Deutsche Stiftung für Recht und Informatik

Transparenzpflichten im Wechselspiel zwischen KI-VO und DSGVO

ERMÖGLICHT DIE KI-VO EINEN BLICK IN DIE BLACKBOX?

Ilan Leonard Selz / Yakin Surjadi
Schürmann Rosenthal Dreyer Rechtsanwälte

Herbstakademie 2023

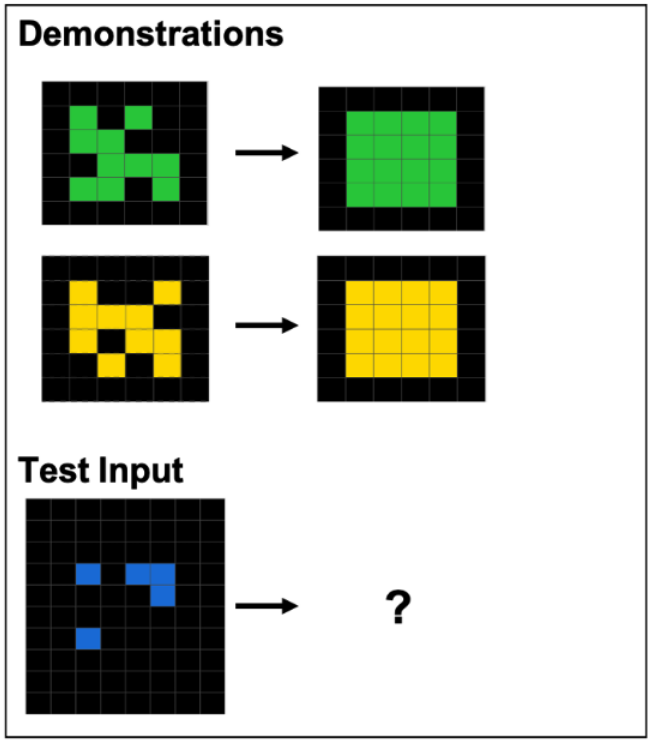
Agenda

I. KI und Blackbox-Problematik
II. Informationspflichten nach KI-VO-E
III. Informationspflichten nach DSGVO
IV. Vergleich und Synergien
V. Fazit

I. KI und Blackbox-Problematik

I. KI und Blackbox-Problematik

- Weites Verständnis von (schwacher) KI:
  - Lösung komplexer Aufgaben
  - Prognosefähigkeit
  - Lernfähigkeit
- Machine Learning = Lernen aus Daten
- Deep Learning = Erkennen von Mustern in diesen Daten mit spezialisiertem Gehirn (Künstliches Neuronales Netz(werk))
- vielschichtiger, undurchsichtiger und komplexer Entscheidungsfindungsprozess
- „Erforschung einer außerirdischen Intelligenz“
- Explainable Artificial Intelligence (XAI)



(a)

System: "You are a helpful assistant that solves analogy making puzzles. Only give the answer, no other words or text."

User: "Let's try to complete the pattern: "

input 1: [0 0 0 0 0 0] [0 3 0 3 0 0] [0 3 3 0 0 0] [0 0 3 3 3 0] [0 3 3 0 3 0] [0 0 0 0 0 0]

output 1: [0 0 0 0 0 0] [0 3 3 3 3 0] [0 3 3 3 3 0] [0 3 3 3 3 0] [0 3 3 3 3 0] [0 0 0 0 0 0]

input 2: [0 0 0 0 0 0] [0 4 0 0 4 0] [0 4 4 4 0 0] [0 4 0 4 4 0] [0 0 4 0 4 0] [0 0 0 0 0 0]

output 2: [0 0 0 0 0 0] [0 4 4 4 4 0] [0 4 4 4 4 0] [0 4 4 4 4 0] [0 4 4 4 4 0] [0 0 0 0 0 0]

input 3: [0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0] [0 0 1 0 1 1 0 0] [0 0 0 0 0 1 0 0] [0 0 0 0 0 0 0 0] [0 0 1 0 0 0 0 0] [0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0]

output 3:

(b)

Moskvichev et al.: The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain, https://doi.org/10.48550/arXiv.2305.07141; 2023

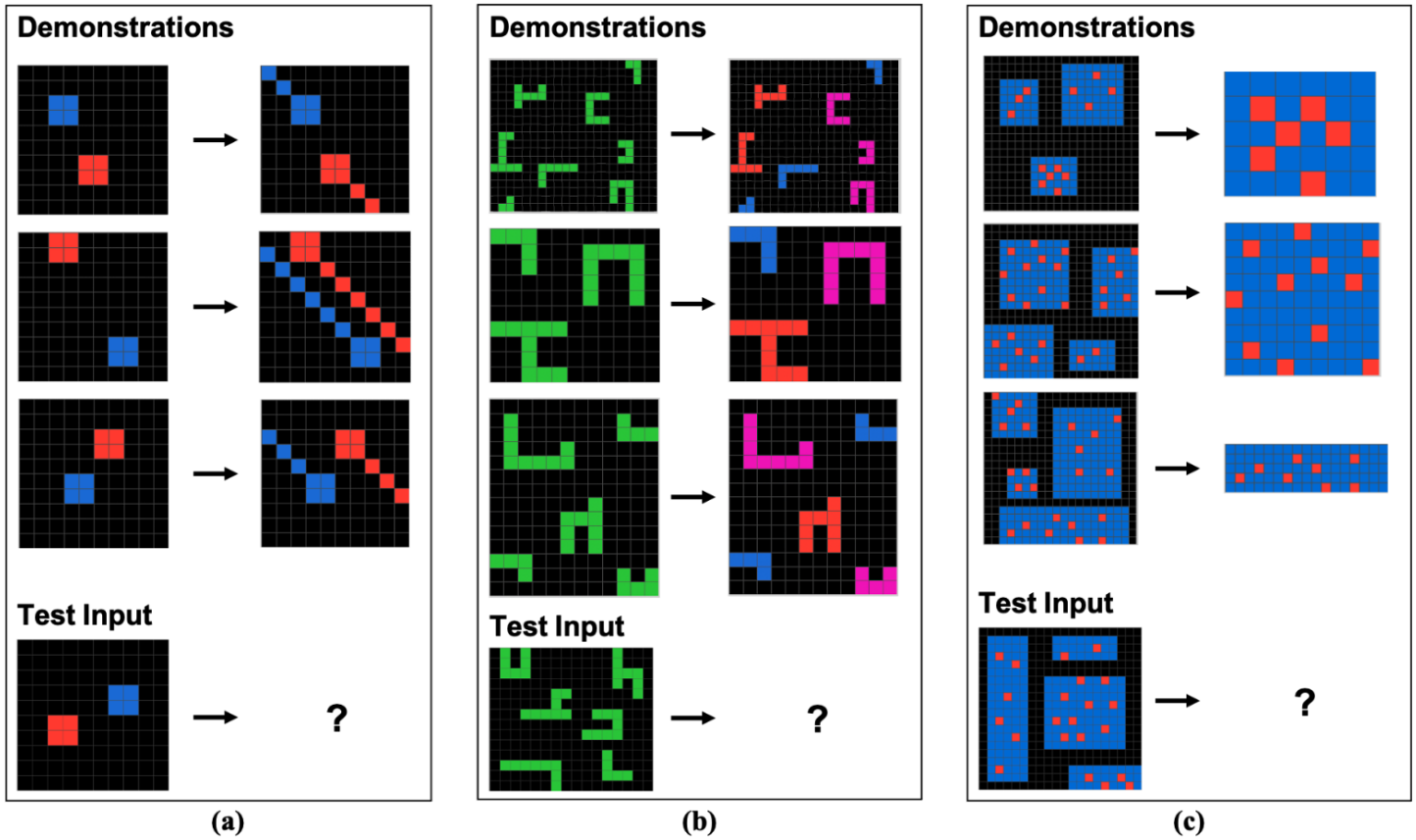


Figure 1: Three sample ARC tasks from [10]. Each task consists of a set of task demonstrations—transformations between colored grids that follow the same abstract rule—and (here) a single test input. The job of the solver is to generate a new grid that results from applying the abstract rule to the test input. (Best viewed in color.)

Moskvichev et al.: The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain, https://doi.org/10.48550/arXiv.2305.07141; 2023

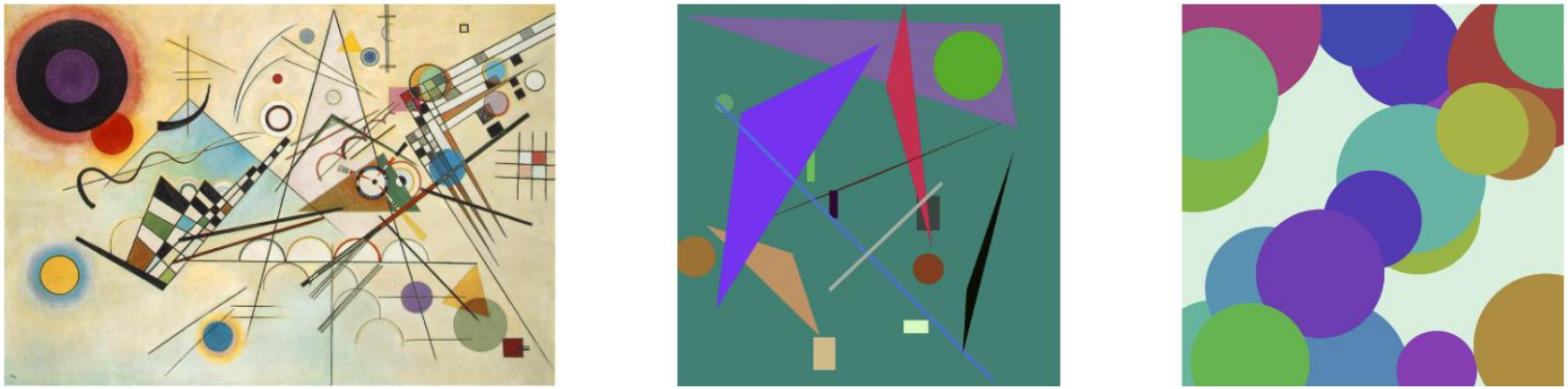


Figure 2.1: The first image is Composition 8, art by Wassily Kandinsky, the second and the third are produced by GPT-4 and ChatGPT respectively with the prompt "Produce Javacript code that creates a random graphical image that looks like a painting of Kandinsky".

Bubeck et al.: Sparks of Artificial General Intelligence: Early experiments with GPT-4, https://doi.org/10.48550/arXiv.2303.12712; 2023

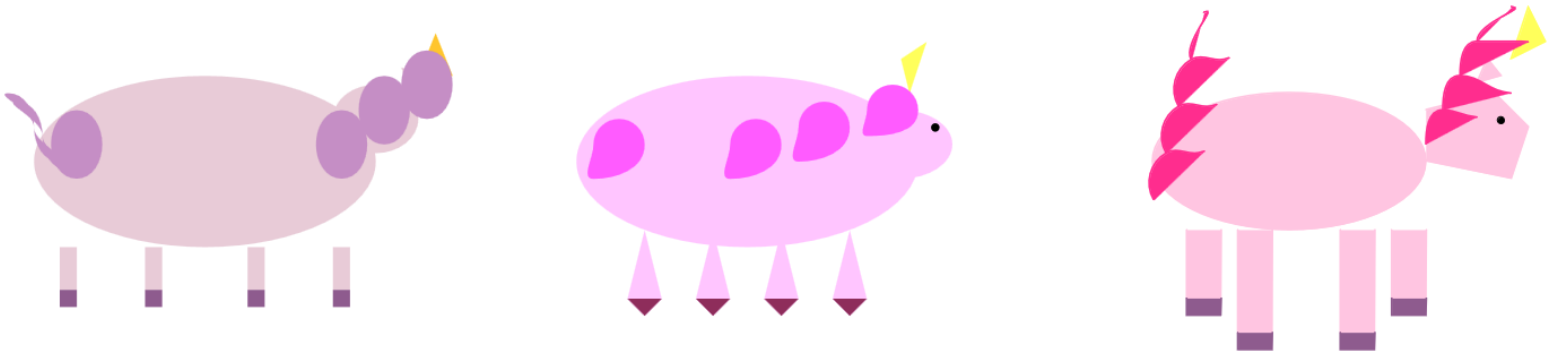
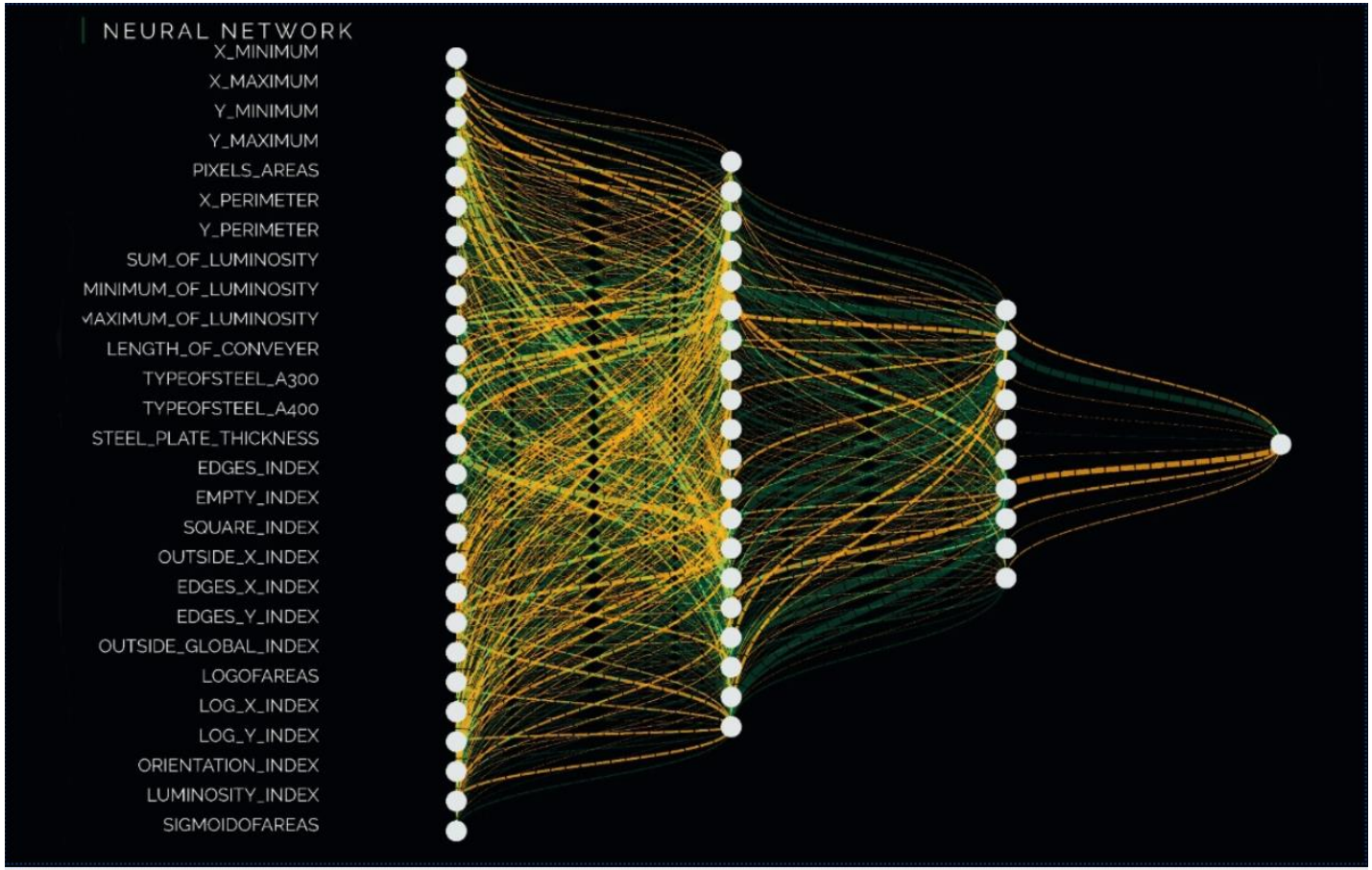


Figure 1.3: We queried GPT-4 three times, at roughly equal time intervals over the span of a month while the system was being refined, with the prompt "Draw a unicorn in TikZ". We can see a clear evolution in the sophistication of GPT-4's drawings.

Bubeck et al.: Sparks of Artificial General Intelligence: Early experiments with GPT-4, https://doi.org/10.48550/arXiv.2303.12712; 2023

Dieser Entscheidungsbaum spiegelt die Entscheidungsfindung eines neuronalen Netzes wider. Es geht um die Klassifizierung: Ist es eine Beule oder ein Kratzer? Die gelben Blattknoten entsprechen der Entscheidung für eine Beule, die grünen repräsentieren die Entscheidung für einen Kratzer.
Foto: Universität Stuttgart/IFF

Vogel: Wie entscheiden Algorithmen? Blick in die Black Box
https://www.uni-stuttgart.de/forschung/forschung-leben/1-2021/blackbox/; 2021



Das neuronale Netz: Die weißen Punkte in der linken Spalte repräsentieren die eingegebenen Daten, der einzelne weiße Punkt rechts das Ergebnis. Was dazwischen passiert, bleibt meistens im Dunkeln.
Foto: Universität Stuttgart/IFF

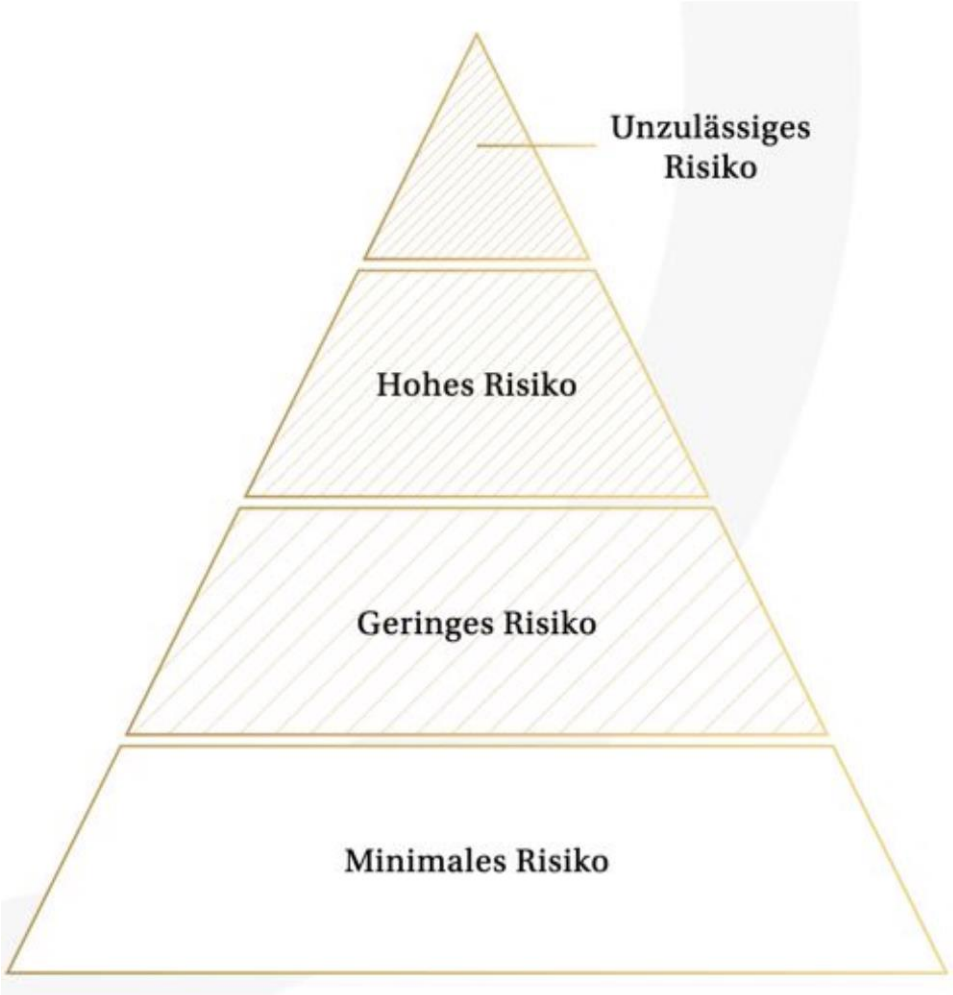
Vogel, Wie entscheiden Algorithmen? Blick in die Black Box
https://www.uni-stuttgart.de/forschung/forschung-leben/1-2021/blackbox/; 2021

II. Informationspflichten nach KI-VO

II. Informationspflichten nach KI-VO

- Weites Begriffsverständnis, vgl. Art. 3 Abs. 1 Nr. 1 KI-VO-E (Parlament): „*ein maschinengestütztes System, das so konzipiert ist, dass es mit unterschiedlichem Grad an Autonomie operieren kann und das für explizite oder implizite Ziele, Ergebnisse wie Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das physische oder virtuelle Umfeld beeinflussen*.“
- Risikobasierter Ansatz
- Primäre Adressaten der KI-VO:
  - Anbieter
  - Betreiber
  - (Nutzer)

II. Informationspflichten nach KI-VO



II. Informationspflichten nach KI-VO

Art. 3 Abs. 1 Nr. 15 KI-VO-E (Parlament):

„Gebrauchsanweisung“ die Informationen, die der Anbieter bereitstellt, um den Betreiber insbesondere über die Zweckbestimmung und die ordnungsgemäße Verwendung eines KI-Systems sowie über die zu treffenden Vorsichtsmaßnahmen zu informieren; einschließlich der besonderen geografischen, verhaltensbezogenen oder funktionalen Rahmenbedingungen, unter denen ein Hochrisiko-KI-System bestimmungsgemäß verwendet werden soll

II. Informationspflichten nach KI-VO

Art. 13 Abs. 1 S. 3 KI-VO-E (Parlament):

Transparenz bedeutet somit, dass zum Zeitpunkt des Inverkehrbringens des Hochrisiko-KI-Systems alle nach dem allgemein anerkannten Stand der Technik verfügbaren technischen Mittel eingesetzt werden, um sicherzustellen, dass die Ergebnisse des KI-Systems vom Anbieter und vom Nutzer interpretierbar sind. Der Nutzer muss in die Lage versetzt werden, das KI-System angemessen zu verstehen und zu nutzen, indem er allgemein weiß, wie das KI-System funktioniert und welche Daten es verarbeitet, sodass der Nutzer die vom KI-System getroffenen Entscheidungen der betroffenen Person gemäß Artikel 68 Buchstabe c erläutern kann.

II. Informationspflichten nach KI-VO

Modalitäten: Art. 13 Abs. 2 KI-VO-E (Parlament)

- Verständlich
- Digitales Format oder auf andere Weise auf einem dauerhaften Datenträger
- Korrekte, eindeutige und möglichst vollständige Information
- Unterstützung von Betrieb und Wartung des KI-System sowie der fundierten Entscheidungsfindung der Nutzer
- in einer hinreichend relevanten, barrierefrei zugänglichen Form
- Bandbreite:
  - PDF
  - Pop-Up / Mouseover in Software
  - ...
  - Vgl. andere Rechtsgebiete: § 3 Abs. 4 Produktsicherheitsgesetz

II. Informationspflichten nach KI-VO

Inhalte: Art. 13 Abs. 3 KI-VO-E (Parlament)

- Namen + Kontaktdaten des Anbieters
- „gegebenenfalls“ die Merkmale, Fähigkeiten und Leistungsgrenzen des Hochrisiko-KI-Systems, einschließlich
  - des Maßes an Genauigkeit, Robustheit und Cybersicherheit
  - aller Umstände, die zu Risiken führen können
  - des Ausmaßes, in dem das KI-System die von ihm getroffenen Entscheidungen erklären kann
- relevante Informationen über Benutzeraktionen, die die Systemleistung beeinflussen können
- erforderliche Wartungs- und Pflegemaßnahmen während der erwarteten Lebensdauer
- Mechanismen, die es den Nutzern ermöglichen Protokolle zu erfassen, zu speichern und auszuwerten
- …

III. Informationspflichten nach DSGVO

III. Informationspflichten nach DSGVO

- Nach Art. 13 Abs. 2 lit. f bzw. Art. 14 Abs. 2 lit. g DSGVO besteht bereits Informationspflicht, die auf KI regelmäßig Anwendung findet:

  „das Bestehen einer automatisierten Entscheidungsfindung einschließlich Profiling gemäß Artikel 22 Absätze 1 und 4 und - zumindest in diesen Fällen - aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer derartigen Verarbeitung für die betroffene Person“

- Was ist umfasst? Grundannahmen, Funktionsweise, Datenkategorien, Kriterien
- Entscheidend: Betroffener muss in der Lage sein, den technischen Entscheidungsprozess grundlegend nachzuvollziehen
- Blackbox muss nicht im Detail erklärt werden

III. Informationspflichten nach DSGVO

- Adressaten: Betroffene
- Besondere Informationspflichten gelten auch bei bloßem Profiling
- Ein Profiling liegt nach der weiten Definition in Art. 4 Nr. 4 DSGVO immer dann vor, wenn es darum geht, *„persönliche Aspekte [...] zu bewerten, [...] zu analysieren oder vorherzusagen.“*
- Modalitäten (Art. 12 DSGVO):
  - in „präziser, transparenter, verständlicher und leicht zugänglicher Form“
  - „klare und einfache Sprache“

IV. Vergleich und Synergien

IV. Vergleich und Synergien

- Verschiedene Ansatzpunkte, aber inhaltliche Überschneidungen
- KI-VO-E enthält konkrete Vorgaben an Informationen, DSGVO fordert Informationen zur „involvierten Logik“
- Anbieter/Betreiber vs. Verantwortlicher/Betroffener
- Gemeinsame Behandlung, um Informationen widerspruchsfrei zu erteilen
- Aber: da Betroffene Durchschnittsverbraucher sind, müssen die Informationen ggf. "heruntergebrochen" werden

V. Fazit

V. Fazit

- In der Wissenschaft zahlreiche, teils komplexe Methoden zur Erklärbarkeit von KI
- Regulatorisch geforderte Erklärbarkeit daher herausfordernd
- Noch keine Best Practice
- Aber: Sowohl KI-VO und DSGVO fordern keine tiefen Einblicke in die Blackbox
- Trotz der Unterschiede zwischen KI-VO und DSGVO können umfangreiche Synergieeffekte genutzt werden
- Die Informationen in der Gebrauchsanweisung nach KI-VO können für die Informationserteilung nach DSGVO nutzbar gemacht werden
- Es muss jedoch der unterschiedliche Adressatenkreis beachtet werden