

Übung 6 – Musterlösung

1. Stellungnahme zum Video

Exemplarische Antwort einer/eines Kursteilnehmenden:

Thema der digitalen Ethik: IT-Literacy

Neben den erschreckenden Möglichkeiten, Menschen durch Informationstechnologien systematisch zu überwachen (Stichwort „Social Scoring in China“), hat mich vor allem das Thema IT-Literacy überrascht. Im Gegensatz zu den anderen genannten Themen der digitalen Ethik (Datenerfassung, Ressourcenknappheit, etc.) bin ich bisher noch nie mit dem Problem in Kontakt gekommen, dass die informatischen Kompetenzen der „Digital Natives“ sehr oberflächlich sind.

Wie in dem Video von Christian Kittl und Martin Kandlhofer angesprochen wird, wachsen zwar die meisten aus „Generation Z“ mit zahlreichen Informationstechnologien auf, haben aber dennoch in den wenigsten Fällen ein tiefgreifendes Verständnis von den Technologien. Die Tatsache, dass ein gigantischer Teil unseres Lebens auf informatischen/technologischen Grundprinzipien beruht, die nur ein Bruchteil der Bevölkerung auch wirklich versteht, ist aus meiner Sicht gefährlich.

Ohne Verständnis von Informationstechnologien bzw. „IT Literacy“ können die Technologien nicht verantwortungsvoll verwendet werden. Wie soll man entscheiden, ob ein KI-gesteuertes Zuhause gefährlich ist, ohne zu verstehen, was ein künstliches neuronales Netz ist?

Wie soll man begreifen, dass Sicherheitsvorkehrungen bei der Datenübertragung enorm wichtig sind, ohne zu verstehen, wie Daten überhaupt transferiert werden?

IT-Literacy schafft demnach nicht nur die Basis für eine erfolgreiche/effiziente Nutzung von Informationstechnologien, sondern insbesondere ein Verständnis für den Zweck, Ursprung und die Gefahren der Technologien.

Wortanzahl: 203

2. KI

Künstliche Intelligenz oder kurz **KI** (= Artificial Intelligence oder kurz AI) ist ein Forschungsreich der Informatik, der enorm an Bedeutung gewonnen hat.

1. Erkläre die Begriffe *Maschinelles Lernen*, *Künstliches neuronales Netz*, *Deep Learning*, *Generative AI*, *ChatGPT*

Maschinelles Lernen: Gruppe von Algorithmen, die in der Lage sind, zu „lernen“, ohne ausdrücklich programmiert zu werden; Mustererkennung und mathematische Methoden sind zentrale Bausteine; Algorithmen müssen trainiert werden, z.B. mit zuvor aufwendig händisch vorbearbeiteten Beispieldaten (überwachtes Lernen) oder ungefilterten Daten (unbeaufsichtigtes Lernen)

Künstliches neuronales Netz: Begriff „Neuronales Netz“ – ursprünglich aus der Neuro-Wissenschaft; es werden künstliche „Neuronen“ zur Informationsverarbeitung eingesetzt; Versuch, die Funktionalität des menschlichen Gehirns abzubilden; geschichtete Strukturen aus miteinander verbundenen Neuronen, die durch Software emuliert und in einer Eingabe- und Ausgabeschicht sowie einer oder mehreren tiefen Schichten zur Merkmalserkennung angeordnet werden.

Deep Learning: Lernstrategie für das Training von Algorithmen; es kommen künstliche neuronale Netze zum Einsatz; Voraussetzung sind große Datenmengen (Trainingsdatensätze) und damit verbunden große Rechenleistungen; Training kann Monate dauern

Generative AI: Auf KI basierende Anwendung, die etwas (z.B. Text, Bild, Audio) erschafft

ChatGPT: Chatbot; generiert Texte auf Basis eines Chat-Prompts; im Hintergrund gibt es ein großes Sprachmodell - Large Language Model (LLM); lernt meist auf Basis von Deep Learning von großen Datensätzen (Texten)

2. Erkläre allgemein, welche **Chancen** sich durch den Einsatz von KI ergeben und gib **zwei konkrete Anwendungsgebiete** an.

Exemplarische Antwort einer/eines Kursteilnehmenden:

2. Da alle KI-Systeme, die heutzutage verwendet werden, schwache KI-Systeme sind, sind diese meist auf spezifische, genau definierte Aufgabenfelder beschränkt. Die Einsatzbereiche sind jedoch sehr vielfältig und weitschichtig. Die drei wichtigsten Sektoren sind dabei der Produktionsbereich, der Dienstleistungssektor und der Militärssektor. Die konkreten Anwendungsgebiete sind dabei z.B. die Optimierung verschiedenster Logistikaufgaben, KI-basierte Übersetzungssysteme, Chatbots oder digitale Assistenten sowie automatisierte Kreditwürdigkeitsprüfung. Chancen ergeben sich durch die Unterstützung der Benutzer*innen sowie einer Zeitersparnis bei repetitiven Tätigkeiten. [1]

[1] <https://link.springer.com/book/10.1007/978-3-658-25561-9>

3. Erkläre allgemein, welche **Risiken** sich durch den Einsatz von KI ergeben und begründe deine Argumente mit **zwei konkreten Beispielen**.

Exemplarische Antwort einer/eines Kursteilnehmenden:

3. Risiken des Einsatzes von KI-Systemen sind beispielsweise falsche Daten, fehlende Transparenz oder unzureichende Nachvollziehbarkeit. [3]
Konkrete Risiken treten vor allem dann auf, wenn KI-Systeme nicht korrekt abgesichert sind und somit Data Breaches oder Data Poisoning bewusst von Angreifern oder auch unbewusst von Entwicklern hervorgerufen werden können. Data Breaches sind dann gefährlich, wenn vertrauliche oder sensible Daten unbefugt offengelegt werden und somit die Privatsphäre verletzt wird. Und Data Poisoning ist gefährlich, da sich die Benutzer*innen auf falsch trainierte KIs verlassen und somit falsche Schlüsse ziehen könnten. [4]

[3] <https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/strategie-bundesregierung.html>

[4] S. Dilmaghani, M. R. Brust, G. Danoy, N. Cassagnes, J. Pecero and P. Bouvry, "Privacy and Security of Big Data in AI Systems: A Research and Standards Perspective," *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019, pp. 5737-5743, doi: 10.1109/BigData47090.2019.9006283.

4. An welchem "Regelwerk" (in Bezug auf KI und Digitale Ethik) hat u.a. Dr. Sarah Spiekermann-Hoff gearbeitet und wozu dient es?

Exemplarische Antwort einer/eines Kursteilnehmenden:

4. Sarah Spiekermann-Hoff arbeitete am IEEE 7000 Ethical System Engineering Standard mit. Dieser ist der weltweit erste Standard zur ethischer Systemtechnik und dient dazu, den Anwender*innen ein Prozessmodell zu bieten, mit dem diese zu ethischen Überlegungen in verschiedenen Phasen der Systeminitilierung, -analyse und -gestaltung animiert werden sollen. [5]

[5] <https://www.wu.ac.at/ec/projects/ieee-p7000-standard/>