SP Grundsatzpapier: Regulierung von KI-Systemen

Einleitung:

Die Entwicklung künstlicher Intelligenz schreitet schnell voran, entsprechende Anwendungen finden in immer mehr Bereichen Anwendung.[1] Innert Sekunden kreieren öffentlich zugängliche Tools wie ChatGPT, Midjourney oder Soundraw Texte, Bilder oder Musikstücke. Sie werden zum Lernen benutzt, erleichtern mitunter Arbeitsprozesse, schaffen Kunst.

Nach anfänglicher Begeisterung über die neue Technologie drängen sich immer mehr gesellschaftliche und politische Fragen auf: Was heisst es für die Zukunft von politischer Meinungsbildung und Mitsprache, wenn fast ohne Aufwand Propaganda in unbegrenzten Mengen getextet werden kann? Welchen Effekt haben die Algorithmen von Plattformen auf die Informationen, die Menschen konsumieren? Welche Auswirkungen hat der Einsatz von grossen Sprach- und Bildmodellen wie ChatGPT oder Midjourney für die Arbeitswelt, etwa für Berufe im Kreativbereich? Was ist mit Urheberrechten, was mit dem Datenschutz, was mit der Rechenschaftspflicht? Wie lassen sich Diskriminierungen verhindern, wenn eine künstliche Intelligenz mit Bias in Rekrutierungsprozessen eingesetzt wird? Wie kann man sich gegen Entscheidungen von KI-Systemen zur Wehr setzen, etwa bei der Job- oder Kreditvergabe? Welche Personen und Organisationen erschaffen, bestimmen und kontrollieren die eingesetzten Systeme, Algorithmen und Metriken? Wie lässt sich der enorm hohe Verbrauch von KI-Systemen an Strom und Ressourcen begrenzen? Gleichzeitig haben algorithmische Systeme auch ein enormes Potential: Wie kann dieses dazu genutzt werden, um konkrete Verbesserungen für die Gesellschaft zu erzielen – etwa, um den ökologischen Herausforderungen unserer Zeit zu begegnen?

[1] In der Definition der OECD handelt es sich bei einem KI-System dabei um ein „maschinenbasiertes System, das für bestimmte von Menschen definierte Ziele Voraussagen machen, Empfehlungen abgeben oder Entscheidungen treffen kann, die das reale oder virtuelle Umfeld beeinflussen». Der Begriff „Künstliche Intelligenz“ ist allerdings mit Vorsicht zu verwenden, da „Intelligenz“ Erwartungen und Befürchtungen weckt. In diesem Papier soll es um eine Vielzahl von Systemen gehen, die automatisierte Entscheide fällen, Empfehlungen vornehmen, Empfehlungen machen oder Vorhersagen treffen. Die *Digital Society Initiative* spricht in ihrem Positionspapier von so genannten „algorithmischen Systemen“, *AlgorithmWatch* und die *Digitale Gesellschaft* wiederum von Automatisierten Entscheidungssystemen (ADM-Systeme). Mit einem weit gefassten Begriffs wie „algorithmische Systeme“ werden auch Anwendungen erfasst, die auf anderen Technologien beruhen. Im Weiteren soll daher mehrheitlich dieser Begriff verwendet werden.

Vor der ungeahnten Mächtigkeit der Technik warnte jüngste ein offener Brief, der bis heute nahezu dreissigtausend Unterschriften gesammelt hat, darunter die von Elon Musk und vielen namhaften KI-Forscher:innen.[2] In ihm wird aufgerufen, die Entwicklung grosser KIs sofort für mindestens sechs Monate auszusetzen, weil Systeme wie das Sprachmodell ChatGPT-4 inzwischen zu mächtig und zu gefährlich geworden seien. Es drohten «fundamentale Risiken für die Gesellschaft und die Menschheit» durch «auf Menschenniveau agierender KI». Bis man sich nicht darauf geeinigt habe, wie das zu regulieren sei, sollten alle KI-Labore auf weitere Forschung verzichten.[3] Von vielen namhaften Forscher:innen, die teils gar im Brief zitiert werden, wurde der Brief und das Institut, das hinter dem Brief steht, aber auch heftig kritisiert.[4][5]

Zur Recht: Während die Befürchtung, KI-generierte Texte und Bilder könnten Informationskanäle mit Unwahrheiten und Propaganda überschwemmen, durchaus berechtigt ist, ist der Brief ansonsten von apokalyptischen Fantasien über die völlige Ersetzung des Menschen durch Maschinen und «den Kontrollverlust über unsere Zivilisation» getragen. Was aber mit grossen Sprachmodellen wie ChatGPT tatsächlich auf uns zukommt, ist weniger eine *technische* Katastrophe bösartiger Computer. Viel konkreter drohen die Sprachmodelle ein *demokratisches* Desaster zu werden – durch die Privatisierung von Sprachtechnologien als zukünftigem Ort politischer Öffentlichkeit. Genau an dieser Stelle kommen Politik und Zivilgesellschaft ins Spiel – und sie sind gefordert.

Die Entwicklungen der letzten Jahre haben gezeigt: ein KI-System wird umso leistungsfähiger, mit je mehr Daten man es füttert. Mit der Grösse steigen aber auch die Kosten exorbitant. Wie eine Studie der Stanford University zu neuen Sprach-KIs zeigt, hat der Wettlauf um immer umfangreichere Modelle inzwischen dazu geführt, dass nur noch eine Handvoll von Firmen im Rennen sind – neben GPT-Entwickler OpenAI sind das Googles Deepmind und Meta. Kleinere, nichtkommerzielle Unternehmungen und Universitäten spielen beim Erreichen immer neuerer Grössenrekorde so gut wie keine Rolle mehr. Damit stehen wir einem neuen Oligopol gegenüber, das Sprachtechnologien in der Hand weniger privatwirtschaftlicher Firmen

[2] https://futureoflife.org/open-letter/pause-giant-ai-experiments/. Am Brief wurde auch Kritik laut. Das initiierende Institut FLI (Future of Life Institute) vertrete umstrittene Ideologien, schüre einen KI-Hype und lenke mit seinem Fokus auf eine zukünftige Superintelligenz von real existierenden Problemen und Herausforderungen ab.
[3] Siehe auch: https://netzpolitik.org/2023/offener-brief-zu-ki-opfer-des-hypes/
[4] https://www.dair-institute.org/blog/letter-statement-March2023
[5] https://www.heise.de/meinung/Welche-Ideologie-hinter-Geoffrey-Hintons-Warnungen-steckt-8988646.html

konzentriert. Wie und mit welchen Daten diese Systeme aber gefüttert werden und was am Schluss dabei rauskommt, ist eine eminent politische Entscheidung. Und diese darf nicht einigen wenigen Firmen überlassen werden, die keiner demokratischen Kontrolle unterliegen und niemandem gegenüber rechenschaftspflichtig sind.

Eine demokratische Kontrolle und Regulierung von KI-Systemen ist angesichts der jüngsten Entwicklung und der Leistungsfähigkeit der neuen Systeme wie GPT4 dringlich. Zahlreiche Regulierungsvorschläge werden derzeit diskutiert: die EU-Kommission hat im März 2023 mit dem „EU AI Act" einen neuen Gesetzesentwurf präsentiert, der die Anwendung von algorithmischen Systemen in Europa regeln soll, die zuständige Kommission hat ihn in einer verschärften Version am 11. Mai 2023 gutgeheissen;[6] in den USA hat die Regierung im Oktober 2022 den AI Bill of Rights zum verantwortungsvollen Umgang mit künstlicher Intelligenz vorgelegt;[7] der Europarat berät derzeit eine Konvention mit dem Titel «On Artificial Intelligence, Human Rights, Democracy and the Rule of Law», die noch dieses Jahr verabschiedet werden soll.[8] Bereits im Oktober 2022 haben zudem Europaparlament und Rat mit dem „Digital Markets Act" und dem „Digital Services Act" weitreichende neue Regularien für die Informations- und Digitalbranche erlassen.

Auch die Schweiz hat erste Schritte hin zu einer Regulierung unternommen. So hat der Bundesrat jüngst einen Bericht und Leitlinien für die Bundesverwaltung verabschiedet.[9] Des Weiteren hat der Bundesrat angekündigt, eine Vernehmlassungsvorlage für die Regulierung von Kommunikationsplattformen (für die Übernahme des EU Digital Services Act) auszuarbeiten.[10] Wie der EU Digital Services Act in der der Schweiz übernommen wird, ist derzeit noch offen. Erste Hinweise deuten auf eine abgeschwächte Übernahme des DSA hin – so taucht etwa der freie Zugang zu Daten für die Forschung nicht in der Vorlage auf, was auch am starken Lobbying der Tech-Unternehmen – insbesondere von Google – liegen dürfte. Auch beim Thema KI zeigt sich die Schweiz bisher zurückhaltend. Noch Anfang 2022 antwortete der Bundesrat auf einen Vorstoss von SP-Nationalrätin Min Li Marti, es brauche kein Gesetz zur

[6] https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence
[7] https://www.whitehouse.gov/ostp/ai-bill-of-rights/
[8] https://coe.int/en/web/artificial-intelligence
[9] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-81319.html
[10] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-94116.html

künstlichen Intelligenz. In Bezug auf die internationalen Regularien will die Schweiz, so die Losung des EDA in seinem KI-Bericht, die internationalen Regeln zur künstlichen Intelligenz «in ihrem Sinne» mitgestalten. Das hiess bisher vor allem eins: eine *laissez-faire* Umsetzung im Sinne der marktwirtschaftlichen Akteure.

Aus sozialdemokratischer Sicht ist eine Regulierung von algorithmischen Systemen fundamental. KI-Systeme tangieren die Grundwerte und Anliegen unserer Partei: Einsatz für Grund- und Menschenrechte, für gute Arbeit und für den Erhalt und die Stärkung einer demokratischen politischen Öffentlichkeit, Kampf gegen Diskriminierung und Parteinahme für die Menschen statt für die Profite der Konzerne. Aus diesem Grund werden in diesem Papier Grundsätze und konkrete Forderung für den Umgang mit algorithmischen Systemen formuliert.

Eckpunkte für eine sozialdemokratische Regulierung von KI und weiteren algorithmischen Systemen:

1. Im Interesse der Menschen, nicht der Konzerne

Bei der technologischen Entwicklung müssen die Interessen der Menschen und des Gemeinwohls im Zentrum stehen. Technische Entwicklung ist kein Selbstzweck und kein Naturereignis, sondern ein Mittel zum Zweck mit dem Ziel, das Leben der Menschen zu verbessern und den chancengleichen Zugang zu Ressourcen zu vereinfachen. Es gibt und gab immer wieder Technologien, die im Interesse der Menschheit stark kontrolliert oder gar eingeschränkt wurden (z.B. Humanmedizin, Nukleartechnologie etc.). Es ist unsere Aufgabe als demokratische Gesellschaft, zu definieren, welche Technologien wir für welche Zwecke anwenden wollen.

2. Grundrechte müssen gewahrt werden / Ethik und Folgenabschätzungen sind zwingend

Eine technologische Entwicklung im Interesse der Menschen heisst auch eine technologische Entwicklung, die die Grundrechte wahrt. Diese müssen sowohl in der Entwicklung wie auch in der Anwendung beachtet werden. Ethische Überlegungen müssen von Anfang in den Prozess einfliessen, die Folgen geprüft und abgeschätzt werden.

3. Einsatz gegen Diskriminierung

Automatisierte Entscheidungssysteme beinhalten immer auch die Gefahr der Diskriminierung.

In verschiedenen Fällen wurde bereits nachgewiesen, dass vermeintlich neutrale und objektive Systeme zu diskriminierenden Verhalten geführt haben.[11] Auch Entscheide von Menschen sind nie frei von Diskriminierung und Vorurteilen, umso wichtiger ist es daher bei automatisierten Entscheidungen, dass Transparenz, Nachvollziehbarkeit, eine Aufsicht und Möglichkeiten zur Einsprache gegeben sind. Sinnvoll ist es im Grundsatz, Diskriminierung allgemein zu regeln, unabhängig davon, ob Menschen oder Maschinen die Entscheidungen treffen. Hier kann der Einsatz von automatisierten Systemen auch Vorteile bieten: sie können untersucht werden und ihr Bias ist nicht durch jahrelange Sozialisierung gefestigt, sondern kann durch Anpassungen am System korrigiert werden. Die Entwicklung von Explainable AI und balancierten Datensätzen können hier weiterhelfen.

4. Technologieneutrale Regulierung

Die Regulierung muss technologieneutral erfolgen, da die Herausforderungen unabhängig von der verwendeten Technologie geregelt werden muss. Die Regelungen können aufgrund der schnellen Entwicklung nur Bestand haben, wenn sie unabhängig von bestimmten Technologien angewendet werden kann. Dennoch können spezifische Einsatzfelder benannt und einzeln reguliert werden (z.b. Predictive Policing).

5. Transparenz und Nachvollziehbarkeit, Deklarationspflicht

Der Einsatz von algorithmischen Systemen muss erkennbar und verständlich sein. Dies gilt insbesondere im Fall einer individuellen Betroffenheit eines automatisierten Entscheids wie beispielsweise im Justizsystem oder in Bewerbungsverfahren. Dabei ist nicht nur die Transparenz entscheidend, sondern auch die Nachvollziehbarkeit der Entscheide. Das bedingt die Offenlegung der Trainingsdaten, sowie den Parametern, die für das Training verwendet wurden, in öffentlichen Verzeichnissen und Registern. Die Systeme müssen zur Überprüfbarkeit zur Verfügung gestellt werden. Beim Einsatz von algorithmischen Systemen durch Behörden müssen Betroffene die Möglichkeit haben, Einspruch zu erheben und eine erneute Beurteilung durch einen Menschen erhalten. Dafür brauchen sie den Zugang zu den notwendigen Informationen. Für den Einsatz von solchen Systemen in der Privatwirtschaft braucht es

[11] Siehe z.B. https://fra.europa.eu/en/publication/2022/bias-algorithm oder https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

ebenfalls eine Rekursmöglichkeit. Für alle Einsatzzwecke, bei denen signifikante Entscheide getroffen werden, braucht es eine vorgeschriebene Evaluation der Systeme, um möglichen diskriminierenden Effekten von solchen Systemen und ihren Einsatz entgegenzuwirken. Öffentliche Forschungsinstitute oder zivilgesellschaftliche Akteure sollen solche Evaluationssysteme bereitstellen und stetig den neusten Erkenntnissen anpassen.

6. Schutz vor Manipulation / Media und Algorithmic Literacy fördern

Die Informations- und Meinungsfreiheit sind verfassungsmässig garantierte Grundrechte und Grundpfeiler der demokratischen Meinungs- und Willensbildung. Eine Einschränkung der Verbreitung von Inhalten darf daher nur mit der grösstmöglichen Vorsicht einhergehen. Gleichzeitig gibt es Gefahren von Manipulation und Fehlinformationen, die sich durch KI-Systeme noch verstärken können. Eine funktionierende Gesellschaft ist daher auch darauf angewiesen, zu wissen, welchen Informationen man vertrauen kann und wie sie zu Stande kommen und weshalb sie gerade angezeigt werden. Das bedeutet, einen Fokus auf die Aufklärung und Stärkung der Medienkompetenz wie auch der Förderung Medien zur Aufrechterhaltung einer breiten und unabhängigen Medienlandschaft zu legen.

Auch bei der Verbreitung von medialen Inhalten spielen KI Systeme eine immer gewichtigere Rolle. KI Systeme bestimmen individuell, wer welches Video oder welchen Beitrag auf Youtube, TikTok, Instagram oder Twitter zu sehen bekommt. Und zwar mit dem Ziel, Menschen möglichst eng und lange an die jeweilige Plattform zu binden und möglichst viel personalisierte Werbung schalten zu können. Algorithmische Selektion wiederum befördert die Entstehung von partikularen, von einander isolierten Teilöffentlichkeiten und Pseudo-Gemeinschaften – ein idealer Nährboden für ressentimentale Politik, wie die zahlreichen rechten Kampagnen der letzten Jahre beweisen. Während sich Meinungen radikalisieren, verschwinden widersprüchliche Inhalte aus dem Feed. In einem besonders frappierenden Fall hat diese Dynamik – befeuert durch den Algorithmus von Facebook – den Genozid an den Rohingya in Myanmar mitbegünstigt.[12]

Wenn Systeme zum Einsatz kommen, die Inhalte durch Algorithmen filtern und gewichten, soll dies offengelegt werden. Nutzer:innen sollen die Möglichkeit haben, die

[12]https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/

Parameter, die bestimmen, was sie zu sehen kriegen, selber festzulegen können. Unter anderem sollen sie auch Beiträge sehen können, die nicht in das Schema passen, welches das System für sie berechnet hat.

7. Sorgfaltspflicht

Algorithmische Systeme können Schaden anrichten und haben dies auch schon konkret getan.[13] Das heisst, dass sich hier zentral Fragen der Sorgfaltspflicht stellen. Hier ist wichtig festzuhalten, dass die Systeme über den ganzen Produktzyklus hindurch reguliert werden sollen. Es braucht Sorgfaltspflichten sowohl für Entwickler:innen als auch für jene, die es einsetzen – ohne dass sie dieser Verantwortung entgehen können z.B. durch einen reinen Haftungsausschluss. Es muss auch klar sein, dass die Sicherheit gewährleistet ist, bevor eine Anwendung auf den Markt kommt und dass die Hersteller und Anbieter für allfällige Schäden haften müssen.

8. KI und die Zukunft der Arbeit

Die maschinellen Kontrolltechnologien des «Tracking, Tracing, Targeting, Ranking, Scoring und Profiling» haben längst die Arbeitswelt erfasst. Die digitale Komponente besteht in Industrie- und Dienstleistungsberufen vor allem in der algorithmischen Arbeitssteuerung, die auf die Kontrolle der Arbeitsprozesse und die Extraktion von Daten zielt.[14] Dadurch sollen Prozesse vereinfacht und automatisiert werden. Der Einsatz von KI-Systeme in der Arbeitswelt darf aber nicht zu einer weiteren Prekarisierung und Dequalifizierung von Arbeit führen. Im Gegenteil muss es vielmehr das Ziel sein, algorithmische System dafür einzusetzen, Arbeit zu erleichtern und die dadurch erzielten Produktivitätsgewinne an die Menschen rück zu verteilen.

Auch heute werden noch immer viele Funktionen, die automatisiert zu sein scheinen, von Menschen erledigt. Zur Moderation von Content greifen die grossen Plattformunternehmen und KI-Entwicklerfirmen auf Klickworker zurück, die unter ausbeuterischen

[13] Vgl. hierzu etwa den Skandal um Kindergeldansprüche in Holland: https://netzpolitik.org/2021/kindergeldaffaere-niederlande-zahlen-millionenstrafe-wegen-datendiskriminierung/

[14] Vgl. Simon Schaupp, Technopolitik von unten. Algorithmische Arbeitssteuerung und kybernetische Proletarisierung, Berlin 2022. Wenn Arbeit immer mehr nur noch im Abarbeiten der Anweisungen von algorithmischen Kontrollsystemen besteht, werden Fachkräfte überflüssig. Aus Sicht der Unternehmen entfällt so nicht nur die langwierige Einarbeitung, es erleichtert auch die Einbindung von prekärer, migrantischer Arbeitskraft.

Arbeitsbedingungen oft traumatisierende Arbeiten erledigen, indem sie tagtäglich Gewalt und Missbrauch von diesen Plattformen markieren und löschen.

9. Unerwünschte und verbotene Anwendungen

Es muss klar geprüft werden, ob es Anwendungen gibt, die nicht wünschbar oder gar klar schädlich sind und daher verboten werden müssen. Mögliche Beispiele sind etwa biometrische Erkennungssysteme (wie z.B. Gesichtserkennung) zu Identifizierungszwecken im öffentlichen zugänglichen Raum. Die Identifizierung kann nicht nur anhand des Gesichtes, sondern auch etwa der Stimme oder dem Gang erfolgen. Weitere Systeme, bei denen ein Verbot (ggf. jeweils mit Ausnahmen) zu prüfen ist: Emotionserkennung (ggf. mit Ausnahme für Menschen mit Beeinträchtigungen), biometrische Kategorisierung anhand sensitiver Daten, bestimmte Anwendungen im Bereich der Polizei und der Migration.

10. Entwicklung von offenen KI Systemen

KI-Systeme können zur Verbesserung der Lebenssituation der Menschen beitragen. Dank Übersetzungstools haben etwa mehr Menschen Zugang zu Informationen in einer Sprache, die sie auch verstehen; Algorithmen helfen beim Steuern von Stromproduktion und -verbrauch; NLP Systeme helfen beim Erkennen und Bekämpfen von *Hate Speech* im Internet. Damit KI-Systeme ihre sozialen und emanzipatorischen Potentiale entfalten können, müssen die Tools aber möglichst einfach und vor allem offen und in guter Qualität verfügbar sind. Öffentliche Forschungsanstalten sollen deshalb die Entwicklung offener Technologien unterstützen und offene Daten kuratieren und zur Verfügung stellen.

Um dem sich abzeichnenden Oligopol in der KI-Entwicklung entgegenzutreten, braucht es eine starke öffentliche Forschung. Nur so lassen sich Transparenz und verifizierbare Fakten gewährleisten. Hochschulen und andere öffentliche Institutionen müssen ausreichend finanzielle Mittel erhalten, um eine federführende Rolle bei der Verbesserung der Transparenz in der Entwicklung von KI-Systemen zu spielen. Derzeit lassen sich bei vielen der heute veröffentlichten Modellen (wie Google's Bard, OpenAI's GPT, usw.) keine verifizierbaren Untersuchungen durchführen, da die Konzerne ihre Daten nicht öffentlich machen und sich dem *Peer-Review-Prozess* entziehen. Die Entwicklung von KI-Systemen soll nicht den Profitinteressen einzelner Unternehmen dienen, sondern der Gesellschaft als Ganzes.

Politische Forderungen

- Vollständige Übernahme des „EU AI Act“
- Rahmengesetz für Regulierung von Algorithmischen Systemen
- Schaffung einer nationalen Ethikkommission für Künstliche Intelligenz
- Allgemeines Antidiskriminierungsgesetz inkl. spezifischer Massnahmen gegen algorithmische Diskriminierung
- Zwingende Folgenabschätzungen im Bereich Technologie, Regulierung, Gleichstellung und Datenschutz
- Pflicht zur Transparenz- und Nachvollziehbarkeit bei Anwendung von Algorithmischen Systemen, Deklarationspflicht, sowohl im öffentlichen wie im privaten Sektor (Register), Offenlegung von Quellcodes, Trainingsdaten und Parametern
- Rekursmöglichkeiten (Recht auf menschliche Begutachtung).
- Schaffung einer Aufsichts- und Kontrollbehörde
- Investition in Forschung und Entwicklung von offenen KI-Technologien
- Forschung in Systeme um Bias in KI Systemen zu erkennen und zu beheben, etwa Explainable AI und kuratierte, repräsentative Datensätze
- Produktehaftungen und Sanktionsmöglichkeiten
- Förderung von Aufklärung und Kompetenz im Umgang mit Algorithmischen Systemen
- Förderung einer breiten und vielfältigen Medienlandschaft
- Beseitigung von Diskriminierung zum Zugang zu neuen Technologien/ Gewährleistung digitale Bildungsgerechtigkeit/ Sicherstellung digitaler Grundkompetenzen aller Menschen
- Klare, nationale einheitliche Regeln zum Umgang mit KI auf allen Bildungsstufen
- Stärkung Jugend- und Konsument:innenschutz (z.B. Anpassung Gesetz über unlauteren Wettbewerb, Anpassung Jugendschutz bei suchterregenden

Onlineplattformen), separate Regulierung von Kommunikationsplattformen

- Verbot von biometrischer Erkennung im öffentlichen Raum
- Regulierungsansätze sollten bereits im Entwicklungsstadium eines KI-Systems greifen
- Gerechte Besteuerung von Tech-Firmen und Verhinderung von Monopolen durch griffiges Wettbewerbsgesetz
- Arbeitszeitverkürzung um den Produktivitätsgewinn durch Automatisierung gerecht zu verteilen

Glossar/Begriffe

AI/KI, Artificial Intelligence, Künstliche Intelligenz

System, das basierend auf einem Input einen Output generiert, der „intelligent“ wirkt. Kann ein System sein, das durch Machine Learning „trainiert“ wurde, oder ein System basierend auf klassischen Algorithmen.

Machine Learning, ML

System, das durch Beispieldaten lernt. Es gibt viele verschiedene Architekturen, die in ihrer Komplexität stark unterschiedlich sind, bsp: Lineare Regression, Bayes-Filter, Neuronale Netzwerke, Transformer, LSTM, Diffusore, Markov-Chain, U-Net, ...

Deep Learning

Grosses Machine Learning System, das mit sehr grossen Datenmengen trainiert wird, was man genau unter Deep Learning versteht, verändert sich in der Zeit. Systeme, die vor einigen Jahren noch Deep Learning waren, sind heute vergleichsweise klein.

LLM, Large Language Models

Grosse Systeme Künstlicher Intelligenz, die darauf trainiert wurden, menschliche Sprache zu verstehen und zu generieren. Sie sind "large" (gross) in Bezug auf die Menge an Trainingsdaten und die Grösse des zugrundeliegenden neuronalen Netzwerks. Beispiele sind die GPT-Familie von Modellen, welche von OpenAI entwickelt wurden.

Supervised Learning

Ein ML-System wird mit einem Datensatz trainiert, der sowohl Input als auch Output Daten enthält. Beispiele sind etwa eine lineare Regression oder ein Classifier.

Unsupervised Learning

Ein System trainiert auf Daten, die keine Labels besitzen. Beispiele sind zum Beispiel ein Clustering-Algorithmus oder ein Text-Transformer.

AGI, Artificial General Intelligence

Künstliche Intelligenz, die neue Fähigkeiten lernen kann, ohne dass die Architektur verändert werden muss. Bsp. GPT-4 lernt Autofahren.

Bias

In diesem Kontext: Machine Learning Systeme werden mit Daten trainiert, in denen bereits ein Bias vorhanden ist. Die Systeme übernehmen dann diesen Bias. Kann sein, weil gewisse Datenpunkte stark untervertreten sind, oder die Label in einem Supervised Learning Model durch menschliche Vorurteile verfälscht sind. Systemische Diskriminierungen werden so in die Modelle übertragen.

Generative Systeme

Grosser Hype um Generative Systeme, wie GPT oder Dall-E, Stable Diffusion, usw.

Vorallem um GPT-4 ein nicht fundierter Hype um „Emergency“, also neues Verhalten, das nicht gelernt wurde. GPT 4 hat kein eigenes Bewusstsein. Der Eindruck entsteht, weil generierte Texte für Menschen so wirken, als wäre Bewusstsein vorhanden.

Emergentes Verhalten lässt sich nicht überprüfen, da Datensatz nicht öffentlich. Die meisten Aufgaben dürften in irgendeiner Art im Datensatz vorkommen, mit dem die System trainiert wurden.

Recommender Systeme

Systeme, die bei Sozialen Netzwerken oder Videoplattformen verwendet werden, um Inhalte

benutzerbasiert anzuzeigen. Basieren auf diversen Signalen, wie Followern, Interaktionen anderer, ähnlicher Benutzer. Twitter hat den Algorithmus dazu veröffenlticht, andere Plattformen verwenden ähnliche Systeme.

Meist eine Kombination aus klassischer Algorithmik und Machine Learning.

Quellen:

https://www.digitale-gesellschaft.ch/uploads/2022/02/Position-der-Digitalen-Gesellschaft-zur-Regulierung-von-automatisierten-Entscheidungssystemen-1.0.pdf

https://algorithmwatch.org/de/regulierung-general-purpose-ai-ki-verordnung/

https://www.dsi.uzh.ch/dam/jcr:3a0cb402-c3b3-4360-9332-f800895fdc58/dsi-strategy-lab-21-de.pdf

https://www.sbfi.admin.ch/sbfi/de/home/bfi-politik/bfi-2021-2024/transversale-themen/digitalisierung-bfi/kuenstliche-intelligenz.html

https://chplusplus.org/automatisiertes-entscheiden-in-der-offentlichen-verwaltung/

Vorstösse SP:

18.4037 Motion Bendahan
Kompetenzzentrum für künstliche Intelligenz in der Bundesverwaltung
https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20184037

20.4338 Postulat Storni

Künstliche Intelligenz. Sicherheitsvorschriften, Transparenz und Information bei Anwendungen von maschinellem Lernen
https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20204388

16.332o Postulat Marti

Bericht zu Chancen und Risiken von künstlicher Intelligenz und Robotik

https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20163320

19.4462 Interpellation Graf-Litscher

Digitale Ethik. Verhinderung von Diskriminierung bei künstlicher Intelligenz

https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20194462

22.1051 ANFRAGE MARTI

Monitoring von Projekten zur künstlichen Intelligenz

https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20221051

21.4406 Postulat Marti

Bericht zur Regulierung von automatisierten Entscheidungssystemen

https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20214406