akademien der wissenschaften schweiz

Alle Dokumente finden Sie hier:
go.akademien-schweiz.ch/atable

Follow-up 4/24 · 03. Dezember 2024

KI regulieren? Mögliche Ansätze für die Schweiz

1. Erkenntnisse

1.1. Fazit Referate (Folien online verfügbar)

Ist die KI von morgen zuverlässiger? Woran aktuell geforscht wird
Philippe Cudré-Mauroux, Professor für Informatik an der Universität Freiburg
Künstliche Intelligenz umfasst eine Reihe verschiedener Technologien, die menschliche Intelligenz simulieren und komplexe Aufgaben übernehmen können. Aktuell besonders präsent sind riesige Sprachmodelle wie GPT4, die mit Hilfe zehntausender Graphikkarten und einer absurden Menge an Daten trainiert wurden – was Investitionen in der Höhe von hunderten Millionen von Dollar benötigt. Die Resultate sind beeindruckend: So liefern diese riesigen Basismodelle Antworten zu Prüfungsfragen, die in vielen Bereichen besser sind als diejenigen einer Mehrheit der menschlichen Prüfungsteilnehmenden.
Doch die Ergebnisse sind leider nicht zuverlässig: Die Modelle sind undurchsichtig und liefern immer nur Wahrscheinlichkeiten, keine Fakten. Sie widerspiegeln stets die Muster in den Trainingsdaten, entsprechend ungenügend sind die Resultate, wenn Eingabedaten zu stark von den Trainingsdaten abweichen. Diese Einschränkungen lassen sich technisch kaum verhindern. Die Forschung arbeitet aber an Ansätzen, die den Umgang damit verbessern. Beispielsweise indem die Resultate erklärbarer werden, der Mensch besser in die Resultatvalidierung eingebunden wird oder verschiedene KI-Methoden zu robusteren Systemen kombiniert werden.

Was bedeutet der «AI Act» der EU für die Schweiz?
Nadja Braun Binder, Professorin für Öffentliches Recht an der Universität Basel
Die KI-Verordnung der EU zielt darauf ab, den europäischen Binnenmarkt zu stärken und menschenzentrierte, vertrauenswürdige KI zu fördern, während sie die Gesundheit, Sicherheit und Grundrechte schützt. Sie wirkt als Produktsicherheitsverordnung und sieht für KI-Systeme mit spezifischem Verwendungszweck vier Risikostufen vor: verbotene Praktiken, Hochrisiko, begrenztes Risiko und minimales Risiko. KI-Systeme mit allgemeinem Verwendungszweck und Basismodelle unterliegen zusätzlichen Transparenz-Pflichten sowie weiteren Pflichten bei systemischem Risiko.
Viele Bestimmungen der KI-Verordnung sind vage und müssen konkretisiert werden, was zum Teil in technischen Normen geschehen wird, die faktisch auch für die Schweiz relevant sein werden.
Eine Übernahme der KI-Verordnung könnte zu Doppelungen führen, würde hinsichtlich der Anerkennung von Konformitätsbewertungsverfahren keine Garantien bieten und könnte die Marktstellung grosser Tech-Konzerne weiter festigen.

Akademien der Wissenschaften Schweiz (a+) · Generalsekretariat
Haus der Akademien · Laupenstrasse 7 · Postfach · 3001 Bern · Schweiz
+41 31 306 92 20 · info@akademien-schweiz.ch · akademien-schweiz.ch @academies_ch
swiss_academies

Chancen von KI kontrolliert nutzen – wie kann dies gelingen?
Thomas Burri, Professor für Europarecht und Völkerrecht an der Universität St. Gallen
Der EU AI Act basiert grundlegend auf der Idee menschlicher Kontrolle von KI. Insbesondere beinhaltet er eine Vorschrift zu menschlicher Aufsicht über KI-Anwendungen, welche in die «Hochrisiko-Kategorie» fallen. Aber ist dies überhaupt möglich und sinnvoll? Thomas Burri unterscheidet zwischen «heisser» (z.B. Landung eines Flugzeuges, braucht sofortige Reaktionen) und «kalter» KI (z.B. Medizinische Analysen). Im Rahmen des Nationalen Forschungsprogramms 77 «Digitale Transformation» untersuchten sie, wie hilfreich die menschliche Kontrolle ist, um ein fliegendes Objekt mit der Hilfe von KI sicher zu landen («heisse» KI). Resultat: Menschen sind nicht besonders gut darin abzuschätzen, ob die KI bei der Problemlösung auf dem richtigen Weg ist oder nicht, und damit, ob ein Eingriff des Menschen zielführend ist. Auch Massnahmen, um die KI nachvollziehbarer zu machen, haben die Quote nicht unbedingt verbessert. Generell wird einem KI-Mensch «Team» mehr Verantwortung anvertraut als KI oder Menschen alleine. Da besteht das Risiko, dass der Mensch zum Sündenbock wird für Fehler der KI. Basierend auf den wissenschaftlichen Resultaten hat Thomas Burri acht Denkanstösse präsentiert, siehe Präsentation.

1.2. Diskussion

Politisch ist die Forderung nach Transparenz naheliegend. Müssen wir uns angesichts der Blackbox-Thematik davon verabschieden?
Wir können grundsätzlich wissen, wie eine KI funktioniert und wie sie trainiert wurde. Aber selbst dann ist es für Anwender:innen unmöglich nachzuvollziehen, wie eine KI nun auf eine bestimmte Lösung gekommen ist. Da ist Transparenz nicht möglich. Hingegen ist eine Deklaration möglich, also dass man angeben muss, wo KI drin ist. Auch bezüglich der fürs Training verwendeten Daten kann Transparenz gefordert werden.
Es stellt sich auch die Frage, wie hilfreich Transparenz ist. Wir nutzen viele Dinge (z.B. Computer) erfolgreich, ohne verstehen zu müssen, wie sie genau funktionieren. Transparenz geht auch mit dem Risiko einher, die Verantwortung abzuschieben, so wie dies heute bei Cookiebannern geschieht. Man flutet die Anwender:innen mit unzumutbar vielen und komplexen Informationen und überträgt ihnen damit die Verantwortung für ihr Handeln. Besser ist es, KI sicherer zu machen bzw. zu überlegen, wie wir die grössten potenziellen Schäden verhindern können.

Der AI Act der EU ist ein Versuch, einen klugen Umgang zu finden. Die Frage stellt sich Ländern weltweit. Ist das nicht Schattenboxen? Die rasante Entwicklung wird uns von den grossen Playern vorgegeben. Wäre es nicht eher eine wettbewerbsrechtliche Frage?
Die Frage ist sehr berechtigt. Aber der Staat hat durchaus Handlungsmöglichkeiten und gerade Big Tech ist durchaus an Regulierungen interessiert. Mit einem pragmatischen Ansatz, der mit den Entwicklungen Schritt hält, können die Risiken eingeschränkt werden. Auch die EU musste den Act anpassen, weil während der Ausarbeitung die grossen Sprachmodelle aufgekommen sind. Wichtig ist zudem die Innovation zu fördern, um eine Vielfalt an Akteuren zu haben. Trotz exzellenter Forschung an den Hochschulen passiert hier in der Schweiz nicht viel, etwa in der Entwicklung von KI für spezifische Anwendungen.

Die Risikobasierung des AI Act der EU scheint sinnvoll zu sein. Damit verknüpft ist aber eine Bewilligungspflicht für riskante Ansätze: kann das funktionieren? Müsste man nicht klarer mit Verboten arbeiten?
Eine Zertifizierung ist herausfordernd, aber möglich. Mit technischen Normen zu Cybersecurity etwa gibt es gute Erfahrungen. Sie gelten international und können Basis für eine Bewilligung sein. Wichtig ist es, auch in der Schweiz eine Zertifizierungsstelle zu haben, um die Kompetenz im Inland zu sichern.

Den Ansatz, KI primär über die diversen Spezialgesetze zu regeln, überzeugt. Gibt es eine Art Kartographie, um zu zeigen, wo heute die KI schon ausreichend adressiert ist bzw. wo es Lücken gibt?

Eine solche Übersicht wäre hilfreich, existiert aber leider (noch) nicht. Grundsätzlich müssen viele Gesetze geprüft werden, ob sie KI angemessen regulieren, und ggf. angepasst werden. Bei den dringendsten Themen, etwa Verkehr, passiert dies bereits. Im Medizinbereich dagegen ist die Regelungsdichte wohl schon ausreichend; hier stellt sich mehr die Frage von Harmonisierungen. Neben dem Anpacken der dringendsten Themen sollte man auch Querschnittsgesetze wie etwa das Datenschutzgesetz angehen.

Müsste die Schweiz den AI Act übernehmen, falls das Rahmenabkommen kommt?
Nein, ziemlich sicher nicht, da es kein bilaterales Abkommen gibt, das den AI Act betrifft. Da der Entwurf des wohl kommenden Vertrages, bzw. der Verträge, aber noch nicht publiziert ist, kann man die Frage noch nicht definitiv beantworten.

Mit der Swiss AI Initiative möchten die ETHs einen eigenen Ansatz entwickeln. Ist das ein Weg, um die Probleme etwa bezüglich Bias oder Nachvollziehbarkeit anzugehen?
Das wird sich zeigen. Zentral ist der Datenzugang. Was sind Daten wert? Wie kommt man, etwa der Staat, zu qualitativ guten Daten inklusive Metadaten? Über die Metadaten finden wir auch eine Lösung die Qualität der Daten aufzuzeigen. Deshalb reguliert auch die EU den Datenzugang, das ist ein wichtiger Ansatz. Die Schweiz hat die Kompetenzen, hier eine Vorreiterrolle einzunehmen. Das Bias-Problem in den Daten wird regulatorisch nicht einfach zu fassen zu sein. Besonders problematisch sind Bias, die sich selbst verstärken. Beispiel: In den USA trauen People of Color dem Gesundheitssystem weniger, weshalb es weniger Daten gibt. Und weil es weniger Daten gibt, werden People of Color schlechter behandelt, das Vertrauen sinkt weiter.

KI bringt neben «Betriebsrisiken» auch soziale Risiken mit. Können Sie dies ausführen?
Ja, die Umwälzungen sind potenziell gross und schnell. Noch vor kurzem haben viele ihren Kindern oder Enkeln gesagt, Programmierer:in sei ein sicherer Job. Ist das immer noch so? Auch vermeintlich sehr sichere Berufsfelder geraten unter Automatisierungsdruck. Im Bereich der Übersetzungen gibt es bereits seit zehn Jahren KI basierte Software und viele Übersetzer:innen verlieren massiv an Aufträgen. Was bedeutet dies für die Aus- und Weiterbildung, wenn Maschinen zunehmend komplexere Aufgaben übernehmen können?

2. Vertiefung

Sämtliche weiterführenden Dokumente finden Sie online unter go.akademien-schweiz.ch/atable:
- Präsentationen
- Kontakte Wissenschaft
- Policy Paper «Der menschliche Umgang mit Künstlicher Intelligenz und deren Regulierung» von Thomas Burri, Markus Christen und weiteren
  https://www.unisg.ch/fileadmin/user_upload/HSG_ROOT/_Kernauftritt_HSG/Universitaet/Schools/LAW/Faculty/Burri/241118_MAIC_screen_de.pdf
- Diverse White Papers zur Regulierung von KI von Nadja Braun Binder und weiteren:
  https://www.itsl.uzh.ch/de/Forschung-und-Beratung/Forschungsprojekte/nachvollziehbare-algorithmen.html
  https://www.youtube.com/watch?v=hvCnTw5gTvc
- Diverse Publikationen zur KI von TA-SWISS, u.a. zu Deepfakes; zur automatischen Stimm-, Sprach- und Gesichtserkennung oder zu ChatGPT
  https://www.ta-swiss.ch/themen#informationsgesellschaft
  Derzeit läuft bei TA-SWISS eine Studienausschreibung zu grossen Sprachmodellen sowie eine Studie zur Nutzung von Gesundheitsdaten.
- Diverse SATW Blogartikel und weitere Beiträge von Expert:innen zum Thema KI Recht und Technik sowie Daten und KI allgemein:
  https://www.satw.ch/de/kuenstliche-intelligenz