

Wie Algorithmen lernen, fair zu sein

Aufgaben zum Video

1. Was passt?

Bearbeite die erste Aufgabe, bevor du das Video abspielst. Wähl die Wörter aus, die zu der jeweiligen Umschreibung passen.

1. ein Computerprogramm, das selbstständig denken und Probleme lösen kann:

2. eine Gruppe, die in der Gesellschaft nur wenig vertreten ist: _____
3. eine schlechte Meinung über jemanden, den man noch gar nicht kennt:

4. die Art, wie man allgemein über jemanden oder etwas denkt: _____
5. eine Reihe von gesammelten Informationen: _____
6. ein Oberbegriff, der zum Beispiel Dinge mit ähnlichen Eigenschaften zusammenfasst: _____

- | | | |
|--------------|---------------|----------------|
| a) Kategorie | b) Minderheit | c) Algorithmus |
| d) Vorurteil | e) Sichtweise | f) Datensatz |

2. Was siehst du im Video?

Schau dir das Video einmal an. Achte genau darauf, was passiert. Wähl alle richtigen Aussagen aus.

Im Video sieht man ...

- a) Gesichter von Menschen, auf denen Linien eingezeichnet werden.
- b) Studentinnen und Studenten im Hörsaal einer Universität.
- c) viele undeutlich gefilmte Internetseiten, über die Rechenkurven gelegt sind.
- d) mehrere Personen, die interessiert auf einen Computerbildschirm schauen.
- e) Fingerabdrücke, die von einer Maschine gescannt werden.
- f) eine Person, deren Fragen von einem Roboter beantwortet werden.
- g) einen Programmiercode auf einem Bildschirm.

3. Was wird im Video gesagt?

Spiel das Video noch einmal ab und hör diesmal genau hin. Wähl die richtigen Antworten aus. Mehrere Lösungen können richtig sein.

1. Die Sprecherin sagt, dass eine KI ...
 - a) Vorurteile gegenüber Minderheiten selbstständig erkennen kann.
 - b) Vorurteile gegenüber einer Gruppe übernimmt, wenn diese kaum im Netz vertreten ist.
 - c) nur eine begrenzte Zeit lang dazulernt.

2. Leon Derczynski sagt, dass ...
 - a) es für sein Team einfach ist, eine Liste mit Kategorien für das KI-Training zu erstellen.
 - b) bestimmte Formen von Hassrede nur betroffenen Menschen auffallen.
 - c) KI nicht in der Lage ist, alle Äußerungen im Netz richtig einzuordnen.

3. Es kommt vor, dass die KI ...
 - a) im Internet Forderungen nach Gerechtigkeit als extrem wahrnimmt.
 - b) die Betreiber von Internet-Seiten warnt, auch wenn Posts gar keine Hassrede enthalten.
 - c) Hassrede ignoriert, weil sie sie als Meinung einer Minderheit wahrnimmt.

4. Die Kopenhagener Forscher ...
 - a) wollen einem Algorithmus beibringen, niemanden im Netz zu diskriminieren.
 - b) finden, dass man den Einsatz von Künstlicher Intelligenz überdenken sollte.
 - c) arbeiten mit Schlagworten, damit die Künstliche Intelligenz Hassrede besser erkennt.

5. Der Algorithmus der dänischen Forscher wurde ...
 - a) 2020 veröffentlicht.
 - b) von Facebook gekauft.
 - c) bisher noch nicht fertiggestellt.

4. Übe die Vokabeln!

Welches Wort passt in welche Lücke? Wähl aus.

Im Netz gibt es immer wieder Beiträge, die bestimmte Gruppen _____ (1). Daher wollen dänische Forscher eine KI _____ (2), die Hassrede erkennt. Um das zu erreichen, _____ (3) die Entwickler Kategorien für verschiedene Arten von Redebeiträgen. Das bedeutet, dass sie sehr große Mengen an Daten _____ (4) müssen. Von ihrer Arbeit werden aber am Ende vermutlich sehr viele Menschen _____ (5).

- a) diskriminieren b) erstellen c) profitieren
d) vergeben e) zusammenstellen

5. Übe die Satzstellung!

Welche Satzstellung ist richtig? Wähl richtig aus.

1. Je mehr Hassrede _____ (es/gibt) im Internet _____ (es/gibt), umso wichtiger _____ (es/ist) _____ (es/ist), Posts zu kontrollieren.
2. Doch oft _____ (es/ist) _____ (es/ist) gar nicht leicht, zu erkennen, wann _____ (es/man) _____ (es/man) mit Hassrede zu tun hat.
3. _____ (Es/Gibt) _____ (es/gibt) in einer Gruppe Menschen, die in ihrem Alltag häufig Diskriminierung erleben, _____ (es/kommt) _____ (es/kommt) vor, dass sie Hassrede leichter erkennen als andere.
4. Vielen _____ (es/fällt) _____ (es/fällt) gar nicht auf, dass bestimmte Ausdrücke abfällig sind, denn hierbei _____ (es/kommt) _____ (es/kommt) auf persönliche Erfahrungen an.

Autor: Philipp Reichert