender öffentlicher
Dienste und Leistungen

- Gesundheitsdienste
- Aufdeckung von Finanzbetrug
- Risikobewertung bei Versicherungen
- Klassifizierung von Notrufen
- Systeme für Triage

Was die EU sagt... zu Hochrisiko-KI-Systemen

Strafverfolgung

- Bewertung des Risikos einer natürlichen Person Opfer von Straftaten zu werden
- Lügendetektor
- Bewertung der Verlässlichkeit von Beweismitteln
- Erstellung von Profilen






Migration, Asyl und Grenzkontrolle

- Lügendetektor
- Bewertung von Einreisrisiken
- Prüfung von Asyl-, Visumsanträgen und Aufenthaltstiteln
- Aufdeckung, Anerkennung oder Identifizierung natürlicher Personen

Rechtspflege und demokratische Prozesse

- Ermittlung und Auslegung von Sachverhalten und Rechtsvorschriften oder alternative Streitbeilegung
- Ergebnis einer Wahl oder eines Referendums oder das Wahlverhalten natürlicher Personen bei der Ausübung ihres Wahlrechts bei einer Wahl oder einem Referendum zu beeinflussen

»Was tun?« spricht Zeus

-  Zuständigkeiten festlegen
-  Anwendungsszenarien definieren, Chancen/Risiken bewerten
-  BR & DSB einbinden
-  KI-(Unternehmens-)Richtlinien einführen
-  Strukturiertes Einführungsverfahren planen & umsetzen

Vielen Dank für eure Aufmerksamkeit

Bei Fragen kommt gerne auf uns zu!