



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Eidgenössisches Departement für  
Wirtschaft, Bildung und Forschung WBF

**Staatssekretariat für Bildung,  
Forschung und Innovation SBFi**  
Forschung und Innovation

---

# Herausforderungen der künstlichen Intelligenz

Bericht der interdepartementalen  
Arbeitsgruppe «Künstliche Intelligenz»  
an den Bundesrat

---

**Mitglieder der interdepartementalen Arbeitsgruppe künstliche Intelligenz:**

Albert Blarer (armasuisse)  
Marcel Buffat (GS-UVEK)  
Christian Busch (SBFI; Projektleitung)  
Daniel Egloff (SBFI)  
Jon Fanzun (EDA)  
Gregor Haefliger (SBFI, Vorsitz)  
Philipp Langer (SBFI)  
Bertrand Loison (BFS)  
Thomas Luder (SIF)  
Arié Malz (GS-EFD)  
Eric Scheidegger (SECO)  
Thomas Schneider (BAKOM)  
Michael Schöll (BJ)  
Peter Schwaar (EZV)  
Michael Stämpfli (BAKOM)  
Verena Weber (GS-WBF)

## Inhaltsverzeichnis

<b>Executive Summary .....</b>	<b>6</b>
<b>1 Auftrag / Ausgangslage .....</b>	<b>17</b>
<b>2 Einordnung des Begriffs «Künstliche Intelligenz» (KI) .....</b>	<b>18</b>
2.1 Ansätze zur Definition künstlicher Intelligenz .....	19
2.2 KI-Systeme versus Methoden des maschinellen Lernens .....	20
2.3 Künstliche Intelligenz als Basistechnologie .....	22
<b>3 Charakteristika / Strukturelemente von KI-Systemen .....</b>	<b>23</b>
3.1 Lernen aus Daten .....	25
3.2 Vorhersagen als Grundlage für (automatisierte) Entscheidungen .....	26
3.3 Menschliche Intelligenz .....	27
3.4 Autonome Handlungsfähigkeit.....	29
3.5 Fehlende Nachvollziehbarkeit .....	31
3.6 Systematische Fehler (Bias) und Scheinkausalitäten .....	32
<b>4 Übergeordnete Aspekte und rechtliche Einordnung .....</b>	<b>34</b>
4.1 Grundsätze der Politik des Bundes im Umgang mit neuen Technologien .....	34
4.2 Autonomie, Verantwortlichkeit und Haftung .....	36
4.3 Nachvollziehbarkeit und Transparenz .....	37
4.4 Bias und Diskriminierung .....	38
4.5 Datenzugang und Datenschutz .....	39
<b>5 Künstliche Intelligenz – Forschung, Entwicklung und Anwendung in der Schweiz .....</b>	<b>40</b>
5.1 KI-Forschungslandschaft Schweiz: Akteursübersicht .....	41
5.2 Leistung von F&E in der Schweiz.....	42
5.3 Herausforderungen im Forschungs- und Innovationsbereich .....	47
<b>6 KI-Themenfelder nach Politikbereichen .....</b>	<b>49</b>
6.1 Internationale Gremien und künstliche Intelligenz.....	50
6.2 Digital Europe Programme .....	53
6.3 Veränderungen der Arbeitswelt.....	56
6.4 Künstliche Intelligenz in Industrie und Dienstleistungen .....	58
6.5 Künstliche Intelligenz in der Bildung.....	61
6.6 Anwendung von künstlicher Intelligenz in Wissenschaft und Forschung.....	64
6.7 Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik .....	66
6.8 Künstliche Intelligenz, Medien & Öffentlichkeit.....	70
6.9 Automatisierte Mobilität und künstliche Intelligenz.....	73
6.10 Künstliche Intelligenz im Gesundheitswesen .....	77
6.11 Künstliche Intelligenz in der Finanzwirtschaft.....	79
6.12 Künstliche Intelligenz in der Landwirtschaft .....	82
6.13 Energie, Klima, Umwelt und künstliche Intelligenz.....	84
6.14 Künstliche Intelligenz in der Verwaltung.....	87
6.15 Weiterentwicklung des allgemeinen Rechtsrahmens im Hinblick auf künstliche Intelligenz....	91
6.16 Einsatz von künstlicher Intelligenz in der Justiz .....	94
6.17 Künstliche Intelligenz, Daten und Immaterialgüterrecht .....	96
<b>Anhang 1: Gesamtübersicht Aktionsfelder .....</b>	<b>99</b>
<b>Anhang 2: Maschinelles Lernen .....</b>	<b>103</b>
<b>Anhang 3: Bibliographie .....</b>	<b>109</b>

## Abbildungsverzeichnis

Abbildung 1: Maschinelles Lernen als Bestandteil von KI-Systemen .....	22
Abbildung 2: Strukturelemente zur Charakterisierung von KI-Anwendungen.....	24

Abbildung 3: Täuschung von KI-Systemen (kontradiktorische maschinelle Lerntechniken) .....	29
Abbildung 4: Fähigkeiten zur Interaktion mit der Umwelt.....	30
Abbildung 5: KI-Projektförderung (Horizon 2020, SNF und Innosuisse): Anzahl Projekte 2015-2018.	41
Abbildung 6: Entwicklung Fördervolumen KI-Forschung .....	42
Abbildung 7: Anzahl KI-Projekte pro Million Einwohner in Horizon 2020, 2015-2018 .....	42
Abbildung 8: Anzahl Publikationen in der NeurIPS nach Organisation, 2017 .....	43
Abbildung 9: Anzahl Weltklassepatente pro Million Einwohner im KI-Bereich im Jahr 2018 .....	45
Abbildung 10: Anzahl KI-Startups pro Million Einwohner weltweit im Jahr 2018.....	46
Abbildung 11: Anzahl KI-Startups pro Million Einwohner in Europa im Jahr 2017 .....	46
Abbildung 12: KI-Struktur in der Schweiz, Subtechnologien und Anwendungsfelder im Jahr 2018.....	47
Abbildung 13: Herausforderungen von KI-Systemen für Unternehmen.....	58
Abbildung 14: Lernansätze, Funktionen und Anwendungsgebiete des maschinellen Lernens.....	103
Abbildung 15: Verschiedene Abstraktionslevel von künstlicher Intelligenz .....	104
Abbildung 16: Vereinfachte Darstellung eines Knotens eines künstlichen neuronalen Netzes.....	105
Abbildung 17: Vereinfachte Darstellung eines künstlichen neuronalen Netzes.....	105
Abbildung 18: Machine Learning/Deep Learning versus regelbasierte Systeme .....	106
Abbildung 19: Illustration eines Deep Learning Modells .....	107

## Tabellenverzeichnis

Tabelle 1: Übersicht prioritäre, bereichsspezifische Aktionsfelder des Bundes .....	12
Tabelle 2: Zentrale technische und rechtliche Herausforderungen von KI .....	34
Tabelle 3: Volumen und Impact der Publikationen im Bereich KI nach Ländern (2011-2015).....	44
Tabelle 4: Initiativen und Aktivitäten im KI-Bereich in der Schweiz (exemplarisch).....	59
Tabelle 5: Spezifische KI-Herausforderungen in Wissenschaft und Forschung .....	64
Tabelle 6: KI-Anwendungen in der Bundesverwaltung .....	88

## Abkürzungsverzeichnis

armasuisse	Bundesamt für Rüstung
ASTRA	Bundesamt für Strassen
BABS	Bundesamt für Bevölkerungsschutz
BAFU	Bundesamt für Umwelt
BAG	Bundesamt für Gesundheit
BAKOM	Bundesamt für Kommunikation
BAV	Bundesamt für Verkehr
BAZL	Bundesamt für Zivilluftfahrt
BEAT	Biometrics Evaluation and Testing
BFE	Bundesamt für Energie
BFI	Bildung, Forschung und Innovation
BFS	Bundesamt für Statistik
BJ	Bundesamt für Justiz
BK	Bundeskanzlei
BLW	Bundesamt für Landwirtschaft
BV	Bundesverfassung der Schweizerischen Eidgenossenschaft
DaziT	Programm zur Modernisierung / Digitalisierung in der Eidgenössischen Zollverwaltung
DEA	Direktion für europäische Angelegenheiten
DEP	Digital Europe Programme
DSG	Bundesgesetz über den Datenschutz
DSGVO	Datenschutz-Grundverordnung der Europäischen Union
DV	Direktion für Völkerrecht
EC	European Commission
EDA	Eidgenössisches Departement für auswärtige Angelegenheiten
EDI	Eidgenössisches Departement des Innern
EFD	Eidgenössisches Finanzdepartement
EFV	Eidgenössische Finanzverwaltung
EMRK	Europäische Menschenrechtskonvention
EPFL	École polytechnique fédérale de Lausanne

## Herausforderungen der künstlichen Intelligenz

ESTV	Eidgenössische Steuerverwaltung
ETH	Eidgenössische Technische Hochschule
EU	Europäische Union
EZV	Eidgenössische Zollverwaltung
F&E	Forschung und Entwicklung
GS	Generalsekretariat
ICT	Information and communications technology
IDAG	Interdepartementale Arbeitsgruppe
IDSIA	Istituto Dalle Molle di Studi sull'Intelligenza Artificiale
IEEE	Institute of Electrical and Electronics Engineers
IGE	Eidgenössisches Institut für Geistiges Eigentum
IKT	Informations- und Kommunikationstechnik
IoT	Internet of Things
ISB	Informatiksteuerungsorgan des Bundes
ITU	Internationale Fernmeldeunion
KI	Künstliche Intelligenz
KVG	Bundesgesetz über die Krankenversicherung
MINT	Mathematik, Informatik, Naturwissenschaft und Technik
ML	Maschinelles Lernen
NCCR	National Centre of Competence in Research
NCS	Nationale Strategie zum Schutz der Schweiz vor Cyber-Risiken
NDB	Nachrichtendienst des Bundes
NFP	Nationale Forschungsprogramme
NOGA	Allgemeine Systematik der Wirtschaftszweige
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
OGD	Open Government Data
ÖV	Öffentlicher Verkehr
RTVG	Bundesgesetz über Radio und Fernsehen
SBFI	Staatssekretariat für Bildung, Forschung und Innovation
SECO	Staatssekretariat für Wirtschaft
SEM	Staatssekretariat für Migration
SNF	Schweizerischer Nationalfonds
StGB	Schweizerisches Strafgesetzbuch
SUPSI	Scuola universitaria professionale della Svizzera italiana
swisstopo	Bundesamt für Landestopografie
UNESCO	Organisation der Vereinten Nationen für Erziehung, Wissenschaft und Kultur
UNO	Organisation der Vereinten Nationen
USI	Università della Svizzera italiana
UVEK	Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation
VBS	Eidgenössisches Departement für Verteidigung, Bevölkerungsschutz und Sport
WBF	Eidgenössisches Departement für Wirtschaft, Bildung und Forschung
WSIS	Weltgipfel zur Informationsgesellschaft
WTT	Wissens- und Technologietransfer
ZHAW	Zürcher Hochschule für Angewandte Wissenschaften

## Executive Summary

### I. Herausforderungen der künstlichen Intelligenz

Neue Anwendungsmöglichkeiten von Methoden künstlicher Intelligenz (KI) zählen aktuell zu den vielversprechendsten Entwicklungen der Digitalisierung. Künstliche Intelligenz hat bereits zahlreiche erfolgreiche Anwendungen ermöglicht, so etwa im Bereich der Bilderkennung, der medizinischen Diagnostik, der Sprachübersetzung oder in der Mobilität. Sie hat als Grundlagentechnologie das Potenzial, sämtliche Wirtschaftsbereiche zu verändern und weist ein erhebliches Innovations- und Wachstumspotenzial auf.

Während die mathematischen Grundlagen künstlicher Intelligenz bereits vor Jahrzehnten entwickelt wurden, haben erst die Verfügbarkeit von enormen Datenmengen und die rasante Entwicklung der Rechenkraft von Computern die sinnvolle und marktfähige Nutzung von Daten mit KI-Methoden ermöglicht. Ein Ende der technologischen Entwicklungen ist derzeit nicht absehbar. Es zeichnen sich bereits viele weitere Anwendungsbereiche von KI ab. Dazu zählen etwa Anwendungen in der Medikamentenentwicklung, der Echtzeitüberwachung von Maschinen und Produktionsprozessen, der Cybersicherheit oder der medizinischen Forschung.

Für die Schweiz ist es wichtig, die Potenziale, die sich durch die neuen Möglichkeiten der künstlichen Intelligenz ergeben, zu nutzen. Dazu gilt es, bestmögliche Rahmenbedingungen zu gewährleisten, damit sich die Schweiz als einer der weltweit führenden innovativen Standorte für Forschung, Entwicklung und Anwendung von KI etablieren und weiterentwickeln kann. Gleichzeitig gilt es auch, die mit der Anwendung von KI verbundenen Risiken zu adressieren und rechtzeitig Massnahmen zu ergreifen.

#### Gute Ausgangslage

Forschung und Entwicklung sind eine wichtige Grundlage für den Erhalt der Wettbewerbsfähigkeit in einem durch rasche technologische Änderung geprägten Umfeld. Der vorliegende Bericht zeigt auf, dass sich die Schweiz im Hinblick auf Forschung und Entwicklung im Bereich künstlicher Intelligenz in einer guten Ausgangslage befindet. Zwar ist der Umfang der KI-Forschung in der Schweiz leicht unterdurchschnittlich (gemessen am Anteil Publikationen pro Einwohner und am relativen Anteil der Publikationen). Bei den führenden Publikationen zählen Schweizer Forschungsinstitutionen jedoch zu den weltweit bedeutendsten Akteuren. Auch in der Entwicklung und Anwendung, gemessen an Patent- und Gründungsaktivitäten, zählt die Schweiz zu den dynamischsten Ländern weltweit.

Trotz dieser guten Ausgangslage bleiben die Herausforderungen angesichts der Geschwindigkeit der Entwicklung sowie des erheblichen Anwendungspotenzials der Technologien der künstlichen Intelligenz gross. Forschung, Innovation und Bildung kommen bei der Bewältigung dieser Herausforderungen eine zentrale Rolle zu. Es muss daher sichergestellt werden, dass die Kompetenzen in diesen Bereichen mit den Entwicklungen Schritt halten und gestärkt werden.

#### Ziel des Berichts

Neben der Geschwindigkeit der Entwicklung neuer KI-Technologien selbst, ergeben sich die wesentlichen **Herausforderungen in erster Linie aus neuen Anwendungsmöglichkeiten**, die mit der künstlichen Intelligenz möglich werden. Dabei muss einerseits gewährleistet werden, dass neue Anwendungen überhaupt ermöglicht werden; denn Forschung und Entwicklung können ihr Potenzial nicht entfalten, wenn für die Anwendung der daraus resultierenden Technologien grosse Hürden bestehen. Andererseits muss der Rechtsrahmen so ausgestaltet werden, dass mögliche unerwünschte Folgewirkungen der Anwendung neuer Technologien abgewendet werden können, ohne dem technologischen Fortschritt allzu sehr hinterherzuhinken. Es ist daher auch zu klären, ob sich aus der konkreten Anwendung von KI Folgen ergeben, die einer Anpassung der Regulierung bedürfen.

Der Bericht nimmt eine Auslegeordnung der relevanten Rahmenbedingungen im Hinblick auf die zunehmende Nutzung von KI vor, beleuchtet die spezifischen Herausforderungen in verschiedenen Anwendungsfeldern über Politikbereiche der gesamten Bundesverwaltung hinweg und diskutiert möglichen Anpassungsbedarf auf Bundesebene.

Der Bericht ist das Ergebnis der Arbeiten der interdepartementalen Arbeitsgruppe künstliche Intelligenz, die das Eidgenössische Departement für Wirtschaft, Bildung und Forschung (WBF) im Herbst 2018 im Auftrag des Bundesrates ins Leben gerufen hat (IDAG KI). Die IDAG KI hat dazu verschiedene thematische Arbeitsgruppen eingesetzt, welche im Rahmen ihrer Arbeiten auch zahlreiche externe Expertinnen und Experten konsultiert haben.

### Strukturelle Elemente von Systemen der künstlichen Intelligenz

Eine allgemein gültige und akzeptierte Definition von künstlicher Intelligenz existiert nicht. Um möglichen Handlungsbedarf für den Bund zu erörtern, stellt der Bericht daher nicht auf die Technologie an sich ab, sondern auf die Frage, welche neuen Anwendungsmöglichkeiten aus dieser Technologie heute oder in absehbarer Zeit entstehen und welche Implikationen diese mit sich bringen.

Im Hinblick darauf kann KI – anstelle einer Definition – durch verschiedene strukturelle Elemente charakterisiert werden, welche mit dem Einsatz von heutigen Anwendungen von KI-Systemen verbunden sind. Der vorliegende Bericht identifiziert namentlich vier zentrale Strukturelemente, die je nach Anwendungsbereich unterschiedlich stark zum Tragen kommen. Demnach sind KI-Systeme in der Lage,

- (1) Daten in Komplexität und Menge in einer Form auszuwerten, die mit anderen Technologien nach heutigem Stand nicht möglich wäre, insbesondere, wenn Algorithmen selbstständig Lernen und dabei in Daten relevante statistische Merkmale finden;
- (2) Vorhersagen als wesentliche Grundlage für (automatisierte) Entscheidungen zu erstellen;
- (3) dadurch Fähigkeiten nachzubilden, die mit menschlicher Kognition und Intelligenz in Verbindung gebracht werden;
- (4) auf dieser Basis weitgehend autonom zu agieren.

Die Elemente sind zwar einzeln in verschiedenen Ausprägungen auch in Nicht-KI-Anwendungen zu finden. Allerdings werden auf Basis von KI gänzlich neue Anwendungen in Kombination dieser Elemente überhaupt erst möglich (z.B. Gesichtserkennung, vollautomatisierte Fahrzeuge).

**Algorithmen des maschinellen Lernens (ML-Algorithmen)** stellen die zentrale und universelle Technologie hinter der erfolgreichen Entwicklung der gegenwärtigen KI-Systeme dar. ML-Algorithmen sind enorm mächtige **Methoden der KI**, deren Funktion allerdings darauf beschränkt ist, Muster in Daten zu erkennen und einfache Vorhersagen zu machen. Im Gegensatz dazu ist ein **KI-System** in der Lage, komplexe Probleme zu lösen, die bisher dem Menschen vorbehalten waren. Dies geschieht, indem komplexe Probleme in eine Reihe einfacher Vorhersageaufgaben unterteilt werden, von denen jede für sich genommen von einem «einfachen» ML-Algorithmus bearbeitet werden kann.

Solche KI-Systeme sind heute jeweils spezifisch für einen bestimmten Anwendungskontext definiert. Anders als ML-Algorithmen, welche zunehmend zu einer universellen Technologie geworden sind, wird daher das domänenspezifische Wissen über die konkreten Anwendungsbereiche, das notwendig ist, um ML-Komponenten in komplexen Anwendungen zu einer vollständigen Lösung zu kombinieren, auf absehbare Zeit nicht automatisiert und bedingt nach wie vor einen erheblichen menschlichen Input.

Es besteht Einigkeit, dass KI zwar Anwendungen ermöglicht, welche kognitiven und perzeptiven menschlichen Fähigkeiten nachempfunden sind und gewisse Aspekte von Intelligenz simuliert; nach

derzeitigem Stand der Technik ist jedoch KI bei weitem nicht mit der menschlichen Intelligenz vergleichbar.

### Spezifische Herausforderungen der künstlichen Intelligenz

Ansätze der künstlichen Intelligenz sind mit bekannten Problemen der Statistik konfrontiert, welche jedoch im Kontext der aktuell gebräuchlichen Methoden verstärkt werden. Namentlich ist bei einigen KI-Methoden nicht mehr **nachvollziehbar**, wie eine bestimmte Vorhersage oder ein bestimmtes Ergebnis zustande kommt oder wieso ein KI-System bei einer konkreten Fragestellung zu einer bestimmten Antwort gelangt ist. Überdies sind Anwendungen auf Basis von KI von der Qualität der Daten und Algorithmen abhängig. So können **systematische Fehler** in Daten oder Algorithmen (z.B. versteckte Einseitigkeiten, etwa, wenn in Daten eine Bevölkerungsgruppe über- oder unterrepräsentiert ist) angesichts der Grösse und Komplexität der verwendeten Daten oft nicht erkannt werden.

Diese Herausforderungen wurden auch von Seiten der für die Erarbeitung des Berichts konsultierten Expertinnen und Experten aus Wirtschaft und Wissenschaft als wesentlich identifiziert. Sie sind in erster Linie zwar technische Herausforderungen, können jedoch in gewissen Anwendungsbereichen auch aus gesellschaftlicher oder rechtlicher Perspektive zu problematischen Resultaten führen, etwa, wenn Personengruppen auf Basis von KI-Entscheidungen unzulässig systematisch diskriminiert werden oder wenn in sensiblen Bereichen das Ergebnis einer Analyse nicht erklärbar ist (z.B. bei Anwendung von KI in der Justiz).

Zwar können diese Probleme teilweise bereits technisch gemildert werden; dies ist jedoch auch mit Nachteilen verbunden. Aktuelle KI-Systeme sind darauf optimiert, eigenständig Zusammenhänge zu erkennen. Eine Erhöhung der Nachvollziehbarkeit geht daher auf Kosten der Leistung von KI-Systemen, was in einigen Fällen wenig wünschbar sein dürfte (z.B. in der medizinischen Diagnostik).

Die **Herausforderungen sind je nach Anwendungsgebiet sehr unterschiedlich**. So ist die fehlende Nachvollziehbarkeit bei einer Empfehlung für Musiktitel prinzipiell unproblematisch. Hingegen beschränkt die nicht oder nur begrenzt nachvollziehbare Einschätzung eines KI-Systems über die Rückfallgefahr einer tatverdächtigen oder verurteilten Person grundlegende Rechte der Betroffenen.

Der Entwurf zur Revision des Datenschutzgesetzes<sup>1</sup> trägt diesen Herausforderungen Rechnung und sieht im Hinblick auf KI-basierte, automatisierte Einzelentscheidungen verschiedene Pflichten für die verantwortliche Person oder Institution vor. So muss eine von einer automatisierten Entscheidung betroffene Person über eine solche Entscheidung informiert werden, wenn diese für sie mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt. Die betroffene Person kann zudem verlangen, dass die Entscheidung von einer natürlichen Person überprüft wird oder dass ihr die Logik mitgeteilt wird, auf der die Entscheidung beruht.

Auch die Fähigkeit von KI-Systemen, zunehmend **autonom zu handeln**, stellt den bestehenden rechtlichen Rahmen auf die Probe. Der Bundesrat hat diesbezüglich bereits mehrfach die bestehenden Regelungen überprüft und ist namentlich im Bereich der zivil- und strafrechtlichen Verantwortlichkeit und des internationalen Privatrechts zum Schluss gekommen, dass diese aufgrund des aktuellen Stands der Technologie derzeit genügen.

### Grundsätzlich geeigneter allgemeiner Rechtsrahmen

Die vorliegende Analyse zeigt, dass der bestehende generelle Rechtsrahmen in der Schweiz derzeit grundsätzlich geeignet ist, mit neuen Anwendungen und Geschäftsmodellen im Bereich KI umzugehen. Die Nutzung neuer und innovativer Technologien erfolgt nicht im rechtsfreien Raum, sondern hat umfänglich das geltende Recht zu respektieren. Die relevanten Rechtsprinzipien sind in der Regel technologieneutral formuliert, so dass sie sich auch auf KI-Systeme anwenden lassen. Der bestehen-

---

<sup>1</sup> Entwurf zur Totalrevision des Bundesgesetzes über den Datenschutz und die Änderungen zum Datenschutz vom 15. September 2017, BBL 2017, 7206 ff.



## Herausforderungen der künstlichen Intelligenz

de Rechtsrahmen erlaubt und begrenzt damit im Grundsatz den Einsatz von KI. Dies gilt insbesondere auch für Diskriminierungen, die auf Basis von KI-Entscheidungen entstehen können. Eine **verantwortungsvolle Nutzung** ist folglich durch die den Rechtsnormen zugrundeliegende Wertordnung definiert und durch deren Beachtung innerhalb der Rechtsordnung gewährleistet.

Es drängen sich somit keine grundlegenden Anpassungen des Rechtsrahmens auf. Angesichts der hohen technologischen Dynamik ist jedoch nicht auszuschliessen, dass sich diese Einschätzung rasch ändern könnte.

## Herausforderungen in Politikfeldern

Auch wenn der allgemeine Rechtsrahmen aus heutiger Sicht grundsätzlich geeignet ist, KI-Anwendungen zu erfassen, können neue Anwendungsmöglichkeiten gleichwohl die bestehende **Regulierung in jeweils spezifischen Politikfeldern** in Frage stellen.

Im Rahmen des Mandats hat die IDAG KI im Sinne einer Momentaufnahme eine Auslegeordnung über KI-bezogene Herausforderungen mit Relevanz für den Bund vorgenommen. Diesbezüglich wurden 17 aus heutiger Sicht relevante Themenbereiche identifiziert, die vordringlich zu betrachten sind.

Die Themenbereiche wurden in Verantwortung des jeweils zuständigen Amtes bearbeitet. Da sich die Herausforderungen des Einsatzes von KI je nach Themenbereich sehr unterschiedlich stellen, waren Abklärungen und Konsultationen in unterschiedlichem Ausmasse notwendig. Während einige Themenbereiche seitens der Ämter bereits ausreichend aufgenommen waren, wurden für andere Themen breit abgestützte Projektgruppen eingesetzt.

Insgesamt wurden sieben grössere, departementsübergreifende Arbeitsgruppen eingesetzt sowie zahlreiche externe Stakeholder und Expertinnen und Experten aus Wissenschaft und Wirtschaft konsultiert. Dabei waren gemäss bundesrätlichem Auftrag die Überlegungen zu einem transparenten und verantwortungsvollen Einsatz von künstlicher Intelligenz zu berücksichtigen.

Im Rahmen der Abklärungen wurden die folgenden 17 Themenbereiche untersucht:

1. Internationale Gremien und künstliche Intelligenz
2. Schweizer Interessenabklärungen an den europäischen KI-Aktivitäten (Digital Europe Programme)
3. Veränderungen der Arbeitswelt
4. Künstliche Intelligenz in Industrie und Dienstleistungen
5. Künstliche Intelligenz in der Bildung
6. Anwendung von künstlicher Intelligenz in Wissenschaft und Forschung
7. Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik
8. Künstliche Intelligenz, Medien & Öffentlichkeit
9. Automatisierte Mobilität und künstliche Intelligenz
10. Künstliche Intelligenz im Gesundheitswesen
11. Künstliche Intelligenz in der Finanzwirtschaft
12. Künstliche Intelligenz in der Landwirtschaft
13. Energie, Klima, Umwelt und künstliche Intelligenz
14. Künstliche Intelligenz in der Verwaltung
15. Weiterentwicklung des allgemeinen Rechtsrahmens im Hinblick auf künstliche Intelligenz
16. Künstliche Intelligenz in der Justiz
17. Künstliche Intelligenz, Daten und Immaterialgüterrecht

## Herausforderungen weitgehend adressiert

In den verschiedenen Politikbereichen stellen die Entwicklungen der künstlichen Intelligenz teilweise grosse Herausforderungen dar. Die Analysen zeigen, dass in vielen Bereichen nach wie vor grosser Klärungs- und Anpassungsbedarf besteht, sich die Akteure in der Schweiz aber bereits weitgehend der Thematik angenommen und reagiert haben. Dies gilt namentlich für die Bildung und Forschung wie auch für die Wirtschaft. In diesen Bereichen wurde eine Vielzahl von Massnahmen aufgegleist, welche die Herausforderungen grundsätzlich angehen. Dies gilt ebenso für die Bundesverwaltung, welche die KI themenspezifisch bereits weitgehend aufgenommen hat.

Der im Bericht neu identifizierte *zusätzliche* Handlungsbedarf bezieht sich einerseits auf zu intensivierende Abklärungen in bestimmten Bereichen, die teilweise auch mit erhöhter Dringlichkeit vorzunehmen sind. Daneben besteht ein weiterer neuer zentraler Handlungsbereich in der besseren Koordination, Vernetzung und Beobachtung angesichts der rasch voranschreitenden technologischen Entwicklungen. Dabei sollen angesichts der raschen Entwicklungen zunächst auf Basis des vorliegenden Berichts strategische Leitlinien auf Ebene des Bundes entwickelt werden. Künstliche Intelligenz ist nicht als isolierte Technologie zu betrachten, sondern als ein wesentlicher Bestandteil der fortschreitenden Digitalisierung von Wirtschaft und Gesellschaft. Die relevanten sektorspezifischen Massnahmen der Departemente und Ämter sind daher in der allgemeinen Digitalisierungsstrategie des Bundes (Strategie «Digitale Schweiz») zu berücksichtigen.

Die im Bericht betrachteten Themen stellen angesichts der raschen Dynamik der technologischen Entwicklungen nur eine **Momentaufnahme** dar. Dies betrifft nicht nur die Auswahl der Themenfelder, sondern auch den absehbaren Anpassungsbedarf der sektoriellen Regulierung.

## Bewährte Politikgrundsätze

Der Bericht bestätigt, dass das zentrale Prinzip eines technologieneutralen Rechtsetzungs- und Regulierungsansatzes sich gerade auch im rasch verändernden technologischen Umfeld von KI, dessen Entwicklung für den Gesetzgeber nur bedingt vorhersehbar ist, bewährt hat. Wo aus Anwendungssicht Regelungsbedarf besteht, soll dieser so ausgestaltet werden, dass alle Technologien gleichermassen adressiert werden.

Der Bund ist daher bestrebt, auch weiterhin eine grundsätzlich technologieneutrale Politik zu verfolgen, welche weitgehend auf eine Förderung spezifischer Technologien verzichtet und technologiespezifische Regulierungen möglichst zu vermeiden versucht. Eine solche Offenheit des Staates gegenüber neuen Technologien erlaubt das optimale Ausschöpfen des Potenzials von neuen Ideen und Innovationen.

Dieser Ansatz stellt aus Sicht des Bundesrates einen wesentlichen Erfolgsfaktor der Schweiz dar. Auf Basis der bewährten Grundsätze will sich die Schweiz daher weiterhin mittels Rechtssicherheit, effizienter Regulierung und guter Reputation als attraktiver Standort für Forschung, Entwicklung und Anwendung von neuen Technologien positionieren.

## II. Aktionsfelder

In den untersuchten Themenbereichen stellt der Bericht eine grosse Anzahl an bereits ergriffenen Massnahmen, Initiativen und aufgegleisten Abklärungen seitens des Bundes dar. Im Rahmen einer Priorisierung wird für die jeweiligen Politikbereiche aufgezeigt, welchen Aktivitäten besondere Beachtung zukommt. Der weitaus überwiegende Teil dieser Massnahmen wird innerhalb bereits bestehender Aktivitäten, klar geregelter Zuständigkeiten und etablierter Verfahren adressiert. Hinzu kommt in einigen Bereichen *zusätzlicher* Abklärungsbedarf, welcher nicht im Rahmen der bestehenden Aktivitäten und Kompetenzen adressiert werden kann.

Nachfolgend werden die Aktionsfelder des Bundes in drei Kategorisierungen kurz dargestellt:

- A) Übergeordnete Aktionsfelder, welche die Thematik der KI in der ganzen Breite adressieren und über die ganze Bundesverwaltung hinweg angegangen werden müssen;
- B) sämtliche bereichs- und themenspezifischen Aktionsfelder in summarischer Übersicht;
- C) zusätzliche Aktionsfelder, für welche etablierte Gefässe und/oder Kompetenzen noch nicht bestehen und für welche daher seitens des Bundesrates mit Kenntnisnahme des Berichts spezifische Aufträge erteilt wurden.

### A) Übergeordnete Aktionsfelder: Sicherstellung des Informations- und Wissensaustausches

Angesichts der raschen Entwicklung von KI-Technologien muss auf übergeordneter Ebene der Informations- und Wissensaustausch wie auch der Dialog mit allen Anspruchsgruppen aufgenommen und geführt werden. Übergeordnet zu den in Anhang 1 und in Kapitel 6 dargestellten themenspezifischen Aktionsfeldern stehen für den Bund in erster Linie die bessere Koordination, Vernetzung und Beobachtung angesichts der rasch voranschreitenden technologischen Entwicklungen im Mittelpunkt:

#### ▪ **Beobachtung der technologischen Entwicklung**

Der vorliegende Bericht ist eine Momentaufnahme. Die weitere Entwicklung der KI-Technologien muss sorgfältig verfolgt werden. Dabei sollen bestehende Monitoringaktivitäten des Bundesamtes für Statistik (BFS) genutzt und allenfalls weiterentwickelt werden. Angesichts der globalen Dimension kann der Dialog jedoch nicht nur auf nationaler Ebene geführt werden. Internationale Foren stellen einen wichtigen Rahmen dar, um grundlegende Fragen rund um die weitere Entwicklung und den Einsatz von KI anzugehen. Zudem soll sich die Schweiz in den relevanten internationalen Organisationen für ein entsprechendes Monitoring durch diese einsetzen.

#### ▪ **Informations- und Wissensaustausch sowie Koordination im internationalen Umfeld**

Die Weiterentwicklung von KI geschieht in einem globalisierten Umfeld. Diese Entwicklungen wie auch die Regulierung lassen sich auf nationaler Ebene nur begrenzt steuern. Zudem stellen sich Gouvernanzfragen einerseits sektorspezifisch sehr unterschiedlich, andererseits verstärken sich die Interdependenzen zwischen vormals voneinander getrennten Politikbereichen. Dies bedingt eine stärkere interdisziplinäre Vernetzung und einen Informations- und Wissensaustausch auf nationaler sowie internationaler Ebene.

Um diesen Informations- und Wissensaustausch zu politischen, gesellschaftlichen, wirtschaftlichen und anderen relevanten Aspekten von KI sicherzustellen, soll ab sofort die vom BAKOM für die Vorbereitung des UNO Weltgipfels zur Informationsgesellschaft (WSIS) ins Leben gerufene, bereits bestehende «**Plattform Tripartite**» genutzt werden. Die Plattform steht allen interessierten Organisationen und Personen offen und verfügt über einen **administrativen Ausschuss aus Vertretern der Bundesverwaltung**, welcher bei Bedarf Positionen des Bundes in internationalen Gremien koordinieren kann. Die «Plattform Tripartite» kann als interdisziplinäres nationales **Kompetenznetzwerk zu KI-Themen und entsprechenden internationalen Gremien und Prozessen** genutzt werden, welches in der Lage ist, Wissen und Erfahrungen auch horizontal zu vernetzen und auf diese Weise kohärente Positionen der Schweiz auf internationaler Ebene zu entwickeln.

#### ▪ **Dialog und Einbettung in die Strategie «Digitale Schweiz»**

Die fortschreitende Digitalisierung hat einen Transformationsprozess ausgelöst. Daraus ergeben sich oft komplexe Fragen, die diskutiert werden müssen. Dies gilt besonders für die Entwicklungen im Rahmen der künstlichen Intelligenz. Deshalb ist es nötig, alle Anspruchsgruppen zu vernetzen und die Zusammenarbeit aller Ebenen der Verwaltung mit der Wirtschaft, der Zivilgesellschaft, der Politik und der Wissenschaft zu fördern.

Künstliche Intelligenz ist nicht als isolierte Technologie zu betrachten, sondern ist ein wesentlicher Bestandteil der allgemeinen Digitalisierung von Wirtschaft und Gesellschaft. Grundsätzliche Fragen der KI-relevanten Politik, die den gesellschaftlichen Dialog erfordern, sind daher in enger Abstimmung mit anderen digitalisierungsrelevanten Themen anzugehen und in den Rahmen der Strategie «Digitale Schweiz» zu integrieren.

## B) Gesamtübersicht bereichsspezifische Aktionsfelder

Die Herausforderungen der künstlichen Intelligenz stellen sich je nach Anwendungsbereich sehr unterschiedlich. Im Zentrum der Weiterführung der vielfältigen Arbeiten des Bundes stehen daher die Aktivitäten in den jeweiligen Politikbereichen im Mittelpunkt. Die gesamte Übersicht über alle bereichsspezifischen Aktionsfelder ist in Tabelle 1 dargestellt. Eine detailliertere Kurzbeschreibung der Massnahmen findet sich in **Anhang 1**. Die ausführliche Darstellung erfolgt jeweils in den Abschnitten zu den einzelnen Politikbereichen in Kapitel 6.

Tabelle 1: Übersicht prioritäre, bereichsspezifische Aktionsfelder des Bundes

<b>Internationale Gremien und KI (BAKOM, EDA)</b>
<ul style="list-style-type: none"> <li>▪ Informations- und Wissensaustausch und Koordination der Positionen des Bundes in internationalen Gremien</li> <li>▪ Stärkung der globalen Gouvernanz im Bereich KI und Integration von KI in die aussenpolitische Strategie 2020-2023</li> <li>▪ Stärkung des Internationalen Genf als Zentrum für digitale Gouvernanz</li> </ul>
<b>Digital Europe Programme (SBFI u.a.)</b>
<ul style="list-style-type: none"> <li>▪ Prüfung einer möglichen Schweizer Beteiligung an Horizon Europe und am Digital Europe Programme</li> </ul>
<b>Veränderungen der Arbeitswelt (SECO)</b>
<ul style="list-style-type: none"> <li>▪ Weiterverfolgung der Auswirkungen der KI auf den Arbeitsmarkt</li> </ul>
<b>KI in Industrie und Dienstleistungen (SECO)</b>
<ul style="list-style-type: none"> <li>▪ Weiterverfolgung der Entwicklungen von KI in Industrie und Dienstleistungen</li> </ul>
<b>KI in der Bildung (SBFI, Kantone und andere relevante Akteure)</b>
<ul style="list-style-type: none"> <li>▪ Gewährleistung der für den Einsatz von KI erforderlichen Kompetenzen auf allen Bildungsstufen</li> <li>▪ Gewährleistung einer transparenten und verantwortungsvollen Anwendung von KI im Bildungsbereich</li> </ul>
<b>Anwendung von KI in Wissenschaft und Forschung (SBFI)</b>
<ul style="list-style-type: none"> <li>▪ Gewährleistung der erforderlichen Kompetenzen in Forschung und Innovation im Rahmen der BFI-Politik</li> </ul>
<b>KI in der Cybersicherheit und Sicherheitspolitik (EDA, armasuisse, NDB, Kompetenzzentrum Cyber, BK, Armee, BABS, WBF, VBS)</b>
<ul style="list-style-type: none"> <li>▪ Prüfung der aussensicherheitspolitischen Auswirkungen von KI</li> <li>▪ Prüfung neuer KI-Bedrohungsformen in Cybersicherheit, Propaganda und Kriegsführung</li> <li>▪ Integration von KI in bestehende Strategien, Aktionspläne und Prozesse zur Stärkung der Abwehr von gegenüber KI-Bedrohungsformen</li> <li>▪ Stärkung der Antizipation des KI-Potenzials in Cybersicherheit und Cyberabwehr durch Zusammenarbeit mit Hochschulen und Industrie sowie durch Forschung und Prüfstände</li> </ul>
<b>KI, Medien &amp; Öffentlichkeit (BAKOM, BK, EDA)</b>
<ul style="list-style-type: none"> <li>▪ Prüfung des Einflusses von Informations-Intermediären und Abklärung der erforderlichen Governance</li> <li>▪ Beobachtung der Entwicklung zum Einsatz von KI im Medienbereich</li> </ul>

<b>Automatisierte Mobilität und KI</b> (ASTRA, BAZL, BAV, swisstopo, VBS (BABS), UVEK)
<ul style="list-style-type: none"> <li>▪ Umsetzung von Massnahmen zum Einsatz von KI in automatisierten Fahrzeugen (Strassen-, Luft-, Schienenverkehr)</li> <li>▪ Umsetzung von Massnahmen und Aufbau der Infrastruktur zum Datenaustausch für KI in Automatisierter Mobilität</li> <li>▪ Erarbeitung der Grundlagen für den Datenschutz im Umfeld der Mobilität</li> <li>▪ Schaffung eines Rechtsrahmens für KI in der Automatisierten Mobilität (Bewilligungen, Zulassungsverfahren) und Förderung der gesellschaftlichen Akzeptanz (Abklärung Fehlertoleranzen)</li> </ul>
<b>KI im Gesundheitswesen</b> (BAG, Swissmedic)
<ul style="list-style-type: none"> <li>▪ Prüfung der Positionierung und gesetzlicher Grundlagen bzgl. Verwendung von Daten, Proben und Biobanken in der Humanforschung</li> <li>▪ Prüfung von zukünftigen Lösungsansätzen in der Arzneimittelentwicklung im Rahmen des Heilmittelgesetzes</li> </ul>
<b>KI in der Finanzwirtschaft</b> (EFD)
<ul style="list-style-type: none"> <li>▪ Weiterverfolgung der Nutzung von KI durch Finanzmarktakteure im Hinblick auf regulatorische Verhaltenspflichten</li> <li>▪ Weiterverfolgung der Entwicklungen aufsichtsrechtlich relevanter operationeller Risiken bei Finanzmarktinstituten</li> <li>▪ Weiterverfolgung der Entwicklungen bei Prämien unter Verwendung von KI bei staatlich geprüften Privatversicherungen</li> </ul>
<b>KI in der Landwirtschaft</b> (BLW)
<ul style="list-style-type: none"> <li>▪ Weiterverfolgung der Auswirkungen von KI in der Landwirtschaft</li> </ul>
<b>Energie, Klima, Umwelt und KI</b> (BFE, BAFU)
<ul style="list-style-type: none"> <li>▪ Weiterverfolgung der Entwicklungen von KI im Bereich Energie</li> <li>▪ Weiterverfolgung der Entwicklungen von KI in den Bereichen Umwelt und Klima</li> </ul>
<b>KI in der Verwaltung</b> (ISB, BFS, Bundesämter mit grossen Datenmengen (u.a. EZV, BFS, ESTV, BLW))
<ul style="list-style-type: none"> <li>▪ Schaffung und Bereitstellung von gemeinsamen Datengrundlagen in der Bundesverwaltung</li> <li>▪ Abklärung der Schaffung eines Kompetenznetzwerkes KI mit Fokus auf technische Aspekte der Anwendung von KI in der Bundesverwaltung</li> <li>▪ Kommunikation über KI relevante Themen in der Bundesverwaltung zum Aufzeigen von Chancen</li> <li>▪ Überprüfung gesetzliche Grundlagen für den Einsatz von KI in der Bundesverwaltung</li> </ul>
<b>Weiterentwicklung des allgemeinen Rechtsrahmens im Hinblick auf künstliche Intelligenz</b> (EDA)
<ul style="list-style-type: none"> <li>▪ Prüfung der Entstehung von KI-spezifischem Völkerrecht und der Auswirkungen auf die Schweiz</li> <li>▪ Verfolgung der Entwicklungen bzgl. der Erkennbarkeit von KI-Systemen in der Interaktion mit Konsumenten</li> </ul>
<b>Einsatz von KI in der Justiz</b> (BJ, EDA)
<ul style="list-style-type: none"> <li>▪ Beobachtung der Entwicklungen bei KI-basierter Entscheid-Unterstützung in Verwaltung und Justiz</li> </ul>
<b>KI, Daten und Immaterialgüterrecht</b> (BAKOM, BJ, IGE, GS EDI/BFS)
<ul style="list-style-type: none"> <li>▪ Weiterführung bestehender Arbeiten zur Datenpolitik des Bundes (u.a. Open Government Data-Strategie)</li> <li>▪ Weiterführung bestehender Arbeiten zum Datenschutz (u.a. Revision Datenschutzgesetz, Prüfauftrag des Bundesrates hinsichtl. Übertragbarkeit, Empfehlungen der Expertengruppe Datenverarbeitung und -sicherheit)</li> <li>▪ Weiterführung bestehender Arbeiten zum geistigen Eigentum (Gewährleistung der Unterstützung der Entwicklungen im Bereich KI)</li> </ul>

### C) Zusätzliche Aktionsfelder

Die Herausforderungen der künstlichen Intelligenz sind weitgehend erkannt und in den verschiedenen Politikbereichen aufgenommen. Dabei besteht in nahezu allen Bereichen angesichts der hohen technologischen Dynamik der Bedarf, die Entwicklungen intensiv zu verfolgen sowie weitere Abklärungen und allenfalls Optimierungen im Rahmen der normalen Tätigkeiten und etablierten Verfahren vorzunehmen. Neben der Vielzahl an bereits aufgenommenen Arbeiten (vgl. Anhang 1) identifiziert der Bericht auch neu den folgenden *zusätzlichen* Abklärungsbedarf, welcher nicht im Rahmen der bestehenden Aktivitäten und Kompetenzen adressiert werden kann:

#### ▪ **Weiterentwicklung des Völkerrechts angesichts der künstlichen Intelligenz**

Internationale Regulierungsbestrebungen im Bereich KI werden aktuell nicht nur von Staaten, sondern von einer Vielzahl von Akteuren, bspw. grossen Technologieunternehmen oder internationalen Organisationen, vorangetrieben. Sie sind bestrebt, sowohl in Form von Industriestandards als auch von ethischen Prinzipien, KI-spezifische Regeln zu schaffen. Wie diese Regeln entstehen, wie sie zu qualifizieren sind, inwiefern dadurch Völkerrecht geschaffen wird und welche Auswirkungen diese auf die Schweiz haben könnten, ist vertieft zu prüfen.

- Der Bericht empfiehlt, dass das EDA (DV) unter Einbezug der betroffenen Departemente dem Bundesrat bis Ende 2020 einen Bericht vorlegt, der vertieft prüft, wie internationale Regeln im KI-Bereich entstehen, wie sie zu qualifizieren sind und inwiefern dadurch Völkerrecht geschaffen wird, und der ggf. Massnahmen vorschlägt, wie sich die Schweiz dazu positionieren soll.

#### ▪ **Künstliche Intelligenz, Medien & Öffentlichkeit**

Intermediäre besitzen das Potenzial, KI-Anwendungen für kommerzielle oder politische Zwecke zu instrumentalisieren oder selbst zu diesen Zwecken instrumentalisiert zu werden. Dadurch kann die öffentliche Meinungs- und Willensbildung, auch im politischen Bereich, beeinflusst werden. Die Thematik ist vertieft zu prüfen und die Möglichkeit, eines schweizerischen Governance-Ansatzes abzuklären.

- Der Bericht empfiehlt, dass das UVEK (BAKOM) in Zusammenarbeit mit der BK dem Bundesrat bis im Frühling 2021 einen Governance-Bericht vorlegt, der spezifische Massnahmen zu dieser Thematik prüft und gegebenenfalls zur Umsetzung vorschlägt.

#### ▪ **Künstliche Intelligenz in der Bundesverwaltung**

Die verwaltungsübergreifende Erfassung von Prozessen sowie ein übergreifender Datenzugriff sind eine der Voraussetzungen, um das Potenzial der KI in der Bundesverwaltung auszuschöpfen. Der überdepartementale Aufbau und Austausch von Wissen und Erfahrung sind für eine wirtschaftliche und koordinierte Entwicklung von KI-Lösungen in der Bundesverwaltung wesentlich. Fragmentierte Lösungen sind dagegen nicht zielführend. Eine gemeinsame Anlaufstelle bzw. ein Kompetenznetzwerk mit speziellem Fokus auf technische Aspekte für die konkrete Anwendung von KI in der Bundesverwaltung könnte eine Lösung darstellen. Dabei sollten die erwähnte Anlaufstelle bzw. das Netzwerk insbesondere beratende Funktionen wahrnehmen.

- Der Bericht empfiehlt, dass das EFD (ISB) in Zusammenarbeit mit dem EDI (BFS) und unter Einbezug der übrigen Departemente und der BK den Mehrwert (u.a. Bedürfnisanalyse) und die Machbarkeit einer gemeinsamen Anlaufstelle bzw. eines Kompetenznetzwerkes mit speziellem Fokus auf technische Aspekte der Anwendung von KI in der Bundesverwaltung prüft. Diese Stelle soll für die Anwendung von KI in der Bundesverwaltung beratende Funktion haben. Mit dem Ziel einer gesamtheitlichen Vernetzung und Technologiebetrachtung sollen auch andere Technologien der digitalen Transformation (bspw. Blockchain, IoT, Big-Data etc.) und die Herausforderungen von deren Anwendung berücksichtigt werden.

### ▪ **Strategische Leitlinien für die KI-relevante Politik**

Angesichts der raschen Entwicklungen und der breiten Diskussionen zum Einsatz von künstlicher Intelligenz, bedarf es auf Ebene des Bundes strategischer Leitlinien. Diese sollen aus dem vorliegenden Bericht abgeleitet werden. Dabei ist künstliche Intelligenz nicht als isolierte Technologie zu betrachten, sondern als ein wesentlicher Bestandteil der fortschreitenden Digitalisierung von Wirtschaft und Gesellschaft.

- ➔ Der Bericht empfiehlt angesichts der von der IDAG KI dargestellten Herausforderungen, dass das WBF (SBFI) in Zusammenarbeit mit dem UVEK (BAKOM) strategische KI-Leitlinien erarbeitet und diese dem Bundesrat bis im Frühjahr 2020 unterbreitet.
- ➔ Der Bericht empfiehlt, dass das UVEK (BAKOM) die KI-Politik als wesentlichen Bestandteil der Strategie «Digitale Schweiz» berücksichtigt. Dazu sollen relevante sektorspezifische Massnahmen der Departemente und Ämter im Rahmen der Strategie «Digitale Schweiz» berücksichtigt werden.





# 1 Auftrag / Ausgangslage

Eine starke Zunahme der Rechenkraft von Computern, die exponentiell wachsende Verfügbarkeit von Daten und die Entwicklung neuer KI-Verfahren des maschinellen Lernens haben in den vergangenen Jahren zu einer Renaissance von Anwendungen der künstlichen Intelligenz geführt. Dabei haben einige beachtliche Entwicklungen deutlich gemacht, dass die zunehmende Nutzung künstlicher Intelligenz das Potenzial hat, Wirtschaft und Gesellschaft stark zu verändern. Bekannte Beispiele für erfolgreiche Anwendungen sind automatisch generierte Sprachübersetzungen, Maschinen, welche mit Menschen Gespräche führen können, autonom verkehrende Fahrzeuge oder Bilderkennungssoftware, welche mittlerweile dem Menschen teilweise überlegen ist. Noch wichtiger als diese Entwicklungen aus Sicht des Endanwenders ist die Bedeutung von künstlicher Intelligenz als Basistechnologie. Dabei handelt es sich um Querschnittstechnologien, die das Potenzial haben, alle Branchen zu durchdringen und eine hohe Produktivitätswirkung auf eine Vielzahl verschiedener Wirtschaftsbereiche auszuüben.

Angesichts der Geschwindigkeit der Entwicklung neuer technologischer Anwendungsmöglichkeiten der künstlichen Intelligenz hat der Bundesrat im Zuge der Aktualisierung der Strategie «Digitale Schweiz» am 5. September 2018 die Chancen und Herausforderungen der künstlichen Intelligenz als eines der zentralen neuen Themen seiner Strategie festgelegt. Auf dieser Grundlage hat er das WBF (SBFI) damit beauftragt, bis Ende 2018 in Zusammenarbeit mit den übrigen Departementen eine Arbeitsgruppe zum Thema künstliche Intelligenz zusammenzustellen. Diese Arbeitsgruppe bekam in erster Linie den Auftrag, den Wissens- und Meinungsaustausch sicherzustellen und die Abstimmung von Positionen der Schweiz im Kontext dieses Themas in internationalen Gremien zu ermöglichen.

Überdies wurde die Arbeitsgruppe damit beauftragt, dem Bundesrat bis Herbst 2019 einen Bericht zu künstlicher Intelligenz zu unterbreiten. Dieser soll unter anderem eine Übersicht über bestehende Massnahmen mit Bezug zu KI aufweisen sowie auf Basis bestehender wissenschaftlicher Erkenntnisse eine Einschätzung zu allfälligen neuen Handlungsfeldern vornehmen. Dabei sollen auch Überlegungen zu einem transparenten und verantwortungsvollen Einsatz von künstlicher Intelligenz gemacht werden.

Mit seiner Entscheidung vom September 2018 hielt der Bundesrat fest, dass auch hinsichtlich der Herausforderungen von künstlicher Intelligenz die Schaffung guter Rahmenbedingungen im Mittelpunkt steht, damit der Einsatz von KI ökonomischen und gesellschaftlichen Mehrwert ermöglichen, zu einer Verbesserung unseres Lebensalltags sowie zu einer Stärkung der Wettbewerbsfähigkeit beitragen kann.

Der Bundesrat bekräftigte darüber hinaus, dass die Schweiz sich auch im Hinblick auf die Entwicklungen der künstlichen Intelligenz im nationalen und internationalen Rahmen für ein Monitoring und eine Evaluation der sich aus dem Einsatz von KI ergebenden Folgen für unser Privat- und Arbeitsleben einsetzt. Zudem sind die Rahmenbedingungen für eine verantwortungsvolle Nutzung so zu gestalten, dass algorithmische Entscheidungssysteme transparent und überprüfbar sind, die Verantwortlichkeiten geregelt sind und die eingesetzten Systeme unsere Werte und Gesetze respektieren.

## Struktur des Berichts

Im vorliegenden Bericht nimmt Kapitel 2 zunächst eine Abgrenzung und Einordnung des Begriffs künstliche Intelligenz vor und stellt die grosse Bedeutung für die Wirtschaft dar (KI als Basistechnologie). Daran anschliessend wird in Kapitel 3 aufgezeigt, welche neuartigen Möglichkeiten mit den Technologien der künstlichen Intelligenz möglich sind. Diese Übersicht zeigt das enorme Potenzial von KI auf. Es werden aber auch die spezifischen Herausforderungen dieser Technologien aufgezeigt. Das Verständnis für die Charakteristika und spezifischen Probleme von KI ist essentiell, um verschiedene gesellschaftliche und rechtliche Probleme, welche oft mit KI in Verbindung gebracht werden, zu verstehen und einzuordnen. Die Einordnung der technischen Herausforderungen aus rechtlicher Sicht und die diesbezüglichen Herausforderungen werden in Kapitel 4 adressiert.

Kapitel 5 zeigt auf, wie die Schweiz hinsichtlich Forschung, Entwicklung und Anwendung von KI im internationalen Vergleich dasteht.

In Kapitel 6 werden die Ergebnisse der thematischen Arbeiten zu den spezifischen Herausforderungen von KI in verschiedenen Politikbereichen des Bundes dargestellt. Es werden jene 17 Themenbereiche dargestellt, welche die IDAG als prioritär identifiziert hat. Dabei wird jeweils zunächst die Relevanz der Anwendung von KI im jeweiligen Themengebiet konkretisiert. In einem zweiten Abschnitt werden die spezifischen Herausforderungen bzgl. KI aufgezeigt, schwerpunktmässig im Kompetenzbereich des Bundes. Es folgt eine Darstellung der laufenden Aktivitäten, die der Bund oder relevante bundesexterne Akteure bereits initiiert haben, um die Herausforderungen zu adressieren. Schliesslich erfolgt eine Bewertung, ob die identifizierten Herausforderungen im jeweiligen Themenfeld durch die laufenden Aktivitäten (oder im Rahmen bestehender Regelungen) bereits ausreichend aufgenommen sind. Dort, wo dies nicht der Fall ist, wird ausgewiesen, welche weiteren Schritte allenfalls erforderlich sind, um die Herausforderungen rechtzeitig anzugehen.

## 2 Einordnung des Begriffs «Künstliche Intelligenz» (KI)

Während wesentliche mathematische Grundlagen künstlicher Intelligenz bereits vor Jahrzehnten entwickelt wurden, haben verschiedene parallele technologische Entwicklungen zu einem Wiederaufblühen des Interesses an künstlicher Intelligenz geführt.

Zum einen wurden im Zuge der breiten Verwendung des Internets und der zunehmenden Nutzung von Sensoren immense **Datenmengen** zugänglich, welche in vielen Anwendungsbereichen meist automatisch und kostenlos anfallen und die als Trainingsdaten für die im Rahmen von KI benutzten Methoden bereitstehen. Die Menge an Trainingsdaten ist mitentscheidend für die Qualität der KI-Anwendungen. Gleichzeitig haben sich in den letzten 10 bis 15 Jahren die Technologien für Datenspeicherung und **Rechenkraft** rasant entwickelt, so dass auch die Verarbeitung enorm grosser Datenmengen erschwinglich geworden ist. Beflügelt von diesen Möglichkeiten haben Weiterentwicklungen neuer KI-Methoden, welche in der Lage sind, eigenständig aus Daten zu lernen (maschinelles Lernen) die sinnvolle und marktfähige Nutzung der Daten und Methoden ermöglicht.

Aus der Kombination der Zugänglichkeit enorm grosser Datenmengen einerseits, und der Fähigkeit von KI-Systemen, funktionale Zusammenhänge selbstständig zu erlernen und zu modellieren andererseits, ergeben sich gänzlich neue Anwendungsmöglichkeiten, die mit herkömmlichen Methoden nicht möglich sind. Solche Entwicklungen haben bereits einige erfolgreiche Anwendungen von künstlicher Intelligenz hervorgebracht, welche in den vergangenen Jahren in verschiedenen Bereichen erfolgreich eingesetzt wurden, so dass KI längst im Alltag vieler Menschen angekommen ist.

Da ein Ende der technologischen Entwicklungen – gerade auch in der Sensorik – derzeit nicht absehbar ist, deuteten sich viele weitere Anwendungsbereiche von KI bereits an. So konnte ein Forscherteam bspw. kürzlich zeigen, dass mithilfe eines KI-Systems Signale, welche das Gehirn an die Gesichtsmuskulatur sendet, dekodiert und auf dieser Basis ganze gesprochene Sätze basierend auf der Gehirnaktivität einer Person erzeugt werden können.<sup>2</sup> In vielen weiteren Bereichen stehen Anwendungen an der Schwelle zum produktiven Einsatz. Dazu zählen etwa Anwendungen in der Medizin, der Pharmaforschung, der Cybersicherheit oder im Finanzwesen.

---

<sup>2</sup> Gopala K. Anumanchipalli, Josh Chartier & Edward F. Chang (2019): "Speech synthesis from neural decoding of spoken sentences"; Nature, Volume 568, pages 493–498. <https://techcrunch.com/2019/04/24/scientists-pull-speech-directly-from-the-brain/>

## 2.1 Ansätze zur Definition künstlicher Intelligenz

Eine allgemein gültige und akzeptierte Definition von künstlicher Intelligenz existiert nicht.<sup>3</sup> Versuche, KI umfassend zu definieren, orientieren sich üblicherweise an der menschlichen Intelligenz. Dabei wird KI etwa als «intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals» verstanden.

Bei einer solchen Definition stellt sich jedoch das Problem, dass (menschliche) Intelligenz zunächst selbst definiert werden müsste. Was jedoch Intelligenz genau ausmacht, ist keineswegs klar. Angegangen wird dieses Problem teilweise dadurch, dass von Intelligenz bereits gesprochen wird, wenn sie als solche erkennbar oder als von einer menschlichen Intelligenz nicht unterscheidbar betrachtet wird. Demgemäss liegt KI beispielsweise vor, wenn «Maschinen Dinge tun, für deren Ausführung man beim Menschen Intelligenz unterstellt»<sup>4</sup>, oder die Handlung eines Systems von Menschen als «intelligent» betrachtet wird.<sup>5</sup>

Differenziertere Definitionen grenzen Intelligenz auf spezifische kognitive Funktionen (z.B. Wahrnehmen, Lernen) ein, oder ergänzen weitere Aspekte, wie z.B. die Fähigkeit zum Lösen komplexer Aufgaben.<sup>6</sup> Bei all diesen Aspekten sind jedoch weitere, nicht-triviale Definitionen notwendig, die klären, was unter diesen, jeweils vorausgesetzten Begriffen verstanden wird.

Es besteht dagegen heute Einigkeit, dass KI nach derzeitigem Stand der Technik bei weitem nicht vergleichbar ist mit der menschlichen Intelligenz. Auch wenn KI-Systeme heute komplexe Aufgaben erledigen können, die bis vor kurzem noch Menschen vorbehalten waren, werden diese jedoch für spezifische Anwendungsgebiete und Fragestellungen entwickelt, und weisen oft deutliche Grenzen auf, wenn sie mit Problemen konfrontiert sind, die mit allgemeinem Menschverstand leicht zu lösen sind.

Unbestritten ist überdies, dass die aktuell prägenden und aufmerksamkeitsregenden Entwicklungen von KI primär einem Teilbereich der Informatik entspringen, namentlich Methoden, bei welchen Computer selbstständig lernen (maschinelles Lernen).<sup>7</sup>

Der vorliegende Bericht untersucht, welcher Anpassungsbedarf sich aus den Entwicklungen der künstlichen Intelligenz in verschiedenen Politikbereichen im Hinblick auf Regulierung und rechtliche Rahmenbedingungen ergibt. Dafür sind Definitionen, welche sich an Attributen menschlicher Intelligenz orientieren, ungeeignet, wie in Abschnitt 3.3 vertieft aufgezeigt wird. Problematisch ist insbesondere, dass mit einer solchen Herangehensweise die Grenzen der künstlichen Intelligenz verwischt werden, was zu irreführender Einschätzung der Technologie führen kann. Zwar können KI-Systeme viele Attribute menschlicher Intelligenz nachahmen (z.B. Lernen, Wahrnehmen), aber solche Fähigkeiten sind nicht mit den menschlichen Pendanten vergleichbar.

<sup>3</sup> Für eine Übersicht über verschiedene Definitionen, siehe OECD (2019): «Artificial Intelligence in Society». <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>

<sup>4</sup> J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

<sup>5</sup> Gemäss dem klassischen Turing-Test etwa ist nicht die Frage relevant, ob Maschinen denken, sondern es ist auf die Fähigkeit einer Maschine abzustellen, ein intelligentes Verhalten zu zeigen, welches dem eines Menschen entspricht oder von diesem nicht unterscheidbar ist. Siehe A. M. Turing (1950). «Computing Machinery and Intelligence» *Mind* 49: 433-460. Zwar haben maschinelle Systeme diesen Test noch nicht vollständig bestanden, aktuelle Anwendungen dürften aber nahe dran sein.

<sup>6</sup> Z.B. "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation." Kaplan Andreas; Michael Haenlein (2018): "Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence", *Business Horizons*, 62(1).

<sup>7</sup> Siehe Matt Taddy (2018): «The Technological Elements of Artificial Intelligence», Chapter in forthcoming NBER book *The Economics of Artificial Intelligence: An Agenda*, Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb, editors; Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016): "Deep Learning", MIT Press, <http://www.deeplearningbook.org>; Agrawal, Gans und Goldberg (2018): «Prediction Judgment, and Complexity»; Yann LeCun, Yoshua Bengio & Geoffrey Hinton (2015): "Deep Learning", *Nature* Vol. 521:436-44, May 2015.

Vor diesem Hintergrund ist für die Erörterung des möglichen Handlungsbedarfs für den Bund, nicht auf die Technologie an sich abzustellen, sondern auf die Implikationen der heute oder in absehbarer Zeit möglichen Anwendungen.

Der vorliegende Bericht geht daher anders vor und verwendet in Kapitel 3 statt einer Definition eine Eingrenzung von künstlicher Intelligenz im Hinblick auf politikrelevante Entwicklungen anhand von strukturellen Elementen, welche KI-Anwendungen heute charakterisieren und die je nach Anwendungsbereich weniger oder stärker zum Tragen kommen. Demnach sind KI-Systeme in der Lage, (1) Daten in Komplexität und Menge in einer Form auszuwerten, die mit anderen Technologien nach heutigem Stand nicht möglich wäre, (2) Vorhersagen als wesentliche Grundlage für (automatisierte) Entscheidungen zu erstellen, (3) dadurch Fähigkeiten nachzubilden, die mit menschlicher Kognition und Intelligenz in Verbindung gebracht werden und (4) auf dieser Basis weitgehend autonom zu agieren.

Bei dieser Einordnung von KI handelt es sich jedoch um eine Momentaufnahme der aktuellen Entwicklungen, welche die derzeitigen Diskussionen prägen. Somit ist nicht auszuschließen, dass sich diese Einschätzungen im Zuge zukünftiger technologischer Entwicklungen und Anwendungsmöglichkeiten neu beurteilt werden müssen.

## 2.2 KI-Systeme versus Methoden des maschinellen Lernens

Der Begriff künstliche Intelligenz ist klar zu unterscheiden von den *Methoden* der künstlichen Intelligenz, namentlich Algorithmen des **maschinellen Lernens (ML)**. Während der Begriff «künstliche Intelligenz» oft in öffentlichen Debatten verwendet wird, verzichten viele Experten gänzlich auf dessen Verwendung und reduzieren KI auf maschinelles Lernen. Die derzeitigen bahnbrechenden Anwendungen von KI basieren primär auf diesem Teilbereich der KI-Forschung, bei welchem Computer in der Lage sind, selbstständig zu lernen.<sup>8</sup>

ML beschreibt Technologien, um (technische) Systeme zu entwickeln, welche in neuen Situationen Vorhersagen/Prognosen (engl. «Predictions») erstellen, indem sie aus vergangenen Erfahrungen lernen.<sup>9</sup> Dabei versuchen auch die heute mehrheitlich genutzten Methoden grundsätzlich eine zu erklärende Variable mit einer Reihe von erklärenden Faktoren zu prognostizieren, resp. vorherzusagen. Somit handelt es sich bei ML im Kern um statistische und algorithmische<sup>10</sup> Methoden, welche eine enge Verwandtschaft zu etablierten und bekannteren statistischen Methoden aufweisen, die ähnliche Funktionen erfüllen (z.B. Regressionsanalysen). Anhang 2 stellt in sehr einfacher Weise dar, wie moderne ML-Algorithmen im Grundsatz funktionieren.

Im Gegensatz zu früheren, regelbasierten KI-Ansätzen versuchen die heutigen, statistischen Verfahren nicht mehr menschliche Regeln abzubilden. Vielmehr werden heute Systeme derart trainiert, dass sie mit einem «trial-and-error»-Ansatz immer wieder Feedback auf die Resultate erhalten und dadurch die Schätzungen so lange anpassen, bis sie die passendsten Zusammenhänge, welche die zu erklärenden Muster möglichst gut erkennen, herausfinden. Allgemein können jedoch KI-Systeme über Lernverfahren hinaus auch andere, z.B. algorithmische, statistische und regelungstechnische Elemente enthalten.

Im Unterschied zu konventioneller Software, die immer nach dem gleichen Schema vorgeht, werden ML-Entscheidungen durch Optimierung getroffen. Maschinelles Lernen führt die Optimierung im All-

---

<sup>8</sup> Auch wenn die aktuellen Entwicklungen wesentlich von ML getrieben sind, umfasst die KI als Teilbereich der Informatik eine Vielzahl weitere Ansätze.

<sup>9</sup> «Artificial intelligence is a set of techniques for developing systems that make predictions in new situations by learning from past experience.» OECD (2019): «Artificial Intelligence in Society».

Diese Definition erklärt auch aus einer Anwendungssicht die jüngsten Fortschritte im Bereich KI zutreffend (Vgl. Agrawal, Gans und Goldberg (2018): «Prediction Judgment, and Complexity»).

<sup>10</sup> Ein Algorithmus ist eine eindeutige Handlungsanweisung zur Lösung eines Problems, welche aus endlich vielen, eindeutig definierten Einzelschritten besteht. Damit können sie zur Ausführung in ein Computerprogramm implementiert werden.

gemeinen «offline» durch. Dies bedeutet, dass ein Modell die beste Prognose für einen bekannten Datensatz voraussagt unter der Annahme, dass sich dies nach dem Training verallgemeinern und auf neue Situationen anwenden lässt. Bei vielen Problemstellungen (wie z.B. bei der Planung) gibt es jedoch zu viele solcher Situationen und es ist nicht möglich, ein Modell vorzuberechnen. In diesem Fall erfolgt die Optimierung während der Anwendung, indem verschiedene Optionen jeweils neu generiert werden und die beste ausgewählt wird (z.B. weitere Produktvorschläge, wenn ein Kunde ein Produkt betrachtet).

Zwar stellen **ML-Algorithmen** die zentrale Technologie hinter der erfolgreichen Entwicklung der gegenwärtigen KI-Systeme dar, sie sollten jedoch als Bausteine in einem grösseren Kontext betrachtet werden. ML-Algorithmen sind enorm mächtige Methoden, deren Funktion allerdings darauf beschränkt ist, Muster in Daten zu erkennen und einfache Vorhersagen zu machen. Im Gegensatz dazu ist ein **KI-System** in der Lage, **komplexe Probleme zu lösen**, die bisher dem Menschen vorbehalten waren. Dies geschieht, indem Probleme in eine Reihe einfacher Vorhersageaufgaben unterteilt werden, von denen jede für sich genommen von einem 'einfachen' ML-Algorithmus bearbeitet werden kann.

Komplexere KI-Systeme werden heute spezifisch für einen bestimmten Kontext entwickelt. Dabei braucht es neben den ML-Methoden auch spezifisches Wissen, um die Struktur zu definieren, welche ein komplexes Problem in bearbeitbare Aufgaben unterteilt und welche das Zusammenspiel der einzelnen ML-Komponenten untereinander regelt. Hinzu kommt, dass Verfügbarkeit und Verwendung geeigneter Daten geregelt werden müssen und das System in der Lage sein muss, sich auch auf Basis des Feedbacks unter realen Bedingungen anzupassen.

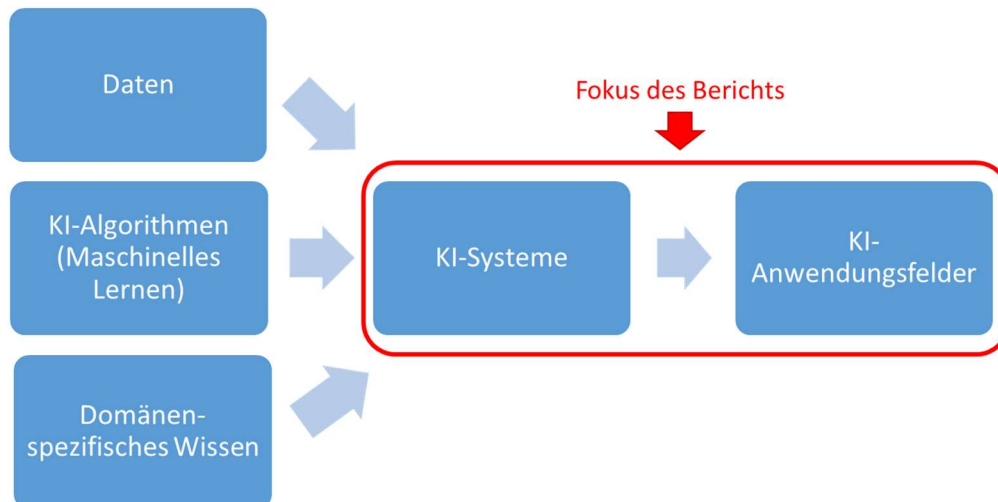
Künstliche Intelligenz kann demnach als **ein vollständig definiertes «Intelligentes System» betrachtet werden**, welches neben den verwendeten Methoden eine Reihe weiterer Elemente umfasst, damit die ML-Methoden in einem spezifischen Kontext sinnvoll eingesetzt werden können.<sup>11</sup>

Der vorliegende Bericht behandelt in erster Linie die Konsequenzen der Anwendung von KI-Systemen in konkreten Anwendungsfeldern (Abbildung 1).

---

<sup>11</sup> Matt Taddy (2018): "The Technological Elements of Artificial Intelligence".

Abbildung 1: Maschinelles Lernen als Bestandteil von KI-Systemen



Quelle: SBFI.

Anders als ML-Algorithmen, welche zunehmend zu einer universellen Technologie geworden sind, wird das Wissen, das notwendig ist, um ML-Komponenten in komplexen Anwendungen zu einer vollständigen Lösung zu kombinieren, auf absehbare Zeit nicht automatisiert und bedingt nach wie vor einen erheblichen menschlichen Input. Domänenspezifisches Wissen braucht es unter anderem auch deshalb, weil es für praktische Probleme in der Regel zu wenige Daten gibt, um ein richtiges Modell zu lernen und die Fehlerraten zu hoch wären.

### 2.3 Künstliche Intelligenz als Basistechnologie

Die ökonomische Literatur betont die hohe Bedeutung von Basistechnologien für das Wirtschaftswachstum, die Produktivität und Beschäftigungsentwicklung. Basistechnologien («General Purpose Technologies») sind Querschnittstechnologien, die eine sehr hohe Produktivitätswirkung auf eine Vielzahl verschiedener Wirtschaftsbereiche aufweisen. Dazu zählen beispielsweise die Elektrizität oder das Internet. Basistechnologien weisen vier Merkmale auf:<sup>12,13</sup>

1. Sie lassen sich in einer Vielzahl von Anwendungsbereichen produktiv nutzen;
2. Preise und Leistungsmerkmale dieser Technologie verändern sich im Zeitablauf stark;
3. Sie ermöglichen Folgeinnovationen für zahlreiche Produkte, Prozesse und Geschäftsmodelle;
4. Es gibt grosse Wechselbeziehungen mit anderen, komplementären Technologien und Folgeentwicklungen.

Für all diese Merkmale gibt es, wie in Kapitel 3 beschrieben wird, Hinweise dafür, dass dies bei KI-Technologien zutrifft. Insbesondere angesichts der vielfältigen Möglichkeiten, wie KI menschliche kognitive und perzeptive Fähigkeiten ergänzen oder substituieren kann, stellen KI-Technologien potenziell Basistechnologien dar. Diese zentrale Bedeutung von KI wird von Experten aus Wissenschaft und Wirtschaft bestätigt, wie in Kapitel 6 (Abschnitte 6.4 und 6.6) gezeigt wird.

<sup>12</sup> Vgl. EFI-Gutachten, 2014 und Jovanovic/Rousseau (2005): «General Purpose Technologies, Handbook of Economic Growth», in: Aghion/Durlauf (Hrsg.), Handbook of Economic Growth, S. 1181–1224, Elsevier.

Iain M. Cockburn, Rebecca Henderson, Scott Stern (2018): «The Impact of Artificial Intelligence on Innovation; An Exploratory Analysis».

Manuel Trajtenberg (2018): «AI as the Next GPT; A Political-Economy Perspective».

<sup>13</sup> Bundesrat (2017): «Bericht über die zentralen Rahmenbedingungen für die digitale Wirtschaft», verfügbar unter <https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirtschaftspolitik/digitalisierung.html>; SBFI (2017): «Herausforderungen der Digitalisierung für Bildung und Forschung in der Schweiz», verfügbar unter <https://www.sbf.admin.ch/sbfi/de/home/das-sbfi/digitalisierung.html>

Entwicklung und Verbreitung von KI in der Wirtschaft, wie auch die Auswirkungen auf Arbeit und Wachstum finden allerdings nicht losgelöst von den allgemeinen Entwicklungen der Digitalisierung statt, sondern sind darin eingebettet.<sup>14</sup> Wichtige Treiber der Digitalisierung sind neben den Fortschritten im Bereich der Software (bzw. KI) die Entwicklungen in der Rechenkraft, der Robotik oder der Sensorik. Ausserdem fördern Entwicklungen in der Prozess- und Speichertechnik oder auch die zunehmende Vernetzung der Information den digitalen Wandel. Diese Entwicklungen stehen in einer engen Interaktion und lassen sich nur schwer voneinander trennen. Die Technologien entwickeln sich nach wie vor rasant. Das Potenzial für weitere grundlegende Veränderungen ist daher nach wie vor beträchtlich.

### 3 Charakteristika / Strukturelemente von KI-Systemen

Grundsätzlich stellen KI-Technologien Instrumente dar, deren Einsatz – wie jener anderen Technologien auch – in der Verantwortung des Nutzers liegt. Gesellschaftliche Herausforderungen können sich jedoch aus den mit diesen Technologien möglichen Anwendungen ergeben. Aus der Optik des Bundes muss sich daher eine Beurteilung eines möglichen Handlungsbedarfs an den Implikationen der heute oder in absehbarer Zeit möglichen Anwendungen orientieren, nicht an den KI-Technologien an sich.

Während die einzelnen Technologien und Methoden von KI noch einigermaßen plausibel eingegrenzt werden können (siehe Kapitel 2), lässt sich KI aus **Anwendungssicht** noch viel weniger exakt von anderen Technologien und Anwendungen abgrenzen. So können bereits einfache Regelsysteme eine Art intelligentes Verhalten aufweisen (z.B. Thermostaten). Etliche Anwendungen können auch auf Basis anderer statistischer Methoden implementiert werden (z.B. Vergabe von Krediten oder Versicherungen). Zudem verändert sich auch die Betrachtung dessen, was als intelligentes Verhalten angesehen wird: Mit zunehmender Leistungsfähigkeit der Maschinen werden Aufgaben, die als intelligent gelten, oft aus der Definition von KI ausgenommen («KI-Effekt»); so werden heute bspw. Schachcomputer in der Regel nicht mehr zu Anwendungen künstlicher Intelligenz gezählt.

Der vorliegende Bericht stützt sich daher nicht primär auf eine Definition ab, sondern charakterisiert Anwendungen künstlicher Intelligenz durch verschiedene strukturelle Elemente, die aufzeigen, was an solchen Anwendungen neuartig ist und wo es daher möglicherweise regulatorischer Anpassungen bedarf (siehe Kasten «Übersicht Strukturelemente»). Dieser Ansatz erlaubt es auch, im Hinblick auf die konkreten Anwendungen die für KI relevanten politischen Handlungsfelder von anderen Politikbereichen (z.B. Datenpolitik/Datenschutz, sektorspezifische Regulierungen) abzugrenzen. Diese Strukturelemente werden nachfolgend erläutert.<sup>15</sup>

#### Übersicht Strukturelemente

Der vorliegende Bericht identifiziert vier zentrale, mit heutigen KI-Systemen verbundene Strukturelemente (vgl. Abbildung 2, blau hinterlegte Felder). Demnach sind KI-Systeme in der Lage,

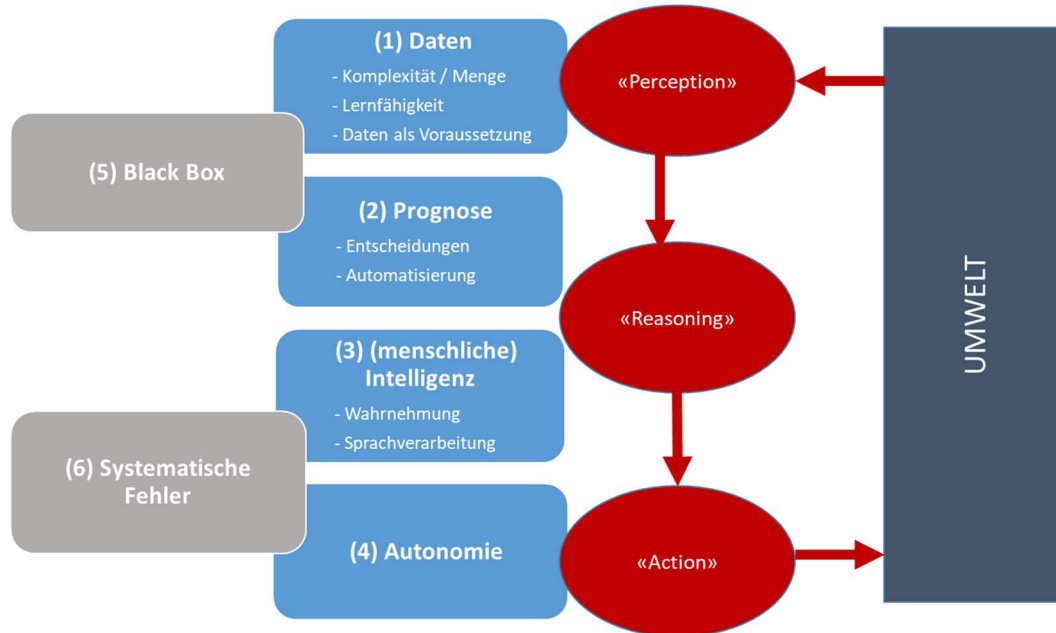
- (1) Daten in Komplexität und Menge in einer Form auszuwerten, die mit anderen Technologien nach heutigem Stand nicht möglich wäre, insbesondere, wenn Algorithmen selbstständig Lernen und dabei in Daten relevante Muster erkennen;
- (2) Vorhersagen als wesentliche Grundlage für (automatisierte) Entscheidungen zu erstellen;

<sup>14</sup> Bericht in Erfüllung der Postulate 15.3854 Reynard «Automatisierung. Chancen und Risiken» vom 16. September 2015 und 17.3222 Derder «Digitale Wirtschaft. Die Arbeitsplätze der Zukunft und Massnahmen für ihre Förderung in der Schweiz identifizieren» vom 17. März 2017, verfügbar unter <https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-68708.html>

<sup>15</sup> Dieser Ansatz stellt eine Synthese aus verschiedenen wissenschaftlichen Quellen dar. Die Strukturierung und Beschreibung wurde zudem mit von swissuniversities vorgeschlagenen KI-Experten diskutiert und konsolidiert, und ist somit auf wissenschaftliche Expertisierung abgestützt.

- (3) dadurch Fähigkeiten nachzubilden, die mit menschlicher Kognition und Intelligenz assoziiert werden;
- (4) auf dieser Basis weitgehend autonom zu agieren.

Abbildung 2: Strukturelemente zur Charakterisierung von KI-Anwendungen



Quelle: SBFI.

Die grosse Stärke aktueller KI-Systeme, enorm grosse Datenmengen auswerten zu können und darin selbstständig Muster und Zusammenhänge erlernen zu können, geht allerdings auch mit technischen Herausforderungen einher; diese Herausforderungen sind ebenso als Strukturelemente aktueller KI-Anwendungen zu werten (grau hinterlegte Felder in Abbildung 2). Nachteilig wirkt sich insbesondere aus, dass

- (5) oft nicht mehr nachvollziehbar ist, wie eine bestimmte Vorhersage oder ein bestimmtes Ergebnis zustande gekommen ist (Black-Box-Problematik);
- (6) fehlerhafte Zusammenhänge in den Daten vielfach nicht mehr erkannt werden können und daher perpetuiert werden (systematische Fehler/Bias).

Diese Strukturelemente kommen je nach Anwendungsbereich weniger oder stärker zum Tragen; so sind bei weitem nicht alle KI-Systeme darauf ausgerichtet, autonom mit der Umwelt zu interagieren. Allerdings ist das Ausmass dieser Fähigkeit (unter Berücksichtigung der übrigen Strukturelemente) bei KI-Systemen deutlich höher. Ebenso wenig sind systematische Fehler bei allen KI-Anwendungen zwingend vorhanden.

Die Strukturelemente sind einzeln in verschiedenen Ausprägungen auch in Nicht-KI-Anwendungen zu finden (bspw. in der Regelungstechnik). Die beschriebenen Merkmale sind zudem differenziert zu betrachten. Gerade die Nachvollziehbarkeit ist auch bei anderen Verfahren und Methoden oft nur für wenige Experten möglich und nicht der breiten Bevölkerung zugänglich. Allerdings stellt KI in vielerlei Hinsicht einen grossen Fortschritt dar, so dass in Kombination dieser Strukturelemente gänzlich neue Anwendungen überhaupt erst möglich werden (z.B. Bild- oder Spracherkennung).

Die neuartigen Möglichkeiten in der Anwendung bringen auch Probleme mit sich. Auch diese sind jedoch in Einzelausprägung aus anderen statistischen Methoden hinlänglich bekannt (etwa statistische Probleme des Bias); in Kombination der verschiedenen Aspekte entstehen jedoch ganz neuartige



Herausforderungen (z.B. die Kombination von fehlender Nachvollziehbarkeit und Manipulationsmöglichkeiten bei vollautomatisiert verkehrenden Fahrzeugen) oder Probleme können stark akzentuiert werden (z.B. unbeabsichtigte Diskriminierungen bei fehlender Nachvollziehbarkeit auf Basis verzerrter Daten).

Die Einschätzung anhand der Strukturmerkmale stellt eine Momentaufnahme dar. Die Entwicklung neuer Verfahren schreitet rasant voran. Zu den aktivsten Forschungsgebieten zählt die Entwicklung von Methoden, um besser zu verstehen, wie bestimmte Resultate zustande kommen. Dabei geht es nicht nur um die Schaffung von mehr gesellschaftlicher Transparenz; eine Verbesserung der Nachvollziehbarkeit liegt nicht zuletzt auch im Interesse der Anwendungen und Anwender selbst.

### 3.1 Lernen aus Daten

Der Begriff «Big Data» bezieht sich in erster Linie auf eine enorm grosse Menge an Daten (quantitative Dimension). Allerdings ist die qualitative Dimension von Daten mindestens ebenso wichtig: Maschinelles Lernen kann mit unstrukturierten Daten umgehen, die für andere Verfahren nicht zugänglich sind. Dies betrifft zum Beispiel Bild- und Sprachinformationen, die lange gar nicht als sinnvoll mit Computern auswertbare Daten angesehen wurden. Seit Jahrzehnten gibt es z.B. Satellitenbilder, die jedoch erst mit neuen Methoden algorithmisch untersucht werden können.<sup>16</sup>

Im Vergleich zu anderen statistischen Methoden erlaubt KI zudem auch ungleich mehr Erklärungsdimensionen (Features) zu verarbeiten, so dass deutlich komplexere Zusammenhänge erkannt werden können, als sie der herkömmlichen Statistik zugänglich sind. Dabei können tausende oder sogar Millionen von Inputdimensionen (z.B. bei der Bildverarbeitung) sinnvoll zu einem Resultat verarbeitet werden. Darüber hinaus existieren Anwendungen, in denen der Mensch nur und erst mithilfe der KI in der Lage ist, eine Situation zu analysieren. So vermag der Mensch sich in drei Dimensionen gut zurechtzufinden, wenn aber die Datenstruktur aus einer (viel) höheren Zahl von Dimensionen besteht, kommt er an seine Grenzen.

Wie die herkömmlichen statistischen Analysemethoden, bei welcher beispielsweise die Stärke des Zusammenhanges zwischen zwei Grössen (Regressionsparameter) im Rahmen der Schätzung ermittelt werden, bestimmen auch Methoden der KI eine geeignete Verknüpfung der Inputdaten selber. Entscheidende Neuerung ist die Qualität des **Lernens aus Daten**. Dies geschieht, indem den Systemen entweder eine Vielzahl an bereits klassifizierten Beispielen vorgegeben wird oder indem ein Regelwerk definiert wird, innerhalb dessen die Maschine durch Ausprobieren selbstständig lernt.

Das bedeutet, dass die KI selber aus den Rohdaten und Informationen lernt, Zusammenhänge zu erkennen. Dies geschieht bei neueren Ansätzen dadurch, dass der Algorithmus eigenständig notwendige Konzepte bildet. Diese sind zwar inspiriert vom menschlichen Denken, aber nur vage vergleichbar mit Konzepten, die der Mensch verwenden würde. So würde ein Mensch ein Auto durch das Vorhandensein von Rädern, Scheiben, Spiegel, Sitzen etc. charakterisieren. Ein Computer, der ein Bild eines Autos auf Basis von einzelnen Pixeln erhält, kann mit solch abstrakten Konzepten nichts anfangen. Stattdessen erarbeitet er sich bspw. anhand einer grossen Zahl von Autobildern das Konzept Auto durch immer wiederkehrende Formen von Kanten, Linien und Konturen (siehe Anhang 2). Der Mensch trägt dazu nur insofern bei, indem er das Ziel vorgibt, hinsichtlich dessen die KI die passenden Parameter finden soll.

Es besteht eine zwingende Abhängigkeit zu Daten, welche in mancher Hinsicht die Achillessehne heutiger KI Systeme (Aufwand für initiales Training, Sicherstellen der Aktualität der trainierten Systeme, Gefahr von Bias basierend auf den Daten) darstellen. Denn KI-Systeme können nicht nur gut mit grossen Datenmengen umgehen, die meisten Methoden benötigen umgekehrt auch grosse Daten-

---

<sup>16</sup> Sendhil Mullainathan and Jann Spiess (2017): "Machine Learning: An Applied Econometric Approach"; Journal of Economic Perspectives - Volume 31, Number 2 - Spring 2017 - Pages 87–106.

mengen, um erfolgreich trainiert zu werden. Dabei nimmt der Nutzen von immer mehr Daten, die zum Trainieren zur Verfügung stehen, nicht ab.

Zudem ist auch die **Qualität der Daten** entscheidend. Einige Algorithmen des maschinellen Lernens, wie sie beispielsweise bei der Bilderkennung angewendet werden, übertreffen die durchschnittlichen menschlichen Fähigkeiten. Um dies zu erreichen, müssen sie jedoch mit grossen Mengen von bereits klassifizierten Bildern trainiert werden. Aktuelle maschinelle Lerntechnologien erfordern somit kuratier- te und passgenaue Daten. Das heisst, sie benötigen Bilder, bei denen bereits bekannt ist, dass sie ein bestimmtes Objekt – allgemeiner ein bestimmtes Muster – abbilden.

In den wenigsten Domänen sind solche Daten jedoch ohne weiteres verfügbar. Die Erstellung und Sammlung von geeigneten Daten ist eines der aktivsten Forschungsgebiete von Wissenschaft und Wirtschaft im Bereich der KI<sup>17</sup> und ermöglicht zunehmend auch die Anwendung von KI mit begrenzten oder bisher unzureichenden Daten. Verschiedene Ansätze versuchen daher neue Formen der synthetischen Generierung von Daten anzuwenden oder vorhandene Daten besser aufzubereiten. Es gibt zudem sehr aktive Forschung im Bereich des maschinellen Lernens, um Methoden zu entwickeln, welche weniger Daten benötigen. Viele reale Probleme sind jedoch zu komplex, um realistisch simuliert zu werden und erfordern daher nach wie vor den Zugang zu realen Daten.

### 3.2 Vorhersagen als Grundlage für (automatisierte) Entscheidungen

Das Ergebnis eines KI-Systems, das auf maschinellem Lernen beruht, ist im Kern immer eine Vorher- sage statistischer Natur, sei es bei der Bild- oder Spracherkennung, bei der Empfehlung für ein Pro- dukt oder der Navigation.<sup>18</sup> Vorhersagen können jedoch vielfältig eingesetzt werden. Insbesondere sind sie die wesentliche Grundlage für Entscheidungen.

Menschen treffen täglich eine Vielzahl von Entscheidungen unter Unsicherheit. Vorhersagen sind eine wesentliche Komponente dieser Entscheidungsprozesse, da sie diese Unsicherheiten reduzieren. Entscheidungsprozesse lassen sich folglich verbessern, wenn die Unsicherheit durch Prognosen re- duziert werden kann.

Viele Problemstellungen lassen sich als Vorhersageproblem formulieren, um die technischen Potenzi- ale von KI zu nutzen. Bei einer medizinischen Diagnose verwendet ein Arzt beispielsweise Informatio- nen (Daten) über die Symptome eines Patienten und ergänzt (prognostiziert) die fehlenden Informatio- nen über die Ursache dieser Symptome. Da Vorhersagen oft die Berücksichtigung umfangreicher Daten bedingt, und die Qualität der Vorhersage mit der Menge an Daten steigt, sind Machine-Learning Verfahren besonders gut geeignet für Systeme, welche auf Basis von Vorhersagen Entscheidungen treffen sollen.

Während Vorhersagen eine wesentliche Komponente von Entscheidungen darstellen, bedingt eine Entscheidung zusätzlich weitere Elemente, insbesondere eine Beurteilung über deren Konsequenzen. So sind KI-Systeme heute in der Lage, gewisse Diagnosen schneller und besser als ausgebildete Ra- diologen zu stellen (z.B. um bestimmte Krebsarten zu erkennen).<sup>19</sup> Allerdings stellen solche Systeme nur Hilfsmittel dar und liefern – wie ein Radiologe – eine Prognose in der Art, dass «basierend auf den Angaben zur Person und der Analyse der Bildgebung ist die Vergrösserung in der Leber mit einer

---

<sup>17</sup> OECD (2019): «Artificial Intelligence in Society», S. 62.

<sup>18</sup> Der englische Begriff “prediction” hat nicht notwendigerweise einen zeitlichen Bezug, sondern bedeutet als Be- griff aus der Statistik, dass eine Variable mithilfe eines durch statistische Methoden ermittelten Zusammen- hangs und auf Basis von gegebenen Werten von erklärenden Variablen prognostiziert werden kann. Im engli- schen wird daher auch «prediction» von «forecast» (für die Zukunft betreffende Prognosen) unterschieden. Zutreffender wäre daher der unübliche Begriff Prädiktion.

<sup>19</sup> Allerdings kommt es darauf an, um welches bildgebendes Verfahren es sich handelt und wie standardisiert die Diagnoseverfahren sind. Im Fall der Brustkrebs-Diagnostik beispielsweise ist die Standardisierung weitgehend gegeben. Im Fall von MRI ist dies jedoch viel schwieriger, weil die Magnete jedes Mal leicht anders sind und ein auf Scanner X trainiertes System somit nicht mit Bildern von Scanner Y gefüttert werden kann.

Wahrscheinlichkeit von 66.6% gutartig, mit einer Wahrscheinlichkeit von 33.3% bösartig und mit einer Wahrscheinlichkeit von 0.1% gar nicht existiert».<sup>20</sup>

Die Interpretation und der Entscheid über die Behandlung liegen jedoch nach wie vor beim behandelnden Arzt. Eine Entscheidung besteht in diesem Beispiel darin, eine Intervention (bspw. eine invasive Untersuchung) zu verordnen oder nicht. Eine solche Entscheidung erfordert eine ganze Reihe weiterer Einschätzungen zu Konsequenzen und Nutzen der Handlung (was sind die Konsequenzen eines invasiven Eingriffs, wenn ein Tumor bösartig oder nicht bösartig ist, oder gar nicht existiert? Was sind die Konsequenzen des Nicht-Eingriffs in allen drei Fällen?).

Der Einsatz von KI kann die Vorhersage deutlich verbessern und somit als Entscheidungsgrundlage einen grossen Nutzen generieren (in dem angeführten Beispiel kann es bspw. dazu führen, dass weniger invasive Untersuchungen durchgeführt werden). Zwar erlauben KI-Systeme, Entscheidungsprozesse zu automatisieren. Zum heutigen Zeitpunkt sind jedoch solche Systeme meist erfolgsversprechender in der Zusammenarbeit mit Menschen und wenn ein Mensch basierend auf dem erhaltenen Input Entscheidungen trifft.<sup>21</sup>

Die Realisierung einer KI-Basisfunktion (z.B. eine Bilderkennung) ist somit nur rudimentär als Entscheidung zu werten. Erst mittels Kombination von KI-Technologien werden Systeme in die Lage versetzt, Entscheidungen (im engeren Sinne) vorzunehmen. Im Beispiel des automatisierten Fahrens kann eine KI-Systemkomponente z.B. ein Strassenschild mit einer Geschwindigkeitsbegrenzung erkennen; andere Komponenten erkennen ein langsames Fahrzeug oder berechnen mit Rückgriff auf Datenbanken die Wahrscheinlichkeit eines Rechtsabbiegens. Dies führt schliesslich zur Entscheidung, langsamer zu fahren, als erlaubt wäre. Eine solche KI-Entscheidung ist aber in verschiedener Hinsicht abhängig von menschlichen Entscheidungen: erstens sind Design-Entscheidungen bei der Kombination von KI-Technologien zu einem KI-System erforderlich. Zweitens fällt ein Mensch den strategischen Entscheid, das KI-System in einem bestimmten Kontext einzusetzen (z.B. anstelle eines menschlichen Entscheiders). Drittens schliesslich, indem entschieden wird, wann und auf welche Weise Menschen die Ergebnisse von KI-Entscheidungen überprüfen sollen.<sup>22</sup>

### 3.3 Menschliche Intelligenz

Zahlreiche Anwendungen von KI bilden Aspekte menschlicher perzeptiver und kognitiver Fertigkeiten ab. So können KI-Anwendungen Bilder klassifizieren, Gesichter erkennen, Sprachen übersetzen, aber auch kreative Tätigkeiten ausüben (komponieren oder Bilder malen). Einige Systeme sind sogar bereits in der Lage in Diskussionen überzeugend zu argumentieren.<sup>23</sup>

Da mit KI komplexe Verhaltensweisen möglich werden, die bislang den Menschen vorbehalten waren, und weil die fortgeschrittensten KI-Technologien überdies vage durch die Art und Weise inspiriert sind, wie das menschliche Gehirn lernt,<sup>24</sup> wird KI oft fälschlicherweise mit einer menschlichen «allgemeinen künstlichen Intelligenz» gleichgesetzt. Dabei wären Maschinen in der Lage, Urteile und Entscheidungen zu fällen, vielfältige Probleme zu lösen, durch Lesen oder Erleben zu lernen, Konzepte zu erstellen, die Welt und sich selbst wahrzunehmen, zu erfinden und kreativ zu sein, auf das Unerwartete in komplexen Umgebungen zu reagieren und zu antizipieren.

<sup>20</sup> Agrawal, Ajay, Gans, Joshua and Goldfarb, Avi (2018): "Prediction, Judgment, and Complexity: A Theory of Decision Making and Artificial Intelligence" in "The Economics of Artificial Intelligence: An Agenda", National Bureau of Economic Research, Inc.

<sup>21</sup> E. Strickland (2019): «IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care», in IEEE Spectrum, vol. 56, no. 4, pp. 24-31, April 2019.

<sup>22</sup> Vgl. TA-SWISS (Hrsg.): «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.

<sup>23</sup> CNN Business (2019): «IBM's fast-talking AI machine just lost to a human champion in a live debate», verfügbar unter <https://edition.cnn.com/2019/02/11/tech/ai-versus-human-ibm-debate/index.html>

<sup>24</sup> Schweizerische Akademie der Technischen Wissenschaften SATW (2019): «Technology Outlook 2019», verfügbar unter <https://www.satw.ch/de/frueherkennung/technologies/>

Auch wenn KI-Systeme in vielen Bereichen kognitive Leistungen erbringen können, die den menschlichen überlegen sind und im Hinblick auf die Kommunikation mit Menschen in den vergangenen Jahren spektakuläre Fortschritte erzielt wurden, besteht Einigkeit, dass eine solche «Allgemeine Künstliche Intelligenz» heutzutage nicht existiert. Es herrschen dagegen grosse Meinungsunterschiede darüber, ob und wann eine solche Allgemeine KI bereit sein könnte.<sup>25</sup>

Die menschliche Intelligenz ist deutlich vielseitiger und basiert auf einem generalisierenden und abstrahierenden Lernen über verschiedene kognitive Funktionen hinweg. Heute mögliche KI-Anwendungen werden demgegenüber für spezifische Anwendungsgebiete und Fragestellungen entwickelt, auch wenn eine gewisse Generalisierung möglich ist («Narrow» KI). Auch die für diesen Bericht beizugezogenen Experten stimmen dieser Interpretation von Künstlicher Intelligenz zu.

Problematisch daran, KI an menschlicher Intelligenz zu messen, ist, dass sie zu einer irreführenden Beurteilung der Technologie führen kann. So zielen KI-Systeme in der menschlichen Interaktion oft erfolgreich darauf ab, den Menschen mit einfachen Mitteln zu täuschen, etwa indem absichtlich menschliche Fehler eingebaut werden. In den Antwortstrategien gehen sie nur scheinbar auf das menschliche Gegenüber ein. Bereits in den 1960er Jahren wurden simple Programme entwickelt, die Versuchspersonen gegenüber kurzzeitig in beeindruckender Weise als menschlich erschienen.<sup>26</sup> Ein im Mai 2018 von Google präsentiertes System nahm während der Präsentation verschiedene überzeugend klingende Anrufe vor, um Termine zu reservieren. Um das System möglichst menschlich klingen zu lassen, wurden von der KI u. a. Denkpausen, absichtliche Ungenauigkeiten und Laute wie „aha“ und „hmm“ eingefügt.

In der Praxis kann dies problematisch sein, wenn Anforderungen an ein System gestellt werden, die mit gesundem Menschenverstand leicht gelöst werden können. Hierbei zeigen sich oft grosse Diskrepanzen, wenn Maschinen Fehler begehen, die einem Menschen nicht unterlaufen würden.<sup>27</sup> Abbildung 3 zeigt, dass durch das Hinzufügen einer unmerklich kleinen Störung (Rauschen) zum Bild eines Pandas - speziell entwickelt, um das Bildklassifizierungsmodell zu täuschen - der Algorithmus mit einer Sicherheit von fast 100% einen Gibbon erkennt. Dies obwohl für den Menschen beide Bilder eindeutig sind (kontradiktorische maschinelle Lerntechniken).<sup>28</sup>

---

<sup>25</sup> Vgl. TA-SWISS (Hrsg.) «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz», unveröffentlichtes Manuskript; OECD (2019): "Artificial Intelligence in Society"; Rodney Brooks (2017). "The Seven Deadly Sins of AI Predictions", MIT Technology Review, October 2017, verfügbar unter <https://www.heise.de/tr/artikel/Essay-Die-sieben-Todsunden-der-KI-Vorhersagen-4003150.html>

<sup>26</sup> Weizenbaum, Joseph (1966): "ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine", verfügbar unter <http://www.cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1966.pdf>

<sup>27</sup> Im «Winograd Schema» müssen bspw. Fragen beantwortet werden, die auf das Sprachverständnis zielen und es keinen einfachen grammatikalischen oder statistischen Test gibt, der die Vieldeutigkeit auflösen könnte:

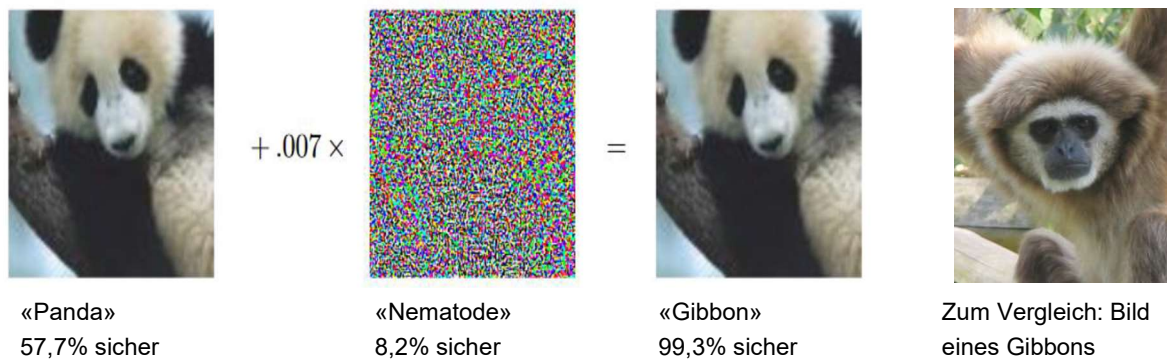
- Die Stadträte lehnten den Demonstranten eine Genehmigung ab, weil sie Gewalt befürchteten. Wer befürchtete Gewalt?
- Die Stadträte lehnten den Demonstranten eine Genehmigung ab, weil sie Gewalt befürworteten. Wer befürwortete Gewalt?

Diese beiden Sätze unterscheiden sich nur in einem Wort, haben aber die entgegengesetzte Antwort. Eine solche Frage zu stellen, hängt nicht von Tricks oder Täuschen ab, sondern davon, etwas über die Welt zu wissen, dass die Menschen verstehen, aber Computer derzeit nicht.

<https://artint.info/2e/html/ArtInt2e.Ch1.S1.SS1.html>

<sup>28</sup> OECD (2019): «Artificial Intelligence in Society», S. 94.

Abbildung 3: Täuschung von KI-Systemen (kontradiktorische maschinelle Lerntechniken)



Quelle: OECD (2019): «Artificial Intelligence in Society».

Aktuelle Berichte über solche Attacken liegen aus der Spracherkennung oder dem automatisierten Fahren vor.<sup>29</sup> Zudem wurden bereits universelle Methoden gefunden, um Täuschungen zu produzieren, die unabhängig vom konkreten Bild oder von der Tonaufnahme funktionieren.

Alle aktuellen KI-Systeme sind anfällig für solche Probleme, die auf dieser Basis gezielt manipuliert werden können. Im Bereich der Cybersicherheit oder des automatisierten Fahrens (unmerkliche Manipulation von Verkehrsschildern) können solche Verwundbarkeiten erhebliche Folgen zeigen.<sup>30</sup> Auch die Finanzindustrie könnte davon betroffen sein, was Betrügern die Möglichkeit eröffnet, mit einer Manipulation ihrer Daten ein anderes Verhalten der KI-Systeme der Finanzhäuser zu bewirken (z.B. ungerechtfertigte Kredite zu vergeben).

Für den vorliegenden Bericht ist eine an die menschliche Intelligenz angelehnte Definition von KI daher wenig zweckmässig. Dies nicht nur, weil der aktuelle technologische Stand keine solche Interpretation einer solchen Intelligenz zulässt, sondern auch weil wissenschaftlich keineswegs klar ist, was menschliche Intelligenz genau ausmacht. Schliesslich ist auch das Imitieren von Fähigkeiten, die normalerweise Menschen zugeschrieben werden, kein neues Phänomen (z.B. Taschenrechner, Schachcomputer).

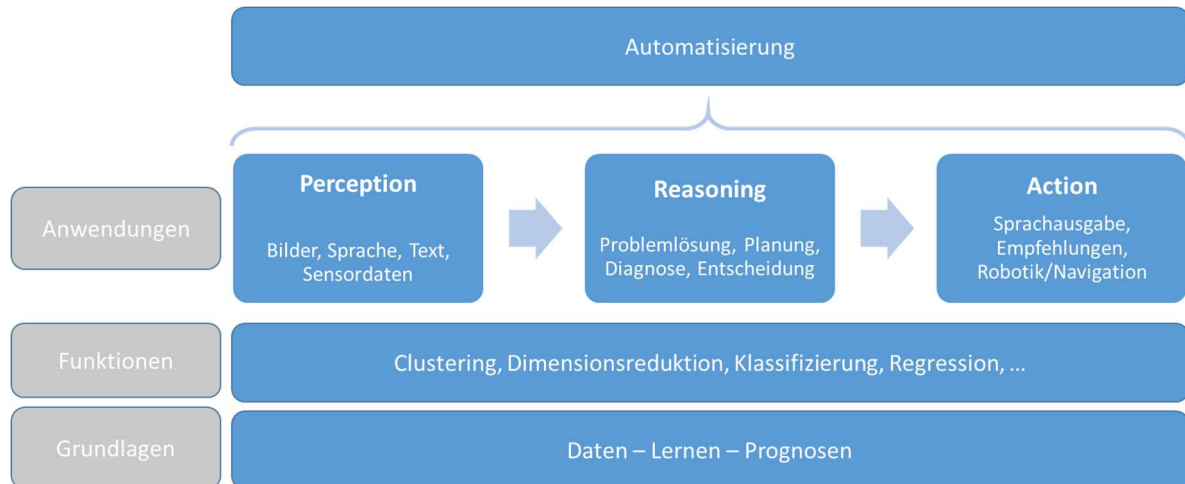
### 3.4 Autonome Handlungsfähigkeit

Auf der Basis der Fähigkeit, grosse und unstrukturierte Daten zu verarbeiten, sowie aus diesen zu lernen und Vorhersagen zu erstellen, ergibt sich die Fähigkeit, die Umwelt «wahrzunehmen». Zusammen mit der Fähigkeit, Vorhersagen in Entscheidungen zu transformieren, etwa zur Lösung von Problemen, ergibt sich, dass KI-Systeme die Fähigkeit besitzen können, in eine vollständige – d.h. von der Wahrnehmung über die Verarbeitung der Information bis zur Handlung – und automatisierte **Interaktion mit der Umwelt** (Menschen oder Maschinen) zu treten (Abbildung 4).

<sup>29</sup> Carlini und Wagner (2018): “Audio Adversarial Examples: Targeted Attacks on Speech-to-Text”, arXiv:1801.01944 [cs.LG], <https://arxiv.org/abs/1801.01944>; Yuan et al. (2018): “CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition”, arXiv:1801.08535 [cs.CR], <https://arxiv.org/abs/1801.08535>; [https://www.ics.uci.edu/~alfchen/yulong\\_ccs19.pdf](https://www.ics.uci.edu/~alfchen/yulong_ccs19.pdf)

<sup>30</sup> Ackerman, E. (2017): “Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms”, IEEE Spectrum, <https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms>; Szegedy et al. (2014): “Intriguing properties of neural networks”, arXiv:1312.6199v4 [cs.CV], <https://arxiv.org/abs/1312.6199v4>; Su, Vargas und Kouichi (2017): “One pixel attack for fooling deep neural networks”, arXiv:1710.08864 [cs.LG], <https://arxiv.org/abs/1710.08864>

Abbildung 4: Fähigkeiten zur Interaktion mit der Umwelt



Quelle: SBFi in Anlehnung an Winston, P. H. (1992): «Artificial Intelligence»; Addison-Wesley.

Auf dieser Basis besteht eine wesentliche neue Dimension somit darin, dass die heute verwendeten (oder in naher Zukunft absehbar einsatzfähigen) KI-Systeme in einer Weise eingesetzt (resp. mit anderen Technologien kombiniert) werden können, die es erlaubt, aus den statistischen Vorhersagen Entscheidungen abzuleiten und anhand dieser – in einem nächsten Schritt – **autonom zu handeln**.

Ziel des vorliegenden Berichts ist es, möglichen Handlungsbedarf in für den Bund relevanten Themenbereichen zu erörtern. Im Hinblick darauf ist die autonome Handlungsfähigkeit ein wesentliches neuartiges Charakteristikum von KI-Systemen, welches zur Abgrenzung von anderen, bereits weiter verbreiteten Technologien hinzugezogen werden kann. Damit einher gehen deutlich weitreichendere Implikationen auch für die Regulierung in verschiedenen Bereichen. Allerdings ist die Berücksichtigung von autonomer Handlungsfähigkeit ebenso wie die anderen Kriterien als alleiniges Merkmal zur Abgrenzung von KI unzureichend (so könnte sonst bereits ein mit einem Sensor ausgestatteter Toaster als KI klassifiziert werden).

Die Eigenschaft der Autonomie kann aber auch dazu dienen, verschiedene KI-Anwendungen selbst im Hinblick auf den Regelungsbedarf zu unterscheiden. So verwenden beispielsweise bestehende Fahrassistenzsysteme teilweise bereits heute künstliche Intelligenz als Hilfsmittel. Ein KI-System, welches dagegen tatsächlich hochgradig autonom fahren kann und somit eine Handlungsfähigkeit unabhängig von einer konkreten Person aufweist, ist dagegen möglicherweise anders zu beurteilen (siehe Abschnitte 4.2 und 6.9).

Eng verbunden mit der autonomen Handlungsfähigkeit sind die Robotik und das Internet der Dinge («Internet of Things»). Bei der Robotik agieren die KI-Systeme physisch mit der Umwelt, wobei KI-Systeme die zentrale Software-Komponente für die Nutzung von Robotik-Systemen in komplexen Anwendungen darstellen. Ein signifikanter Teil der potenziellen Autonomie von KI-Systemen basiert zudem auf der Zunahme der Kommunikation zwischen Maschinen selbst (Internet der Dinge). Allerdings ist die Autonomie von KI nicht auf die physische Interaktion mit der Umwelt beschränkt, was etwa die automatisierten Handelssysteme im Finanzbereich aufzeigen. Diese Technologiebereiche sind eng mit der KI verbunden, weisen aber auch eigene Problemstellungen und Herausforderungen auf.<sup>31</sup>

Die Frage, wo eine autonome Handlung beginnt und wie dies rechtlich zu qualifizieren ist, ist derzeit noch offen. Ebenso bedarf es einer gesellschaftlichen Debatte darüber, wie gleiche Tätigkeiten von Mensch und Maschine in bestimmten Bereichen gleich oder ungleich behandelt werden sollen. Empirische Untersuchungen, die im Rahmen einer laufenden Studie der Stiftung für Technologiefolgen-Abschätzung TA-Swiss durchgeführt worden sind, zeigen deutliche Unterschiede hinsichtlich der Be-

<sup>31</sup> Vgl. TA-SWISS (Hrsg.) «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.

urteilung von Verantwortlichkeit auf, wenn KI-Systeme in Entscheidungsprozesse involviert werden. So vertraut man beispielsweise einem KI-System zur Fake-News-Erkennung deutlich mehr als einem KI-System zur Beurteilung von Stellenbewerbern. Entscheidet eine Person darüber, ob ein Mensch oder ein KI-System eine bestimmte Entscheidung vornehmen soll, und geschieht dann ein Fehler durch den Menschen bzw. die KI, dann halten zudem KI-Skeptiker die Person für deutlich stärker verantwortlich, wenn sie die Entscheidung der KI überlassen haben. Dagegen machen KI-Experten diese Person deutlich stärker verantwortlich, wenn sie einem Menschen die Entscheidung übertragen hat.

### 3.5 Fehlende Nachvollziehbarkeit

Viele der aktuell erfolgreichsten KI-Methoden stellen eine «Black Box» dar; d.h. es ist nicht mehr erklärbar und nachvollziehbar, wie eine bestimmte Vorhersage oder ein bestimmtes Ergebnis zustande kommt oder wie ein KI-System bei einer konkreten Fragestellung zu einer spezifischen Antwort gekommen ist.<sup>32</sup> Selbst wenn es gelänge, den Zusammenhang zwischen den Eingabedaten und dem Resultat zu erkennen, wäre eine Erklärung der Art, dass das Resultat der KI-Berechnung nichtlinear von 1000 Eingabegrößen abhängt, welche je mit einem individuellen Gewicht in die Berechnung eingehen, zwar richtig, aber für den Benutzer nicht sonderlich brauchbar.

Mit dem Black-Box-Problem verbunden sind weitere Herausforderungen. So können Zusammenhänge in Daten erkenntlich werden, ohne dass eine zugrundeliegende Theorie vorhanden ist. Damit kann ein Modell funktionieren ohne dass sich begründen lässt, wie oder warum es für eine bestimmte Art von Problem wirksam ist. Auch ist die Leistung eines Modells angesichts der Black-Box-Problematik nur schwer vorherzusagen.<sup>33</sup> Zudem können solche Modelle nicht mehr wie klassische Software umfassend getestet werden. All diese Probleme können den erfolgreichen Transfer eines Modells von der Entwicklung in die Realität erschweren. Allerdings sind in erster Linie bestimmte ML-Algorithmen des Deep Learning betroffen (siehe Anhang 2 für eine Erklärung). Dagegen führen beispielsweise bestimmte Formen des überwachten Lernens bei Klassifikationsaufgaben zu Ergebnissen, bei denen der Prozess der Klassifikation durchaus einsichtig ist.

Je nach Anwendungsgebiet ist die fehlende Nachvollziehbarkeit mehr oder weniger problematisch. Nutzt man AI für automatisierte Übersetzungen, dürfte das Problem wenig relevant sein, zumal die Bestimmung der Qualität einer Übersetzung relativ schnell ersichtlich ist. Doch wiederum im Beispiel des automatisierten Fahrens ist das Risiko, dass sich ein System in einer Situation völlig unerklärlich verhält, deutlich gravierender.<sup>34</sup> Insbesondere wenn solche Fehler selten auftauchen, sind sie nur schwer zu antizipieren. Dies stellt auch ganz neue Anforderungen auf regulatorischer Ebene. So müsste ein KI-gesteuertes Fahrzeug neben der heute geltenden technischen Zulassung auch im Hinblick auf die Fähigkeit des KI-Fahrroboters zum sicheren automatisierten Betrieb des Verkehrsmittels evaluiert werden (Kapitel 6.9).

Aber auch zahlreiche weitere Themen, die im vorliegenden Bericht erörtert werden, sind mit der Herausforderung der Nachvollziehbarkeit konfrontiert; dazu zählt zum Beispiel die Frage, wie ein von Algorithmen kreierte Medikament beurteilt werden soll, oder wie der Einsatz von KI im Justizsystem mit dem Anspruch auf rechtliches Gehör in Einklang gebracht werden kann.

---

<sup>32</sup> MIT Technology Review (2017): "The Dark Secret at the Heart of AI",

<https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>

Im Unterschied zu «klassischen» Computeralgorithmen nutzen die neuen ML-Technologien – insbesondere die *deep neural networks* – andere Programmieretechniken. Anstelle von klaren Software-Strukturen, die für den Programmierer zumindest prinzipiell nachvollziehbar sind, wird zwar ein neuronales Netz durch den Programmierer vorgegeben, dessen Konnektivität und Gewichtung der Verbindungen verändern sich aber über enorm viele Trainings-Zyklen (ein Bilderkennungs-Algorithmus wird z.B. mit Millionen von Bildern trainiert). Vgl. TA-SWISS (Hrsg.) «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.

<sup>33</sup> TA-SWISS (Hrsg.) «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.

<sup>34</sup> MIT Technology Review (2017): "The Dark Secret at the Heart of AI".

### 3.6 Systematische Fehler (Bias) und Scheinkausalitäten

Die fehlende Nachvollziehbarkeit der Ergebnisse von KI kann weitere, bekannte Probleme der Statistik akzentuieren: Scheinkausalitäten und systematische Fehler («Bias»).

**Systematische Fehler** entstehen, wenn die geschätzten Werte systematisch von «wahren» Wert abweichen. Dies kann z.B. entstehen, wenn die Stichprobe aus einer nicht repräsentativen Bevölkerungsgruppe besteht. Ein systematischer Fehler kann aber auch durch das nicht korrekte Anwenden oder Limitationen der verwendeten Schätzmethode entstehen. Im Gegensatz zu den zufälligen Fehlern wird der Effekt des systematischen Fehlers mit wachsender Zahl der Untersuchungseinheiten (z.B. mit einer Vergrößerung des Stichprobenumfangs) grösser.

Eine **Scheinkausalität** besteht, wenn zwischen zwei Grössen zwar ein statistisch gemessener Zusammenhang (Korrelation) besteht, ohne dass aber ein ursächlicher (kausaler) Zusammenhang vorhanden ist. Scheinkausalitäten können entstehen, wenn zwei Grössen von einer gemeinsamen Ursache abhängen, die nicht berücksichtigt wird<sup>35</sup> oder wenn eine Korrelation rein zufällig ist (Koinzidenz).<sup>36</sup>

Da Scheinkausalitäten alleine schon wegen den Regeln des Zufalls in grossen Datenmengen existieren, und KI das grösste Potenzial bei umfangreichen Datenmengen ausspielen kann, ist zu erwarten, dass sich das Problem im Rahmen von KI-Anwendungen akzentuieren wird. Dies insbesondere, wenn KI eigenständig nach Mustern in sehr grossen Datenmengen suchen soll. Anders als der Mensch ist eine KI-Anwendung aber nicht grundsätzlich in der Lage, auf Basis theoretischer und intuitiver Überlegungen eine Scheinkausalität von einer echten Kausalität zu unterscheiden und damit mögliche Probleme zu erkennen.<sup>37</sup>

Dabei machen sich offenbar ein Grossteil (nach Schätzung einer Studie etwa die Hälfte) der KI-Systeme aus menschlicher Sicht naive Lösungsstrategien zunutze.<sup>38</sup> So bewerten bspw. einige KI-Systeme Bilder anhand des Kontextes; dabei werden Bilder als «Zug» klassifiziert, wenn Schienen vorhanden sind. Die eigentliche Aufgabe, Züge zu erkennen, lösen diese Systeme nicht - auch wenn die Mehrzahl der Bilder korrekt klassifiziert werden. Andere stützen ihre Klassifikationsentscheidung auf Artefakte, die während der Präparation der Bilder entstanden sind und mit dem zu bewertenden Bildinhalt nichts zu tun haben.

Es sollte also verhindert werden können, dass Eigenschaften der Muster erlernt werden, die zwar in den Trainingsdaten mit dem Ergebnis in irgendeiner Weise korreliert sind, die aber in anderen Situationen nicht zur Entscheidung herangezogen werden können.<sup>39</sup> Der Einsatz solcher, falsch trainierter Systeme in der medizinischen Diagnostik oder in sicherheitskritischen Bereichen könnte enorme Gefahren bergen.

In alltäglichen Anwendungen kann sich das Problem des Bias überdies darin äussern, dass sich in den Daten Einseitigkeiten oder Befangenheiten verbergen, die dann das Verhalten des Algorithmus prägen. Diese Aspekte werden oft im Hinblick auf Überlegungen zu **Fairness und Diskriminierung**

---

<sup>35</sup> Beispiel: Die Rechenfähigkeit von Kindern korreliert mit deren Armlänge. Der Grund für die Korrelation ist lediglich, dass ältere Kinder sowohl besser rechnen können als auch längere Arme haben als jüngere Kinder.

<sup>36</sup> So korrelieren in Deutschland die Anzahl der jährlichen Anwaltsprüfungen mit der Waldfläche; beide nehmen zu, es ist jedoch blosser Zufall, dass sich die Grössen ungefähr im Gleichschritt entwickeln.

<sup>37</sup> Es gibt verschiedene ML-Ansätze zur Bestimmung kausaler Abhängigkeiten. Allerdings sind diese aktuell noch im Forschungsstadium.

<sup>38</sup> Sebastian Lapuschkin, Stephan Wäldchen, Alexander Binder, Grégoire Montavon, Wojciech Samek & Klaus-Robert Müller (2019): «Unmasking Clever Hans predictors and assessing what machines really learn», Nature Communications, volume 10, Article number: 1096 (2019).

<sup>39</sup> Das Verzerrung-Varianz-Dilemma ist ein zentrales Problem beim überwachten Lernen. Idealerweise möchte man ein Modell wählen, das sowohl die Gesetzmässigkeiten in den Trainingsdaten genau erfasst, als auch sich auf ungesehene Testdaten generalisieren lässt. Jedoch ist es in der Regel unmöglich, beides gleichzeitig zu tun (Vgl. Scott Fortmann-Roe (2012): «Understanding the Bias-Variance Tradeoff», <http://scott.fortmann-roe.com/docs/BiasVariance.html>)



von KI-Anwendungen diskutiert. So zeigte offenbar eine Google-Nachfrage nach «professional hair», mehrheitlich die Frisuren weisser Frauen und die Suche nach «unprofessional hair» jene schwarzer Frauen.<sup>40</sup> Bekannt ist auch ein Beispiel, wonach Amazon auf die Verwendung eines algorithmischen Kandidatenauswahlsystems verzichtete, nachdem festgestellt wurde, dass das System auf der Grundlage verzerrter Lerndaten männliche Kandidaten bevorzugte.

Es ist wichtig, die technische und normative Sicht auseinanderzuhalten. Bias in statistischer Sicht ist nicht gleichzusetzen mit Bias im Sinne eines Vorurteils oder einer Voreingenommenheit zugunsten oder gegen eine Sache, Person oder Gruppe im Vergleich zu einer anderen, in der Regel in einer Weise, die als ungerecht angesehen wird.

In praktischer Nutzung ist der historische Bias meist ausschlaggebend. Oft ist dieser allerdings nicht als eigentlicher Fehler eines KI-Systems anzusehen, sondern ist vielmehr eine rechtliche Herausforderung. Wenn die Kreditwürdigkeit bestimmter Bevölkerungsgruppen historisch schlecht war, dann erkennen die Systeme aus technischer Sicht keinen Fehler in der Fortführung der Historie.<sup>41</sup>

Das Bias-Problem ist bei KI besonders relevant, weil weder Programmierer noch Anwender in der Lage sein dürften, versteckte Einseitigkeiten in Trainings-Datensätzen zu identifizieren, wenn diese aus Millionen von Datenpunkten bestehen. Werden solche KI-Anwendungen weit verbreitet eingesetzt, besteht die Gefahr, dass Personen auf Basis von KI-Entscheidungen unzulässig systematisch diskriminiert werden.

Zwar lassen sich im Hinblick auf Fairness- oder Diskriminierungsaspekte Vorgaben machen. So lassen sich Algorithmen derart konstruieren, dass sie bestimmte Datencharakteristika (z.B. Informationen zu Geschlecht oder sozialem Status) systematisch ignorieren. Allerdings beeinflusst dies dann die Genauigkeit des Algorithmus.<sup>42</sup> Mathematische Überlegungen zeigen, dass sich gewisse Anforderungen an Algorithmen (z.B. bezüglich Genauigkeit und Fairness) prinzipiell nicht gleichzeitig erreichen lassen.<sup>43</sup> Zudem kann es sein, dass unerwünschte Merkmale (unbeabsichtigt) aus anderen Merkmalen, mit denen sie korreliert sind, abgeleitet werden können.

Das Problem der Fairness von Algorithmen ist komplex, denn gegebene rechtliche Normen müssen in eine für Computerprogramme verständliche Sprache übersetzt werden. Das Problem ist, dass die Idee «Mitgliedern verschiedener Gruppen die gleiche Behandlung anzubieten», mehrere mögliche Interpretationen hat. KI-Systeme können beispielsweise so programmiert werden, dass Mitglieder verschiedener Gruppen (z.B. Männer und Frauen) die gleiche Wahrscheinlichkeit auf eine positive Vorhersage (z.B. ein guter Kandidat für eine Stelle) haben. Man könnte jedoch ebenso verlangen, dass sich die Klassifizierungsfehler erster und zweiter Ordnung (x wird fälschlicherweise der Gruppe G zugeordnet bzw. nicht zugeordnet,) über diskriminierungsrelevante Gruppen nicht unterscheiden dürfen. So könnte verlangt werden, dass ein KI-System zur Beurteilung von Anstellungen gut geeignete Frauen gleichwahrscheinlich fälschlicherweise aussondert wie gut geeignete Männer – denn solche Klassifizierungsfehler wird es immer geben.

---

<sup>40</sup> The Guardian (2018): "Women must act now, or male-designed robots will take over our lives", <https://www.theguardian.com/commentisfree/2018/mar/13/women-robots-ai-male-artificial-intelligence-automation>

<sup>41</sup> Berkeley Haas (2018): "Minority homebuyers face widespread statistical lending discrimination, study finds", <https://newsroom.haas.berkeley.edu/minority-homebuyers-face-widespread-statistical-lending-discrimination-study-finds/>

<sup>42</sup> TA-SWISS (Hrsg.) «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.

<sup>43</sup> Dies hängt mit dem Problem zusammen, dass bei einer Klassifikation die Minimierung des Fehlers einer fälschlichen Zuordnung (z.B. eine E-Mail wird als Spam klassifiziert wird, obwohl sie eigentlich kein Spam ist) auf Kosten der der fälschlichen Nicht-Zuordnung geht (E-Mail wird nicht als Spam klassifiziert wird, obwohl sie eigentlich Spam ist).

## 4 Übergeordnete Aspekte und rechtliche Einordnung

Die im vorangehenden Kapitel beschriebenen, mit dem Einsatz von KI verbundenen technischen Herausforderungen haben je nach Anwendung konkrete rechtliche Implikationen. Eine Bewertung seitens des Bundes muss daher zwingend die Frage stellen, inwiefern diese Probleme rechtlich zu qualifizieren sind. Auch dies kann nicht generell beantwortet werden, sondern muss sich aus dem konkreten Anwendungsfall ableiten. So sind bspw. an ein KI-System, welches Restaurants vorschlägt, nicht die gleichen Anforderungen hinsichtlich Transparenz und Nachvollziehbarkeit zu stellen, wie an eines, das im Justizsystem eingesetzt wird oder an eines, welches sich ohne menschliche Fahrer im Strassenverkehr bewegt.

Dennoch lassen sich allgemeine Aussagen machen, wie die mit KI-Anwendungen einhergehenden Herausforderungen unser bestehendes Rechts- und Wertesystem tangieren. Abschnitt 4.1 wird daher zunächst aufzeigen, dass **der geltende Rechtsrahmen und die grundsätzlichen Prinzipien des Bundes** im Umgang mit neuen Technologien auch im Rahmen des Einsatzes künstlicher Intelligenz Anwendung finden. Abschnitte 4.2 bis 4.4 greifen dann die spezifischen technischen Herausforderungen der Methoden der künstlichen Intelligenz (zunehmende Autonomie, fehlende Nachvollziehbarkeit, systematische Fehler) auf und zeigen auf, wie diese aus **rechtlicher Sicht** einzuordnen sind (vgl. Tabelle 2).

Tabelle 2: Zentrale technische und rechtliche Herausforderungen von KI

Technische Herausforderung		Rechtliche Herausforderung
Autonome Handlungsfähigkeit	→	Verantwortlichkeit und Haftung
Black-Box-Problematik	→	Transparenz/Nachvollziehbarkeit
Bias/Systematische Fehler	→	Mögliche Diskriminierung
Datenverfügbarkeit und -qualität	→	Datenzugang und Datenschutz

Quelle: SBFJ.

### 4.1 Grundsätze der Politik des Bundes im Umgang mit neuen Technologien

Der Bundesrat hat sich angesichts rascher technologischer Entwicklungen wiederholt dazu geäußert, wie der Bund grundsätzlich mit neuen und innovativen digitalen Technologien umgeht.<sup>44</sup> Diese Grundsätze liegen auch dem vorliegenden Bericht zugrunde:

#### (i) Bottom-Up Ansatz

Die Politik soll für optimale, innovationsfreundliche Rahmenbedingungen sorgen, welche die Entfaltung neuer Technologien ermöglichen, während die Entscheidung über die Wahl spezifischer Technologien den individuellen Akteuren überlassen bleibt. Auch im Rahmen der Forschungs- und Innovationsförderung verzichtet der Bund weitgehend auf eine Förderung spezifischer Technologien.

#### (ii) Anwendungssicht

Bei der Beurteilung von neuen Technologien stehen die Anwendungen und deren Auswirkungen im Vordergrund; die Überprüfung von Regulierungen hat sich nicht an der Technologie an sich zu orientieren, sondern hat dort anzusetzen, wo es im Hinblick auf **konkrete Anwendun-**

<sup>44</sup> Bericht des Bundesrates «Rechtliche Grundlagen für Distributed Ledger Technologie und Blockchain in der Schweiz»; Digitalisierung; «Eine Industriepolitik für die Schweiz», Bericht des Bundesrats vom 16.04.2014 in Erfüllung des Postulats Bischof.

gen von künstlicher Intelligenz Lücken oder Risiken für fundamentale Rechte von betroffenen Personen gibt. Dabei sollen in erster Linie Hindernisse beseitigt werden, die einer erfolgreichen Anwendung beim Einsatz neuer Technologien im Weg stehen. Gleichzeitig sollen negative Folgewirkungen der Anwendung gemildert oder verhindert werden.

**(iii) Technologieneutralität**

Die Schweiz verfolgt einen prinzipienbasierten und technologieneutralen Rechtsetzungs- und Regulierungsansatz, der jedoch bei Bedarf auch Ausnahmen zulässt; dabei sind die Regeln möglichst wettbewerbsneutral zu gestalten. Die rechtlichen Rahmenbedingungen sollen nicht auf einzelne Technologien ausgerichtet sein, sondern vergleichbare Aktivitäten und Risiken grundsätzlich, das heisst wo immer möglich und sinnvoll, gleich behandeln.

**(iv) Marktversagen**

Der Staat soll im marktwirtschaftlichen Bereich aus Effizienzgründen regulatorisch nur intervenieren, wenn die Markteffizienz im Vergleich zum Status quo gesteigert werden kann. Der Markt regelt grundsätzlich das Problem der Koordination der unterschiedlichen wirtschaftlichen Aktivitäten und Interessen der Akteure am besten. Am Ausgangspunkt für staatliche Massnahmen sollten daher die Marktversagen stehen. Wenn kein Marktversagen vorliegt – oder andere überwiegende öffentliche Interessen vorliegen – und der Einsatz von KI im Rahmen privatwirtschaftlicher Tätigkeiten erfolgt, ist grundsätzlich von einer Regulierung abzusehen.

**(v) Rechtliche Zulässigkeit**

Aus dem bottom-up Ansatz ergibt sich, dass der Einsatz von KI – ebenso wie die Wahl anderer Technologien – per se noch keinen staatlichen Handlungs- oder Regulierungsbedarf begründet. Systeme künstlicher Intelligenz sind – wie jede andere Technologie auch – in erster Linie ein Hilfsmittel, deren Anwendung durch Private im Grundsatz zulässig ist. Die Nutzung neuer und innovativer Technologien erfolgt aber nicht im rechtsfreien Raum, sondern hat vollumfänglich das geltende Recht zu respektieren. Die Regulierungsfrage stellt sich insbesondere, wenn KI-Anwendungen Grund- und Menschenrechte tangieren, ein Marktversagen droht oder den Bereich des staatlichen Handelns betreffen.

**(vi) Besondere Beachtung der Grund- und Menschenrechte**

Grundrechte sind elementare Rechte des Individuums und zentrale Wertgehalte und Ordnungsprinzipien des Rechtsstaats. Daher sind den Anliegen der Grundrechte in der gesamten Rechtsordnung Nachachtung zu verschaffen und es ist auf deren volle Verwirklichung hinzuwirken. Dies wird neben der Geltendmachung von subjektiven Rechtsansprüchen gegenüber dem Staat insbesondere auch durch die programmatische Schicht der Grundrechte erreicht. Danach obliegt es den rechtsetzenden Organen, möglichst freiheitliche und rechtsgleiche Bestimmungen zu erlassen, die den Grundrechten Rechnung tragen.<sup>45</sup> Zudem hat der Staat dafür zu sorgen, dass die Grundrechte, soweit sie sich dazu eignen, auch unter Privaten wirksam werden (Art. 35 Abs. 3 BV). Neben den Grundrechten relevant sind die völkerrechtlich garantierten Menschenrechte, wie sie beispielsweise auf europäischer Ebene in der Konvention zum Schutze der Menschenrechte und Grundfreiheiten vom 4. November 1950 (EMRK; RS 0.101) verankert sind.

KI kann praktisch auf alle Grund- und Menschenrechte Auswirkungen zeitigen. Werden Grund- und Menschenrechte tangiert oder erweist sich die geltende Rechtsordnung als ungenügend, besteht ein regulatorischer Handlungsbedarf. Dabei stellt sich die Frage, ob regulatorische (Mindest-) Vorschriften erlassen, institutionelle Vorkehrungen getroffen und rechtsstaatliche Kontrollen eingeführt werden müssen oder ob auf ethische oder technische Richtlinien abgestellt werden kann. Im Allgemeinen ist sicherzustellen, dass KI sich zum Wohle der Gesellschaft entwickelt und dafür eingesetzt wird.

---

<sup>45</sup> Regina Kiener/Walter Kälin/Judith Wytenbach, Grundrechte, 3. Auflage, 2018.

**(vii) Notwendige gesetzliche Grundlage für staatliches Handeln**

Der Staat (Verwaltung, Justiz) darf grundsätzlich künstliche Intelligenz als Hilfsmittel verwenden, auch wenn dies die Rechtsstellung von Personen betrifft, sofern für das staatliche Handeln in seiner konkreten, auch technischen Ausprägung die notwendigen gesetzlichen Grundlagen bestehen. Besondere Sensibilität besteht in diesem Bereich hinsichtlich der Beachtung der Grund- und Menschenrechte, da sie sich in erster Linie an staatliche Organe richten.

Nach Ansicht des Bundesrates ist es somit grundsätzlich nicht an den Behörden zu entscheiden, welche Technologie sich in welchem Ausmass durchsetzen wird. Gerade in einem sich rasch verändernden technologischen Umfeld, dessen Entwicklung für den Gesetzgeber nur bedingt vorhersehbar ist, hat sich dieser Ansatz bewährt. Erstens bietet er eine hohe Flexibilität. Zweitens wird dadurch das Ziel der Wettbewerbsneutralität unterstützt. Drittens entschärft ein technologieneutraler Ansatz das potenzielle Problem, dass Gesetzgebungsprozesse oft dem technologischen Fortschritt hinterherhinken. Damit soll allerdings nicht ausgeschlossen werden, dass es Ausnahmereiche gibt, in denen eine spezifische rechtliche Anpassung im Hinblick auf KI-Technologien angezeigt ist.

Auf Basis dieser Grundsätze sollen optimale Rahmenbedingungen geschaffen werden, so dass die Schweiz sich als attraktiver Standort für Anwendungen von neuen Technologien positionieren kann. Dabei muss der sich abzeichnende digitale Transformationsprozess bereichsübergreifend vernetzt und im Dialog mit den allen Anspruchsgruppen angegangen werden, damit die damit verbundenen Chancen genutzt und die Herausforderungen erfolgreich bewältigt werden können. Dies hat der Bundesrat in seiner Strategie «Digitale Schweiz» als Grundsatz festgeschrieben. Eine solche Offenheit des Staates gegenüber allen Anspruchsgruppen und vor allem neuen Technologien erlaubt das Ausschöpfen des Potenzials von neuen Ideen und Innovationen.

## **4.2 Autonomie, Verantwortlichkeit und Haftung**

Die Fähigkeit von KI-Systemen, zunehmend autonom agieren zu können, stellt den bestehenden rechtlichen Rahmen auf die Probe. Die Frage der rechtlichen Bewertung der Autonomie ist derzeit vor allem für den Einsatz von KI in der Robotik relevant. Dabei stellt sich aus zivilrechtlicher Sicht die Frage, wer den durch ein solches System verursachten Schaden zu tragen hat, bzw. dafür haftet. Das schweizerische Haftpflichtrecht hat einen sehr allgemeinen Charakter und ist dadurch technologie-neutral und flexibel. Die allgemeinen Haftungsregeln lassen sich auch auf den Einsatz neuer Technologien anwenden, wobei Roboter grundsätzlich als Sachen betrachtet werden. Haftbar gemacht werden kann stets nur eine natürliche oder juristische Person und nicht die Maschine. Die Haftung für den Betrieb autonomer Informatiksysteme muss immer an die Handlung oder Unterlassung einer deliktstfähigen Person anknüpfen, und zwar selbst dann, wenn die Maschine ohne direkte Aufsicht der Person agiert.<sup>46</sup>

Zudem können autonom agierenden KI-Systemen aus heutiger Sicht Attribute, die für eine Zuweisung von Verantwortlichkeit im Rechtsverkehr Voraussetzung sind, nur schwerlich zugeschrieben werden. Maschinen scheinen weder in der Lage zu sein, vorsätzlich (d.h. mit Wissen und Willen), fahrlässig (d.h. ohne die Folgen ihrer Handlung aus pflichtwidriger Unvorsichtigkeit zu beachten) oder schuldhaft (d.h. persönlich vorwerfbar) zu handeln, noch eine Urteilsfähigkeit (d.h. subjektive Einsichtsfähigkeit, Willensbildungsfähigkeit und Willensumsetzungsfähigkeit) zu entwickeln.

Aufgrund des aktuellen Stands der Technologie erachtet der Bundesrat die bestehenden Regelungen derzeit für genügend. Dass deren Anwendung auf Roboter Verantwortlichkeitslücken ergäben, hat sich bislang nicht gezeigt.<sup>47</sup> Dies gilt derzeit auch im Bereich des Strafrechts. Tatsächlich können Vergehen, die unter Einsatz von Robotern verübt werden, wie jede andere Straftat, die eine Person mithil-

<sup>46</sup> vgl. 15.3446 Ip. FDP-Liberale Fraktion.

<sup>47</sup> vgl. z.B. 18.3445 Ip. Marchand-Balet mit Bezug auf selbstfahrende Fahrzeuge und Bericht des Bundesrates in Erfüllung des Postulats Leutenegger Oberholzer 14.4169 «Auto-Mobilität» bzgl. Sachverhalten mit Auslandsbezug.

fe eines Gegenstands verübt, verfolgt werden. Somit besteht hier nach derzeitigem Stand der Dinge keine Rechtslücke, die der Gesetzgeber schliessen müsste.

Diese Einschätzung schliesst jedoch nicht aus, dass sich früher oder später die Frage nach spezifischem Regelungsbedarf stellen wird. Ein Blick in die Vergangenheit zeigt, dass es sich dabei keineswegs um eine neue Situation handelt: So hat der technische Fortschritt immer wieder neue Gefahrenquellen mit sich gebracht, beispielsweise Motorfahrzeuge, Kernenergie oder gentechnisch veränderte Organismen. In all diesen Fällen hat der Gesetzgeber mit der Einführung einer speziellen Gefährdungshaftung reagiert. Diese hat zur Folge, dass eine durch die neue Technologie verursachte Schädigung einer bestimmten Person zugerechnet wird, die dann verschuldensunabhängig für Schädigungen einstehen muss. Wer von der neuen Technologie profitiert, soll auch deren Risiken übernehmen.<sup>48</sup>

### 4.3 Nachvollziehbarkeit und Transparenz

Da bei Entscheidungen auf Basis von KI-Systemen oft nicht nachvollziehbar ist, wie diese zustande gekommen sind, sind aus rechtsstaatlicher Sicht Vorkehrungen zu treffen, wie die Transparenz bei solchen Entscheidungen gefördert werden kann.

Eine Form der Erklärbarkeit ist denn auch im Revisionsentwurf zum Datenschutzgesetz vorgesehen: Der Datenverantwortliche (bei dem es sich um eine Privatperson oder ein Bundesorgan handeln kann) muss die betroffene Person über jede Entscheidung informieren, die ausschliesslich aufgrund einer automatisierten Bearbeitung personenbezogener Daten (einschliesslich Profiling) getroffen wird und die rechtliche Auswirkungen auf die betroffene Person hat oder sich erheblich auf sie auswirkt (Art. 19 des Entwurfs<sup>49</sup>). Sofern die betroffene Person dies verlangt, muss der Datenverantwortliche ihr Gelegenheit geben, ihren Standpunkt geltend zu machen. Die betroffene Person kann verlangen, dass die Entscheidung durch eine natürliche Person überprüft wird. Darüber hinaus muss der Datenverantwortliche eine Folgenabschätzung vornehmen, wenn die vorgesehene Bearbeitung personenbezogener Daten möglicherweise ein erhöhtes Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person mit sich bringt (Art. 20). Besondere Pflichten gelten für automatisierte Einzelentscheidungen von Bundesbehörden.

Macht die betroffene Person ihr Auskunftsrecht geltend, muss der Datenverantwortliche ihr in jedem Fall mitteilen, dass eine automatisierte Einzelentscheidung getroffen wurde und auf welcher Logik diese Entscheidung beruht (Art. 23, Abs. 2, Ziffer f). Ähnliches sieht auch die Datenschutz-Grundverordnung der EU vor (vgl. insbesondere Art. 15 Ziff. 1 Bst. h DSGVO).

Artikel 19 und 23 des DSG-Revisionsentwurfs sind nicht anwendbar, wenn Menschen in die Entscheidungsfindung eingreifen und wenn die künstliche Intelligenz lediglich als Entscheidungshilfe dient. Unter menschlichem Eingreifen ist die Überprüfung durch eine natürliche Person zu verstehen, die die Situation selbst bewertet und die von dem durch die Maschine gelieferten Ergebnis abweichen kann. Ein mögliches Beispiel für eine automatisierte Entscheidung wäre die Verhängung einer Busse für eine Geschwindigkeitsübertretung ausschliesslich auf Grundlage von Fotografien des Kennzeichens und der Person am Steuer sowie automatischer Hinzuziehung von Daten aus dem Fahrzeugregister und Nutzung eines Gesichtserkennungstools.

Spezielle Anforderungen an die Nachvollziehbarkeit bestehen auch bei nicht-automatisierten Einzelentscheidungen von Behörden, welche mithilfe von KI getroffen werden und die Rechtsstellung einer Person betreffen. Aus dem verfassungsrechtlichen Anspruch auf rechtliches Gehör leitet sich die Pflicht der Behörden ab, ihre Entscheide zu begründen. Stützt sich eine Behörde für ihren Entscheid somit auf KI ab, ist es unerlässlich, dass das System über die von ihm berücksichtigten Informationen und

---

<sup>48</sup> vgl. 17.3040 Po. Reynard.

<sup>49</sup> BBL 2017 7217 ff.; vgl. auch 17.059 Datenschutzgesetz. Totalrevision und Änderung weiterer Erlasse zum Datenschutz, <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaefte?AffairId=20170059>.

Kriterien, die getroffenen Annahmen und die massgebenden Beweggründe für das Ergebnis Aufschluss gibt.

Die Nutzung künstlicher Intelligenz durch eine staatliche Stelle kann ein Gesetz im formellen Sinn erfordern, wenn personenbezogene Daten bearbeitet werden und möglicherweise erheblich in die Grundrechte der betroffenen Person eingegriffen wird. Daher sieht der DSGVO-Revisionsentwurf eine formelle Rechtsgrundlage vor, wenn schützenswerte Daten bearbeitet werden, wenn ein Profiling durchgeführt wird oder wenn der Zweck oder die Art und Weise der Datenbearbeitung zu einem schwerwiegenden Eingriff in die Grundrechte der betroffenen Person führen können (Art. 30 Abs. 2).

Aus den oben genannten Gründen geht hervor, dass die Anforderungen an die Nachvollziehbarkeit steigen, je schwerer der Eingriff in die Grund- und Menschenrechte der betroffenen Person ausfällt. So sollte beispielsweise eine durch KI generierte Empfehlung für Musiktitel weniger problematisch sein, während Entscheidungen von KI-Systemen, welche etwa die Rückfallgefahr einer tatverdächtigen oder verurteilten Person evaluieren, grundlegende Rechte stark tangieren können.

Eine weitere Herausforderung stellt sich, wenn Unternehmen KI in der Interaktion mit Kunden einsetzen, z.B. mittels Chatbots. Diese können vielseitig eingesetzt werden, um Fragen der Verbraucher zu beantworten, ihnen zu helfen oder sie zu beraten. Da es möglich ist, mit einem Chatbot wie mit einem Menschen zu sprechen, ist es für den Verbraucher unter Umständen nicht erkennbar, dass es sich um eine Maschine handelt. Gemäss der OECD-Empfehlung zur künstlichen Intelligenz vom 22. Mai 2019 beinhaltet ein verantwortungsvoller Ansatz für künstliche Intelligenz die Information von betroffenen Personen über eine solche Interaktion mit Systemen der künstlichen Intelligenz. Wenn die Konsumenten diesbezüglich nicht vorher und systematisch über die Interaktion mit Systemen der künstlichen Intelligenz informiert wurden, könnte in der Schweiz das Bundesgesetz gegen den unlauteren Wettbewerb (UWG) zu Anwendung gelangen (vgl. Kapitel 6.15).

### 4.4 Bias und Diskriminierung

Systematische Fehler in den Daten oder Algorithmen von KI-Systemen können dazu führen, dass in Anwendungen bestimmte Personengruppen systematisch diskriminiert werden. Insbesondere wenn die betreffenden Daten einen diskriminierenden Bias enthalten, besteht eine hohe Wahrscheinlichkeit, dass dieser durch das Algorithmensystem reproduziert wird. Dies ist zwar kein Alleinstellungsmerkmal von KI, denn alle auf Daten oder Algorithmen basierenden Entscheidungen können, soweit sie sich auf vorhergehende menschliche Entscheidungen stützen, die gleichen Vorurteile weitertragen und vervielfachen, die auch der menschlichen Entscheidungsfindung zugrunde lagen. Allerdings kann sich das Problem mit der Black-Box-Problematik und der Möglichkeit, Entscheidungsverfahren zu automatisieren, wesentlich verschärfen.

Angesichts der neuartigen Möglichkeiten, grosse Datenmengen auswerten zu können, können bei der Verwendung von KI zudem neue Probleme auftreten. So kann es unerwünscht oder rechtlich untersagt sein, gewisse Informationen in Entscheidungen zu berücksichtigen. Ein KI-System kann aber unter Umständen die relevanten Informationen dennoch in scheinbar unabhängigen oder neutralen Daten identifizieren (Stellvertreterdaten). So könnte in einem Bewerbungsverfahren zwar auf die Frage nach einer Schwangerschaft einer Bewerberin verzichtet werden, aber die Information könnte aus anderen Daten gewonnen werden.<sup>50</sup>

Die Vorhersagen der Algorithmen können die Diskriminierung sogar noch verstärken: Hat die Polizei in der Vergangenheit beispielsweise mehr Personen mit Migrationshintergrund festgenommen und ihre Aufmerksamkeit daher auf bestimmte Quartiere konzentriert, in denen viele Menschen mit Migrationshintergrund leben, ist die Wahrscheinlichkeit hoch, dass sich der betreffende Bias durch diese Aufmerksamkeit reproduziert. Das Ergebnis sind mehr Festnahmen als in anderen Quartieren oder

---

<sup>50</sup> Die Studie «Discrimination, artificial intelligence and algorithmic decision-making» des Europarats nennt ein Beispiel, in dem aufgrund der Einkäufe einer Frau erkannt wurde, dass sie schwanger war.

Bevölkerungsgruppen, die weniger aufmerksam beobachtet werden, wodurch sich die Vorstellung weiter verfestigt, die betreffenden Personengruppen oder Quartiere stellen ein erhöhtes Kriminalitätsrisiko dar («Feedback-Schleife»).

Es kann nicht ausgeschlossen werden, dass der Einsatz künstlicher Intelligenz auch zu neuen Arten der Diskriminierung führt. Stützt sich beispielsweise ein Vorhersagesystem auf die Tatsache, dass 80 Prozent der Bewohner eines bestimmten Quartiers ihre Rechnungen verspätet begleichen und lehnt ein Unternehmen den Bewohnerinnen und Bewohnern des Quartiers auf Grundlage dieser Ergebnisse jeden Kredit ab, benachteiligt dies auch die 20 Prozent der Bevölkerung, die ihre Rechnungen rechtzeitig bezahlen.

Die Studie «Algorithmen und Menschenrechte»<sup>51</sup> des Europarats empfiehlt, dass bei der Beurteilung, ob ein Algorithmus diskriminierende Behandlung fördert oder im Gegenteil verhindern kann, auf die juristische Unterscheidung zwischen direkter und indirekter Diskriminierung abzustützen ist. Das Schweizer Recht, insbesondere das Gleichstellungsgesetz<sup>52</sup>, kennt diese Unterscheidung. Von direkter Diskriminierung spricht man, wenn eine Person ihre Entscheidung auf die Zugehörigkeit einer Person zu einer bestimmten Gruppe stützt, die in der Vergangenheit Ziel von Diskriminierungen wurde. Von indirekter Diskriminierung spricht man, wenn sich die Entscheidung auf scheinbar neutrale Faktoren stützt, die im Ergebnis jedoch bestimmte Bevölkerungsgruppen benachteiligen, die Ziel von Diskriminierungen sind oder waren (beispielsweise der Faktor Teilzeitarbeit, wodurch Frauen benachteiligt werden).

Das Schweizer Recht ermöglicht es, diesen Risiken in gewissem Rahmen zu begegnen, auch in Beziehungen zwischen Privatpersonen. Die bestehenden Gesetze sind grundsätzlich auch für Diskriminierungen, die aus KI-Systemen resultieren, anwendbar. Das Gleichstellungsgesetz, das Diskriminierungen aufgrund des Geschlechts bei der Einstellung und in privatrechtlichen und öffentlich-rechtlichen Arbeitsverhältnissen verbietet, betrifft beispielsweise auch KI-basierte Auswahlprozesse.

Die im Rahmen der DSGVO-Revision vorgesehenen Massnahmen (siehe Kapitel 4.3 oben) können hier präventiv und korrigierend wirken: Die betroffene Person hat beispielsweise die Möglichkeit, bei einer automatisierten Einzelentscheidung ihren Standpunkt geltend zu machen und kann verlangen, dass die Entscheidung durch eine natürliche Person überprüft wird.

Besondere Aufmerksamkeit erfordern die Diskriminierungsrisiken bei Anwendung künstlicher Intelligenz im öffentlichen Sektor, wenn es zum Beispiel um die Entwicklung oder um die Durchsetzung des Rechts geht. Dem könnte man durch Risikoanalysen beim Entwerfen der entsprechenden Systeme sowie durch regelmässige Beobachtungen und Beurteilungen gerecht werden.

Es gilt zu bedenken, dass es hinsichtlich des Diskriminierungsrisikos nicht das Gleiche ist, ob KI für ein Schachspiel eingesetzt wird oder mit KI geeignete Bewerberinnen bzw. Bewerber für eine Arbeitsstelle auszuwählen. Und letztlich darf auch nicht vergessen gehen, dass künstliche Intelligenz, sofern sie gut umgesetzt ist, auch menschliche Vorurteile erkennen und verhindern kann und somit zu weniger diskriminierende Entscheidungen beitragen kann.

## 4.5 Datenzugang und Datenschutz

Insbesondere für die wissenschaftliche Forschung, aber auch für praktische Anwendungen von KI in Wirtschaft und Gesellschaft ist die Verfügbarkeit und der Zugang zu Daten zentral. Der Bund verfügt zu diesem Zweck über eine Datenpolitik. Diese will in erster Linie den Zugang zu Daten, namentlich zu frei verfügbaren Daten (Open Data) als Rohstoff einer digitalen Wirtschaft und Gesellschaft fördern,

<sup>51</sup> Etude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires (Studie über die Bedeutung der Menschenrechte im Rahmen von Techniken zur automatisierten Datenverarbeitung und eventuelle gesetzliche Implikationen), Studie des Europarats DGI (2017)12, S. 29 ff.

<sup>52</sup> SR 151.1.

und zeitgemässe und kohärente Rechtsgrundlagen und Rahmenbedingungen zu schaffen, mit denen sich die Schweiz als attraktiver Standort für eine Wertschöpfung durch Daten positioniert. Die Datenpolitik definiert aber auch den Rechtsrahmen innerhalb dessen Daten zulässig gesammelt, verknüpft und ausgewertet werden können.

Gleichzeitig muss der Bund die Sicherheit und den Schutz von Daten gewährleisten. Dies betrifft einerseits den Schutz von persönlichen Daten im Rahmen des Datenschutzes, andererseits aber auch den Schutz des geistigen Eigentums, der bei der Aufbereitung oder der Verwendung von Daten tangiert werden kann, beispielsweise, wenn urheberrechtlich geschützte Texte oder Bilder für KI-Anwendungen zur Verfügung stehen sollen.

Datenschutzverletzungen sind kein auf KI beschränktes Problem; KI kann dieses jedoch verschärfen. Durch die wachsende Anzahl digitaler Fussabdrücke, können heute mithilfe von Algorithmen ohne aktive Befragung detaillierte Persönlichkeitsprofile erstellt werden. Insbesondere die Möglichkeit, personenbezogene Daten aus verschiedenen Quellen zu kombinieren, bietet ein enormes Potenzial für Anwendungen künstlicher Intelligenz.

Beim **Immaterialgüterrecht** geht es zum einen um adäquate Schutzrechte für KI-Systeme, die teilweise mit Daten trainiert werden, die ihrerseits rechtlichen Bestimmungen unterliegen. Andererseits können KI-Systeme «Neues» erschaffen, was die Frage stellt, ob Werke der Literatur und Kunst sowie Erfindungen, die durch den Einsatz von KI entstanden sind, durch Urheberrechte bzw. Patente geschützt werden können und wem die Rechte gegebenenfalls zustehen.<sup>53</sup>

Das **Patentrecht** lässt in der Schweiz nach herrschender Meinung nur natürliche Personen als Erfinder im Rechtssinne zu (allenfalls, je nach Auslegung noch juristische Personen). Dagegen ist die Anerkennung von KI-Systemen aufgrund der ihnen zumindest gegenwärtig noch mangelnden Rechtsfähigkeit ausgeschlossen.

Zum anderen ist zu klären, wie das **Urheberrecht** damit umgehen soll, dass viele Formen von KI enorme Mengen von Daten für den Trainingsprozess benötigen, die zumindest teilweise urheberrechtlich geschützt sind. Die Daten müssen zur Verwendung durch KI in der Regel vervielfältigt werden, was grundsätzlich eine Urheberrechtsverletzung darstellt. Für die Weiterentwicklung von KI könnte dies eine erhebliche Hürde darstellen.

Die Regulierung in diesen Bereichen muss daher Zielkonflikte zwischen diesen unterschiedlichen Ansprüchen berücksichtigen. Auf die übergreifenden Herausforderungen in diesem Themenfeld wird in Kapitel 6.17 des vorliegenden Berichts eingegangen. Zusätzlich dazu können sich weitere, themenspezifische Fragen zu Datenzugang und Datensicherheit stellen. Diese werden in den jeweiligen Themenfeldern aufgenommen.

## 5 Künstliche Intelligenz – Forschung, Entwicklung und Anwendung in der Schweiz

Dieses Kapitel enthält eine kurze Analyse des gegenwärtigen Stands der künstlichen Intelligenz (KI) in der Schweiz. Nach der Darstellung der wichtigsten Akteure auf diesem Gebiet (Forschungslandschaft Schweiz) werden drei Themen anhand quantitativer Indikatoren untersucht: Wo steht die Schweiz im Hinblick auf Forschung, Entwicklung und Anwendung von künstlicher Intelligenz.<sup>54</sup>

---

<sup>53</sup> TA-SWISS (Hrsg.) «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.

<sup>54</sup> Es ist anzumerken, dass die Beurteilung von KI-bezogenen Technologieentwicklungen schwierig ist: Die Grenzen zwischen KI und anderen Technologien sind verschwommen und verändern sich ständig.



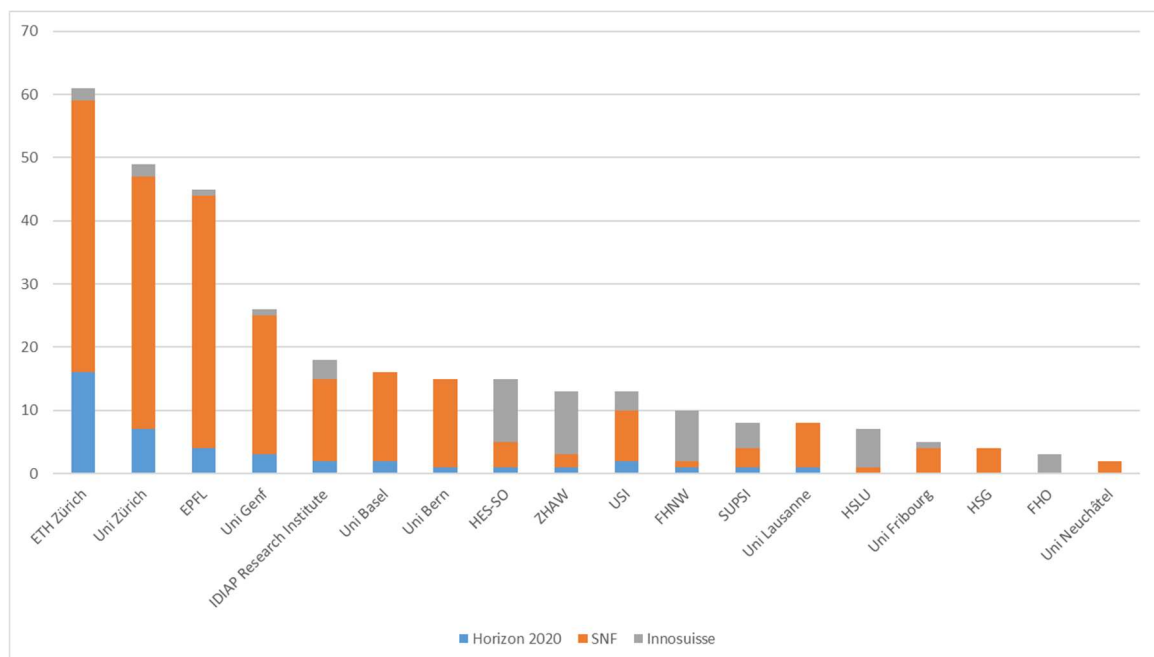
## 5.1 KI-Forschungslandschaft Schweiz: Akteursübersicht

Die Schweiz hat sich zu einem wichtigen Standort für KI-Technologie entwickelt. Dabei zeigt sich eine vielfältige und dynamische Forschungslandschaft, die geprägt ist von einigen bekannten und seit längerer Zeit etablierten Forschungszentren. Dazu zählen das Istituto Dalle Molle di Studi sull'Intelligenza Artificiale IDSIA (in Abbildung 5 ist das IDSIA teilweise in USI oder SUPSI zugeordnet) und das IDIAP Research Institute (ehemals Institut d'intelligence artificielle perceptive) wie auch die Zentren im ETH-Bereich. Der ETH-Bereich baut überdies derzeit seinen strategischen Fokusbereich Datenwissenschaften deutlich aus, wozu die ETH Zürich und die EPFL ein gemeinsames Swiss Data Science Center SDSC aufgebaut haben. Aber auch die Fachhochschulen und die übrigen Universitäten sind im Bereich KI aktiv. Das Datalab, ein virtueller Zusammenschluss verschiedener Departemente der ZHAW, pflegt beispielsweise eine enge Zusammenarbeit mit der Industrie. Die hohe Qualität der Forschung ist nicht zuletzt auch für die in der Schweiz im Bereich KI tätigen Firmen von grosser Bedeutung. Private Initiativen wie die Swiss Group of Artificial Intelligence and Cognitive Science SGAICO ergänzen daher die Initiativen seitens der Hochschullandschaft und bringen Forscher und Anwender zusammen und fördern Wissensvermittlung, Vertrauensförderung und Interdisziplinarität.

Auch die Forschungsförderung hat sich der Thematik KI angenommen. Der Bund investiert beispielsweise über den SNF in das NFP 75 – Big Data, das NFP 77 – Digitale Transformation sowie in den NCCR Robotics. Über Innosuisse wird das Nationale thematische Netzwerk Swiss Alliance for Data-Intensive Services (Data + Service) unterstützt, welches Unternehmen bei der Entwicklung neuer Produkte und Dienstleistungen unterstützt und damit den Wissenstransfer fördert, sowie über das Impulsprogramm «Fertigungstechnologien», in welchem ebenfalls KI zur Anwendung kommt.

Auf Basis der Datenbanken von SNF, Innosuisse sowie der EU (Horizon 2020) lässt sich eine grobe Akteursübersicht über die Forschungsinstitutionen der Schweiz darstellen. Abbildung 5 zeigt anhand der Anzahl von SNF, Innosuisse und im Rahmen von Horizon 2020 geförderten **Projekte**, dass die Schweizer KI-Forschung breit aufgestellt und über alle Landesteile abgestützt ist. Neben den beiden ETH sind auch die Universitäten und Fachhochschulen in der KI-Forschung aktiv. Letztere engagieren sich vor allem im WTT, was der Wissensbasis in die Wirtschaft massgeblich zugutekommt.

Abbildung 5: KI-Projektförderung (Horizon 2020, SNF und Innosuisse): Anzahl Projekte 2015-2018

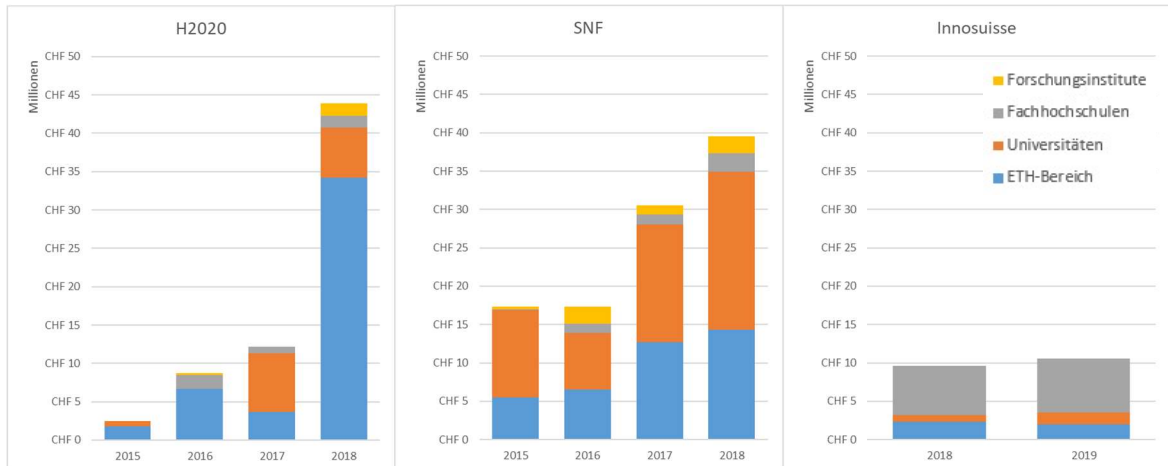


Quelle: Auswertung SBFI auf Basis der Datenbanken von SNF (P3), Innosuisse (2018 bis Q2 2019) und EU; die Datenbanken wurden nach denselben KI-Stichwörtern durchsucht.

## Herausforderungen der künstlichen Intelligenz

Betrachtet man die **Fördervolumen** zeigt sich ein deutlicher Anstieg in den letzten Jahren (Abbildung 6). Dies gilt namentlich für die von der EU geförderte Forschung. Es zeigt sich bei der Verteilung der Fördergelder auch ein klares Muster. Während der ETH-Bereich bei den EU-Projekten dominiert, findet die SNF-Förderung vor allem an den Universtäten Niederschlag. Die Innosuisse arbeitet im Bereich KI vor allem mit Fachhochschulen zusammen, wobei die Fördervolumen insgesamt deutlich geringer sind.

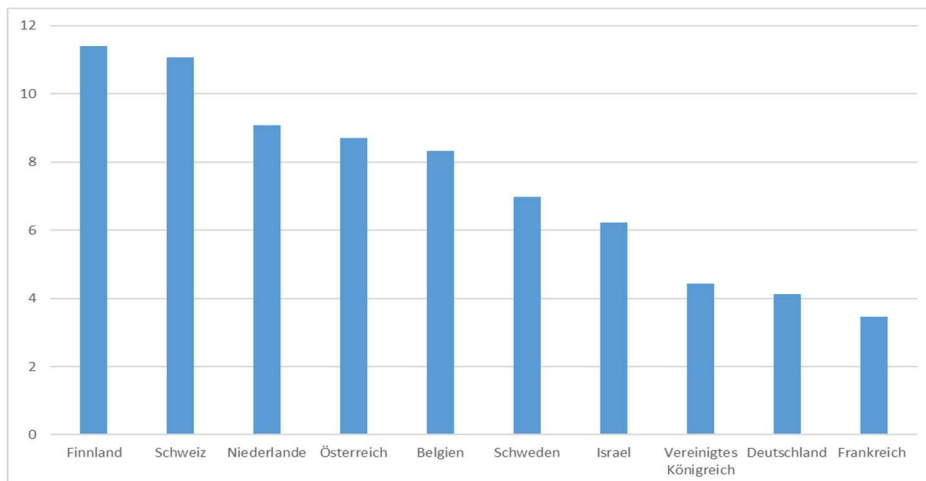
Abbildung 6: Entwicklung Fördervolumen KI-Forschung



Quelle: Auswertung SBFi auf Basis der Datenbanken von SNF (P3), Innosuisse (2018 bis Q2 2019) und EU.

Der internationale Vergleich im Rahmen von Horizon 2020 zeigt, dass die Schweiz gemessen an der Anzahl KI-Projekte pro Million Einwohner gut dasteht (Abbildung 7). Dies kann als Zeichen gewertet werden, dass die einheimische KI-Forschungslandschaft international konkurrenzfähig ist.

Abbildung 7: Anzahl KI-Projekte pro Million Einwohner in Horizon 2020, 2015-2018



Bemerkung: Dargestellt sind ausgewählte Länder.

Quelle: Auswertung SBFi auf Basis der Datenbanken von EU.

## 5.2 Leistung von F&E in der Schweiz

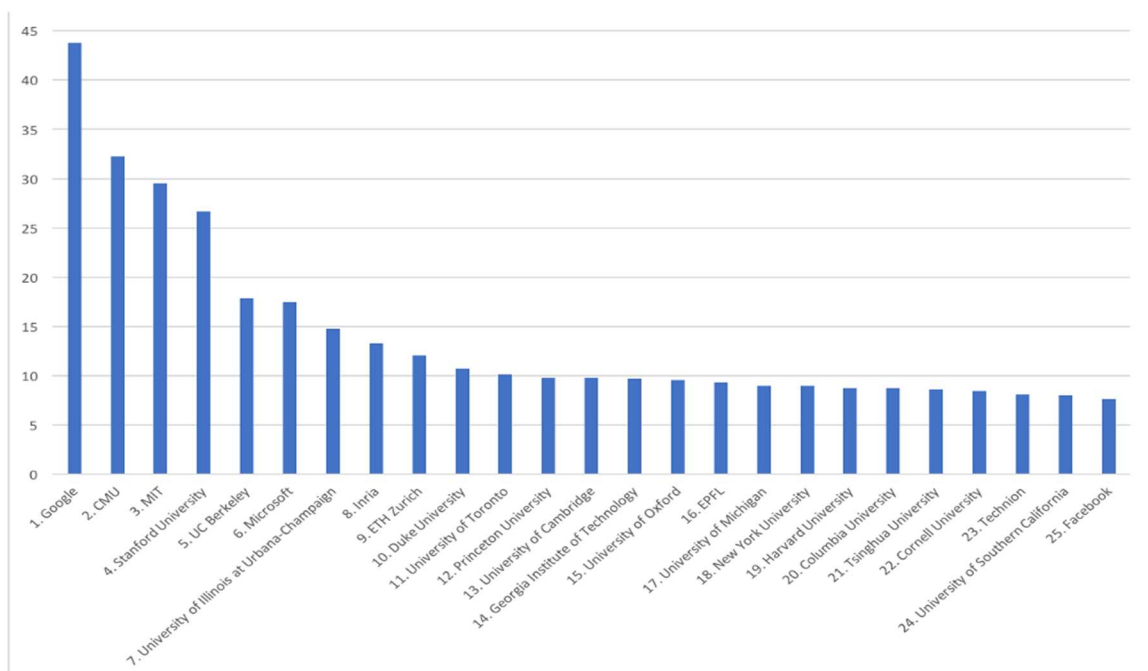
### 5.2.1 KI-Forschung in der Schweiz

Die Eidgenössischen Technischen Hochschulen rangieren im QS World University Ranking 2019 auf Platz acht bzw. neun der bedeutendsten Computer Science- und Informatik-Schulen der Welt (ohne

Abbildung). Neben den Forschungsk Kooperationen mit Universitäten und Unternehmen in der Schweiz tragen die internationalen Kontakte zu diesem Spitzenplatz bei. So arbeitet das Institute for Machine Learning der ETH Zürich mit amerikanischen Spitzenuniversitäten, der EPFL und dem Max-Planck-Institut für intelligente Systeme zusammen. Technologiekonzerne wie Google, Facebook, IBM oder Microsoft arbeiten in der Schweiz nicht nur eng mit der Forschung zusammen, sondern stellen auch intensiv Talente der Hochschulen ein. Neben der ETHZ und der EPFL sind weitere Universitäten in der Schweiz im Bereich Computer Science höchst erfolgreich, wodurch die Unternehmen Spitzenkräfte rekrutieren und Forschungsk Kooperationen mit Universitätslabors eingehen können.

Spezifisch im Bereich KI kann die Qualität der Institutionen anhand der Anzahl Publikationen in der «Conference on Neural Information Processing Systems» (NeurIPS), der renommiertesten Konferenz in diesem Bereich, dargestellt werden. Auch hier zeigt sich, dass die beiden ETH zu den weltweit renommiertesten Hochschulen zu zählen sind (Abbildung 8).

Abbildung 8: Anzahl Publikationen in der NeurIPS nach Organisation, 2017



Quelle: Conference on Neural Information Processing Systems (NeurIPS), siehe auch

<https://medium.com/@chuvpilo/whos-ahead-in-ai-research-insights-from-nips-most-prestigious-ai-conference-df2c361236f6>

Für Forschende ist die Veröffentlichung von Artikeln in wissenschaftlichen Zeitschriften der wichtigste Weg, um ihre Erkenntnisse weiterzugeben. Die statistische Auswertung wissenschaftlicher Publikationen liefert somit einen guten Überblick über die entsprechende Forschungstätigkeit in einem Land.

Betrachtet man die Volumen an wissenschaftlichen Publikationen aus der Schweiz im Bereich KI insgesamt, so zeigt sich, dass die Forschung hinsichtlich Umfang von den grossen Ländern dominiert wird, **die Qualität der Schweizer Forschung aber sehr hoch ausfällt**. Laut «Times Higher Education» kamen im Zeitraum 2011 bis 2015 mehr als 41 000 Publikationen zum Thema Künstliche Intelligenz aus China, das sind fast doppelt so viele wie aus den USA (25 500). Dahinter folgen Japan (11 700) auf dem dritten, das Vereinigte Königreich (10 100) auf dem vierten und Deutschland (8000) auf dem fünften Platz. Die Schweiz zählt 1700 Publikationen zu künstlicher Intelligenz im Zeitraum 2011 bis 2015. In Pro-Kopf-Betrachtung steht die Schweiz deutlich besser da, fällt aber hinter Länder wie Singapur, Hong Kong oder Niederlande zurück (ohne Abbildung). Doch auch wenn China bei der Zahl der Publikationen weit vorne liegt, nimmt das Land hinsichtlich der nach Disziplin gewichteten **Zitierungsrate (Impact)** nur den 34. Platz ein. Nach diesem Indikator, der eher die Qualität und die Be-

achtung der Publikationen in der wissenschaftlichen Gemeinschaft misst, liegt die Schweiz mit einem Impact von 2,71 weltweit an erster Stelle, gefolgt von Singapur (2,24) und Hongkong (2,00).

Tabelle 3: Volumen und Impact der Publikationen im Bereich KI nach Ländern (2011-2015)

Rang (Impact)	Land	Impact (nach Bereich gewichtete Zitierungen; 1=Durchschnitt)	Anzahl Publikationen
1	Schweiz	2.71	1'685
2	Singapur	2.24	2'432
3	Hong Kong	2.00	2'205
4	USA	1.79	25'471
5	Italien	1.74	6'221
6	Niederlande	1.71	2'458
7	Australien	1.69	5'227
8	Deutschland	1.66	7'957
9	Belgien	1.64	1'537
10	Grossbritannien	1.63	10'120

Bemerkung: Dargestellt sind nur die zehn Länder mit dem grössten Impact

Quelle: Elsevier (2018): «AI report».

Die Entwicklung von Techniken des maschinellen Lernens (machine learning) spielt in der KI-Forschung eine entscheidende Rolle. Seit einigen Jahren ist eine erhebliche Zunahme der **wissenschaftlichen Publikationen zum maschinellen Lernen** zu beobachten. An der Spitze der Forschungsanstrengungen auf diesem Gebiet stehen die Vereinigten Staaten mit mehr als 3,3 Mal mehr vielzitierten Publikationen als China auf Platz zwei und 5,6 Mal mehr als das Vereinigte Königreich auf dem dritten Rang. Die Schweiz erreicht beim maschinellen Lernen Platz 14 in der Rangliste und liegt damit gleichauf mit Ländern wie Singapur oder den Niederlanden. Die gute Platzierung in diesen internationalen Rankings belegt die hervorragende Qualität der Schweizer Forschung auf dem Gebiet der KI.

### 5.2.2 KI-Entwicklung in der Schweiz (Patentanalyse)

Patentanmeldungen sind ein Mass für die technische und wirtschaftliche Nutzung von Forschungsergebnissen. Ihre Anzahl kann in diesem Sinne als Indikator für die Entwicklungstätigkeit angesehen werden.

Laut OECD kommen mehr als 60 Prozent aller Patentanmeldungen im Bereich KI aus nur drei Ländern: Japan, Südkorea und den Vereinigten Staaten. China erreicht mit 10,4 Prozent den vierten Platz. Alle anderen Länder bleiben jeweils unter 5 Prozent. Die Schweiz erreicht mit **0,4 Prozent** aller weltweit angemeldeten KI-Patente den **16. Platz**.

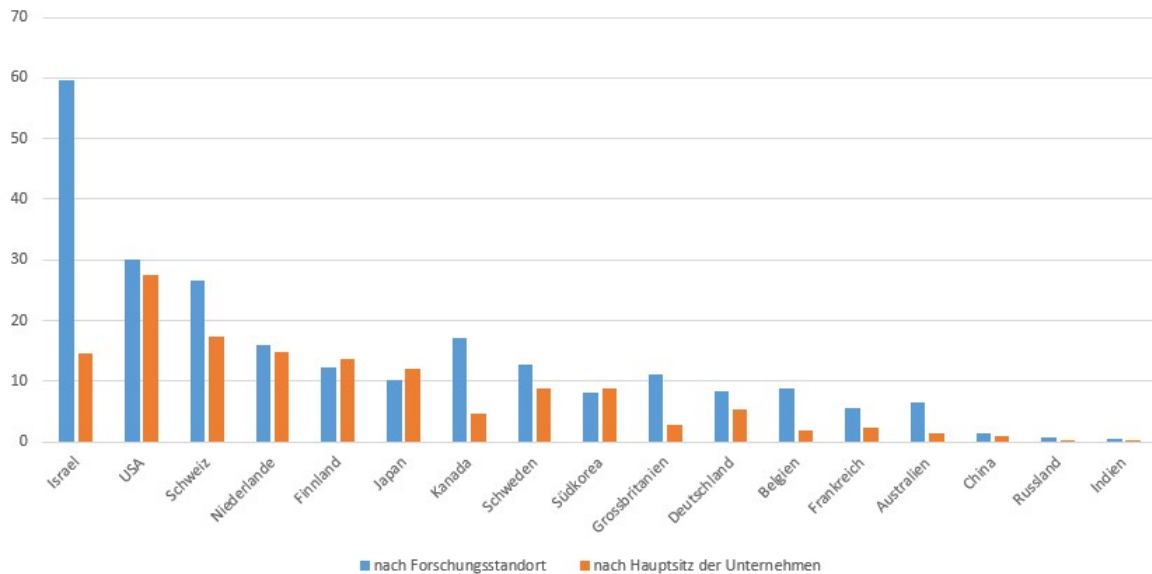
Betrachtet man die 2000 Unternehmen, die in Forschung und Entwicklung am aktivsten sind (und auf die zusammengenommen 75 Prozent aller durch die fünf grossen Patentämter geschützten Erfindungen entfallen), ist festzustellen, dass diejenigen unter ihnen mit Sitz in Japan, Südkorea, Taiwan oder China für rund 70 Prozent aller Patentanmeldungen im Bereich KI verantwortlich sind. Auf die in den USA ansässigen Unternehmen entfallen 18 Prozent (Platz 3). Die Schweiz nimmt den **13. Platz** ein. Die etwas bessere Platzierung der Schweiz im Vergleich zum vorherigen Indikator könnte darauf zurückzuführen sein, dass hierzulande viele Grossunternehmen ihren Hauptsitz haben und dass multinationale Unternehmen ihre Patentanträge oftmals von ihrem Hauptsitz aus einreichen.<sup>55</sup>

<sup>55</sup> OECD (2017): Science, Technology and Industry Scoreboard 2017 – The Digital Transformation, verfügbar unter <https://www.oecd.org/sti/oecd-science-technology-and-industry-scoreboard-20725345.htm>

Wenn man nicht die absoluten Zahlen betrachtet, sondern nach **Anzahl Patente pro Million Einwohner** vergleicht, steht die Schweiz im internationalen Vergleich deutlich besser da. Auch die Betrachtung der Qualität der Patente verändert die Position der Schweiz: So werden in China zwar enorm viele Patente eingereicht, der technische Neuerungsgehalt und die ökonomische Verwertbarkeit sind jedoch deutlich tiefer als in vielen fortgeschrittenen Volkswirtschaften, einschliesslich der Schweiz.

Die untenstehende Grafik zeigt die eingereichten Weltklassepatente<sup>56</sup> mit Bezug zu KI in Relation zur Bevölkerungsgrösse, welche entweder nach dem Forschungsstandort oder nach dem Hauptsitz der Firmen eingeteilt wurden. Die Schweiz nimmt hinter Israel und den USA den 3. Platz ein.

Abbildung 9: Anzahl Weltklassepatente pro Million Einwohner im KI-Bereich im Jahr 2018



Quelle: Darstellung SBFI auf Basis von EconSight (2019): «Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung».

Insgesamt ist in der Schweiz mit Blick auf die KI-Forschung in der Grundlagen- und angewandten Forschung damit eine **solide Basis vorhanden, um den Wissens- und Technologietransfer (WTT) zu realisieren**.

### 5.2.3 Anwendung von KI in der Schweiz

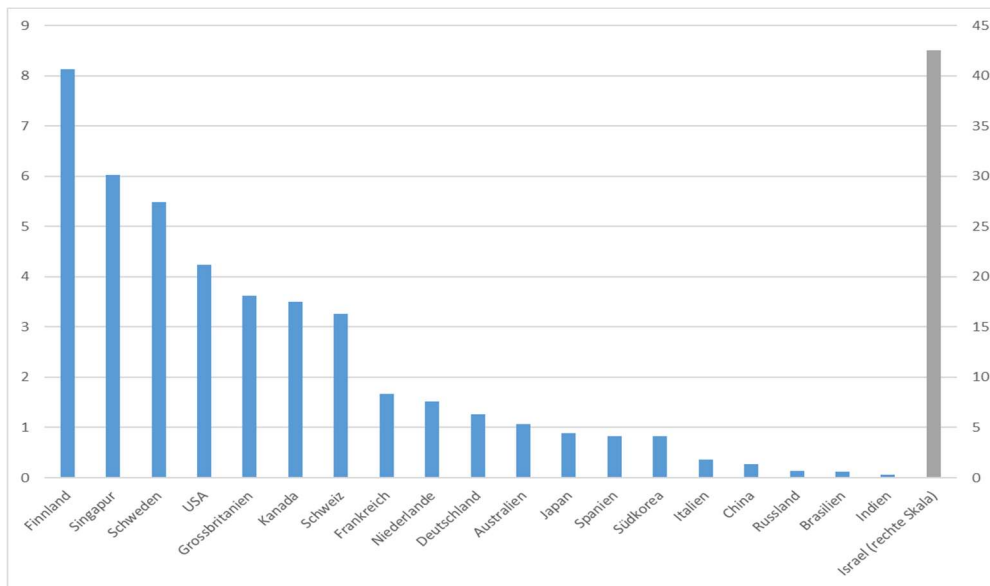
Um einen Eindruck zu erhalten, wie dynamisch die Anwendung von KI-Innovationen erfolgt, kann die Startup-Tätigkeit interessante Hinweise liefern. Es zeigt sich beispielsweise, dass die USA im internationalen Vergleich sehr viele KI-Startups hervorbringen. Damit gelingt es den USA ausgesprochen gut, Ergebnisse aus der Grundlagenforschung über den Kanal der Unternehmensgründung in die Wirtschaft zu transferieren. Gemäss einer Studie von Roland Berger sind fast 40 Prozent aller KI-Startups in den USA ansässig.<sup>57</sup>

Europa liegt hinter den USA zurück, die Schweiz befindet sich im Mittelfeld der betrachteten Länder. Wird die Bevölkerungsgrösse berücksichtigt, ist die Schweiz im Hinblick auf die Anzahl Startups jedoch in der Spitzengruppe vertreten (Abbildung 10). Jedoch wertet die Studie von Roland Berger verschiedene nationale Quellen aus, weshalb die Anzahl nur bedingt vergleichbar ist (die hohe Anzahl Israels dürfte daher die Position im internationalen Vergleich überschätzen).

<sup>56</sup> Weltklassepatente, d.h. die besten zehn Prozent der Patente in einer bestimmten Technologiesparte, gemessen an der Marktabdeckung und der Zitierung durch andere Patente.

<sup>57</sup> vgl. Roland Berger (2018): "Artificial Intelligence: A strategy for European startups", [https://www.rolandberger.com/publications/publication\\_pdf/roland\\_berger\\_ai\\_strategy\\_for\\_european\\_startups.pdf](https://www.rolandberger.com/publications/publication_pdf/roland_berger_ai_strategy_for_european_startups.pdf)

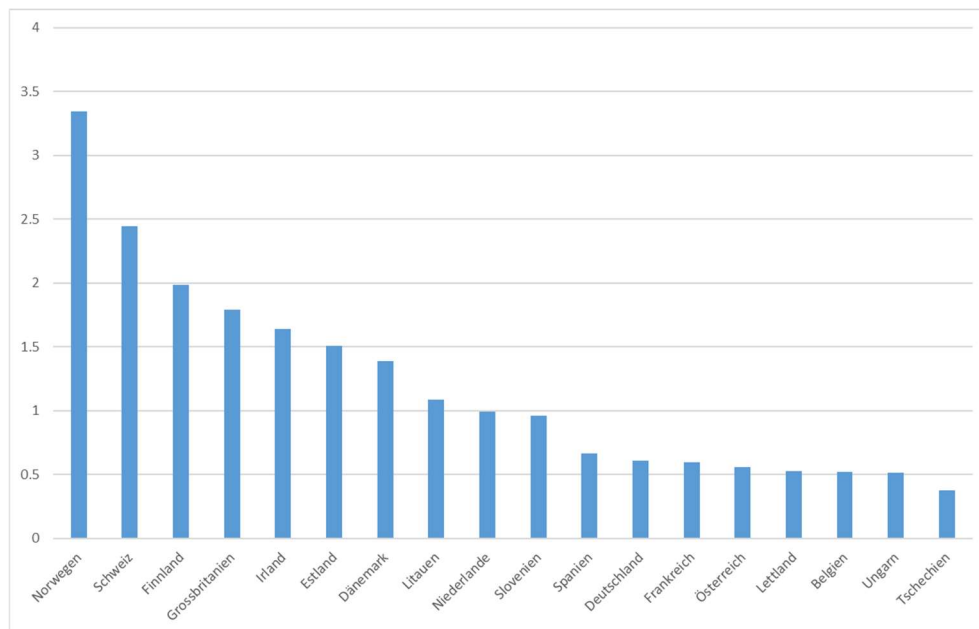
Abbildung 10: Anzahl KI-Startups pro Million Einwohner weltweit im Jahr 2018



Quelle: Berechnung SBFI auf Basis von Roland Berger (2018): «Artificial Intelligence: A strategy for European startups».

Auch bei einer weiteren Studie von ASGAR (2017), welche nur die europäischen Länder umfasst, zählt die Schweiz im Hinblick auf die Gründungstätigkeit im KI-Bereich zu den dynamischsten Ländern (Abbildung 11) und kann – sowohl absolut als auch pro Kopf – im europäischen Vergleich einen vorderen Platz einnehmen.

Abbildung 11: Anzahl KI-Startups pro Million Einwohner in Europa im Jahr 2017



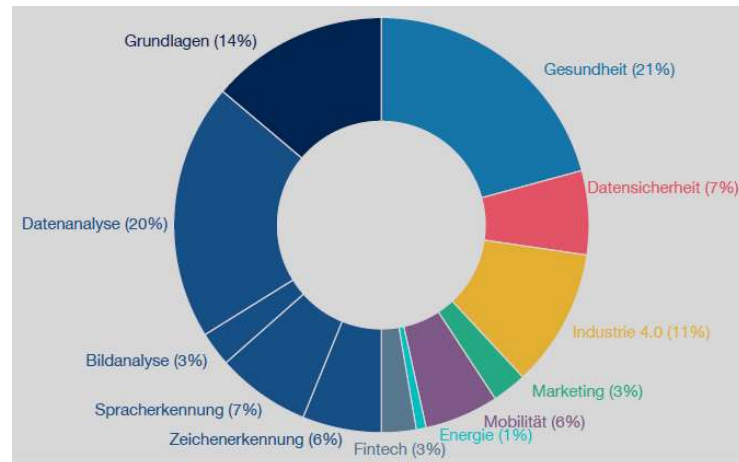
Quelle: Berechnung SBFI auf Basis von ASGAR (2017): «The European Artificial Intelligence Landscape».

KI findet Anwendung in nahezu allen Bereichen und Branchen. Dazu gehören Verkehr, Gesundheit, Sicherheit (digital und physisch), Finanzen, Marketing und Bildung. Betrachtet man jedoch die Schwerpunkte der Entwicklung, so zeigt sich gemäss einer von EconSight durchgeführten Studie, dass KI in der Schweiz überdurchschnittlich häufig im Gesundheitswesen entwickelt wird.<sup>58</sup> Die Anzahl

<sup>58</sup> Vgl. EconSight (2019): Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung, <https://www.econsight.ch/artificial-intelligence/>

der Patentanmeldungen in diesem Bereich steigt im Übrigen bereits seit rund zehn Jahren stark an. Zwei weitere **volumenstarke Bereiche im Schweizer Patent-Portfolio** sind die Datenanalyse und die Industrie 4.0. Gemessen an der Wirtschaftsstruktur sind somit einige grosse Branchen auch im Bereich KI sehr aktiv (Pharma, Medtech, Industrie), während andere grosse Branchen eher unterdurchschnittlich vertreten sind (Finanz).

Abbildung 12: KI-Struktur in der Schweiz, Subtechnologien und Anwendungsfelder im Jahr 2018



Quelle: EconSight (2019): «Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung».

### 5.3 Herausforderungen im Forschungs- und Innovationsbereich

Auch wenn sich die Schweiz, wie die vorangegangenen Abschnitte gezeigt haben, hinsichtlich der Qualität der Forschung und der Dynamik der Innovationstätigkeit im Bereich künstlicher Intelligenz in einer guten Lage befindet, bleiben die Herausforderungen gross. Neuartig am derzeitigen Strukturwandel ist die **Geschwindigkeit und Breite der Durchdringung** der Technologien der Digitalisierung in verschiedenen Sektoren. Dabei verändern Digitalisierungstechnologien insgesamt zunehmend auch bestehende Schlüsseltechnologien in anderen Bereichen – auch in Bereichen, in welchen die Schweiz führend ist. Forschung und Bildung, kommen in der Bewältigung dieser Herausforderungen eine zentrale Rolle zu. Zudem ist der Wissens- und Technologietransfer (WTT) für die rasche Umsetzung der Erkenntnisse aus der Grundlagenforschung zur Gewährleistung der Wettbewerbsfähigkeit in allen Branchen entscheidend.

Den Technologien der KI kommt innerhalb der Digitalisierung eine **besondere Rolle** zu. Sie erlauben das Automatisieren von Aufgaben, für welche bisher menschliche Fähigkeiten der Wahrnehmung, Schlussfolgerung und Interaktion nötig waren. Allerdings findet die Entwicklung und Verbreitung von KI nicht losgelöst von weiteren technologischen Entwicklungen statt; sie ist in den **allgemeinen Prozess der Digitalisierung eingebettet**. Wichtige Treiber der Digitalisierung sind neben den Fortschritten im Bereich Software bzw. KI die Entwicklungen in der Robotik, der Sensorik und der additiven Fertigung. Ausserdem fördern Entwicklungen in der Prozess- und Speichertechnik oder auch die zunehmende Vernetzung der Information den digitalen Wandel. Diese Treiber stehen in einer engen Interaktion und lassen sich nur schwer voneinander trennen. Der Bund verfolgt daher in der Forschungs- und Innovationspolitik eine **ganzheitliche Sicht** im Hinblick auf die Digitalisierung.

In dieser breiten Perspektive sind in verschiedenen Bereichen der Digitalisierung, darunter auch in der Forschung, **punktuellen Schwächen** auszumachen.<sup>59</sup> Namentlich im Hinblick auf jene Forschungska-

<sup>59</sup> vgl. SBFI (2017): «Herausforderungen der Digitalisierung für Bildung und Forschung in der Schweiz», [https://www.sbf.admin.ch/dam/sbfi/de/dokumente/webshop/2017/bericht-digitalisierung.pdf.download.pdf/bericht\\_digitalisierung\\_d.pdf](https://www.sbf.admin.ch/dam/sbfi/de/dokumente/webshop/2017/bericht-digitalisierung.pdf.download.pdf/bericht_digitalisierung_d.pdf)

pazitäten, welche notwendig sind, um die Digitalisierung in ihrer ganzen Breite auf höchstem Niveau abzudecken fällt die Schweiz innerhalb der IKT-bezogenen Forschungsfelder im internationalen Vergleich in wichtigen Forschungsbereichen ab. Es fallen darunter die Beschaffung, Verarbeitung, Speicherung, Verwaltung und Verbreitung von Informationen (zentrale Bestandteile von «Big Data») und der Einsatz digitaler Technologien in der Kommunikation von Geräten und Systemen (zentrale Aspekte des «Internets der Dinge», resp. der «Industrie 4.0»). Hier verlief die Zunahme der Forschungsleistung in der Schweiz im Vergleich zu führenden Ländern deutlich unterdurchschnittlich, so dass die Schweiz an Terrain verloren hat.

Vor diesem Hintergrund hat das WBF im Jahr 2017 die Herausforderungen der Digitalisierung für Bildung und Forschung in der Schweiz untersucht und den „Aktionsplan Digitalisierung im BFI-Bereich in den Jahren 2019 und 2020“ lanciert. Ziel des Aktionsplans war die Stärkung der Kompetenzen in Bildung und Forschung. Die Massnahmen in den verschiedenen Aktionsfeldern sind weitgehend bereits umgesetzt oder befinden sich in Umsetzung.

Der Aktionsplan hatte angesichts der hohen Geschwindigkeit der technologischen Entwicklungen den Zweck einer raschen Initiierung verschiedener Aktivitäten in spezifischen Bereichen und in enger Zusammenarbeit mit den betroffenen Akteuren. Da sich das Schweizer System, welches der Autonomie der Akteure eine grosse Bedeutung beimisst, jedoch grundsätzlich sehr gut bewährt hat, war die im Aktionsplan erfolgte Finanzierung als zeitlich befristete Anschubfinanzierung ausgestaltet worden. Die Massnahmen werden daher im Rahmen der BFI-Periode 2021–2024 konsolidiert und in Eigenregie der Akteure weitergeführt.

Zahlreiche Industrieländer, darunter Kanada und Finnland, haben KI als eine entscheidende Zukunftstechnologie identifiziert. Die **nationalen Strategien zur Förderung von KI unterscheiden sich jedoch stark**.<sup>60</sup> Es zeigt sich, dass jedes Land ganz eigene Schwerpunkte legt, um KI im privaten und öffentlichen Sektor voranzutreiben. So will China bis 2030 führend im Bereich der KI werden. Mit der Bereitstellung von beachtlichen finanziellen Ressourcen für Forschung und Entwicklung beabsichtigt das Land den Abstand zum bisherigen KI-Marktführer USA zu verringern. Demgegenüber setzt Finnland den Fokus ganz auf KI-Anwendung und auf Kooperation zwischen Wirtschaft, Wissenschaft und Verwaltung.

Die Schweiz verfügt über eine Innovationspolitik, welche nicht im Rahmen einer einzigen umfassenden «Innovationsstrategie», sondern dezentral und innerhalb mehrerer eigenständiger, nach Bedarf aber sachlich koordinierter Politikbereiche aufgegliedert ist. Diese Organisation gesteht den einzelnen Akteuren einen hohen autonomen Handlungsspielraum zu, ermöglicht andererseits gerade damit aber auch unterschiedliche, passgenaue Antworten auf neue Herausforderungen und Chancen – wie das Thema der Digitalisierung zeigt.<sup>61</sup> Dieser Ansatz gilt grundsätzlich auch für KI als Teilbereich der Digitalisierung.

Obwohl die Schweiz über exzellente Forschungseinrichtungen verfügt, substanzielle Mittel im Bereich Digitalisierung einsetzt und gute Instrumente zur Förderung von F&E in digitalen Technologien zur Verfügung stehen, muss der Kompetenzaufbau im BFI-Bereich angesichts der hohen technologischen Dynamik und wachsender Konkurrenz weiter gestärkt werden. Hierzu sind in der Schweiz die KI-relevanten Aktivitäten im BFI-Bereich zu intensivieren. Dies geschieht in erster Linie in eigener Regie durch die öffentlichen und privaten Akteure, aber auch im Rahmen der Nutzung der etablierten Förderinstrumente des Bundes für den Kompetenzaufbau und den WTT.

Im Rahmen der bestehenden Digitalisierungsgefässe und –organe ist somit innerhalb der BFI-Botschaft verstärkt zu gewährleisten, dass die Akteure in Wissenschaft, Bildung und WTT auf die Herausforderungen der Digitalisierung vorbereitet sind und sich den Herausforderungen der künstlichen Intelligenz im Rahmen ihrer jeweiligen Digitalisierungsaktivitäten annehmen.

<sup>60</sup> OECD (2019): «Artificial Intelligence in Society».

<sup>61</sup> Siehe hierzu auch Bundesrat (2018): «Gesamtschau der Innovationspolitik», Bericht in Beantwortung des Postulats Derder 13.3073 vom 13. März 2013.



## 6 KI-Themenfelder nach Politikbereichen

Der Bundesrat beabsichtigte mit dem Auftrag zur Einsetzung der interdepartementalen Arbeitsgruppe künstliche Intelligenz (IDAG KI) in erster Linie die Sicherstellung des Wissens- und Meinungsaustausches und die Abstimmung von Positionen der Schweiz in internationalen Gremien. Entsprechend waren sämtliche Departemente in der vom WBF (SBFI) eingesetzten Arbeitsgruppe vertreten.

Im Rahmen des Mandats hat die IDAG KI eine Auslegeordnung über KI-bezogene Herausforderungen mit Relevanz für den Bund vorgenommen. Diesbezüglich wurden 17 relevante Themenbereiche identifiziert, die aus Sicht des Bundes prioritär zu betrachten sind.

Die Themenbereiche wurden in eigener Verantwortung des jeweils zuständigen Fachamtes bearbeitet. Da sich die Herausforderungen der Anwendung von KI je nach Themenbereich sehr unterschiedlich stellen, bestand die Notwendigkeit für eine breite Konsultation in unterschiedlichem Ausmass. Während einige Themenbereiche seitens der Ämter bereits ausreichend adressiert waren, mussten für andere Themen breit abgestützte Projektgruppen eingesetzt werden.

Insgesamt wurden sieben grössere, departementsübergreifende Arbeitsgruppen eingesetzt sowie zahlreiche externe Stakeholder und Experten aus Wissenschaft und Wirtschaft konsultiert. Gemäss bundesrätlichem Auftrag wurden schwerpunktmässig auch Überlegungen zu einem transparenten und verantwortungsvollen Einsatz von künstlicher Intelligenz in die Arbeiten integriert.

Die Ergebnisse der Arbeiten der zuständigen Fachämter, resp. der Arbeitsgruppen werden im Folgenden zusammenfasst. Dabei folgen alle Abschnitte einer einheitlichen Struktur:

- Ein erster Abschnitt konkretisiert zunächst die **Relevanz** der Anwendung von KI im jeweiligen Themengebiet.
- In einem zweiten Abschnitt werden die **spezifischen Herausforderungen** bzgl. KI aufgezeigt, schwerpunktmässig im Kompetenzbereich des Bundes.
- Die Darstellung der laufenden Aktivitäten versucht anschliessend aufzuzeigen, welche **Massnahmen** der Bund oder relevante bundesexterne Akteure bereits initiiert haben, um die Herausforderungen zu adressieren.
- Schliesslich erfolgt eine **Bewertung**, ob die identifizierten Herausforderungen im jeweiligen Themenfeld durch die laufenden Aktivitäten (oder im Rahmen bestehender Regelungen) bereits ausreichend adressiert sind. Dort, wo dies nicht der Fall ist, erfolgt schliesslich eine erste Auslegeordnung, welche **weiteren Schritte** allenfalls erforderlich sind, um die Herausforderungen rechtzeitig anzugehen.

## 6.1 Internationale Gremien und künstliche Intelligenz<sup>62</sup>

### 6.1.1 Übersicht

Die aktuellen Entwicklungen und Herausforderungen der KI sind zum grossen Teil globaler Natur. KI ist daher zu einem zentralen und vieldiskutierten Thema auf der internationalen Agenda geworden. Das Interesse an internationaler Kooperation beruht einerseits auf der Notwendigkeit, Ressourcen für Forschung und Entwicklung zu bündeln und den Zugang zu grossen Datenquellen («Big Data») zu sichern. Andererseits ist die digitalisierte Welt zunehmend mobil und Daten, Produkte und Dienstleistungen überschreiten die Landesgrenzen. Es mehren sich darum die Rufe nach ethischen Prinzipien und internationalen Normen, um das positive Potenzial von KI bestmöglich zu nutzen und gleichzeitig die Risiken zu erkennen und zu bewältigen. Immer mehr internationale Organisationen (z.B. UNO, UNESCO, OECD, Europarat und die EU) sowie auch technische Gremien wie die IEEE Standards Association befassen sich mit Fragen rund um den Einsatz von KI. Diskutiert wird insbesondere, welche Rahmenbedingungen und Gouvernanz-Modelle auf internationaler Ebene geschaffen werden müssten, damit KI für die Menschen transparent und nachvollziehbar ist, ethische Grundwerte respektiert und Verantwortlichkeiten geklärt sind und der Strukturwandel für unsere Gesellschaften verkräftbar wird. Dabei soll auf bestehende Regeln und Standards – z.B. in den Bereichen Menschenrechte, Datenschutz und verantwortungsvoller Unternehmensführung – aufgebaut werden.

### 6.1.2 Herausforderungen

Traditionelle globale multilaterale Regulierungsprozesse sind oft schwerfällig und können den rasanten Entwicklungen im Bereich KI kaum gerecht werden. Aufgrund von unterschiedlichen Auffassungen der Mitgliedsländer bezüglich der Rolle der Staaten sind viele UNO-Institutionen in Fragen der globalen digitalen Gouvernanz oft blockiert. Zudem verfügen einige wenige globale Technologieunternehmen über gewaltige Daten- und finanzielle Ressourcen und sind daran, auch im KI-Bereich ihre Marktmacht auszubauen. Der Handlungsspielraum der nationalen Regierungen und der internationalen Organisationen wird zunehmend in Frage gestellt. Für kleine Staaten wie die Schweiz wird es in diesem Kontext schwierig, eigenständige nationale Regelungen international durchzusetzen. Die Bedeutung der internationalen Kooperation nimmt entsprechend zu.

In den letzten Jahren haben einige im Bereich KI führende Unternehmen für sich selbst Prinzipien im Umgang mit KI erarbeitet. In den internationalen Debatten mehren sich jedoch die Stimmen, dass Selbstregulierung durch die Industrie nicht (mehr) ausreicht, um einen transparenten, nachvollziehbaren und verantwortungsvollen Einsatz von KI zu gewährleisten. Zudem leiden die von transnational operierenden Technologieunternehmen geführten Prozesse ethischer Standardsetzung an einem Demokratiedefizit. Vor diesem Hintergrund steigt das Interesse an neuen, «smarten» Regulierungsansätzen. Gefragt werden mehr und mehr dynamische, flexible, interdisziplinäre und dezentral ausgerichtete Gouvernanz-Strukturen sowie die Schaffung demokratisch legitimierter internationaler Standards. Aus Sicht der Schweiz ist es dabei insbesondere zentral, alle relevanten Anspruchsgruppen – nebst den Staaten auch die Privatwirtschaft, die Zivilgesellschaft und die technischen Experten – aus aller Welt in die politischen Entscheidungsprozesse einzubeziehen und bei deren Umsetzung wirksam in die Pflicht nehmen zu können.

Wie genau KI auf internationaler Ebene reguliert werden soll, wird kontrovers diskutiert. Während die EU und einige ihrer Mitgliedstaaten – auch vor dem Hintergrund der Europäischen Datenschutzgrundverordnung – für eine rechtsverbindliche Regulierung plädieren, machen andere auf die Gefahren von zu viel oder zu strikter Regulierung für das Innovationspotenzial von KI sowie auch für die Meinungs- und Medienfreiheit aufmerksam. Einige internationale Organisationen (z.B. Europarat, OECD und EU) haben erste (rechtlich unverbindliche) Normen zu KI erarbeitet (siehe Details im ausführlichen Bericht der Projektgruppe). Ob es in kurz- oder mittelfristiger Zukunft zu einer internationalen Einigung auf verbindliche Normen, die über die Ebene von Grundprinzipien hinausgehen kommen wird, ist derzeit noch genauso offen wie die Frage, welche Institutionen und Prozesse langfristig für die Gouvernanz von KI entscheidend sein werden.

<sup>62</sup> Für eine ausführliche Fassung siehe Bericht der Projektgruppe «Internationale Gremien und künstliche Intelligenz», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

### 6.1.3 Bestehende Aktivitäten

Für die Schweiz als kleines, hochentwickeltes und vernetztes Land ist es zentral, die Debatte um die globale Gouvernanz von KI aktiv mitzugestalten. Die Schweiz ist daher in den relevanten Gremien und Prozessen aktiv engagiert. Dies betrifft einerseits etablierte Organisationen wie der UNO, OECD, ITU, UNESCO, EU und des Europarates, andererseits aber auch von kleinere, neuere Gruppen und Think Tanks, welche substantiell gute Arbeiten leisten. So hat sich die Schweiz unter anderem für die Lancierung des Hochrangigen Panels des UNO-Generalsekretärs zur Digitalen Kooperation, welches auch das Thema KI einschliesst, eingesetzt, dessen Arbeit unterstützt und die Ergebnisse wesentlich mitgeprägt.<sup>63</sup> Inhaltlich bringt sich die Schweiz mit ihren freiheitlichen, rechtsstaatlichen und demokratischen Werten sowie ihrem Fachwissen in den Diskurs ein und engagiert sich insb. dafür, dass beim Einsatz von KI grundlegende und etablierte Werte und Normen wie die Menschenrechte respektiert und alle relevanten Anspruchsgruppen einbezogen werden. Gleichzeitig plädiert die Schweiz für einen ausbalancierten Regulierungsansatz, der Innovation ermöglicht, und setzt sich dafür ein, dass den Verschiedenheiten der Policy-Ansätze der Staaten so weit wie möglich Rechnung getragen wird. Dabei vertraut die Schweiz in ihrer eigenen Forschungs- und Innovationspolitik stark auf ihr bewährtes partizipatives System, das sich weitgehend darauf fokussiert, gute Rahmenbedingungen zu setzen, die den Akteuren in Forschung und Wirtschaft Freiräume lassen, eigene Lösungen und Themenschwerpunkte zu finden. Die Schweiz verfügt mit dem Internationalen Genf über einen Standort, der viele Voraussetzungen erfüllt, zu einem Zentrum für globale Gouvernanz von KI zu werden. Genf steht jedoch in Konkurrenz zu anderen Metropolen, die zum Teil über deutlich mehr Ressourcen verfügen.

### 6.1.4 Bewertung und Handlungsbedarf

Das Engagement und die grundsätzliche Positionierung der Schweiz in internationalen Gremien zu KI gilt es weiterzuführen. Gleichzeitig besteht Potenzial, dass sich die Schweiz in einigen Bereichen künftig noch stärker engagieren und Genf als Zentrum der globalen KI-Gouvernanz etablieren könnte.

<p><b>Aktionsfeld 1: Informations- und Wissensaustausch und Koordination der Positionen des Bundes in internationalen Gremien</b></p> <p>Die Weiterentwicklung von KI geschieht in einem globalisierten Umfeld. Dabei stellen sich Gouvernanzfragen einerseits sektorspezifisch unterschiedlich, andererseits verstärken sich die Interdependenzen zwischen vormals getrennten Politikbereichen. Diese lassen sich auf nationaler Ebene nur begrenzt steuern und bedingen eine stärkere interdisziplinäre Vernetzung, einen Informations- und Wissensaustausch und einen Dialog mit allen Anspruchsgruppen, national und international.</p>	
<p>Nutzung der «Plateforme Tripartite» als interdisziplinäres nationales Kompetenznetzwerk zu KI-Themen sowie für die Koordination der Positionen des Bundes in internationalen Gremien im Bereich KI</p>	<p>Um den Dialog und Informations- und Wissensaustausch zu politischen, gesellschaftlichen, wirtschaftlichen und anderen relevanten Aspekten von KI sicherzustellen, soll ab sofort die vom BAKOM für die Vorbereitung des UNO Weltgipfels zur Informationsgesellschaft (WSIS) ins Leben gerufene, bereits bestehende «Plateforme Tripartite» genutzt werden. Die Plattform steht allen interessierten Organisationen und Personen offen und verfügt über einen administrativen Ausschuss aus Vertretern der Bundesverwaltung, welcher bei Bedarf Positionen des Bundes in internationalen Gremien koordinieren kann. Die «Plateforme Tripartite» kann als interdisziplinäres nationales Kompetenznetzwerk zu KI-Themen genutzt werden, welches in der Lage ist, Wissen und Erfahrungen auch horizontal zu vernetzen und auf diese Weise kohärente Positionen der Schweiz auf internationaler Ebene zu entwickeln. Die Tabelle der Schweizer Vertretungen in KI-relevanten Gremien wird regelmässig aktualisiert.</p> <p><b>Zuständigkeit:</b> BAKOM                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

<sup>63</sup> Für eine vollständige Übersicht der Vertretungen der Schweiz in KI-relevanten internationalen Gremien sei auf den Text und die Tabelle im ergänzenden Bericht verwiesen.

<b>Aktionsfeld 2: Globale Gouvernanz</b> Es zeigen sich Lücken im Gouvernanz-System für die globale digitale Welt und im Bereich KI.	
1) Stärkung der globalen Gouvernanz	Die globale Gouvernanz ist zu stärken, indem die Schweiz die Entwicklung neuer Gouvernanz-Modelle, Prozesse und Strukturen aktiv unterstützt und fördert und die Umsetzung der Empfehlungen des Hochrangigen Panels des UNO-Generalsekretärs zur digitalen Kooperation mitgestaltet.  <b>Zuständigkeit:</b> BAKOM, EDA <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
2) Integration KI in Aussenpolitische Strategie 2020 - 2023	Neue Technologien wie KI haben auch aussenpolitische Implikationen. Ein verstärktes Engagement in der globalen Gouvernanz von KI ist im Rahmen der Erarbeitung der aussenpolitischen Strategie 2020-2023 zu diskutieren.  <b>Zuständigkeit:</b> EDA, BAKOM <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<b>Aktionsfeld 3: Internationales Genf</b> Genf verfügt über ideale Voraussetzungen, um sich als Zentrum für globale Gouvernanzfragen bzgl. KI zu etablieren, steht jedoch in Konkurrenz zu anderen, zum Teil ressourcenstärkeren Metropolen.	
1) Bessere Vernetzung und Zusammenarbeit der KI-relevanten Akteure	Die Schweiz verstärkt ihre Anstrengungen zur Förderung einer besseren Vernetzung und Zusammenarbeit der KI-relevanten Akteure und zur (Weiter-)Entwicklung von globalen Gouvernanzstrukturen für KI in Genf.  <b>Zuständigkeit:</b> BAKOM, EDA <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
2) Prüfung zur verstärkten Zusammenarbeit für «AI for Good Summit»	BAKOM, EDA und die Genfer Behörden erörtern mit der ITU Möglichkeiten der Zusammenarbeit zur strategischen Weiterentwicklung von deren «AI for Good Summit».  <b>Zuständigkeit:</b> BAKOM, EDA <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
3) Stärkung der Geneva Internet Platform	Das Potenzial der Geneva Internet Platform (GIP) wird noch besser genutzt und als Instrument im Dienste der Schweiz weiter gestärkt.  <b>Zuständigkeit:</b> EDA, BAKOM <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
4) Stärkung Internationales Genf als Zentrum für digitale Gouvernanz in Aussenpolitische Strategie 2020 - 2023	Ein verstärktes Engagement zur Stärkung des internationalen Genfs als Zentrum für digitale Gouvernanz inkl. KI ist im Rahmen der Erarbeitung der neuen aussenpolitischen Strategie 2020-2023 zu diskutieren.  <b>Zuständigkeit:</b> EDA, BAKOM <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## 6.2 Digital Europe Programme

### 6.2.1 Übersicht

Das Digital Europe Programme (DEP) ist ein neues Förderprogramm der EU, welches ab 2021 offiziell mit einer Laufzeit bis Ende 2027 startet. Das DEP soll aus fünf Pfeilern bestehen **i)** Hochleistungsrechnen **ii)** Künstliche Intelligenz **iii)** Cybersicherheit und Vertrauen **iv)** Fortgeschrittene digitale Kompetenzen **v)** Einführung, optimale Nutzung digitaler Kapazitäten und Interoperabilität.

An dieser Stelle wird auf den zweiten Pfeiler des Digital Europe Programmes (DEP) eingegangen, da im vorliegenden Bericht lediglich das Thema künstliche Intelligenz behandelt wird. Ziel des zweiten Pfeilers des DEP ist es, Kernkapazitäten im Bereich der KI, darunter Datenressourcen und Algorithmenbibliotheken, aufzubauen, sie für alle Unternehmen und öffentlichen Verwaltungen zugänglich zu machen und die Vernetzung zwischen bestehenden KI-Erprobungs- und Versuchseinrichtungen in den europäischen Mitgliedstaaten zu fördern. So müssen für die Entwicklung von KI-Technik grosse, qualitativ hochstehende Datensätze verfügbar sein. Die EC schlägt vor, dass die EU in den nächsten Jahren gemeinsam mit den Mitgliedstaaten und dem Privatsektor bis zu 1 Mrd. EUR in die Schaffung eines gemeinsamen europäischen Datenraums investiert, der Innovatoren, Unternehmen und dem öffentlichen Sektor Daten für die Weiterverwendung leicht zugänglich macht.<sup>64</sup> Weiter soll zur Förderung der Zusammenarbeit zwischen Wissenschaft und Industrie in Europa und zur Festlegung einer gemeinsamen strategischen Forschungsagenda im Bereich der KI langfristig eine neue öffentlich-private Forschungs- und Innovationspartnerschaft eingerichtet werden. Ein weiterer wichtiger Schritt bei der Markteinführung modernster Technologien besteht in deren Erprobung und Prüfung unter realen Bedingungen. Konkret bedeutet dies, dass im Rahmen des DEP rund 1,5 Mrd. EUR für die Einrichtung weltweit führender Test- und Versuchsstandorte für KI-gestützte Produkte und Dienstleistungen in ganz Europa bereitgestellt werden sollen.

### 6.2.2 Herausforderungen

Für die Schweiz stellt sich die Frage nach einer möglichen Beteiligung am gesamten Digital Europe Programme oder an Teilen davon, sowie nach der Einbettung in die nationale Forschungs- und Innovationslandschaft. Zu beachten ist, dass das DEP über Komplementaritäten und Synergien mit einer Reihe weiterer vorgeschlagener Förderinitiativen der EU im Zeitraum 2021-2027, insbesondere mit dem EU-Forschungsrahmenprogramm «Horizon Europe» verfügt. Die Schweiz beteiligt sich seit 1987 an Projekten des EU-Forschungsrahmenprogramms (FRP), seit 2004 im Rahmen der Bilateralen Verträge I als assoziierter Staat. Eine Fortführung der Beteiligung ab 2021 ist vorgesehen, nachdem die aktuelle Generation (8. FRP resp. «Horizon 2020») Ende 2020 ausläuft. Eine mögliche Beteiligung am Digital Europe Programme wird aus den vorgenannten Gründen der engen Verbindung zum Rahmenprogramm der EU für Forschung und Innovation «Horizon Europe» in der Botschaft zur Beteiligung der Schweiz am Horizon Europe-Paket behandelt.

Neue Technologien im ICT-Bereich sollen weiterhin im FRP *erforscht und entwickelt* werden. Die *marktnahe Forschung, Innovation und Umsetzung in marktfähige Produkte* sowie die Verbreitung und Akzeptanz strategischer digitaler Kapazitäten und Infrastrukturen in Bereichen von öffentlichem Interesse und in der Privatwirtschaft soll jedoch zukünftig im Digital Europe Programme erfolgen. Aus diesem Grund muss sich die Schweiz überlegen, ob sie die in Horizon Europe geleisteten Forschungs- und Innovationsarbeiten und die dadurch gewonnenen Erkenntnisse nicht zusammen mit den anderen EU-Staaten weiter zu marktfähigen Produkten entwickeln möchte. Dies würde durch eine Beteiligung am Digital Europe Programme erfolgen, welches Nicht-EU-Mitgliedstaaten wie der Schweiz für eine Beteiligung offensteht.

Weitere Herausforderungen bestehen vor allem darin, dass die Verhandlungen der zuständigen Institutionen auf EU-Ebene zur Ausgestaltung des Digital Europe Programmes zu den Teilnahmebedingungen und zum Budget noch laufen und aufgrund der Wahlen des Europäischen Parlaments einerseits und des Brexit andererseits verzögert werden. Wie, ob und in welcher Form sich die Schweiz

<sup>64</sup> Europäische Kommission (2018): «Künstliche Intelligenz für Europa», <https://ec.europa.eu/transparency/regdoc/rep/1/2018/DE/COM-2018-237-F1-DE-MAIN-PART-1.PDF>

allenfalls am DEP beteiligen kann, ist derzeit daher noch nicht vollständig klar. Dasselbe gilt auch für Horizon Europe.

### 6.2.3 Bestehende Aktivitäten

Die Schweiz und die Europäische Kommission (EC) prüfen derzeit die Unterzeichnung der Europäischen KI-Deklaration, welche bereits von 24 europäischen Staaten im April 2018 unterzeichnet wurde. Die Deklaration umreist die wichtigsten Bereiche der KI, in welchen die Unterzeichnerstaaten zusammenarbeiten wollen, um die Wettbewerbsfähigkeit Europas in der Erforschung und Nutzung der KI zu sichern und soziale, wirtschaftliche, ethische und rechtliche Fragen anzugehen. Die Vorarbeiten für das DEP im Bereich KI haben auf europäischer Ebene bereits begonnen. Die europäische KI-Strategie basiert auf den wissenschaftlichen und industriellen Stärken Europas und sieht eine ethische, sichere und hochmoderne KI vor, die den Menschen in den Mittelpunkt der Entwicklung stellt. Im Dezember 2018 hat die EC einen koordinierten Aktionsplan für KI vorgelegt, der aus mehreren konkreten Massnahmen besteht, die z.T. bereits im Rahmen des aktuellen Forschungsrahmenprogramms Horizon 2020 umzusetzen sind. Die Schweiz hat bei der Erarbeitung dieses Aktionsplans mitgewirkt. Im Juni 2018 hatte die EC eine hochrangige Expertengruppe zu KI eingesetzt und mit der Erarbeitung von ethischen Richtlinien sowie Politik- und Investmentempfehlungen zu KI betraut. Basierend auf diesen Arbeiten wurde im April 2019 nun eine Pilotphase gestartet, um zu überprüfen, ob die im Konsens erarbeiteten Richtlinien in der Praxis auch umsetzbar sind.

#### Schweizer Interessenlage

Das SBFI hat unter Konsultation von anderen Bundesstellen<sup>65</sup> und der ZHAW im Februar 2019 eine Umfrage bei Schweizer Stakeholdern durchgeführt, um deren Positionen zum DEP und ihr Interesse an einer allfälligen Teilnahme der Schweiz am DEP zu sondieren. An der Online-Befragung teilgenommen haben insgesamt 150 Institutionen. Die vollständigen Ergebnisse zu dieser Umfrage veröffentlichte das SBFI in einem separaten Bericht.<sup>66</sup> Wichtig zu erwähnen ist, dass es sich bei dieser Umfrage lediglich um die Interessenabwägung von potentiellen Teilnehmer in der Schweiz handelt. Für eine allfällige Teilnahme der Schweiz am DEP müssen separat davon als nächster Schritt die genauen Modalitäten einer Teilnahme (Vollasoziiierung, Teilasoziiierung, Drittstaat) auf technischer Ebene abgeklärt werden. Eine allfällige Teilnahme am DEP hängt weiter vom europapolitischen Gesamtkontext, den Verhandlungsergebnissen mit der EU und von der Finanzlage der Schweiz ab.

Für den zweiten DEP-Pfeiler sieht die Interessenlage der Schweizer Stakeholder wie folgt aus: 84 der insgesamt 150 befragten Institutionen und Experten haben angegeben, dass sie an einer allfälligen Beteiligung am Pfeiler «künstliche Intelligenz» interessiert wären. Von diesen 84 Institutionen sind 61 aus dem öffentlichen Sektor und 23 aus dem privaten Sektor. Dies zeigt auf, dass das Thema sowohl für den öffentlichen wie auch den privaten Sektor von grosser Relevanz ist. Auffallend ist auch, dass sich sowohl kleinere als auch grössere Institutionen und Organisationen für KI interessieren. Insgesamt haben über 75%<sup>67</sup> der an Pfeiler II interessierten Institutionen oder Forschenden bestätigt, dass sie sich an Aktivitäten zu KI beteiligen würden. Die Bereitschaft, sich tatsächlich an den Forschungsaktivitäten im KI-Bereich auf europäischer Ebene zu beteiligen, ist vor allem im öffentlichen Sektor sehr gross mit quasi 100%, der private Sektor ist hingegen etwas zurückhaltender. Eine nationale Direktfinanzierung («projektweise Beteiligung») durch das SBFI oder eine andere Instanz, anstelle einer Förderung durch die EU, finden nur 66% erstrebenswert. Aufschlussreich ist auch, dass die befragten Institutionen bereit wären, durchschnittlich ca. 35% der Projektkosten selber zu tragen.

<sup>65</sup> SECO, BAKOM, armasuisse und DEA.

<sup>66</sup> SBFI (2019): «Ergebnisse der Umfrage zum Digital Europe Programme (DEP)», [https://www.sbf.admin.ch/dam/sbfi/de/dokumente/2019/07/ergebnisse-dep.pdf.download.pdf/bericht\\_dep\\_d.pdf](https://www.sbf.admin.ch/dam/sbfi/de/dokumente/2019/07/ergebnisse-dep.pdf.download.pdf/bericht_dep_d.pdf)

<sup>67</sup> Im Bericht zur Umfrage wurden jeweils zwei Werte berechnet. Ein ungewichteter Wert, bei welchem die Anzahl Antworten angeschaut wurde und ein gewichteter Wert, bei welchem die von den Antworten abgedeckten Anzahl Personen miteinbezogen wurde.

### 6.2.4 Bewertung und Handlungsbedarf

Da auf europäischer Ebene unter Horizon 2020 bereits eine breitangelegte Initiative im Bereich künstliche Intelligenz läuft, sollten auf nationaler Ebene Doppelspurigkeiten nach Möglichkeit vermieden werden. So befürworten beispielsweise die befragten Stakeholder die Beteiligung der Schweiz an den meisten der vorgeschlagenen Handlungsfelder im Rahmen des europäischen koordinierten Aktionsplans zur KI. Auch wenn die Teilnahmebedingungen für die Schweiz noch nicht restlos geklärt sind, erwarten die Stakeholder eine Beteiligung an Horizon Europe und aufgrund der Umfrageergebnisse scheint eine Beteiligung am zweiten Pfeiler des DEP von den nationalen Stakeholdern gewünscht zu werden. Daher sollten nationale Forschungsinitiativen im Bereich KI komplementär zu den europäischen Initiativen ausgestaltet werden. Dies vor allem im Hinblick auf eine effiziente Nutzung der finanziellen und personellen Ressourcen und mit dem Ziel eines daraus resultierenden maximalen Outputs. Das SBFI wird dies als federführendes Amt in geeigneter Form und in Zusammenarbeit mit DEA und EFV bei der Vorbereitung der nächsten Schritte berücksichtigen.

<p><b>Aktionsfeld 1: Schweizer Beteiligung an Horizon Europe und am Digital Europe Programme</b>                  KI ist ein Kernthema des nächsten Rahmenprogramms für Forschung und Innovation «Horizon Europe» und des «Digital Europe Programme».</p>	
<p>Prüfung Beteiligung «Horizon Europe» und «Digital Europe Programme»</p>	<p>Das SBFI klärt unter Einbezug der weiteren betroffenen Bundesämter (namentlich DEA) ab, ob und wo eine Kooperation mit der EU sowie eine Mitwirkung bei ihren Aktivitäten im Bereich KI angestrebt werden soll.</p> <p><b>Zuständigkeit:</b> SBFI u.a.      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen und Botschaften</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

## 6.3 Veränderungen der Arbeitswelt

### 6.3.1 Übersicht

Wie in Kapitel 2.3 beschrieben, wird der künstlichen Intelligenz (KI) das Potenzial einer Basistechnologie zugesprochen, die – vergleichbar mit der Dampfmaschine oder der Elektrifizierung – ganze Branchen und Volkswirtschaften durchdringen und umgestalten kann. Es ist unbestritten, dass dieser technologische Wandel wichtige Impulse für die wirtschaftliche Entwicklung liefert. Gleichzeitig wird KI als Quelle neuer Risiken wahrgenommen. Ein wiederkehrendes Thema in der Diskussion sind mögliche negative Auswirkungen auf die Beschäftigung. Verschiedentlich wird die Befürchtung geäußert, dass KI dereinst in der Lage sein wird, die menschliche Intelligenz zu ersetzen und auf diese Weise praktisch alle Tätigkeiten automatisierbar zu machen. Welche Chancen und Risiken ergeben sich für den Schweizer Arbeitsmarkt aus den Entwicklungen im Bereich der KI? Welches sind die Erfolgsfaktoren für eine erfolgreiche Anpassung an den technologischen Wandel?

### 6.3.2 Herausforderungen

Im Hinblick auf die Entwicklungen am Arbeitsmarkt findet die Entwicklung und Verbreitung von KI nicht losgelöst von anderen Treibern des Wandels statt. Es gibt etwa derzeit keine Hinweise, dass KI den Arbeitsmarkt fundamental in anderer Weise verändern wird, als dies bisherige technologische Entwicklungen und namentlich die Technologien der Digitalisierung getan haben.

Der Einsatz von neuen Technologien beeinflusst die gesamtwirtschaftliche Beschäftigungsentwicklung über verschiedene Wirkungskanäle. Technologischer Fortschritt generell und die Entwicklung digitaler Technologien ermöglichen Produktivitätssteigerungen und Kosteneinsparungen. Lohnt sich die Automatisierung von Arbeitsschritten aus betriebswirtschaftlicher Sicht, werden diese mittel- bis längerfristig durch die neue Automatisierungstechnologie substituiert. Laufen diese Substitutionsprozesse sehr schnell ab oder passen sich die auf dem Arbeitsmarkt angebotenen Kompetenzprofile nicht oder nur langsam an die Arbeitsnachfrage an, sind auf gesamtwirtschaftlicher Ebene Qualifikationsmismatches und technologisch bedingte Arbeitslosigkeit mögliche Folgen.

Neben Substitutionseffekten üben Automatisierungstechnologien eine komplementäre Wirkung auf bestehende Stellen aus. Komplementäre Technologien werten das Tätigkeitsportfolio von Arbeitskräften auf und steigern die Produktivität der Arbeit, was sich wiederum in steigenden Löhnen widerspiegeln kann. KI eröffnet das Potenzial, dass vermehrt auch Berufe mit hohen Anforderungen an Urteilsvermögen und kritisches Denken von dieser komplementären Wirkung profitieren können.

Zudem wirkt technologischer Fortschritt als Antrieb für die gesamtwirtschaftliche Nachfrage. Produktivitätssteigerungen und tiefere Produktionskosten führen in der Regel zu sinkenden Produktpreisen, was wiederum das verfügbare Realeinkommen der Konsumenten erhöht. Eine steigende Nachfrage führt zu Produktionserhöhungen und einem höheren Bedarf an Arbeitskräften. Die Bezifferung dieses Gesamtnachfrageeffekts ist schwierig, da das Ausmass massgeblich von Faktoren wie der Preissensitivität der Nachfrage oder der Konsumneigung der Haushalte abhängt. Schliesslich geht auch von der Entwicklung neuer Technologien und Anwendungen selbst ein beschäftigungssteigernder Effekt aus.

Gemäss aktuellen Schätzungen werden in den OECD-Ländern 14 Prozent der Stellen als stark automatisierungsgefährdet eingeschätzt.<sup>68</sup> Die Wahrscheinlichkeit einer negativen Beschäftigungsentwicklung auf aggregierter Ebene stuft die OECD dabei als gering ein. Herausforderungen sieht die OECD derzeit vor allem auch darin, dass sich aufgrund des Strukturwandels bei weiteren 32 Prozent der Stellen die Tätigkeitsstruktur in den kommenden Jahrzehnten massgeblich verändern könnte.

### 6.3.3 Bestehende Aktivitäten

Der Bundesrat hat sich in der jüngeren Vergangenheit vertieft mit Fragen der Auswirkungen der Digitalisierung auf die Schweizer Volkswirtschaft bzw. den Schweizer Arbeitsmarkt auseinandergesetzt.<sup>69</sup>

<sup>68</sup> Vgl. OECD (2019): «The Future of Work - Employment Outlook 2019», verfügbar unter [https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019\\_9ee00155-en](https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019_9ee00155-en)

<sup>69</sup> Zu erwähnen sind der Bericht über die zentralen Rahmenbedingungen der digitalen Wirtschaft vom 11. Januar 2017 und der Bericht über die Auswirkungen der Digitalisierung auf Beschäftigung und Arbeitsbedingungen vom 8. November 2017. Auch zu den Herausforderungen der Digitalisierung für Bildung und Forschung legte



Insgesamt befindet sich die Schweiz in einer guten Ausgangslage, um auch von der KI profitieren zu können. Gemäss Bundesrat ist es wichtig, den attraktiven Wirtschaftsstandort Schweiz mit stabilen wirtschaftlichen Rahmenbedingungen, einer stabilitätsorientierten Geldpolitik, einer anpassungsfähigen Arbeitsmarktregulierung mit einer gut funktionierenden Sozialpartnerschaft sowie einer aktivierenden Arbeitsmarktpolitik zu wahren. Entscheidend ist ferner, die vergleichsweise starke Innovationskraft der Schweizer Unternehmen, welche unter anderem durch die hohe Qualität des Schweizer Bildungssystems mit seiner ausgesprochenen Durchlässigkeit gefördert wird, aufrechtzuerhalten.

Die weitere Entwicklung ist sorgfältig zu beobachten. Der Bundesrat hat deshalb im November 2017 ein Monitoring der Auswirkungen des digitalen Wandels auf den Arbeitsmarkt beschlossen. Die Resultate des Monitorings sollen nach fünf Jahren (bis Ende 2022) in einem Bericht zusammengefasst werden und ein zeitnahe Erkennen neuer Herausforderungen ermöglichen.

### 6.3.4 Bewertung und Handlungsbedarf

Obwohl der digitale Wandel ein seit längerem laufender Prozess ist, wird in den entwickelten Ländern kein Rückgang der Gesamtbeschäftigung aufgrund neuer Technologien festgestellt. Weggefallene Stellen konnten bisher stets in anderen Bereichen kompensiert oder gar überkompensiert werden. Der Bundesrat geht in seinem Bericht zu den Auswirkungen der Digitalisierung auf Beschäftigung und Arbeitsbedingungen davon aus, dass auch der mit der Digitalisierung verbundene Strukturwandel zu neuen Beschäftigungsmöglichkeiten und einem gesamtwirtschaftlichen Beschäftigungsanstieg führen wird. Aufgrund des in den letzten zwanzig Jahren beobachteten Wandels ist jedoch zu erwarten, dass sich die Schweizer Arbeitswelt auch in den kommenden Jahrzehnten markant verändern wird, dies sowohl nach Branchen, Berufen als auch Tätigkeiten. Der Bundesrat hat in diesem Bericht auch festgehalten, dass trotz den bedeutenden Fortschritten der KI aktuell kein überdurchschnittlich rascher Strukturwandel auf gesamtwirtschaftlicher Ebene zu beobachten ist. Auch für die kommenden Jahre ist eher mit einer schrittweisen Entwicklung im Rahmen eines laufenden Strukturwandels als mit einer disruptiven technischen Revolution zu rechnen.

Für die Schweizer Volkswirtschaft geht es darum, das Potenzial der neuen Technologien zur Erhöhung der Produktivität und des Wachstums zu nutzen. Die KI wird dabei an Bedeutung gewinnen. Der technologische Wandel eröffnet dem auf Innovation und wertschöpfungsintensive Produkte ausgerichteten Wirtschaftsstandort Schweiz die Chance, weiter an Wettbewerbsfähigkeit zu gewinnen und so für die hiesigen Arbeitskräfte die bestehenden attraktiven Erwerbsmöglichkeiten zu erhalten bzw. neue zu schaffen. Hierfür ist es entscheidend, dass die Erfolgsfaktoren zur Bewältigung des strukturellen Wandels erhalten bleiben.

Damit angesichts der anhaltenden technologischen Entwicklung kein Qualifikations-Mismatch entsteht, müssen sich die Qualifikationen und Kompetenzen der Beschäftigten schnell genug an die veränderten Bedürfnisse des Arbeitsmarkts anpassen. Daher sind eine zeitnahe Ausrichtung der Bildung und eine entsprechende Anpassung der Bildungsgänge an die veränderten Anforderungen unabdinglich. Lebenslanges Lernen in Form von kontinuierlicher Aus- und Weiterbildung wird weiter an Bedeutung gewinnen. Der Bundesrat hat diese Herausforderung erkannt und entsprechende Massnahmen – wie die Förderung der Grundkompetenzen am Arbeitsplatz – bereits eingeführt (siehe Abschnitt 6.5). Vor diesem Hintergrund sind in Bezug auf die Auswirkungen der künstlichen Intelligenz auf den Arbeitsmarkt derzeit keine neuen Massnahmen erforderlich.

Aktionsfeld 1: Auswirkungen der KI auf den Arbeitsmarkt	
Weiterverfolgung der Entwicklungen des Schweizer Arbeitsmarkts	<p>Das SECO wird die Herausforderungen beobachten und aufkommende Fragen im Rahmen der bestehenden Kompetenzen bearbeiten.</p> <p><b>Zuständigkeit:</b> SECO                      <b>Status:</b> Beobachtung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

er einen Bericht vor, ebenso wie zu den Auswirkungen der Digitalisierung auf das Steuerwesen und die Finanzierung der Sozialversicherungen. Des Weiteren führte er mit dem «Digitalen Test» eine umfassende Umfrage durch, um zu ermitteln, wo die bestehende wirtschaftspolitisch relevante Gesetzgebung die Digitalisierung unnötig behindert.

## 6.4 Künstliche Intelligenz in Industrie und Dienstleistungen<sup>70</sup>

### 6.4.1 Übersicht

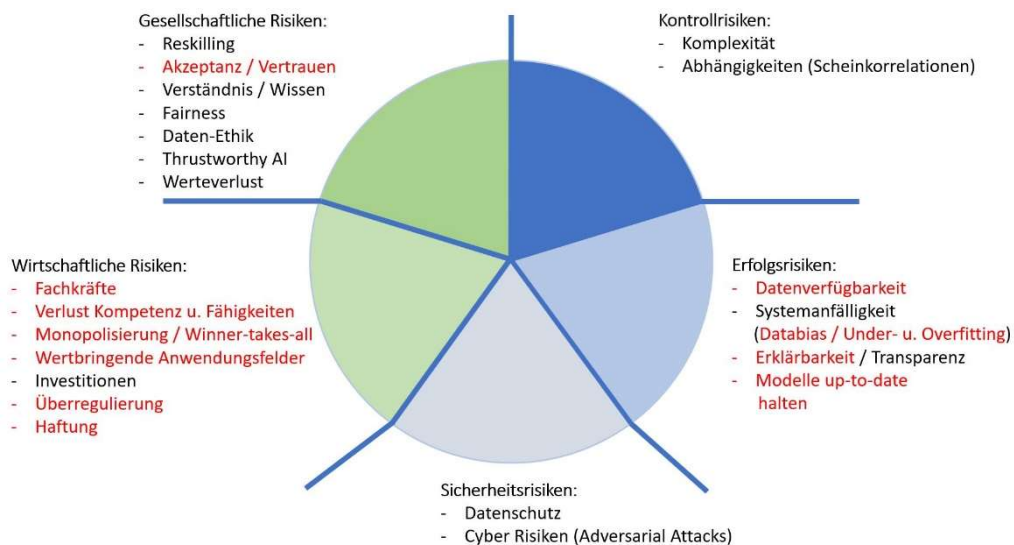
Im Kontext der vierten industriellen Revolution eröffnet KI neue Möglichkeiten, um aus dem Potenzial digitaler Vernetzung und der einhergehenden Datenflut einen Mehrwert zu schöpfen. Auf der Kosten- seite ermöglicht es den Firmen und Dienstleistern, die KI-Anwendungen anzuwenden, den internen Aufwand zu reduzieren. Neben der Anwendung von KI in Bereichen, in denen Kosten reduziert werden können, birgt eine durch KI verbesserte Kundenerfahrung für die Unternehmen zusätzliches Potenzial. Auf der Einnahmenseite ermöglichen KI-Anwendungen den Firmen und Dienstleistern verbesserte oder neue Produkte und Dienstleistungen.

Die Anwendung der KI in der Industrie und im Dienstleistungssektor steht erst am Anfang. Aktuell haben in der Schweiz viele Projekte Pilotcharakter und Firmen investieren in KI, um Erfahrungen zu sammeln. Die KI übernimmt wahrnehmungsbezogene Aufgaben, die bislang nur der Mensch erfüllen konnte (z.B. repetitive Aufgaben) oder Aufgaben innerhalb der Prozessoptimierung oder -automatisierung. Prozesse können durch KI effizienter (Zeit und Kosten) und skalierbar durchgeführt werden.

### 6.4.2 Herausforderungen

Die Herausforderungen für Unternehmen können aus Sicht der konsultierten Expertinnen und Experten einerseits in technische Herausforderungen (wie Kontrollrisiken oder Sicherheitsrisiken), andererseits in gesellschaftliche (z.B. Akzeptanz / Vertrauen) sowie wirtschaftliche (u.a. Fachkräfte, Überregulierung) kategorisiert werden. Abbildung 13 zeigt die vielfältigen Risiken für Unternehmen, wobei jene spezifischen Herausforderungen «rot markiert» sind, die firmenintern meist eine hohe Priorität haben.

Abbildung 13: Herausforderungen von KI-Systemen für Unternehmen



Quelle: SATW (2019): «Künstliche Intelligenz in Industrie und Dienstleistungen».

### 6.4.3 Bestehende Aktivitäten

In erster Linie organisieren sich Wirtschaftsakteure in zahlreichen Initiativen, welche sich direkt oder indirekt mit KI beschäftigen, selbst (einige Beispiele sind in Tabelle 4 dargestellt). Über diese Aktivitäten hinaus verfolgen **Firmen** selbst eigene Projekte zu KI.<sup>71</sup> Diese adressieren u.a. Aspekte wie ethi-

<sup>70</sup> Der Abschnitt zu Bedeutung und Herausforderungen von KI in den Industrie- und Dienstleistungssektoren wurde im Auftrag des SBFi durch die SATW erarbeitet. Dazu wurden Experten aus Forschung und Wirtschaft aus verschiedenen Bereichen befragt. Der Textbeitrag spiegelt die Meinung der Experten wider. Für eine ausführliche Fassung siehe SATW (2019), «Künstliche Intelligenz in Industrie und Dienstleistungen», Bericht im Auftrag des Staatssekretariats für Bildung, Forschung und Innovation, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

<sup>71</sup> Zu nennen sind Microsoft, Google, PwC, D-One und weitere.

sche Richtlinien, die verantwortungsvolle Nutzung von KI oder ähnliche Themen, um das Vertrauen in Produkte und Dienstleistungen, bei denen KI angewendet wird, zu fördern.

Tabelle 4: Initiativen und Aktivitäten im KI-Bereich in der Schweiz (exemplarisch)

<b>Swiss Alliance for Data-Intensive Services</b>	In dieser Initiative erarbeitet die Expertengruppe «Data Ethics» einen «Ethical Codex for Data-Based Value Creation», welcher das Vertrauen in die Technologie unterstützen soll.
<b>Swiss Group of Artificial Intelligence and Cognitive Science</b>	Fachgruppe der Schweizer Informatik Gesellschaft, welche den Austausch zwischen Forschern, Anwendern und Interessierten im Bereich KI fördert.
<b>SwissCognitive</b>	Initiative von zahlreichen Firmen aus Industrie und aus dem Dienstleistungssektor, welche als Austausch- und Netzwerkplattform fungiert und sich vollkommen dem Thema KI widmet.
<b>Digitalswitzerland</b>	Schweizweite Initiative bestehend aus über 150 Mitgliedern mit der Vision, die Schweiz als führenden Innovationstandort zu festigen.
<b>Industrie 2025</b>	Schweizer Brancheninitiative, die betroffene Akteure über das Thema Industrie 4.0 informiert, sensibilisiert, vernetzt und fördert.
<b>Swiss Smart Factory</b>	Plattform am Switzerland Innovation Park Biel/Bienne für interdisziplinäre Fragestellungen rund um Industrie 4.0.

Quelle: SATW (2019), «Künstliche Intelligenz in Industrie und Dienstleistungen».

Der Einsatz von KI geht mit der generellen Entwicklung der Digitalisierung einher. Der Bund unterstützt mit seiner **Strategie «Digitale Schweiz»** die landesweite Digitalisierung in sämtlichen Bereichen. Ausserdem wurden im Rahmen des **«Aktionsplans Digitalisierung im BFI-Bereich»** eine Vielzahl von Massnahmen zur Stärkung der digitalen Kompetenzen in Bildung, Forschung und Innovation lanciert. Der Fachkräftemangel bei Berufen, die im Kontext der Digitalisierung zentral sind, wird seitens des Bundes bereits in vielfältiger Hinsicht angegangen, so etwa im Rahmen der 2018 abgeschlossenen **Fachkräfteinitiative**.

#### 6.4.4 Bewertung und Handlungsbedarf

Die wichtigsten Herausforderungen für die Anwendung von KI in den Industrie- und Dienstleistungssektoren sind aus Expertensicht durch die Wirtschaft selbst zu bewältigen. Dies betrifft insbesondere die technischen Herausforderungen beim Einsatz von KI. Die Wirtschaft hat diese Herausforderungen im Rahmen zahlreicher Initiativen aufgegriffen, welche durch öffentliche Initiativen unterstützt werden. Dies beinhaltet auch Initiativen, die auf das Schaffen von Vertrauen für KI-Technologien und den verantwortungsvollen Nutzen von KI abzielen.

Die im Kontext dieses Berichts befragten Experten sehen bezüglich KI in Industrie und im DL-Sektor einige übergeordnete Herausforderungen, welche den Bund betreffen (eine ausführliche Darstellung ist im Bericht SATW (2019) «KI in Industrie und Dienstleistungen» zu finden). Diese beziehen sich jedoch weitgehend nicht ausschliesslich auf KI, sondern auf die Digitalisierung insgesamt. Entsprechend wurden sie zum Teil bereits im Zuge der Reaktion auf die Digitalisierung und den digitalen Transformationsprozess erkannt und aufgegriffen.

Der Anregung der Experten, dass der Bund ein **gemeinsames nationales Zielbild über KI** formulieren soll, wird dahingehend entsprochen, dass als Folgearbeit auf Ebene des Bundes strategische Leitlinien erarbeitet werden. Zudem wird die KI-relevante Politik im Rahmen der Strategie «Digitale Schweiz» als prioritäres Thema aufgenommen. Dabei können auch relevante private Initiativen berücksichtigt werden. Um dem Fachkräftemangel entgegen zu wirken, sollen gemäss den Experten die Rahmenbedingungen für die Anstellung von **in der Schweiz ausgebildeten ausländischen Fachkräften** verbessert werden. Dazu beitragen kann u.a. die derzeitige Umsetzung der **Motion 17.3067 Dobler**, welche zum Ziel hat, die Voraussetzungen für die Anstellung von in der Schweiz ausgebildeten Studierenden aus Drittstaaten zu erleichtern. Es besteht somit insgesamt derzeit kein Bedarf für neue Aktionsfelder.

**Aktionsfeld 1: KI in der Wirtschaft**

Die Anwendung von KI hat das Potenzial zu erheblicher Effizienzsteigerungen in Produktion und Service und ermöglicht eine stärkere Individualisierung in Richtung massgeschneiderte Kundenlösungen.

Weiterverfolgung der Entwicklungen von KI in Industrie und Dienstleistungen

Das SECO wird vor diesem Hintergrund die Herausforderungen bezüglich KI in Industrie und im DL-Sektor beobachten und aufkommende Fragen im Rahmen der bestehenden Kompetenzen bearbeiten.

**Zuständigkeit:** SECO

**Status:** Beobachtung im Rahmen bestehender Kompetenzen

**Zusätzlicher Handlungsbedarf:** Nein

## 6.5 Künstliche Intelligenz in der Bildung<sup>72</sup>

### 6.5.1 Übersicht<sup>73</sup>

**Auswirkungen der künstlichen Intelligenz auf Lehren und Lernen:** KI bringt enorme Chancen zur Verbesserung von Lehr- und Lernprozessen mit sich. Bisher ist KI in der Praxis noch relativ wenig verbreitet, dies kann sich jedoch schnell ändern.<sup>74</sup> KI schafft zusätzliche Möglichkeiten für das automatische Korrigieren von Übungen oder das Auswerten von Tests. Mittels «Learning Analytics» ermöglicht sie die Nachverfolgung und Analyse des Lernverhaltens. Dadurch kann die Ausbildung besser an die individuellen Bedürfnisse der Lernenden angepasst und künftige Misserfolge können vorhergesagt werden. Für detaillierte Informationen hinsichtlich Verbesserungen im Bereich Lehren und Lernen siehe Kapitel 1 des ergänzenden Kurzberichts «Künstliche Intelligenz in der Bildung».

### 6.5.2 Herausforderungen

**Erforderliche Kompetenzen für den Einsatz und die Herstellung intelligenter Systeme:** Künstliche Intelligenz wird sowohl im privaten als auch im beruflichen Bereich immer mehr Teil unseres Lebens sein. Dies hat Folgen für die Kompetenzen, die die Bürgerinnen und Bürger besitzen müssen, um in einer digitalisierten Gesellschaft leben und arbeiten zu können. Die in Zukunft erforderlichen Kompetenzen sind jedoch schwer vorherzusagen. Sicher ist, dass durch KI neue Berufe entstehen werden und andere wahrscheinlich verschwinden werden. Zudem werden immer mehr digitale Kompetenzen (Grund- und vertiefte Kenntnisse) verlangt. Die Digitalisierung ist eine Gefahr für die am wenigsten qualifizierten Berufe, deren repetitive Tätigkeiten leicht zu automatisieren sind. Dabei ist zu beachten, dass «soft skills», oder bereichsübergreifende Kompetenzen wie Neugier, Kreativität, Teamfähigkeit, Empathie, Führungskompetenz oder auch Problemlösefähigkeiten, zunehmend gefragt sein werden, da diese heute noch nicht von Maschinen übernommen werden können.

Aufgrund der Digitalisierung werden in nahezu allen Sektoren (grundlegende oder vertiefte) digitale Kompetenzen notwendig sein; die KI erfordert jedoch verstärkte Kompetenzen in gewissen spezifischen Bereichen. So benötigen beispielsweise die Branchen, die mit KI arbeiten, ein sehr hohes Kompetenzniveau im Bereich Algorithmen und Mathematik. Darüber hinaus kommt es verstärkt auf kritisches Denken und ethische Argumentationsfähigkeit an. In Branchen, die die intelligenten Systeme lediglich einsetzen, steht das intuitive Verstehen der Algorithmen im Vordergrund, um deren Möglichkeiten und Grenzen einschätzen zu können. Zu unterstreichen ist, dass sich die KI auf alle Wirtschaftssektoren auswirken und sowohl handwerkliche als auch hochqualifizierte Berufe wie zum Beispiel Rechtsanwältinnen bzw. -anwälte betreffen könnte. Für Näheres zur Entwicklung der Kompetenzen siehe Kapitel 2.1 des ergänzenden Kurzberichts.

Die Chancengleichheit und der Einbezug von Frauen in die Forschung und Produktion sowie bei der Anwendung von KI gilt es zu beachten. Es sollen mehr Mädchen und Frauen für den MINT-Bereich motiviert und darin ausgebildet werden.

**Herausforderungen für die pädagogische Praxis:** Die tatsächlichen Auswirkungen von Systemen mit künstlicher Intelligenz sind bislang begrenzt, da KI in der Praxis heute nur in recht geringem Umfang eingesetzt wird. Doch es geht bei der Anwendung im Bildungsbereich um die Verwertung sensibler Informationen aus dem persönlichen Bereich. Besonders relevant sind daher Probleme der fehlenden Nachvollziehbarkeit von Ergebnissen oder des in den Algorithmen enthaltenen Bias. Für Näheres zu den pädagogischen Herausforderungen siehe Kapitel 2.2 des ergänzenden Kurzberichts «Künstliche Intelligenz in der Bildung».

Es sei darauf hingewiesen, dass die Rolle, die die Schulen in der Bildung sowie in pädagogischer und gesellschaftlicher Hinsicht spielen, in keiner Weise durch den Einsatz von KI in Frage gestellt wird. Die Lehrkraft wird nicht an Bedeutung verlieren. Im Gegenteil: Sie muss sich in Zukunft lediglich weniger mit dem Korrigieren von Aufgaben oder anderen administrativen Tätigkeiten befassen, die dann von

<sup>72</sup> Für eine ausführliche Fassung siehe Staatssekretariat für Bildung, Forschung und Innovation (2019), «Künstliche Intelligenz in der Bildung», verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

<sup>73</sup> Dieser Text basiert wesentlich auf einem Beitrag, der von Pierre Dillenbourg, Prof. ordinaire en technologies de formation à l'EPFL, im Auftrag des SBFI verfasst wurde.

<sup>74</sup> Tuomi, I., «The Impact of Artificial Intelligence on Learning, Teaching, and Education. Policies for the future», Eds. Cabrera, M., Vuorikari, R & Punie, Y., Publications Office of the European Union, Luxembourg, 2018.

Maschinen erledigt werden können, und kann sich ganz der Vorbereitung des Unterrichts oder der gezielten Unterstützung der einzelnen Schülerinnen und Schüler widmen.

**Ethische Auswirkungen der Anwendung von KI im Bildungsbereich:** Durch die zunehmende Nutzung von Daten ergeben sich auch ethische Fragestellungen. Wer soll in Zukunft wieviel Daten von wem zu welchem Zweck sammeln und bearbeiten dürfen? Wie soll die Datenkultur im Bildungswesen aussehen? Von welchen Prinzipien soll sie geleitet werden? Eine zentrale Aufgabe eines sicheren, integrierten Systems standardisierter Datensammlung und -auswertung im Bildungsbereich ist das Vertrauen in diese Nutzung und damit die Akzeptanz der betroffenen Anspruchsgruppen dafür zu stärken. Dieses Vertrauen zu schaffen ist eine notwendige Grundvoraussetzung dafür, dass Erkenntnisse aus allfälligen Datenanalysen als hilfreich und gerecht wahrgenommen werden<sup>75</sup>.

**Lücken in der pädagogischen Forschung:** Um die in Lehr- und Lernprozessen einzusetzenden intelligenten Systeme zu verbessern und die oben genannten technischen und ethischen Probleme zu lösen, sind weitere Forschungsarbeiten erforderlich. Wichtig ist auch die Zusammenarbeit zwischen den KI-Fachleuten und den Forschenden im Bildungssektor, da nur so den Besonderheiten dieses Bereichs Rechnung getragen werden kann.

### 6.5.3 Bestehende Aktivitäten

Für die Auswirkungen von KI auf die Kompetenzen und auf das Bildungssystem ist es zentral, die Entwicklungen nicht isoliert, sondern als integralen Bestandteil der allgemeinen Entwicklungen der Digitalisierung zu betrachten. Insbesondere da es schwer ist, vorauszusagen, wie sich die Kompetenzanforderungen in Zukunft entwickeln werden und welche Fähigkeiten im Konkreten benötigt werden, um mit den Tools der künstlichen Intelligenz umzugehen.

Kantons- und Bundesbehörden sowie die Bildungseinrichtungen auf allen Ebenen des Bildungssystems sind sich der Herausforderungen der Digitalisierung bewusst. Die Entwicklungen werden intensiv beobachtet, damit Risiken rechtzeitig begegnet werden und das Potenzial von KI voll ausgeschöpft werden kann. Zahlreiche Strategien, Initiativen und Massnahmen sind bereits geplant oder befinden sich in Umsetzung. Diese Strategien betreffen die Digitalisierung im Bildungsbereich ganz allgemein. Es ist jedoch klar, dass sie auch darauf abzielen, dass KI verantwortungsbewusst eingesetzt wird und sichergestellt ist, dass angemessene Kompetenzen in diesem Bereich vermittelt werden. Kapitel 3 des ergänzenden Kurzberichts zeigt die Strategien und Massnahmen auf, die auf allen Ebenen des Bildungssystems umgesetzt werden. Darüber hinaus enthält es eine nicht abschliessende Liste von KI-relevanten Aktivitäten, die bereits jetzt auf den einzelnen Ebenen des Bildungssystems stattfinden.

Es ist festzuhalten, dass das Thema Kompetenzen und Bildung bzgl. KI auf internationaler Ebene sehr präsent ist. So decken die KI-Empfehlungen der OECD vom Mai 2019 auch die erforderlichen Kompetenzen («skills») ab. Eines der Ziele des «Coordinated Plan on Artificial Intelligence»<sup>76</sup> der Europäischen Kommission vom Dez. 2018 ist die Anpassung von Bildungsplänen und -systemen, um die Gesellschaft besser auf KI vorzubereiten. Auch die durch die EU am 26. Juni 2019 veröffentlichten «Policy and investment recommendations for trustworthy Artificial Intelligence» verweisen auf diese Kompetenzen. Schliesslich ist die Förderung von Kompetenzen in Feldern wie KI auch ein wichtiger Aspekt des «Digital Europe Programme» (DEP) für die Jahre 2021-2027 (siehe Kapitel 6.2). Das SBFI beteiligt sich an verschiedenen Arbeitsgruppen der Europäischen Kommission, wie der «Working Group Digital Education» oder der «Working Group VET», die sich auch mit KI beschäftigen. Die Schweiz ist somit gut über die KI-Aktivitäten ihrer europäischen Nachbarn im Bildungsbereich informiert. Die EU und ihre Mitgliedsstaaten haben ebenfalls erkannt, wie wichtig diese Technologie ist und diskutieren derzeit deren Potenziale und Herausforderungen. Allerdings haben bis heute nur wenige Länder (z. B. Finnland) spezielle KI-Strategien oder Initiativen im Bereich Bildung ins Leben gerufen.

### 6.5.4 Bewertung und Handlungsbedarf

Bund und Kantone arbeiten heute auf Grundlage ihrer jeweiligen Kompetenzen im Rahmen des Koordinationsausschusses «Digitalisierung in der Bildung» eng zusammen, um gute Rahmenbedingungen

<sup>75</sup> educa.ch (2019): «Daten in der Bildung – Daten für die Bildung. Grundlagen und Ansätze zur Entwicklung einer Datennutzungspolitik für den Bildungsraum Schweiz». Bern.

<sup>76</sup> Mitteilung «Coordinated Plan on Artificial Intelligence», COM(2018) 795.

für die Digitalisierung im Bildungsbereich zu schaffen. Im Hochschulbereich arbeiten Bund und Kantone über die Gremien der Schweizerischen Hochschulkonferenz (SHK) zusammen. Die Herausforderungen der Digitalisierung für den Bildungsbereich sind von den entsprechenden Gremien aufgenommen worden, welche bereits eine Vielzahl von Massnahmen lanciert haben. Diese Arbeiten umfassen auch die Thematik KI. Im besonderen Bereich der Hochschulen hat die dafür zuständige Schweizerische Hochschulkonferenz (SHK) im Mai 2019 gestützt auf die strategische Planung von swissuniversities die gesamtschweizerischen hochschulpolitischen Prioritäten für die Periode 2021-2024 festgelegt. Dazu gehören u.a. auch die Förderung des Nachwuchses in den MINT-Bereichen und ganz speziell die Förderung der Ausbildung von IKT-Fachkräften sowie die Stärkung der Digital Skills der Absolvierenden und des wissenschaftlichen Personals. Derzeit stellt sich kein Bedarf für weitere, über diese Massnahmen hinausgehenden Abklärungen oder Gremien. Dabei ist darauf hinzuweisen, dass die Digitalisierung einer der Schwerpunktbereiche des Leitbilds «Berufsbildung 2030» ist und zu den wichtigen Querschnittsbereichen der BFI-Botschaft 2021–2024 gehört. In diesem Zusammenhang gibt es auch Überlegungen spezifisch zum Thema künstliche Intelligenz. In den bestehenden Gefässen, sowie im Dialog mit den Akteuren, sollen die folgenden Themen nach Bedarf aufgenommen bzw. weiterverfolgt werden.

**Aktionsfeld 1: Gewährleisten angemessener Kompetenzen**

Die erforderlichen Kompetenzen für den Einsatz der KI müssen sowohl in der Schule als auch während des gesamten Lebens im Rahmen von Weiterbildungen erworben werden, um zu vermeiden, dass es zu einer Spaltung der Gesellschaft zwischen Menschen mit diesen Kompetenzen und dem Rest der Bevölkerung kommt. Dabei sind die «soft skills» (oder bereichsübergreifenden Kompetenzen) ebenso grundlegend wie das allgemeine Verständnis von Algorithmen. Das Bildungssystem muss auch den angemessenen Erwerb spezifischer Kompetenzen zur Produktion von KI-Systemen sicherstellen und folglich zur Ausbildung von KI-Spezialistinnen und -Spezialisten beitragen.

Es sollen mehr Mädchen und Frauen für den MINT-Bereich motiviert und darin ausgebildet werden.

Vermittlung der für den Einsatz künstlicher Intelligenz erforderlichen Kompetenzen auf allen Bildungsstufen	Das SBFI sorgt dafür, dass auf den durch das SBFI verantworteten Ebenen des Bildungssystems (z. B. Berufsbildung) die für den Einsatz künstlicher Intelligenz erforderlichen Kompetenzen vermittelt werden. Hinsichtlich der übrigen Ebenen des Bildungssystems arbeitet das SBFI innerhalb der Gremien, in denen es vertreten ist, eng mit den Kantonen zusammen.  <b>Zuständigkeit:</b> SBFI, Kantone und andere relevante Akteure <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
-------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Zusätzlicher Handlungsbedarf:** Nein

**Aktionsfeld 2: Gewährleisten einer transparenten und verantwortungsvollen Anwendung von KI im Bildungsbereich**

Der Einsatz künstlicher Intelligenz in Lehr- und Lernprozessen wirft eine Reihe reglementarischer Fragen hinsichtlich des Zugriffs auf die während der Ausbildung erhobenen Daten sowie die Sammlung und Verwendung dieser Daten auf. Um von den Vorteilen von KI profitieren zu können, ist ein transparenter und verantwortungsvoller Einsatz im Bildungsbereich unerlässlich. Um dieses Ziel zu erreichen, müssen die Risiken hinsichtlich Datensicherheit und Datenschutz, Schutz der Privatsphäre und ethischer Aspekte minimiert werden.

Transparente und verantwortungsvolle Anwendung von KI im Bildungsbereich	Das SBFI analysiert innerhalb der Organe, in denen es vertreten ist, in enger Zusammenarbeit mit den Kantonen gegebenenfalls erforderliche Massnahmen, um den transparenten und verantwortungsvollen Einsatz von KI im Bildungsbereich sicherzustellen.  <b>Zuständigkeit:</b> SBFI, Kantone und andere relevante Akteure <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
--------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Zusätzlicher Handlungsbedarf:** Nein

## 6.6 Anwendung von künstlicher Intelligenz in Wissenschaft und Forschung<sup>77</sup>

### 6.6.1 Übersicht

In der Forschung ist KI in allen Disziplinen breit anwendbar, bietet einen neuen Zugang zu bestehendem Wissen und birgt grosse Potenziale für neue Entdeckungen. So ermöglicht KI zum Beispiel Live-Untersuchungen von Prozessen, die dem Menschen sonst nicht zugänglich wären. Im Wissenschaftsbereich wird KI eingesetzt, um Daten zu generieren, zu analysieren oder Methoden zu verbessern. Auch machen KI-unterstützte Übersetzungsprogramme fremdsprachige Publikationen einem grösseren Kreis zugänglich und KI-unterstützte Formen der Literaturlauswertung führen zu kürzeren Review-Durchlaufzeiten. Zudem wird wissenschaftliche Arbeit effizienter und produktiver, da die Zahl nötiger Versuche sinkt und sich Prozesse schneller modellieren lassen.

### 6.6.2 Herausforderungen

Aus Expertensicht bestehen bezüglich eines verstärkten Einsatzes von KI in Wissenschaft und Forschung zahlreiche Herausforderungen. Diese sind einerseits technischer Natur, betreffen andererseits aber auch übergeordnete rechtliche und gesellschaftlichen Fragen:

Tabelle 5: Spezifische KI-Herausforderungen in Wissenschaft und Forschung

<b>Erklärbarkeit</b>	«Explainable AI» wird immer wichtiger in Wissenschaft und Forschung.
<b>Vertrauen in KI-Systeme</b>	Der Autonomiegrad von KI-Systemen ist je nach Anwendungsbereich zu definieren.
<b>Fehleranfälligkeit</b>	Verzerrungen («bias») sind nur schwer zu detektieren.
<b>Ausbildung</b>	Ist zentral, um KI-Systeme zu verstehen und Resultate richtig zu interpretieren.
<b>Private Akteure</b>	Sind wichtig für die Forschung, aber Abhängigkeiten von Software-Tools sind möglich.
<b>Datenschutz</b>	Möglichst passgenaue Richtlinien zum Schutz von Daten und Privatsphäre sind nötig.
<b>Datenakquisition</b>	Richtlinien sind nötig, welche Daten und (online) KI-Tools genutzt werden dürfen.
<b>Geistiges Eigentum</b>	Fragen zu Datenschutz, Datenverfügbarkeit und geistigem Eigentum sind zu klären.
<b>Zugänglichkeit</b>	Anpassung des Verlagswesens auf die KI-spezifischen Bedürfnisse (z.B. Zugriff auf Publikationen und Analyse mit Text Mining) sind nötig.
<b>Infrastrukturen</b>	Die für den Kompetenzaufbau im Bereich KI erforderlichen Infrastrukturen müssen finanziert werden können.

Quelle: SATW (2019): «Künstliche Intelligenz in Wissenschaft und Forschung».

### 6.6.3 Bestehende Aktivitäten

Die Schweiz hat eine Vielzahl von Institutionen und Initiativen, welche sich intensiv mit der KI-Forschung beschäftigen und die Entwicklung der KI-Technologien in den vergangenen Jahrzehnten

<sup>77</sup> Dieser Abschnitt zu «KI in Wissenschaft und Forschung» wurde im Auftrag des SBFI durch die SATW erarbeitet. Die SATW hat sieben Experten aus verschiedenen Forschungsdisziplinen anhand eines strukturierten Interviews befragt. Der auf dieser Basis verfasste Bericht wurde im Anschluss mit von swissuniversities vorgeschlagenen KI-Experten diskutiert und konsolidiert. Finalisiert wurde der vorliegende Abschnitt durch das SBFI. Der Textbeitrag spiegelt die Meinung der Experten wider. Für eine ausführliche Fassung siehe SATW (2019), «Künstliche Intelligenz in Wissenschaft und Forschung», Bericht im Auftrag des Staatssekretariats für Bildung, Forschung und Innovation, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)



stark mitgeprägt haben. Auch international werden Kompetenzen und Wissen für KI aufgebaut, Talente gefördert und der Wissenstransfer intensiviert. In diese Initiativen sind auch Akteure aus der Schweiz stark involviert und übernehmen teilweise eine führende Rolle. Kapitel 5 nennt einige Beispiele bestehender Aktivitäten.

Die Anwendung von KI bei der **Bearbeitung von Forschungsdaten und -publikationen** setzt insbesondere deren einfache und möglichst freie Verfügbarkeit voraus. Die Schweizer Hochschulen haben hierfür Schritte in den Bereichen «Open Access to Publication» und «Open Research Data» eingeleitet. Sie haben mit dem SNF eine nationale **Open-Access-Strategie** erarbeitet mit dem Ziel, dass bis 2024 alle mit öffentlichen Geldern finanzierten wissenschaftlichen Publikationen frei zugänglich sind (bei SNF-finanzierten Projekten bereits ab 2020). Auf europäischer Ebene ist die **BEAT-Plattform** eine europäische elektronische Computerinfrastruktur für Open Science.

#### 6.6.4 Bewertung und Handlungsbedarf

Derzeit stellt sich für den Bund aus Sicht der Experten kein Bedarf für weitere Massnahmen oder Gremien. In den bestehenden Gefässen, sowie im Dialog mit den Akteuren, sollen die folgenden Themen nach Bedarf aufgenommen bzw. weiterverfolgt werden.<sup>78</sup>

**Weiterentwicklung und Anwendung von KI-Methoden:** Die Herausforderungen von KI sind durch die Wissenschaft und Hochschulen direkt anzugehen, was auch bereits intensiv getan wird. Dabei ist die Interdisziplinarität und der Austausch bezüglich traditionellen und «neuen» KI-Methoden zu verstärken. Personen an den Hochschulen können hierfür intern und extern als «Infopoints KI» agieren. Forschungsschwerpunkte, neue Initiativen (wie beispielsweise die «Digital Society Initiative» der Universität Zürich) definieren die Hochschulen selbst. Hier kann der Bund keine Vorgaben machen.

**Übergeordnete Fragen in Kompetenz des Hochschulbereichs:** Übergeordnete Fragen sind durch die bestehenden, koordinierenden (politischen und akademischen) Gremien im Hochschulbereich zu klären, u.a.: (i) Dialog über die Herausforderungen und die Empfehlungen bezüglich KI; (ii) Prüfung des Bedarfs und ggf. Entwicklung einer Strategie für Regeln zum Einsatz von KI; (iii) Klärung Rolle privater Akteure und datenbezogene Fragen (Owner-/Stewardship); (iv) Bereits aufgenommen seitens Hochschulen: Zugänglichkeit (Open Science/Open Access) und Infrastrukturen (Repositorien); (v) Ethikkommissionen können bei übergeordneten Fragen eine Rolle spielen.

**Verstärkter Kompetenzaufbau bezüglich KI:** Die Schweiz verfügt bereits über gute Instrumente zur Förderung von KI. Ein isoliertes Förderprogramm wäre kaum sinnvoll, da viele Disziplinen betroffen wären und sich eine Förderung nicht fokussieren liesse. Die Stärkung allgemeiner digitaler Kompetenzen ist im Rahmen der bestehenden Gefässe und Kompetenzen sicherzustellen und wurde auf Ebene Bund bereits im «Aktionsplan Digitalisierung im BFI-Bereich» und im Rahmen der offenen, kompetitiven Instrumente des Bundes aufgenommen. In der strategischen Planung der Hochschulen für die Jahre 2021-2024 wurden Digitalisierung und KI als zentrale Schwerpunkte definiert.

Aktionsfeld 1: Kompetenzen in Forschung und Innovation	
Forschung und Innovation kommen in der Bewältigung der Herausforderungen von KI eine zentrale Rolle zu. Die Aufrechterhaltung der Kompetenzen auf höchstem Niveau ist daher sicherzustellen.	
Gewährleistung der Forschungs- und WTT-Kompetenzen im Rahmen der BFI-Politik	<p>Die Massnahmen des «Aktionsplans Digitalisierung im BFI-Bereich» sind im Rahmen der BFI-Periode 2021–2024 aufgenommen und werden in Eigenregie der Akteure weitergeführt. Innerhalb der bestehenden Gefässe und -organe ist in der BFI-Politik zudem zu gewährleisten, dass die Akteure in Wissenschaft und WTT für die Herausforderungen der künstlichen Intelligenz vorbereitet sind und sich dieser im Rahmen ihrer jeweiligen Digitalisierungsaktivitäten und Zuständigkeiten annehmen.</p> <p><b>Zuständigkeit:</b> SBFI                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<sup>78</sup> Die Empfehlungen der Experten betreffen auch die Bereiche Daten, Recht und Geistiges Eigentum; für die diesbezüglichen Aspekte sei auf die entsprechenden Abschnitte des Berichts verwiesen.

## 6.7 Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik<sup>79</sup>

### 6.7.1 Übersicht

Aus sicherheitspolitischer Sicht gibt es drei Themenfelder, in denen KI eine wesentliche und zunehmende Rolle spielt: (i) Aussensicherheitspolitik und internationale Gouvernanz; (ii) Streitkräfte und Wandel der Kriegführung; (iii) Nachrichtendienste und innere Sicherheit.

Im KI-Bereich hat auf globaler Ebene ein technologischer Wettlauf eingesetzt, der die strategische Stabilität und internationale Sicherheit herausfordert und belastet; dazu gehört auch die Weiterentwicklung von Waffensystemen mit zunehmender Autonomie, welche die Kriegführung stark verändern werden. Sicherheitspolitisch relevante Fragen wirft auch das Ungleichgewicht zwischen staatlichen und privaten Akteuren auf. Die führenden Technologieunternehmen verfügen bezüglich KI über einen erheblichen Wissens-, Daten- und Anwendungsvorsprung. Sicherheitspolitische Fragen werden schliesslich im Bereich der Überwachung und Kriminalitätsbekämpfung aufgeworfen; auch hier eröffnen KI-Technologien neue Möglichkeiten.

### 6.7.2 Herausforderungen

KI hat auf die Sicherheitspolitik und auf die Cybersicherheit negative wie auch positive Auswirkungen. KI kann noch gezielter für rasche, präzise Cyberangriffe eingesetzt werden (z.B. für Spionagezwecke, Phishing), aber auch für Desinformation, Propaganda oder Waffensysteme. KI kann zudem zur Verstärkung der Verteidigungsmöglichkeiten eingesetzt werden, wie das heute bereits der Fall ist (z.B. frühzeitige Identifizierung von Cybervulnerabilitäten). Auch können neue Angriffsmuster und -vektoren besser und rascher erkannt und Sicherheitselemente möglichst in Echtzeit (z.B. Malwareschutz, Firewalls) kontrolliert und gesteuert werden.

**Aussensicherheitspolitik und internationale Gouvernanz:** Dabei geht es darum, Bedeutung und Einfluss von KI auf die Aussensicherheitspolitik und internationalen Beziehungen zu untersuchen. Die Fragen lauten: (i) Inwieweit beeinflussen KI-Systeme die internationale strategische Stabilität? (ii) Inwieweit führen KI-Systeme zu einem Verlust der internationalen Eskalationskontrolle? (iii) Welche Herausforderungen bringt KI für die Rüstungskontrolle?

**Streitkräfte und Wandel der Kriegführung:** Dabei geht es darum, den Einfluss von KI auf das Kriegs- und Konfliktbild zu antizipieren und mögliche Konsequenzen auch für die eigenen Verteidigungsfähigkeiten zu ziehen. Die Fragen lauten: (i) Wie beeinflusst KI die militärische Innovation und Weiterentwicklung von Fähigkeiten? (ii) Welche Auswirkungen hat KI auf die militärischen Entscheidungsprozesse und was sind die Konsequenzen davon (Beschleunigung, Effizienzsteigerung der Führungsunterstützung und der Aufklärung, Einbezug der Akteure, Auswirkungen auf Kriegsvölkerrecht, Umgang mit Unsicherheit, stabile/nichtstabile Lösungen)? (iii) Welche Auswirkungen hat KI auf asymmetrische Konfliktformen?

**Nachrichtendienste und innere Sicherheit:** Dabei geht es darum, den Einfluss von KI auf Sicherheitsinstrumente des Staates im Inneren zu analysieren und deren Chancen und Risiken zu bewerten. Die Fragen lauten: (i) Wie wirkt sich KI auf die Tätigkeit staatlicher Nachrichtendienste aus? (ii) Inwieweit fördert KI die Wirkung von Propaganda und Desinformation? (iii) Wie wirken sich KI-Systeme auf kriminelle Aktivitäten im Cyber-Bereich aus? Welche neuen Möglichkeiten und Chancen bietet KI für die nachrichtendienstlichen Tätigkeiten?

### 6.7.3 Bestehende Aktivitäten

Es bestehen heute auf Stufe Bund bereits Instrumente, mit denen sicherheitspolitisch relevante Aspekte bezüglich KI und deren Auswirkungen abgedeckt und angegangen werden:

**Nationale Strategie zum Schutz der Schweiz vor Cyberisiken (NCS):** Die NCS enthält laufende und geplante Aktivitäten und Projekte, die zur Stärkung des Schutzes vor Cyber-Risiken beitragen und auch einen Bezug zu KI haben. Insbesondere mit den Instrumenten MELANI (Melde- und Analysestel-

<sup>79</sup> Für eine ausführliche Fassung siehe Bericht der Projektgruppe «Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

le Informationssicherung) und GovCERT (Governmental Computer Emergency Response Team) sind beim ISB Fähigkeiten vorhanden, um auch neue Risiken im Cyberbereich (inkl. KI) zu analysieren.

**Nationale Strategie zum Schutz kritischer Infrastrukturen (SKI):** Die SKI-Strategie enthält 17 Massnahmen, deren Ziel es ist, den Schutz von kritischen Infrastrukturen zu verbessern und dadurch die Verfügbarkeit von wichtigen Gütern und Dienstleistungen (z.B. Informations- und Kommunikationsdienste) zu gewährleisten. Im Rahmen der Strategie können auch neue, KI-basierte Risiken für die Verfügbarkeit von kritischen Dienstleistungen (z.B. Cyber Supply Chain Risks), wie auch Chancen für einen besseren Schutz (z.B. KI-gestützte Monitoring- und Entscheidungsfindungsprozesse) erfasst werden.

**Aktionsplan Cyberdefence (APCD):** Mit diesem, zum Teil vertraulichen Aktionsplan verstärkt das VBS seine Cyberfähigkeiten systematisch. Neben dem Eigenschutz geht es vor allem darum, die Cyberaspekte des Nachrichtendienstgesetzes (NDG) und des Militärgesetzes (MG) umzusetzen und in der Lage zu sein, Betreiber kritischer Infrastrukturen, welche unter Cyberangriff sind, zu unterstützen.

**Cyberdefence-Campus (CYD-Campus):** Seit Anfang 2019 besteht der im APCD definierte CYD-Campus bei armasuisse W+T. Der CYD-Campus ist eine Plattform zur Antizipation und Früherkennung und zum Monitoring von neuen Technologien, darunter auch KI-Entwicklungen. Der Campus basiert auf einer engen Zusammenarbeit mit den Hochschulen (z.B. ETH) und der Wirtschaft. Das Schweizer Drohnen- und Robotik-Zentrum (SDRZ) untersucht in nationalen und bi- resp. multilateralen Projekten Chancen und Risiken des kombinierten Einsatzes von Robotik und KI für die Sicherheit der Schweiz.

**Internationale Zusammenarbeit:** Das EDA hat seine technologischen und cyberspezifischen aus- sen- und sicherheitspolitischen Aktivitäten verstärkt (u.a. Schaffung des Büros für Cyber-Aussen- und Cyber-Sicherheitspolitik). Auf technischer Ebene pflegt das govCERT des ISB die internationale Zusammenarbeit, um Informationen zur Vorfallobewältigung auszutauschen. Der Nachrichtendienst des Bundes pflegt intensive Kontakte mit Partnerdiensten von Ländern, die ebenfalls als Ziel oder Standort von Angriffsinfrastrukturen von Cyberangriffen betroffen sind. Der Einfluss von KI wird bei dieser raschen, zielgerichteten Zusammenarbeit regelmässig diskutiert. Die Armee kooperiert im Cyberdefence-Bereich bilateral insbesondere mit Nachbarstaaten, wobei KI auch Thema des Austausches ist. Auch multilateral wird die Kooperation verstärkt, z.B. mit der Beteiligung am *Cooperative Cyber Defence Center of Excellence (CCDCOE)* in Tallinn, das sich auch mit dem Thema KI befasst.

#### 6.7.4 Bewertung und Handlungsbedarf

<p><b>Aktionsfeld 1: Aussensicherheitspolitische Auswirkungen</b>                  Der Einsatz von KI wirft Fragen zur strategischen, internationalen Stabilität auf. KI wird als Instrument der politischen Einflussnahme und der Machtprojektion genutzt. Sie prägt das Ungleichgewicht zwischen staatlichen und privaten Akteuren und das allgemeine Wettrüsten.</p>	
<p>Prüfung von aussenpolitischen Implikationen bei Einsatz von KI-gestützten Systemen</p>	<p>Das EDA prüft die aussenpolitischen Implikationen, die mit dem Einsatz von KI-gestützten Systemen einhergehen. Dabei ergeben sich aussenpolitische Fragestellungen in den Bereichen Regulierung, Rolle des Staates, Normen, sowie Menschenrechte; betroffen ist auch die erschwerte Rüstungskontrolle.</p> <p><b>Zuständigkeit:</b> EDA; armasuisse (VBS)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

**Aktionsfeld 2: Bedrohungsformen und Doktrin**

Der Einsatz von KI wirft Fragen zu neuen Bedrohungsformen und zum erhöhten Bedrohungspotenzial von Systemen und Lösungen im militärischen und im Sicherheitsbereich auf, insbesondere der kritischen Infrastrukturen. Der Einfluss von KI und die Veränderung der Bedrohung müssen in den folgenden Domänen überwacht und abgeschätzt werden.

<p>1) Prüfung Cybersicherheit im Rahmen neuer Bedrohungsformen angesichts der Nutzung von KI</p>	<p>Cybersicherheit: Wie wird KI zur Durchführung komplexer Angriffe genutzt und wie wird dadurch die Bedrohungslage verschärft? Welche Risiken entstehen durch den Einsatz von KI in Kombination mit der zunehmenden Vernetzung (IoT)?</p> <p><b>Zuständigkeit:</b> NDB; armasuisse (VBS); Kompetenzzentrum Cyber (EFD); EDA</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>2) Prüfung Propaganda und Informationsoperationen angesichts der Nutzung von KI</p>	<p>Beeinflussungs- und Informationsoperationen: Wie werden Propaganda und Desinformation (z.B. bei Wahlen) angewendet?</p> <p><b>Zuständigkeit:</b> BK; Kompetenzzentrum Cyber (EFD); NDB; armasuisse (VBS); EDA</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>3) Prüfung der Kriegsführung und der Fähigkeiten unterhalb der Kriegsschwelle angesichts der Nutzung von KI</p>	<p>Kriegsführung: Wie werden entscheidunterstützende Systeme mit KI den militärischen Entscheidungsprozess verändern? Wie werden die heutigen relevanten Akteure im Entscheidungsprozess künftig einbezogen, inklusive Spezialisten des Kriegsvölkerrechts? Wie werden Waffen mit zunehmender Autonomie (LAWs<sup>80</sup>) an der Schnittstelle zwischen Robotik und KI die Kriegsführung verändern?</p> <p>Unterhalb der Kriegsschwelle: Welcher Einfluss hat KI auf Aufklärung, La-gebild, Command &amp; Control (C2I) und Wirkung im hybriden Umfeld mit nicht-staatlichen Akteuren?</p> <p><b>Zuständigkeit:</b> Armee; GS-VBS; NDB; armasuisse (VBS); EDA</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>

**Zusätzlicher Handlungsbedarf:** Nein

**Aktionsfeld 3: Fähigkeiten und Kapazitäten**

Der Einsatz von KI prägt sowohl einzelne neue Bedrohungsformen wie auch optimierte Schutz- und Abwehrmöglichkeiten. Die schon definierten Strategien und Aktionspläne (NCS, SKI, APCD) sollen KI als kritischen Entwicklungsfaktor berücksichtigen.

<p>1) Verstärkte Integration und Nutzung von KI-Lösungen in Streitkräften und im nachrichtendienstlichen Bereich</p>	<p>Das VBS verfolgt die Integration und Nutzung von KI-Lösungen in den militärischen Prozessen (Planung, Führung, Logistik) auf allen Ebenen und bei den Waffensystemen (inkl. LAWs). Es fördert die Entwicklung von Fähigkeiten für Schutzmassnahmen und Agilität durch den Einsatz von KI-Technologien in den heutigen Sicherheitsinstrumenten der Schweiz.</p> <p><b>Zuständigkeit:</b> Armee; NDB; armasuisse (VBS)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
----------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<sup>80</sup> Lethal Autonomous Weapons sind Waffensysteme, die entwickelt wurden, um militärische Ziele (Personen, Anlagen) ohne weitere menschliche Einwirkung auszuwählen und anzugreifen.

<p>2) Prüfung Upgrade-Möglichkeiten beim Akquisitionsprozess von kritischen Systemen</p>	<p>Der Akquisitionsprozess von kritischen Systemen soll Upgrade-Möglichkeiten durch potenzielle künftige KI-Funktionalitäten berücksichtigen.</p> <p><b>Zuständigkeit:</b> armasuisse; Armee (VBS)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>3) Vermehrte Berücksichtigung von KI-Komponenten bei Lieferanten- und Subakkordanten von kritischen Systemen</p>	<p>Die Übersicht der gesamten Lieferanten- und Subakkordanten (Cyber Security Supply Chain) von kritischen Systemen (der Armee und der militärischen Verwaltung) soll auch unter Berücksichtigung von KI-Komponenten etabliert werden.</p> <p><b>Zuständigkeit:</b> NDB; armasuisse; Armee (VBS)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>4) Regelmässige Überprüfung des Technologiestands bei Betreibern von kritischen Infrastrukturen</p>	<p>Der Technologiestand der wichtigsten Partner (Betreiber kritischer Infrastrukturen) wird regelmässig überprüft, um die Interoperabilität im Rahmen von KI-Einsatzmöglichkeiten zu gewährleisten.</p> <p><b>Zuständigkeit:</b> armasuisse; BABS (VBS)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

<p><b>Aktionsfeld 4: Antizipation durch Zusammenarbeit, Forschung und Prüfstände</b> Die enge Zusammenarbeit mit den Hochschulen und der Industrie im Rahmen der sozialwissenschaftlichen und technischen Forschung ist ein Schlüsselfaktor für eine erfolgreiche Wahrnehmung des KI-Potenzials im Cybersicherheit- und Cyberabwehrbereich.</p>	
<p>1) Verstärkte Zusammenarbeit mit leitenden Bildungs- und Forschungsinstituten</p>	<p>Das VBS, das WBF und das EDA stärken im Rahmen ihrer Kompetenzen die Zusammenarbeit mit leitenden Bildungs- und Forschungsinstituten und steigern ihre Antizipationsfähigkeiten durch Forschungsprojekte (Förderung einer translationellen Forschung).</p> <p><b>Zuständigkeit:</b> VBS, WBF, EDA</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>2) Berücksichtigung der Entwicklungen im KI Bereich beim Cyberlagebild</p>	<p>Das integrale Cyberlagebild berücksichtigt Entwicklungen im KI Bereich.</p> <p><b>Zuständigkeit:</b> NDB; GS-VBS; armasuisse (VBS); Kompetenzzentrum Cybersicherheit (EFD)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>3) Verstärkte Teilnahme an internationalen Gremien und Forschungsinitiativen im KI Bereich</p>	<p>Durch gezielte Teilnahmen an internationalen Gremien und Forschungsinitiativen kann sich die Schweiz als tragender Partner im KI Bereich positionieren.</p> <p><b>Zuständigkeit:</b> VBS, WBF, EDA</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p>4) Überprüfung der Notwendigkeit, der Bedeutung und des Potentials einer schweizerischen KI-Prüfstelle</p>	<p>Das EFD und das VBS erstellen in Zusammenarbeit mit dem EDA und dem WBF eine Studie zum Bedarf eines nationalen KI-Prüfstands (Aktuelles Umfeld und künftige Herausforderung, Handlungsbedarf und Empfehlungen, inkl. Vor- und Nachteile eines nationalen KI-Prüfstands).</p> <p><b>Zuständigkeit:</b> Kompetenzzentrum Cybersicherheit (EFD); GS-VBS; armasuisse (VBS) in Zusammenarbeit mit EDA und WBF</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

## 6.8 Künstliche Intelligenz, Medien & Öffentlichkeit<sup>81</sup>

### 6.8.1 Übersicht

Öffentlichkeit ist die gesellschaftliche Sphäre, in der sich Menschen über kollektive Belange austauschen.<sup>82</sup> In der Öffentlichkeit werden Informationen, Werte und Ideen sowie Vorstellungen über die Gesellschaft verbreitet und diskutiert. Eine demokratische Gesellschaft ist deshalb auf eine funktionierende Öffentlichkeit als gesellschaftlichen «Raum des Politischen» angewiesen, in dem Debatten über gesellschaftliche Werte und Ziele sowie über die Möglichkeiten und Pfade ihrer Erreichung offen stattfinden können.

Lange Zeit erfolgte die Herstellung von Öffentlichkeit und damit eine wesentliche Voraussetzung der Meinungsbildung zu politischen Themen vorwiegend durch ein mehr oder wenig stabiles Gefüge von **Massenmedien** (Fernsehen, Radio, Zeitungen und Zeitschriften). Diese massenmediale Herstellung von Öffentlichkeit hat sich nicht nur an die allgemeine schweizerische Rechtsordnung zu halten, sondern folgt darüber hinaus auch bekannten, transparenten und gesellschaftlich verhandelbaren Regeln: In der Schweiz allgemein im Medienbereich durch die von der Branche selbst definierten und über den Schweizer Presserat durchgesetzten «Standesregeln»<sup>83</sup>, im Radio- und Fernsehbereich zusätzlich durch die Vorgaben des Bundesgesetzes über Radio und Fernsehen vom 24. März 2006 (RTVG; SR 784.40).

Im Laufe der letzten zehn Jahre haben verschiedene Onlineangebote die herkömmliche Herstellung von Öffentlichkeit strukturell verändert: Suchmaschinen (z.B. Google), soziale Netzwerkplattformen (z.B. Facebook), Multimedia-Plattformen (z.B. Youtube) und Micro-Blogging-Dienste (z.B. Twitter), sind zunehmend relevante Akteure in der öffentlichen Kommunikation. Ihnen gemeinsam ist, dass sie zwar selber keine oder kaum Inhalte produzieren, jedoch als Vermittler zwischen Inhalteproduzent und Konsument auftreten. Daher werden diese Dienstleister von der Wissenschaft als **«Intermediäre»**<sup>84</sup> bezeichnet. Diese selektieren, priorisieren, filtern, aggregieren und verbreiten Informationen. Sie entscheiden zum Beispiel darüber, was Nutzerinnen und Nutzer sehen und was sie nicht sehen. Ähnlich wie traditionelle Massenmedien beeinflussen damit auch Intermediäre die Wahrnehmung der Welt, die Meinungsbildung und menschliches Verhalten. Ihre Auswahllogiken unterliegen jedoch anderen, nicht-journalistischen und nicht transparenten Regeln. Dies birgt Chancen und Risiken für die Kommunikationsgrundrechte.

Massenmedien und Intermediäre greifen unter anderem bei der Auswahl und Distribution von Informationen auf Künstliche Intelligenz (KI) zurück. Die Strukturierung von Öffentlichkeit, die Schaffung sozialer Realität und die Meinungsbildung von Bürgerinnen und Bürgern werden damit bereits heute wesentlich von KI-basierten Diensten mitgestaltet.

### 6.8.2 Herausforderungen

**Herstellung von Öffentlichkeit durch Massenmedien unter Rückgriff auf Künstliche Intelligenz:** Algorithmischer Journalismus und KI-gestützte Anwendungen kommen in allen Prozessbereichen (Aggregation, Produktion und Distribution) von Journalismus zum Einsatz. Und sie können für Inhalte jeglicher Art verwendet werden.<sup>85</sup> Auch in der Schweiz greifen klassische Massenmedien sowie die Nachrichtenagentur Keystone-SDA in ihrer journalistischen Arbeit auf Software zurück, die sich Algorithmen und Künstlicher Intelligenz bedient.<sup>86</sup> Im Bereich der Informationsproduktion beispielsweise nutzen die Nachrichtenagentur Keystone-SDA mit «Lena» und Tamedia mit «Tobi» je unterschiedliche Software, die unter Rückgriff auf die KI-Anwendung «Natural Language Generation» (NLG) Artikel in Deutsch und Französisch verfassen können. Bekannt ist auch der Einsatz von Algorithmen bei den

<sup>81</sup> Für eine ausführliche Fassung siehe Bericht der Projektgruppe «Künstliche Intelligenz, Medien & Öffentlichkeit», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

<sup>82</sup> vgl. Dreyer und Schulz 2019, S. 6.

<sup>83</sup> vgl. Schweizer Presserat: Journalistenkodex.

<sup>84</sup> Auch «Informationsintermediäre», vgl. zur Begrifflichkeit z.B. Dreyer und Schulz 2019, S. 5.

<sup>85</sup> Daten, Text, Bild, Audio, Video; vgl. für eine Zusammenstellung der Anwendungsformen Goldhammer et al. 2019, S. 16.

<sup>86</sup> z.B. Tamedia, NZZ, Ringier Axel Springer, La Liberté; vgl. Goldhammer et al. 2019, S. 21–26.

internationalen Nachrichtenagenturen Agence France-Presse (FR), Austria Presse Agentur (AT), PA Press Association und Thompson Reuters (beide GB), Associated Press und Bloomberg (beide USA).

**Herstellung von Öffentlichkeit durch Intermediäre unter Rückgriff auf Künstliche Intelligenz:** Intermediäre stellen (wie auch klassische Massenmedien) verfügbares Wissen in einer Weise bereit, «die theoretisch vielfältig und objektiv, aber auch einseitig und falsch sein kann»<sup>87</sup>. Je grösser die Anteile auf dem Nutzermarkt, desto grösser sind dabei auch die (je nachdem positiven oder negativen) Potenziale der Intermediäre.

Mit Blick auf die Risiken lässt sich festhalten, dass die Vielfalt des Informationsangebots eines Nutzers oder einer Nutzerin eingeschränkt werden kann. Entweder wenn durch «maschinelles Lernen» eine inhaltliche Vorauswahl getroffen wird und dem Nutzer dadurch in erster Linie stets ähnliche Inhalte angezeigt werden («**Filter Bubbles**»), oder weil solche Verfahren es den Nutzerinnen und Nutzern einfacher machen, unerwünschte Sichtweisen auszuschliessen («**Echo chambers**»). Im Bereich der automatisierten Inhalte-Erkennung wie zum Beispiel bei der Bildererkennung können Fehlbewertungen dazu führen, dass legale Inhalte gelöscht oder ausgefiltert werden («**Zensur**»).

Intermediäre können aber auch direkt in die politische Meinungsbildung eingreifen; zum Beispiel durch das Schalten von anonymer zielgruppenspezifischer politischer Werbung (sogenannten «**Dark Ads**») oder wenn sie mit Aufforderungen zum Wählen experimentieren.<sup>88</sup> Zudem können algorithmische Systeme von «externen» dritten Akteuren ausgenutzt werden: So können zum Beispiel «**Social Bots**» dazu eingesetzt werden, die Multiplikatorfunktionen von Plattformen zur Stärkung von Partikularinteressen oder zur Verschlechterung der Sichtbarkeit oppositioneller Ansichten zu nutzen.<sup>89</sup> Darauf hat auch der Bundesrat in seinem Bericht «Rechtliche Basis für Social Media» hingewiesen.<sup>90</sup> Zu nennen sind in diesem Zusammenhang zudem auch «shit storms» und «hate speech».<sup>91</sup>

Es ist sicher falsch anzunehmen, dass über soziale Medien wahrgenommene Falschinformationen sofort die Einstellung oder sogar das (Wahl- und Abstimmungs-)Verhalten eines Rezipienten verändern würden. Und wie für die Meinungsbildung über Medien gilt auch für jene über Intermediäre, dass Misstrauen in Eliten, Unzufriedenheit mit der Politik, ökonomische Ungleichheiten oder kultureller Ausschluss einen entscheidenderen Einfluss auf die individuelle Wahlentscheidung haben dürfte.<sup>92</sup> Zudem weisen empirische Untersuchungen darauf hin, dass die Nutzung von Social Media die Vielfältigkeit von Diskursen momentan insgesamt eher fördert. Es scheint auch nicht, dass die «gesellschaftliche Hauptthemen-Agenda» durch Intermediäre fragmentiert würde und es gibt derzeit keine belastbaren Hinweise für die Existenz von Filterblasen.<sup>93</sup>

Dennoch ist es evident, dass Intermediäre das (theoretische) Potenzial besitzen, KI-Anwendungen für kommerzielle oder politische Zwecke zu instrumentalisieren oder selbst zu diesen Zwecken instrumentalisiert zu werden. Dadurch kann die öffentliche Meinungs- und Willensbildung, auch im politischen Bereich, beeinflusst werden.<sup>94</sup> Intermediäre folgen in der Regel nicht gemeinwohlorientierten Zielen (z.B. Meinungsvielfalt), sondern sind auf individuelle und ökonomische Interessen ausgerichtet. Zudem birgt ihre kommunikative Macht auch Risiken für die Offenheit gesellschaftlicher Kommunikationsprozesse.<sup>95</sup> Der Bundesrat hat sich in der Vergangenheit bereits zu Fragen der zivilrechtlichen Verantwortlichkeit von Intermediären geäussert. Fragen insbesondere der Rechtsdurchsetzung im Zusammenhang mit Intermediären sind Gegenstand von überwiesenen Vorstössen.<sup>96</sup> Gleichwohl bleiben

<sup>87</sup> Dreyer und Schulz 2019, S. 7, vgl. auch ; Lobigs und Neuberger 2018.

<sup>88</sup> vgl. Fichter 2018.

<sup>89</sup> vgl. Gillespie 2017.

<sup>90</sup> vgl. Schweizerischer Bundesrat 2017.

<sup>91</sup> vgl. Jarren 2018b, S. 36.

<sup>92</sup> vgl. Livingstone 2019; Eidgenössische Medienkommission EMEK 2019.

<sup>93</sup> vgl. Dreyer und Schulz 2019, 11; 16-19.

<sup>94</sup> Dieser Bericht fokussiert auf die Tätigkeiten und Potenziale von Intermediären, die direkt mit der Herstellung von Öffentlichkeit zusammenhängen. Weitere Themen, die mit den Aktivitäten von Intermediären in Zusammenhang stehen, werden hier ausgeklammert und zum Teil von anderen Projektgruppen bearbeitet; so zum Beispiel Fragen im Zusammenhang mit der Datennutzung.

<sup>95</sup> vgl. Saurwein et al. 2017.

<sup>96</sup> Vgl. aus jüngerer Zeit nur Bericht des Bundesrates Die zivilrechtliche Verantwortlichkeit von Providern» vom 11. Dezember 2015, abrufbar unter: <<https://www.ejpd.admin.ch/ejpd/de/home/aktuell/news/2015/2015-12-110.html>> sowie die Motionen 18.3379 RK-S «Zugriff der Strafverfolgungsbehörden auf Daten im Ausland» und 18.3306 Glättli «Rechtsdurchsetzung im Internet stärken durch ein obligatorisches Zustellungsdomizil für grosse kommerzielle Internetplattformen».

grundsätzliche Fragen der rechtlichen und gesellschaftlichen Verantwortung von Intermediären offen. Diese gewinnen angesichts der raschen Veränderungen im Prozess der Herstellung von Öffentlichkeit und der zunehmenden Nutzung von Intermediären durch die Bevölkerung an Aktualität und Tragweite.

### 6.8.3 Bestehende Aktivitäten

**Herstellung von Öffentlichkeit durch Massenmedien unter Rückgriff auf Künstliche Intelligenz:** Nach aktuellem Kenntnisstand sind in der Schweiz in Bezug auf Massenmedien und Künstliche Intelligenz weder spezifische regulatorische Aktivitäten noch entsprechende Forderungen gesellschaftlicher Akteure bekannt.

**Herstellung von Öffentlichkeit durch Intermediäre unter Rückgriff auf Künstliche Intelligenz:** Aus verschiedenen Ländern sind regulatorische Aktivitäten im Bereich der Intermediäre bekannt. In Deutschland sollen künftig auch «Medienplattformen» und «Medienintermediäre» in den Medienstaatsvertrag aufgenommen werden und Social Bots einer Kennzeichnungspflicht unterliegen. In Frankreich wurde ein «Gesetz zur Bekämpfung von Falschinformationen» verabschiedet und im Vereinigten Königreich ist ein Gesetz für soziale Netzwerke mit dem Ziel geplant, ein «sicheres Internet» zu garantieren.

In seinem Bericht zur rechtlichen Basis für Social Media kam der Bundesrat 2017 zum Schluss, «dass derzeit keine zusätzlichen Regulierungsaktivitäten in Bezug auf Social Media ausgelöst werden müssen»<sup>97</sup>. Zwar anerkannte er die Gefahr, die von «Fake News» und «Social Bots» in Bezug auf die Beeinflussung der demokratischen Meinungsbildung ausgehe. Gleichzeitig wollte er aufgrund der insgesamt unübersichtlichen Lage zu diesem Zeitpunkt die Frage nicht beantworten, ob eine staatliche Regulierung angezeigt sei. Stattdessen setzte er auf Selbstregulierung der Branche und Beobachtung der nationalen und internationalen Entwicklungen.<sup>98</sup>

### 6.8.4 Bewertung und Handlungsbedarf

<b>Aktionsfeld 1: Schweizerische Governance im Bereich Intermediäre</b> Angesichts des erheblichen Einflusses von Intermediären ist es angebracht, die Thematik vertieft zu prüfen und einen schweizerischen Governance-Ansatz abzuklären.	
Erarbeitung Governance-Bericht im Bereich Intermediäre	Dem Bundesrat soll bis im Frühling 2021 ein Governance-Bericht vorgelegt werden, der Massnahmen prüft und gegebenenfalls vorschlägt. <b>Zuständigkeit:</b> BAKOM / BK <b>Status:</b> Prüfung zuhanden Bundesrat
<b>Zusätzlicher Handlungsbedarf:</b> Ja	
<b>Aktionsfeld 2: Beobachtung der Entwicklung im Bereich Medien</b> Es stellen sich Fragen bezüglich Nachvollziehbarkeit, Transparenz und Verantwortung aufgrund des Einsatzes von KI im Medienbereich.	
Verfolgen der Entwicklungen zum Einsatz von KI im Medienbereich	Regulatorische Ansätze im Ausland (vgl. z.B. Entwurf der deutschen Bundesländer für einen «Medienstaatvertrag») sind ebenso zu verfolgen wie wissenschaftliche Debatten zu Nachvollziehbarkeit/Transparenz und Verantwortung beim Einsatz von KI im Medienbereich. Dies kann im Rahmen der laufenden Aktivitäten der Verwaltung erfolgen. <b>Zuständigkeit:</b> BAKOM / EDA <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<sup>97</sup> Schweizerischer Bundesrat 2017, S. 52.

<sup>98</sup> ebd., S. 52.



## 6.9 Automatisierte Mobilität und künstliche Intelligenz<sup>99</sup>

### 6.9.1 Übersicht

Der Einsatz von Künstlicher Intelligenz (KI) in der Automatisierten Mobilität ermöglicht es, das grosse Potenzial der Automatisierung<sup>100</sup> zur Verbesserung der Verkehrssicherheit und Steigerung der Effizienz des Verkehrssystems z.B. hinsichtlich Kapazität, Auslastung, Nachhaltigkeit und Finanzierung besser zu erschliessen. Das gilt sowohl für Verkehrsmittel selbst als auch für Verkehrsinfrastrukturen insgesamt und die darauf aufbauenden Zentralsysteme. Der Einsatz von KI optimiert dabei eine grosse Anzahl vernetzter (automatisierter) Verkehrsmittel unter Zeit-, Komfort-, Kosten-, Umwelt- oder Erlebnisgesichtspunkten. Es können gleichermassen individuelle Präferenzen und gemeinschaftliche Ziele verfolgt werden. KI wird bereits heute in der Mobilität eingesetzt z.B. in Fahrzeugen als Bestandteil von Fahrerassistenzsystemen zur Umfelderkennung oder in der Verkehrsinfrastruktur zur Verkehrsflussmessung. Aber ein automatisiertes Fahrzeug mit nur deterministischen Algorithmen kann nicht auf Unerwartetes in komplexen Situation reagieren, also beispielsweise eine Kreuzung in der Innenstadt mit allen möglichen Verkehrsteilnehmern des Individualverkehrs und des ÖV queren. Dazu sind neu selbstlernende Algorithmen erforderlich.

### 6.9.2 Herausforderungen

**Verkehrs- und Betriebssicherheit in einem KI-unterstützten Gesamtverkehrssystem:** In einem vernetzten automatisierten Gesamtverkehrssystem wird das Sicherheitsniveau nicht nur durch das Verkehrsmittel selbst, sondern auch durch andere technische Systeme und betriebliche Abläufe beeinflusst. Umfassende Sicherheitsmassnahmen sind deshalb auch für das Gesamtsystem erforderlich, was eine gesamtheitliche Beurteilung der Sicherheit voraussetzt. Die zukünftigen Rollen und Aufgaben, mithin die jeweilige Verantwortung von Staat und Privatwirtschaft sind offen, insbesondere bezüglich Zulassung von automatisierten Verkehrsmitteln oder Cybersecurity.

Eine Grundvoraussetzung für funktionssichere automatisierte Verkehrssysteme ist, dass die KI ihre Leistungsfähigkeit anhand realitätsnaher Daten systematisch verbessern kann. Die Definition von Regeln zur Erstellung, Erhebung und Zurverfügungstellung solcher Daten kann den Zugang künftig erleichtern sowie ermöglichen und den Schutz vor Manipulation erhöhen. Ob KI-gesteuerten automatisierten Verkehrsmitteln die gleiche Fehlertoleranz wie bei von Menschen gesteuerten zugestanden werden soll, ist gesellschaftlich weitgehend noch offen und birgt politische Hürden.

**Datenschutz und Durchsetzung gesetzlicher Vorgaben im Verkehr:** Der Einsatz von KI bietet vielfältige Möglichkeiten zur Auswertung der weitgehend ohne Zusatzkosten verfügbaren Daten in der Mobilitätswelt, die nahezu alle Aktivitäten ihrer Nutzer erfasst und speichert. Dabei stellen sich Fragen von Eingriffen in die persönliche Freiheit und in den Datenschutz. Ihre Klärung kann nicht dem Markt überlassen werden, sondern muss gesellschaftlich breit ausgehandelt werden. Zudem ist offen, wie weit der Staat diese Daten zur Durchsetzung von gesetzlicher Vorgaben verwenden soll respektive darf.

**Effizientes Verkehrs- und Mobilitätsmanagement:** Es wird davon ausgegangen, dass in Zukunft KI die Verkehrsmittel steuert. Angestrebt wird dabei eine gesellschaftlich-politische und wirtschaftliche Balance hinsichtlich der Regeln, nach denen die KI individuelle Wünsche der Verkehrsteilnehmenden sowie kollektive Bedürfnisse der Gesellschaft optimiert. In der vernetzten Automatisierten Mobilität steigt die Menge der verfügbaren Daten exponentiell: einerseits wegen der hohen Leistungsfähigkeit der KI-basierten Steuerungssysteme, andererseits wegen neu möglichen Mobilitätsdiensten wie Sharing. Darum sind verschärfte Diskussionen um Datenschutzbelange und Nutzungsrechte an diesen Daten wahrscheinlich.

---

<sup>99</sup> Für eine ausführliche Fassung siehe den Bericht der Projektgruppe «Automatisierte Mobilität und künstliche Intelligenz», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

<sup>100</sup> In der Schweiz wie auch in Europa benutzen die Behörden den Term «automatisiert» anstelle von «autonom», um hervorzuheben, dass automatisierte Verkehrsmittel vernetzt sein müssen. Siehe auch Bericht des Bundesrates «Automatisiertes Fahren», Dez 2016 oder «Europe on the Move III» u.a. COM/2018/283.

**Rechtlicher Status automatisierter Systeme in der Mobilität:** Systeme der Automatisierten Mobilität werden auch negative Folgen nach sich ziehen. So werden beispielsweise hoch- und vollautomatisierte Fahrzeuge in Unfälle verwickelt werden. Der Bundesrat hat sich mit diesen Fragen bereits befasst und sieht derzeit keinen Handlungsbedarf (vgl. Abschnitt 4.2).<sup>101</sup> Hingegen soll die Entwicklung insbesondere international weiterhin aufmerksam verfolgt werden.

### 6.9.3 Bestehende Aktivitäten

Das **ASTRA** für den Strassenverkehr und das **BAV** für den Bahnverkehr verfolgen die Entwicklungen im Bereich der automatisierten Mobilität, berichten darüber und bewilligen Versuche. Sie evaluieren basierend auf den Erkenntnissen Regulierungsbedarf für die Ermöglichung von Automatisierter Mobilität und der Kontrolle deren Risiken. Das **BAKOM** sorgt für stabile und leistungsfähige Telekommunikationsinfrastrukturen, welche eine zentrale Grundvoraussetzung für ein KI-basiertes Mobilitätssystem bilden. Das BAKOM verfolgt die nationalen und internationalen Entwicklungen, namentlich auch im Bereich «Internet of Things (IoT)». Das BAV setzt seine Massnahmenpläne für Multimodale Mobilitätsdienstleistungen um. Das **BAZL** nimmt sich den Herausforderungen beim Einsatz von unbemannten Luftfahrzeugen an.

### 6.9.4 Bewertung und Handlungsbedarf

Grundsätzlich gleichen die beschriebenen Herausforderungen denen der Digitalisierung – mit KI in der Automatisierten Mobilität werden sie aber weiter verschärft. Sie werden gegenwärtig von allen Beteiligten kontrovers diskutiert. Die öffentliche Hand muss zusätzliche Ressourcen einsetzen, um mit diesen Entwicklungen Schritt halten zu können. Dies aus gesamtgesellschaftlicher Sicht mit dem Ziel, die Chancen der neuen Technologien zu nutzen und die Risiken frühzeitig zu minimieren.

Für die derzeit absehbaren grundlegenden Fragen zum Einsatz von KI in der Automatisierten Mobilität sind die zuständigen Ämter sensibilisiert. Verschiedene Arbeiten dazu laufen. Im Zusammenhang mit KI gibt es grundsätzlich keinen zusätzlichen Handlungsbedarf. Hingegen kann aus den laufenden Massnahmen und der laufenden Entwicklung neuer Handlungsbedarf erwachsen.

#### Aktionsfeld 1: Einsatz von KI in automatisierten Fahrzeugen

Der Einsatz von KI ist in automatisierten Fahrzeugen nötig, damit sie im parallelen Betrieb mit konventionellen Fahrzeugen (von Menschen gesteuert) auf der bestehenden Infrastruktur sicher eingesetzt werden können.

1) Koordination bestehende Arbeiten bezüglich automatisierten Fahrzeugen	ASTRA und BAV verfolgen diesbezüglich die nationale und internationale Entwicklung und koordinieren die Arbeiten bezüglich automatisierten Fahrzeugen ( <i>siehe hierzu den Bericht des Bundesrates «Automatisiertes Fahren – Folgen und verkehrspolitische Auswirkungen» (Dez 2016)</i> ). <b>Zuständigkeit:</b> ASTRA / BAV <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
2) Erarbeitung Luftverkehrsmanagement für Unmanned Aircraft Systems (UAS)	Erarbeitung eines speziellen Luftverkehrsmanagements für UAS (Unmanned Aircraft Systems). <b>Zuständigkeit:</b> BAZL <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
3) Abklärungen für Pilotprojekt zur Unterstützung der Steuerung des Bahnrespektive des Strassenverkehrs	Durchführung von Abklärungen für eine allfällige Pilotanwendung (Künstliche Intelligenz in den Betriebszentralen / im Verkehrsmanagement). <b>Zuständigkeit:</b> BAV / ASTRA <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<sup>101</sup> Antworten auf 18.3445 Ip. Marchand-Balet «Automatisierte Fahrzeuge und Haftung. Wann wird die Gesetzgebung in der Schweiz angepasst?» und 17.3040 Po. Reynard «Die Schaffung einer Rechtspersönlichkeit für Roboter prüfen».

<b>Aktionsfeld 2: Zwingender Datenaustausch für KI in Automatisierter Mobilität</b> Der Datenaustausch zwischen allen Verkehrsteilnehmern und der Infrastruktur wird forciert, damit KI in der automatisierten Mobilität optimal funktionieren kann.	
1) Umsetzung bestehender Massnahmenpläne	Umsetzung der bestehenden Massnahmenpläne, insbesondere bzgl. Multimodalen Mobilitätsdienstleistungen und Aufbau einer Verkehrsdatenplattform (siehe Berichte «Multimodale Mobilitätsdienstleistungen; Massnahmenpläne: Mobilitätsdaten und Öffnung Vertrieb weiterer Mobilitätsanbieter ausserhalb des öV» (Dez 2018) und «Bereitstellung und Austausch von Daten für das automatisierte Fahren im Strassenverkehr» (Dez 2018)).  <b>Zuständigkeit:</b> BAV / ASTRA <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
2) Aufbau «Verkehrsnetz Schweiz» zur Geolokalisierung	Aufbau eines «Verkehrsnetzes Schweiz» zur Geolokalisierung.  <b>Zuständigkeit:</b> swisstopo <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
3) Abklärungen für Pilotprojekt mobile breitbandige Sicherheitskommunikation (MSK)	Durchführung von Abklärungen für ein allfälliges Pilotprojekt für eine mobile breitbandige Sicherheitskommunikation (MSK).  <b>Zuständigkeit:</b> VBS (BABS) <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<b>Aktionsfeld 3: Datenschutz in Automatisierter Mobilität</b> Die vielfältigen Nutzungsmöglichkeiten der KI im erweiterten Umfeld der Mobilität werden unter gleichzeitiger Einhaltung der datenschutzrechtlichen Vorgaben gewahrt.	
Koordinationsnetzwerk mit dem EDÖB	Der Bund koordiniert eine enge regulatorischer Begleitung. Das ASTRA und das BAV klären die weiteren Schritte ab, insbesondere bezüglich Erreichen von Mobilitätszielen und Durchsetzung gesetzlicher Regelungen.  <b>Zuständigkeit:</b> ASTRA / BAV <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<b>Aktionsfeld 4:                      Rechtssetzung und gesellschaftliche Akzeptanz von KI in Automatisierter Mobilität</b> Schaffung eines Rechtsrahmens für KI in der Automatisierten Mobilität samt Förderung der gesellschaftlichen Akzeptanz.	
1) Bewilligungen für automatisiertes Fahren	Mit Bewilligungen für automatisiertes Fahren werden die Akzeptanz von KI in der Automatisierten Mobilität gefördert und Erkenntnisse, insbesondere für die Revision des Strassenverkehrsgesetzes (SVG) und des Eisenbahngesetzes (EBG) gewonnen.  <b>Zuständigkeit:</b> ASTRA / BAV <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
2) Koordination Zulassungsverfahren für automatisierte Fahrzeuge	Die Ämter des UVEK koordinieren Zulassungsverfahren für automatisierte Fahrzeuge.  <b>Zuständigkeit:</b> UVEK <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen

## Herausforderungen der künstlichen Intelligenz

3) Abklärung Fehlertoleranzen von KI	Das ASTRA und das BAV klären ab, welche Fehlertoleranzen KI in der Automatisierten Mobilität zugestanden werden sollen. <b>Zuständigkeit:</b> ASTRA / BAV <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
4) Koordinationsnetzwerk im Bereich Recht und Internationales	Die Ämter des UVEK stellen die rechtliche und internationale Zusammenarbeit sicher. <b>Zuständigkeit:</b> UVEK <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## 6.10 Künstliche Intelligenz im Gesundheitswesen

### 6.10.1 Übersicht

Die Digitalisierung verändert auch das Gesundheitswesen rasant: noch nie gab es so viele Möglichkeiten, Gesundheitsdaten zu erheben, zu sammeln und auszuwerten. Die exponentielle Rechenleistung und neue Ansätze der Informationsverknüpfung und Verarbeitung erleichtern die Auswertung dieser Datenmengen wesentlich. Gesundheitsdaten wird ein grosses Potenzial zugeschrieben, welches – richtig extrahiert – nicht nur für die medizinische Forschung relevant sein kann, sondern auch zu einer effizienten und individuell optimalen Gesundheitsversorgung sowie zu einer Verbesserung der öffentlichen Gesundheit beitragen kann. Künstliche Intelligenz (KI) und immer komplexere Algorithmen bergen auch für den Gesundheitsbereich grosse Chancen. Durch die verschiedenen Ansätze der KI werden Verbesserungen auf allen Ebenen prophezeit: hinsichtlich der Versorgungsqualität, des Angebots von Gesundheitsleistungen, des persönlichen Nutzens am Patienten/Patientin sowie der Kosteneffizienz.

### 6.10.2 Herausforderungen

Die Herausforderungen liegen einerseits in der korrekten Einschätzung der Chancen, welche die Digitalisierung für das Gesundheitssystem mit sich bringt. Andererseits gilt es, die möglichen Risiken früh zu erkennen und zu eliminieren oder zumindest zu minimieren.

Die *Chancen* der Verarbeitung und Auswertung von Gesundheitsdaten, der Umsetzung der Erkenntnisse in die Praxis und der Weiterentwicklung der Konzepte und Ansätze der datengetriebenen Medizin liegen unter anderem

- in einer verbesserten Prävention und Gesundheitsförderung durch eine frühere und umfassendere Identifizierung von diversen Risikofaktoren und die Entwicklung von Massnahmen für eine Veränderung des Gesundheitsverhaltens,
- in einer effizienteren und mehr patientenzentrierten Versorgung durch die Vermeidung von ineffektiven oder gar schädlichen Therapien und im Einsatz neuer, individuell ausgerichteter Therapien (z. B. in der Onkologie),
- in einer Verbesserung der epidemiologischen Überwachung und zur Vorhersage von Epidemien,
- in einer Erhöhung der Arzneimittelsicherheit durch die Verbesserung der Überwachung und eine Reduktion der Nebenwirkungen durch den gezielteren Einsatz von Arzneimitteln sowie durch neue Diagnostika,
- in einer Verbesserung des Kosten-Nutzen-Verhältnisses im Gesundheitssystem durch Ermittlung der Überversorgung und Gestaltung einer bedarfsgerechten Versorgung sowie durch Effizienzsteigerung.

Die häufigsten *Risiken* der datengetriebenen Medizin betreffen den Daten- und Persönlichkeitsschutz. Je umfassender Datensätze mit Personendaten sind, desto schwieriger lassen sie sich anonymisieren. Zwar lassen Forschungsergebnisse, die aus solchen Datensätzen gewonnen werden, in der Regel keine Rückschlüsse auf die untersuchten Personen zu. Dennoch besteht immer die Gefahr, dass - aufgrund von Sicherheitslücken oder mutwillig - Personendaten in falsche Hände geraten. Hinzu kommt, dass mit der raschen Weiterentwicklung der Analysetechniken (z.B. Hochdurchsatzsequenzierung des Genoms) sowie mit neuen digitalen Technologien und den Möglichkeiten zur Verknüpfung von Datensätzen die Möglichkeit entsteht, Anonymisierung und Verschlüsselung von Daten mit geringem Aufwand zu unterlaufen. Ein absoluter Schutz dieser Daten ist aus technischer Sicht nicht möglich. Die ungewollte Weitergabe von Gesundheitsdaten birgt Risiken, die vor allem die Bereiche Identität, Privatsphäre, Eigentums- und Persönlichkeitsrechte sowie Diskriminierung tangieren.

Einige Faktoren können sich auch negativ auf die Güte/Korrektheit der gewonnenen Erkenntnisse auswirken und stellen damit ebenfalls Risiken dar. Im Fall von Künstlicher Intelligenz oder *machine-learning*-Systemen im Bereich der Datenanalyse sind die zugrundeliegenden Algorithmen und Entscheidungs-Regeln oft evolutiv entstanden und entgehen einer Analyse oder Überprüfung von aussen (*black box*). Eine so generierte Erkenntnis kann somit nicht ohne Weiteres durch Analyse der Entstehung auf ihre Richtigkeit überprüft werden. Lernfähige Systeme können auch „falsch“ lernen.

### 6.10.3 Bestehende Aktivitäten

Das BAG befindet sich hinsichtlich der datengetriebenen Medizin in einem Spannungsfeld: auf der einen Seite hat es für den Schutz der Persönlichkeit und des Selbstbestimmungsrechts von Patientinnen und Patienten sowie Gesunden Sorge zu tragen. Auf der anderen Seite setzt der Bundesrat in seinen gesundheitspolitischen Prioritäten darauf, die Versorgungsqualität zu erhöhen sowie die Gesundheit durch Effizienzsteigerungen bezahlbar zu halten. Für alle Aspekte bergen Digitalisierung im Allgemeinen und die datengetriebene Medizin sowie die künstliche Intelligenz grosse Chancen, bringen aber auch Risiken mit sich. Das BAG anerkennt, dass eine generelle Verbesserung der Standardisierung und der Verfügbarkeit von Gesundheitsdaten die Forschung mit personenbezogenen Gesundheitsdaten verbessern und fördern würde und damit auch den Nutzen der Daten im Hinblick auf die Versorgung und die Gesundheitspolitik erhöhen könnte – gleichzeitig gilt es jedoch die informationelle Selbstbestimmung der betroffenen Personen, den Datenschutz und die Datensicherheit zu stärken.

### 6.10.4 Bewertung und Handlungsbedarf

Die oben erwähnten Herausforderungen dürften das schweizerische Gesundheitswesen und die Gesundheitspolitik in Zukunft vermehrt beschäftigen. Das BAG wird vor diesem Hintergrund die Auswirkungen der KI auf die Medizin und die Gesundheitsversorgung beobachten und gegebenenfalls dem Bundesrat Vorschläge zur Anpassung der relevanten bundesrechtlichen Grundlagen unterbreiten.

Aktionsfeld 1: Humanforschung	
1) Prüfung gesetzliche Grundlagen im Bereich Weiterverwendung von Proben und Daten sowie Biobanken	<p>Im Rahmen der Evaluation des Humanforschungsgesetzes HFG (Ende 2019) wird abgeklärt, ob die gesetzlichen Grundlagen hinsichtlich Weiterverwendung von Proben und Daten angesichts der aktuellen Entwicklungen zeitgemäss sind; sowie hinsichtlich Biobanken, ob der Schutz der Beteiligten, die Freiheit der Wissenschaft und die öffentliche Gesundheit gesichert sind.</p> <p><b>Zuständigkeit:</b> BAG                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
2) Einheitliche Positionierung im Bereich Datennutzung und Big Data	<p>Hinsichtlich der Nutzung von Gesundheitsdaten für die Forschung (insbesondere im Bereich Big Data) wird sich das BAG deutlicher positionieren und mit anderen Bundesstellen abstimmen (z.B. im Rahmen der Strategie Digitale Schweiz), damit die Patientinnen und Patienten bzw. Bürgerinnen und Bürger ihre Rechte bzgl. Privatsphäre, Datenhoheit bestmöglich wahren und über deren Verwendung aufgeklärt bestimmen können.</p> <p><b>Zuständigkeit:</b> BAG                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

Aktionsfeld 2: Heilmittelgesetz HMG	
Die KI wird ihr Potenzial für die Verbesserung der Gesundheitsversorgung erst dann voll entfalten können, wenn sie nicht nur im Rahmen von Forschungsprojekten, sondern im klinischen Prozess zur Anwendung gelangt. Das HFG deckt nur einen Teil ab, Das HMG wird wohl langfristig wichtiger werden.	
Prüfung von zukünftigen Lösungsansätzen in der Arzneimittelentwicklung	<p>Das BAG und Swissmedic prüfen, mit welchen zukunftsorientierten Lösungsansätzen dem präzisionsmedizinischen Trend in der Arzneimittelentwicklung angemessen begegnet werden kann.</p> <p><b>Zuständigkeit:</b> BAG / Swissmedic <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## 6.11 Künstliche Intelligenz in der Finanzwirtschaft

### 6.11.1 Übersicht

Die Finanzbranche setzt die Methoden der künstlichen Intelligenz zum Beispiel bei der Erkennung von unrechtmässigen Kreditkartentransaktionen schon seit längerem ein, die Anwendungsgebiete<sup>102</sup> dehnen sich nun aber aus. Der Nutzen der KI besteht darin, dass die Kosten sinken, weil sich arbeitsintensive Abläufe in erheblichem Ausmass automatisieren oder beschleunigen lassen. Zudem könnten Produkte entstehen, die dem Kunden nützlicher, einfacher, günstiger oder stärker auf ihn zugeschnitten erscheinen. Genauere Kenntnisse über den Kunden kann in der Tendenz präzisere Entscheide zur Folge haben. Die KI dürfte die Allokation von Ressourcen verbessern, denn wenn ein gegebener Franken einem schlechteren Vorhaben verwehrt wird, steht er einem besseren Projekt zu Verfügung.

### 6.11.2 Herausforderungen

Die Finanzindustrie ist wegen des notwendigen Kundenschutzes und Stabilitätsanforderungen stärker als viele andere Branchen reguliert. Die Wahrscheinlichkeit, dass Anwendungen einer neuen Technologie mit der bestehenden Regulierung kollidiert oder aber explizit in diese eingearbeitet werden muss, ist deshalb grösser. Daraus folgt, dass der Einsatz der KI in der Finanzwirtschaft genauer zu verfolgen ist als in anderen Wirtschaftszweigen. Das generelle Phänomen, dass die komplexeren Varianten der KI-Methoden die Resultate im Einzelfall kaum erklären können, tritt auch in den Anwendungen im Finanzsektor auf. Das kann dazu führen, dass erstens Fehlentscheidungen eines Algorithmus nicht erkannt werden und zweitens ein Finanzdienstleister eine gesetzliche Verhaltens-, Verantwortlichkeits-, Informations- oder Rechenschaftspflicht<sup>103</sup> gegenüber einem Kunden nicht zu erfüllen vermag.

Falls die Daten, die in der Lernphase einer KI-Anwendung benutzt werden, eine **Diskriminierung** widerspiegeln, wird das System diese übernehmen und anwenden. Der Grund für eine Diskriminierung können Einzelfallentscheidungen, aber auch andere und sogar zufällige Artefakte sein. Gegebenenfalls leidet nur das Finanzinstitut unter der negativen Auswirkung, zum Beispiel, wenn deswegen die Investitionen in bestimmte Anlagekategorien fälschlicherweise zu hoch oder zu tief ausfallen (Fehlallokation). Regulatorisch bedeutsam wäre hingegen, wenn ein Finanzdienstleister die Kunden falsch beriete, die gesetzlichen Eignungs- und Angemessenheitsprüfungen fehlerhaft durchführte oder einen Versicherten ohne juristischen oder versicherungstechnischen Grund benachteiligte.<sup>104</sup>

Der Umstand, dass KI-Algorithmen bereits wegen unmerklich kleiner Veränderungen an den Eingangsdaten zu anderen Ergebnissen und Prognosen gelangen können<sup>105</sup>, kann missbraucht werden. Er eröffnet Drittpersonen die Möglichkeit, mit einer für Menschen kaum sichtbaren **Manipulation der Daten** ein anderes Verhalten der KI-Systeme der Finanzhäuser zu bewirken. Es ist in Zukunft nicht ausgeschlossen, dass die KI in betrügerischer Absicht mit unerkannten Datenänderungen beispielsweise dazu gebracht werden kann, ungerechtfertigte Kredite zu vergeben, Sponsorengelder zu sprechen, Überweisungen auszuführen, Vorgänge inkorrekt zu verbuchen, Versicherungsrisiken als zu tief einzuschätzen und Kundenidentitäten im elektronischen Verkehr falsch zuzuordnen. Solches wäre als ein operationelles Risiko aufsichtsrechtlich relevant.

---

<sup>102</sup> Z.B. Identität des Kunden feststellen; E-Mails katalogisieren; Kunden-Chatbot; individueller Vorschlag von Anlage-, Bank- oder Versicherungsprodukten; Zuordnung des Kunden in eine Serviceklasse; individuelle Kreditwürdigkeit eines Kunden; Versicherungsprämie; tagesgenaue Prognose der Kontobewegungen; relevante Abschnitte in Dokumenten finden; Kursprognose; illegitime Transaktionen anhand von Anomalien erkennen.

<sup>103</sup> Beispielsweise die Pflicht gemäss FIDLEG, Art. 15, Abs. 2, die Gründe für jede Empfehlung, die zum Erwerb oder zur Veräusserung eines Finanzinstruments führen, zu dokumentieren und dem Kunden zugänglich zu machen. Siehe auch den Erwägungsgrund 71 zum Artikel 22 der EU-Datenschutzgrundverordnung mit dem Recht auf Erklärung.

<sup>104</sup> Finanzdienstleistungsgesetz FIDLEG, Art. 6 bis 16. Aufsichtsverordnung AVO Art. 117.

<sup>105</sup> Su, Vargas und Kouichi (2017): "One pixel attack for fooling deep neural networks", arXiv:1710.08864 [cs.LG], <https://arxiv.org/abs/1710.08864>

Der Konflikt zwischen dem **Schutz von personenbezogenen Kundendaten** und dem Datenbedarf der KI-Systeme könnte bei Banken geringer ausfallen als im Versicherungswesen, da Bankprodukte zwar individueller werden, aber weniger als ein Versicherungsprodukt von einer Flut von Personendaten profitieren. Bei einem gelockerten Datenschutz oder einem freiwilligen Herausgeben von Daten besteht die Möglichkeit, dass Versichertenprämien in Zukunft von zum Beispiel Einkäufen von Lebensmitteln und Büchern, von der Nutzung des Internets und elektronischer Geräte, vom Reiseverhalten, vom täglichen Muster des Arbeitsweges und von vielem mehr abhängt.<sup>106</sup>

Die KI eröffnet den Versicherern im Zusammenspiel mit grossen Datenmengen die Möglichkeit, die höheren von den tieferen Risiken präziser zu unterscheiden. Das Resultat sind sogenannte risikogerechte, auf die spezifischen Eigenschaften eines zu versichernden Risikos abgestimmte Versicherungsprämien. Zwar benutzen die Versicherungsunternehmen dazu bereits seit langem ausgefeilte statistische Methoden. Aber erst mit der KI-Technik lässt sich der Prämientarif auf eine grosse Zahl von Datenmerkmalen und ihre nichtlinearen Zusammenhänge abzustützen. In der *Sozialversicherung*, z.B. der obligatorischen Krankenpflegeversicherung nach KVG, ist eine Solidarität zwischen den Versicherungsnehmern bezüglich der Prämienhöhe in der Regel gesellschaftlich erwünscht<sup>107</sup>, eine vollständig risikogerechte Prämie wäre damit nicht verträglich. Im *Privatversicherungswesen*, abgesehen von den Risiken in der beruflichen Vorsorge und in der Krankenzusatzversicherung, greift der Staat nicht in die Bildung der Versicherungsprämie ein. Dem individuellen Preis jedes Risikos, sofern er denn korrekt bestimmt werden kann, steht aufsichtsrechtlich nichts entgegen. Bemerkenswert ist, dass die Prämienindividualisierung das Versicherungskonzept des Risikoausgleichs im Kollektiv *nicht* beeinträchtigt. Die Risikodiversifikation in der Gefahrengemeinschaft bleibt vollumfänglich bestehen, denn diese entsteht nicht durch das Einebnen der Prämie, sondern durch die Unabhängigkeit und Zufälligkeit der Schadeneintritte bei den einzelnen versicherten Risiken.

Versicherer kommen mit der KI nicht nur als Anwender in Berührung, sondern auch durch das Versichern von Firmen, welche die KI einsetzen oder KI-Systeme verkaufen, und von Produkten, die eine KI enthalten. Dabei stellen sich einige der (haftpflicht-)rechtlichen Fragen, die bereits an einleitender Stelle erwähnt werden, erneut.

### 6.11.3 Bestehende Aktivitäten

Das EFD verfolgt die Entwicklungen der KI in der Finanzwirtschaft, ohne bisher KI-spezifische regulatorische Schritte als notwendig erachtet zu haben. Die Finanzmarktaufsicht beaufsichtigt nach den Finanzmarktgesetzen die unterstellten Beaufsichtigten, z.B. Versicherungsunternehmen und Banken.

### 6.11.4 Bewertung und Handlungsbedarf

Die erwähnten Herausforderungen, z.B. das tiefe Niveau an Erklärbarkeit, gewisse Ausprägungen einer möglichen Diskriminierung, das Missbrauchspotenzial mit modifizierten Eingabedaten als operationelles Risiko, Fragen des Datenschutzes insbesondere in der Versicherungsbranche, dürften sowohl den Gesetzgeber als auch die Aufsicht in Zukunft mindestens punktuell beschäftigen. Das EFD wird vor diesem Hintergrund die Einsatzgebiete der KI bei den Finanzdienstleistern weiterverfolgen und aufkommende Fragen im Rahmen ordentlicher Regulierungsrevisionen bearbeiten. Eine KI-spezifische Änderung der Finanzmarktregulierung ist aktuell nicht vorgesehen.

---

<sup>106</sup> Lebensmittel: Gemüse in bar und Süssigkeiten elektronisch zu bezahlen, könnte die Versicherungsprämie erhöhen, weil die KI eines Versicherers nur letzteres registrieren wird. / Bücher: Aus dem Leseverhalten lassen sich möglicherweise Informationen über den sozialen Status, das Zahlungsvermögen und die Persönlichkeit gewinnen. / Mobiltelefone: (i) Die Daten der Beschleunigungssensoren und Gyroskope der Smartphones verraten, ob der Träger Treppen steigt oder den Lift benutzt. (ii) Aus Bewegungsdaten könnte auf Schlafmangel geschlossen werden. (iii) Flugpassagiere sind wegen der Flughöhe nur schwach gegen die kosmische Strahlung abgeschirmt, Versicherer könnten dies in Verbindung mit der Gesundheit bringen.

<sup>107</sup> Beispielsweise, weil die Versicherungsdeckung für alle Personen gesellschaftlich gewünscht ist oder gewisse Tätigkeiten, Regionen oder Bevölkerungsgruppen mit erhöhtem Risiko staatlich unterstützt werden sollen.



**Aktionsfeld 1: Verhaltenspflichten**

Es ist nicht ausgeschlossen, dass zukünftige Anwendungen der KI im Finanzwesen mit bestehenden regulatorischen Verhaltenspflichten der Finanzmarktakteure kollidieren.

Weiterverfolgung der Entwicklungen bei Verhaltenspflichten	Das EFD verfolgt die Entwicklung der KI-Anwendung in Finanzwirtschaft und einen potentiellen Bedarf einer Anpassung der Verhaltenspflichten. Vorläufig sind keine Massnahmen geplant. <b>Zuständigkeit:</b> EFD <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Zusätzlicher Handlungsbedarf:** Nein

**Aktionsfeld 2: Operationelle Risiken**

Anwendungen der KI können bei Finanzmarktinstituten zu aufsichtsrechtlich relevanten operationellen Risiken führen.

Weiterverfolgung der Entwicklungen operationeller Risiken bei Finanzmarktinstituten	Das EFD verfolgt die Entwicklung der KI-Anwendungen. In der Regulierung sind die operationellen Risiken hinreichend generell beschrieben, eine Regulierungsänderung erscheint deshalb nicht notwendig. <b>Zuständigkeit:</b> EFD <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
-------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Zusätzlicher Handlungsbedarf:** Nein

**Aktionsfeld 3: Bestimmung der Versicherungsprämie in der Privatversicherung**

Ein unbedachter Umgang mit den Methoden der KI sowie Artefakte in den Daten könnte versicherungsmathematisch ungerechtfertigte Diskriminierungen in der Prämienstruktur hervorrufen.

Weiterverfolgung der Entwicklungen bei Versicherungsprämien in der Privatversicherung	Staatlich geprüft werden Privatversicherungsprämien bei den Risiken in der beruflichen Vorsorge und in der Zusatzversicherung zur sozialen Krankenversicherung. Die Aufsichtsbehörde hat dabei auch die Aufgabe, die Versicherten gegen Missbrauch zu schützen. Eine Regulierungsänderung erscheint nicht notwendig. <b>Zuständigkeit:</b> EFD <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
---------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Zusätzlicher Handlungsbedarf:** Nein

## 6.12 Künstliche Intelligenz in der Landwirtschaft

### 6.12.1 Übersicht

Auch in der Landwirtschaft wird KI vermehrt eingesetzt. Die Verbesserung der Genauigkeit kognitiver Computertechnologien wie der Bilderkennung verändert die Landwirtschaft, die sich traditionell auf die Augen und Hände von erfahrenen Landwirten verlassen hat, um z.B. das richtige Obst zum richtigen Zeitpunkt zu pflücken. So können Ernte-Roboter, die mit KI-Technologie ausgestattet sind, in Echtzeit Entscheidungen treffen, welche Handlungen selbstständig auszuführen sind – basierend auf Daten von Kameras und Sensoren. Diese Art von Roboter ist zunehmend in der Lage, (selbst komplexere) landwirtschaftliche Aufgaben zu übernehmen, die bis anhin menschliche Arbeit erfordert haben.

Verschiedene Forschungsinstitutionen und Technologie-Start-ups nicht zuletzt auch in der Schweiz sind dabei, innovative Lösungen zu untersuchen und entwickeln, um KI in der Landwirtschaft vermehrt zu nutzen. Dabei wird KI-Technologie unter anderem in folgenden Anwendungsgebieten eingesetzt:

**Agrar-Roboter:** Solche Roboter übernehmen zur Entlastung der Landwirte wesentliche landwirtschaftliche Aufgaben wie die Ernte von Nutzpflanzen oder die Bekämpfung von Schadpflanzen. Im Vergleich zu menschlichen Arbeitskräften sind diese (selbstlernenden) Roboter zunehmend schneller und produktiver. So hat das Schweizer Start-up [ecoRobotix](#) eine erste vollständig autonome Maschine für eine umweltschonendere und wirtschaftlichere Unkrautbekämpfung basierend auf Bilderkennung entwickelt.

**Überwachung von Nutzpflanzen und Böden:** Algorithmen aus den Bereichen «Computer Vision» und «Deep Learning» werden genutzt, um Daten zum Zustand von Nutzpflanzen und Böden zu erheben und zu verarbeiten. Die steigende Verfügbarkeit von Satellitendaten ermöglicht dabei eine immer bessere Überwachung. Der Nutzen für den Landwirt liegt dabei z.B. im gezielteren Einsatz von Wasser, Saatgut, Dünge- und Pflanzenschutzmitteln sowie beim Festlegen des optimalen Erntezeitpunktes. Ein Forschungsprojekt an der [ETH Zürich](#) zum Zustand der landwirtschaftlich genutzten Flächen in der Schweiz kombiniert dazu Fernerkundung mit maschinellem Lernen zur Bildinterpretation.

**Überwachung von Nutztieren:** Das selbstlernende, auf KI basierte System [SmartBow](#) aus Österreich erlaubt, mittels fortlaufender Ortung von Kühen und deren Ohrbewegungen individuell auf das Aktivitäts- und Fressverhalten der Tiere zu schliessen. Die Daten werden online sofort ausgewertet und der Landwirt erhält in Echtzeit Unterstützung in Form von Grafiken und Alarmen, die auffälliges Verhalten, Abkalbetermine oder Brünstigkeit der Tiere anzeigen. Dieses System ist in der Schweiz bei [Agroscope](#) im Einsatz.

**«Predictive Analytics»:** Die Verwendung von Modellen aus dem Bereich des maschinellen Lernens ermöglicht es, die Auswirkungen von Umweltfaktoren wie sich ändernde Wetterbedingungen z.B. auf den Ernteertrag zu verfolgen oder gar vorherzusagen. Der Landwirt profitiert von solchen modellbasierten Vorhersagen durch eine gesteigerte Produkterentabilität und eine ressourcenschonendere Produktion. Das Schweizer Start-up [Gamaya](#) bietet derartige Lösungen basierend auf hyperspektral Daten und «Big Data»-Analysen an.

**Saatgut- und Züchtungsforschung:** Die Verwendung von KI erlaubt die visuelle Keimlingsbeurteilung und Samenidentifikation via phänotypische Charakterisierung. Mittels KI-basierter Analysetools und Algorithmen können so für neue Sorten Vorhersagen von Pflanzeigenschaften auf Basis des gesamten Genoms und grosser Phänotypen-Datensätze gemacht werden. Die Firma [NRGene](#) aus Israel bietet eine solche «Cloud»-Plattform an. Unter Berücksichtigung von Interaktionseffekten zwischen Genotyp und Umweltfaktoren wird es dadurch möglich, dem Landwirt standortspezifische Sortenempfehlungen abzugeben. Ein ähnlicher Nutzen ergibt sich in der Tierzucht.

### 6.12.2 Herausforderungen

Neben den vielfältigen Anwendungsmöglichkeiten gibt es auch Herausforderungen, was den Einsatz von KI-Technologie in der Landwirtschaft anbelangt. Dazu gehören eine lokal noch ungenügend ausgebaut digitale Infrastruktur, eine oft noch mangelhafte agronomische Datengrundlage, eine zögerliche Aufnahmebereitschaft seitens der Landwirte und teilweise hohe Investitionskosten für die potenziellen Nutzer. Dazu kommen rechtliche Unsicherheiten, wie z.B. beim Umgang mit Agrardaten und dem Einsatz autonomer (unbemannter) Flug- und Fahrzeuge.

## Herausforderungen der künstlichen Intelligenz

Im Sinne eines transparenten und verantwortungsvollen Einsatzes von KI müssen auch Bedenken berücksichtigt werden hinsichtlich der Auswirkungen auf den Strukturwandel (Zukunft kleinbäuerlicher Betriebe, Arbeitsplatzverlust) und der Abhängigkeit von multinationalen Technologiekonzernen. Auf der anderen Seite hat die KI das Potenzial, zur Steigerung der Wettbewerbsfähigkeit und der Nachhaltigkeit der Schweizer Landwirtschaft sowie zur administrativen Vereinfachung und effizienteren Erreichung agrarpolitischer Ziele beizutragen.

### 6.12.3 Bestehende Aktivitäten

Das Bundesamt für Landwirtschaft BLW verfolgt die Entwicklungen der KI in der Landwirtschaft laufend. Es hat dazu ein Business Intelligence Competence Center eingerichtet, das im Bereich «Big Data»-Analysen und künftig auch «Predictive Analytics» tätig ist. Zum Umgang mit digitalen Daten und Anwendungen wurde im Jahr 2018 die [Charta zur Digitalisierung der Schweizer Land- und Ernährungswirtschaft](#) lanciert, deren Umsetzung durch die [Gemeinschaft der Unterzeichnenden der Charta](#) es weiter zu fördern gilt. Wie die obengenannten Beispiele zeigen, verfügt die Schweiz über eine lebendige Start-up Szene. Der Bundesrat ist bestrebt, die Rahmenbedingungen für Start-ups kontinuierlich zu verbessern, damit die Schweiz für Jungunternehmen weiterhin attraktiv bleibt.

### 6.12.4 Bewertung und Handlungsbedarf

Aktionsfeld 1: Auswirkungen der KI in der Landwirtschaft	
Weiterverfolgung der Entwicklungen der KI in der Landwirtschaft	<p>Die oben erwähnten Herausforderungen dürften die Schweizer Landwirtschaft in Zukunft beschäftigen. Das BLW wird vor diesem Hintergrund die KI in der Landwirtschaft weiterverfolgen und aufkommende Fragen im Rahmen bestehender Kompetenzen bearbeiten.</p> <p><b>Zuständigkeit:</b> BLW                      <b>Status:</b> Beobachtung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## 6.13 Energie, Klima, Umwelt und künstliche Intelligenz

### 6.13.1 Übersicht

Ab dem Jahr 2050 soll die Schweiz netto keine Treibhausgasemissionen mehr ausstossen. In den nächsten 30 Jahren muss u.a. der Energiesektor deshalb weitgehend umgestellt und wesentlich effizienter und unabhängiger von fossilen Energieträgern werden. Parallel dazu müssen der zu hohe Ressourcenverbrauch und der anhaltende Verlust an Biodiversität gestoppt werden. Der grösste Teil der derzeitigen Umweltbelastung<sup>108</sup> wird durch die Faktoren Ernährung, Wohnen und Mobilität verursacht, sowohl in der Schweiz als auch im Ausland. Die Herausforderungen für die Transformation der Systeme in den Bereichen Energieversorgung, Ernährung, Wohnen und Mobilität sind gross und KI wird als Schlüsseltechnologie eine wichtige Rolle spielen, um diesen Herausforderungen zu begegnen.

Mit Hilfe von KI-Technologien kann beispielsweise die Nachfrage nach Energie, Lebensmitteln oder Konsumgütern wesentlich genauer vorausgesagt werden als bisher. Dies nutzt der Produktion und dem Handel, da der Einsatz von KI das Potenzial besitzt, Verluste in der Produktionsplanung, fehlgeleitete Transporte oder zu grosse Lager zu reduzieren – und dies über komplexe, zeitkritische und globale Lieferketten hinweg. Ausserdem können Informationen zur Verfügbarkeit von Rohstoffen, dem Zustand von Ökosystemen an den Produktionsstandorten (oder generell zu ökologischen und sozialen Aspekten) in die Produktionsabläufe integriert und grundsätzlich jedem Kunden zur Verfügung gestellt werden. KI kann in der Folge auch auf der Nachfrageseite jeden Konsumenten und Marktteilnehmer darin unterstützen, angepasst an individuelle Konsumbedürfnisse nicht nur die preisgünstigste, sondern auch die umweltfreundlichste Produktvariante zu finden und zu beziehen.

KI wird auch eine zentrale Rolle im **Energiesektor** spielen, der historisch gewachsen und bisher eher wenig digitalisiert ist. Sie kann von grossem Nutzen sein in der Transformation von einem zentral organisierten Gesamtsystem zu einem dezentralen, auf erneuerbaren Energien basierenden und flexiblerem System, in dem die verschiedenen Energieinfrastrukturen (Strom, Gas und Wärme) notwendigerweise enger gekoppelt sind. Mit KI können bei der Netzplanung optimierte Ergebnisse erzielt und Prognosen zu Verbrauch und Produktion deutlich verbessert werden. Es ist dies ein Beitrag zur Sicherstellung der Koordination eines wegen fluktuierender, dezentraler Einspeisungen und der steigenden Anzahl von Akteuren immer komplexer werdenden Energiesektors. Nicht zuletzt kann KI dabei helfen, den Energieverbrauch zu reduzieren. Digitalisierung und KI-Technologien können so – bei gleichzeitig hoher Versorgungssicherheit – die Dekarbonisierung unterstützen und helfen, die Kosten der Transformation so gering wie möglich zu halten.<sup>109</sup>

### 6.13.2 Herausforderungen

KI-Technologien sind im Kontext der **Umweltziele** ein zweischneidiges Schwert: Aufbau und Nutzung dieser Technologien benötigen bereits heute grosse Mengen an Energie und Rohstoffen und ein weiteres Wachstum des Umweltverbrauchs durch KI-Technologien ist absehbar. Die Frage der umweltverträglichen Herstellung, Wiederverwertung und Entsorgung der benötigten Infrastruktur und Geräte im Sinne der Kreislaufwirtschaft wird sich zunehmend stellen. Auch wenn mit Hilfe von KI grosse Potentiale zur ökologischen Optimierung von Prozessen, Produkten und Märkten erschlossen werden können, ist das heute noch nicht der Fall. KI-Technologien können sowohl über die gezielte Entwicklung von Anwendungen im Umweltbereich gefördert werden oder indirekt über die Bereitstellung von grossen Mengen an Umweltdaten (Push-Ansatz). Diesen Ansatz verfolgt insbesondere die EU-Kommission mit dem satellitengestützten Erdbeobachtungsprogramm «Copernicus».

Im Sinne eines Pull-Ansatzes kann KI als Schlüsseltechnologie einen wichtigen Beitrag dazu leisten, Ziele und Anforderungen an den Umweltverbrauch der genannten Systeme in den Bereichen Energieversorgung, Ernährung, Wohnen und Mobilität zu erreichen, indem sie der damit einhergehenden

<sup>108</sup> Bericht des Bundesrates (2018): «Umwelt Schweiz 2018», verfügbar unter <https://www.bafu.admin.ch/bafu/de/home/dokumentation/umweltbericht/umweltbericht-2018.html>

<sup>109</sup> Bundesamt für Energie (2019): «Digitalisierung im Energiesektor - Dialogpapier zum Transformationsprozess», verfügbar unter <https://www.bfe.admin.ch/bfe/de/home/versorgung/digitalisierung.html>

Komplexität begegnet. Dafür müssten die Ziele jedoch klar deklariert und allenfalls auch vorgegeben werden.

Im **Energiesektor** ist die Sachlage hinsichtlich KI noch unübersichtlich. Die Energiestrategie 2050 setzt wichtige Rahmenbedingungen sowie Push- und Pull-Mechanismen zum Umbau des Energiesektors. Inwiefern diese Digitalisierung und insbesondere die Entwicklung und Anwendung von KI begünstigen ist jedoch noch unklar. Machine Learning Ansätze beispielsweise, obwohl schon länger verfügbar, finden noch kaum Anwendung. Stattdessen wird noch eher auf konventionelle Methoden bei Planung und Betrieb der Energieinfrastrukturen gesetzt. So bleibt denn zu analysieren, wie die Entwicklung von KI im regulierten, kleinteiligen und durch eher wenig Effizienzdruck geprägten Energiesektor vorangetrieben werden kann. Ein offener Energiemarkt würde weitere Anreize für innovative, auf KI basierte Lösungen und Dienstleistungen setzen. Push-Ansätze für Digitalisierung und die Entwicklung von KI sind in der Regulierung noch eher wenig etabliert und die Dateninfrastruktur ist kaum leistungsfähig genug, um KI in der Fläche zu ermöglichen. Auch sind die Auswirkungen von KI auf die Versorgungssicherheit noch unklar. Nicht zu unterschätzen sind die Herausforderungen bezüglich Energieeffizienz. So ist noch offen, wie KI diese unterstützen kann und ob dies allenfalls Implikationen für den Datenschutz hätte.

### 6.13.3 Bestehende Aktivitäten

Die Push-Förderung von KI-Technologien im **Umweltbereich** fokussiert im Moment sehr stark auf die Auswertung grosser Mengen von Umweltdaten (z.B. Satellitendaten). Die Verfügbarkeit von Satellitenbildern ist in den letzten Jahren massiv angestiegen. So liefert das europäische Erdbeobachtungsprogramm Copernicus seit 2017 mehrmals pro Woche eine komplette Bildaufnahme der Schweiz und mehrmals pro Monat ein komplettes Bild der gesamten Erdoberfläche. Die Auswertung dieser Datenmengen ist ohne KI-Technologie unmöglich. Das Anwendungspotential ist breit und kann sowohl national (z.B. Erkennung von Feldfrüchten im Ackerbau) wie auch international (z.B. Detektion von Waldbränden und Brandrodung) eingesetzt werden. Auch die Daten aus nationalen Umweltmessnetzen werden seit einigen Jahren gezielt verfügbar gemacht und unterstützen das Training von KI-Anwendungen. Auf Seiten der Pull-Förderung von KI stehen die langfristigen umweltpolitischen Ziele im Vordergrund. Das Ziel einer klimaneutralen Schweiz bis im Jahr 2050 ist ein grosser Anreiz, um die Potenziale von KI zu entwickeln und für die Zielerreichung zu nutzen. Im Sinne dieser Pull-Förderung werden zurzeit im Auftrag des Bundesamtes für Umwelt in einer Studie Chancen und Risiken der Digitalisierung für die Umweltpolitik analysiert. Ergebnisse aus dieser Studie werden Ende 2019 erwartet.

Im **Energiebereich** umfasst die Energiestrategie 2050 leichte Pull-Ansätze (Ziele) sowie einen ersten wichtigen Baustein für den Aufbau der Dateninfrastruktur, ohne die KI nicht eingesetzt werden kann: Bis Ende 2027 werden intelligente Messsysteme (sogenannte Smart Meter) im Strombereich eingeführt. Sie ermöglichen eine digitale und feingranulare Datenerfassung der Stromproduktion bzw. des Verbrauchs. In einem zweiten Schritt wird die Etablierung einer digitalen Plattform (Datahub) diskutiert, die den Datenaustausch effizienter gestalten und Daten einfacher verfügbar machen soll.<sup>110</sup> Standardisierte maschinenlesbare Schnittstellen (API) spielen dabei eine wichtige Rolle. Eine Plattform und maschinenlesbare Schnittstellen könnten zum Kern der Dateninfrastruktur avancieren. Darüber hinaus ist ein Dialog zur Digitalisierung des Energiesektors vorgesehen, in welchem Fragen zu den Rahmenbedingungen für den Einsatz von KI aufgegriffen werden können. Solche Rahmenbedingungen, die spezifisch die nutzbringende Anwendung von KI zum Ziel haben, könnten insbesondere im Rahmen sogenannter «Regulatory Sandboxes» getestet werden, insofern in spezifischen Bereichen neue Anwendungen aufgrund von bestehenden Regulierungen nicht möglich sind.<sup>111</sup> Darüber hinaus bietet das Bundesamt für Energie das «Pilot- Demonstrations- und Leuchtturmprogramm», in dessen Rahmen auch KI Projekte unterstützt werden könnten.<sup>112</sup>

<sup>110</sup> Bundesamt für Energie (2019): «Digitalisierung im Energiesektor - Dialogpapier zum Transformationsprozess».

<sup>111</sup> Derzeit untersucht das Bundesamt für Energie im Rahmen einer Studie internationale Erfahrungen zu Regulatory Sandboxes und ihre Anwendung im Energiebereich.

<sup>112</sup> Bundesamt für Energie (2019): Pilot-, Demonstrations- und Leuchtturmprogramm, <https://www.bfe.admin.ch/bfe/de/home/forschung-und-cleantech/pilot-demonstrations-und-leuchtturmprogramm.html>

### 6.13.4 Bewertung und Handlungsbedarf

KI kann als Schlüsseltechnologie einen wichtigen Beitrag leisten, um die ökologischen Anforderungen zu erreichen, die an Systeme für Energieversorgung, Ernährung, Wohnen und Mobilität gestellt werden. Die Potentiale werden bisher erst ungenügend genutzt. Dies zeigt sich auch daran, dass der Bereich Energie im Patentportfolio der Schweiz nur eine sehr kleine Rolle einnimmt (siehe Kapitel 5, Abbildung 11). Der Bedarf an Know-How und Kapital für die Nutzung dieser KI-Potentiale ist gross. Der Zugang zu Umweltdaten und zu einer ausreichenden Rechenkapazität für KI-Anwendungen sollten gewährleistet sein.

Die KI befindet sich im Energie- und Umweltsektor noch in einem frühen und dynamischen Stadium. Im Energiesektor müssen die derzeitigen Anreize und Infrastrukturen kritisch reflektiert und hinsichtlich KI geprüft werden. Dies umfasst die Ausgestaltung des Energiemarktes und der Netzregulierung, die kapitalintensive Investitionen eher begünstigt, Transparenz und eine leistungsfähige Dateninfrastruktur. Auch ein Wissensdefizit hinsichtlich der Anwendungsfelder und Nutzen von KI sowie ein Mangel an Fachpersonal können Faktoren sein. Es gilt daher, zunächst die Barrieren zu analysieren und die Grundlagen zu verbessern, zum Beispiel indem die Datenverfügbarkeit und die Transparenz verbessert werden. Weiter gilt es, Wissen aufzubauen und zu verbreiten sowie zu sensibilisieren. Vor diesem Hintergrund ist eine sektorielle KI-Strategie zu prüfen. Dabei wären konkrete Anwendungsbeispiele und Piloten (zum Beispiel zur Identifizierung regulatorischer Hemmnisse) zu reflektieren.

<b>Aktionsfeld 1: KI im Energiesektor</b>	
Die Anwendung von KI hat das Potenzial zu erheblichen Effizienzsteigerungen in der Energieversorgung. Sie wird den Ausbau der Erneuerbaren Energien, Energieeinsparungen und so den Klimaschutz unterstützen, der insgesamt steigenden Komplexität begegnen und eine wichtige Rolle im Betrieb und in Bezug auf Versorgungssicherheit wahrnehmen.	
Weiterverfolgung der Entwicklungen im Bereich «Energie»	Das BFE wird vor diesem Hintergrund die Herausforderungen bezüglich KI in der Energieversorgung beobachten, Grundlagen erarbeiten, Kompetenz aufbauen und Barrieren regulatorischer und technischer Natur reduzieren. <b>Zuständigkeit:</b> BFE <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

<b>Aktionsfeld 2: KI in Umwelt und Klima</b>	
KI ist eine Schlüsseltechnologie, um die ökologischen Anforderungen an die Systeme Ernährung, Wohnen und Mobilität erfüllen zu können. Zu diesem Zweck sollen die dafür notwendigen Daten (zum Beispiel Verfügbarkeit Rohstoffe, Zustand Ökosysteme an Produktionsstandorten oder Informationen in Produktionsabläufen) möglichst einfach zur Verfügung gestellt und wo immer möglich in die Informationsflüsse der Lieferketten und Märkte integriert werden. Ein weiteres Wachstum des Umweltverbrauchs durch KI-Technologien ist absehbar. Fragen der umweltverträglichen Herstellung, Wiederverwertung und Entsorgung der benötigten Infrastruktur und Geräte im Sinne der Kreislaufwirtschaft werden sich zunehmend stellen.	
Weiterverfolgung der Entwicklungen im Bereich «Umwelt und Klima»	Das BAFU sorgt dafür, dass Umweltinformationen wenn möglich als offene und digitale Datensätze zur Verfügung stehen und für KI-Anwendungen verwendet werden können. Es begleitet die umweltrelevanten Herausforderungen bezüglich KI auch aus Perspektive der Kreislaufwirtschaft und stellt die nötigen Grundlagen zur Verfügung. <b>Zuständigkeit:</b> BAFU <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## 6.14 Künstliche Intelligenz in der Verwaltung

### 6.14.1 Übersicht

KI kann die Verwaltung auf allen Ebenen entlasten, die Kundenorientierung sowie die Servicequalität verbessern und zur Wirtschaftlichkeit beitragen. Der Einsatz von KI in der Verwaltung ermöglicht es, Daten, welche man früher nicht maschinell verarbeiten konnte, künftig schnell, effizient und rund um die Uhr zu verarbeiten. In fünf Bereichen können bereits Anwendungen von KI in der Verwaltung eingesetzt werden: [1] Erkennung von Text, [2] Erkennung von Bild & Video, [3] automatische Übersetzungshilfen (für Verwaltungsdokumente), [4] automatische Auswertung von Tonaufnahmen sowie [5] Interaktion via Chatbots.

### 6.14.2 Herausforderungen

Ein erfolgreicher Einsatz von KI in der Verwaltung auf Ebene Bund, Kantone und Gemeinden bedingt verschiedene Voraussetzungen: Grosse Datenmengen sind erforderlich, damit Systeme künstlicher Intelligenz angewendet werden können. Innerhalb der Bundesverwaltung bestehen grosse Datenmengen u.a. bei der EZV (DaziT), der ESTV, dem BFS oder dem BLW. Deshalb sind diese Verwaltungseinheiten auch führend bei der Einführung von KI. Ein weiterer Erfolgsfaktor ist der Datenaustausch: Dieser muss zwischen den drei Staatsebenen und innerhalb der rund 70 Verwaltungseinheiten der Bundesverwaltung geregelt und einfach möglich sein. Insbesondere sind aufgrund des Datenschutzes dem Datenfluss personenbezogener Daten Grenzen gesetzt; dieser Datenfluss muss verbessert werden, soll KI ihr gesamtes Potenzial entfalten können.

KI wird die Dienstleistungen gegenüber allen Kunden der Verwaltung verbessern können. Der Mensch wird aber bei Verwaltungsentscheiden, welche die Rechtsposition des Adressaten oder die Leistungen des Staates betreffen, zumindest kurz- bis mittelfristig letzte Entscheidungsinstanz bleiben müssen: Transparenz, Nachvollziehbarkeit und Kontrolle der Verwaltungsergebnisse müssen gewahrt bleiben, was gemäss Technologiestand (noch) nicht möglich ist. KI ist eine junge Forschungsrichtung mit wenig Erfahrungswerten und teilweise unbekanntem Fehlerquoten. Verwaltungsaktivitäten hingegen erlauben nur eine sehr geringe Fehlerrate.

Während mittlerweile ein breites Angebot an Basissystemen künstlicher Intelligenz als proprietäre wie auch als open-Source Lösungen erhältlich ist, sind Modellierung und Adaption dieser Systeme Schlüsselherausforderungen. Dafür braucht es Expertenwissen und Erfahrung, die aktuell in der Bundesverwaltung nur punktuell zur Verfügung stehen und auch schwierig zu erhalten sind. Ein dezentraler Aufbau dieser Kompetenz und deren Managements ist deshalb nicht zielführend. Schliesslich gilt es, die Bedenken und Zukunftsängste der Mitarbeiter/-innen innerhalb der Verwaltung zu adressieren und KI als Chance und nicht als Gefahr zu positionieren.

### 6.14.3 Bestehende Aktivitäten

Funktionsfähige Anwendungen in der Bundesverwaltung, welche operativ breit genutzt werden können, liegen aktuell noch nicht vor oder sind noch in einem frühen Stadium (Studie bzw. Prototyp).

Tabelle 6 stellt die derzeit bekannten Projekte dar (Stand März 2019).

Tabelle 6: KI-Anwendungen in der Bundesverwaltung

Projekt	Verwaltung	Beschreibung
Programm DaziT Projekt Data Analytics	EZV	Das Projekt stellt sicher, dass die EZV für die gezielte Nutzung von Data Analytics zukünftig optimal aufgestellt ist. Anwendungsfälle sind Firmenprofile, Risikoanalyse und Warenschmuggel.
Programm DaziT Evaluation einer Chatbot Lösung	EZV	Um die Aufwände, die durch Grenzübertritte entstehen (insbesondere im Mitarbeiterbereich) zu senken, wird ein sehr hoher Automatisierungsgrad der Prozesse angestrebt. Zurzeit wird eine Chatbot Lösung evaluiert.
Arealstatistik Deep Learning (ADELE)	BFS	Das Projekt versucht durch den Einsatz von KI, die Luftbildinterpretation zumindest teilweise zu automatisieren, um Veränderungen zu identifizieren und zu klassifizieren. Hier eignet sich «Deep Learning», weil grosse Mengen an Trainingsdaten vorhanden sind.
Automatisierung der NOGA-Kodierung (NOG-Auto)	BFS	Anhand der Machine-Learning-Methoden sollen die Kodierung der im BFS bereits vorhandenen Daten automatisiert werden (gegenwärtig noch manuell vorgenommen).
Machine Learning SoSi	BFS	Gruppierung typischer Verlaufsmuster bezüglich Leistungsbezügen im System der sozialen Sicherheit und Erwerbsarbeit sowie Schätzung der Gruppenzugehörigkeit unter Anwendung eines Machine-Learning-Ansatzes werden untersucht und auf ihre Zuverlässigkeit hin geprüft.
Plausibilitätsprüfung mit Machine Learning	BFS	Dieses Projekt beabsichtigt die Plausibilitätsprüfungen im BFS anhand von Machine Learning Algorithmen zu erweitern, zu beschleunigen und gleichzeitig die Datenqualität zu steigern.
Pilotversuch einer arbeitsmarktoptimierten Kantonsverteilung von Asylsuchenden mit Machine Learning	SEM	In einem mengenmässig und zeitlich begrenzten Pilotversuch wird eine auf einem Machine Learning Algorithmus basierende arbeitsmarktoptimierte Kantonsverteilung von Asylsuchenden getestet. <sup>113</sup>

Quelle: ISB.

Gemäss einer Umfrage zum Einsatz von künstlicher Intelligenz bei den Kantonen arbeiten auch die Kantone auf allen Ebenen und in diversen Anwendungsbereichen an der Konzeption bzw. Implementierung von KI Anwendungen.<sup>114</sup> So setzt die Luzerner Polizei eine KI-Lösung für Videoauswertungen bei Straftaten ein. Auch prüfen mehrere Kantone den Einsatz von Chatbot-Lösungen zur Benutzerunterstützung (u.a. Kt. St. Gallen, Freiburg und Luzern). Die Sozialversicherungsanstalt des Kantons St. Gallen (SVA) hat beispielsweise im Kontakt mit ihren Kunden einen Chatbot getestet. Einwohner/-innen können über den Chatbot in einem Dialog ihren Anspruch auf die Prämienverbilligung abklären.

Ein internationaler Vergleich zeigt: Viele andere (Industrie-) Länder arbeiten auch an KI Strategien, die den Einsatz von KI in der Verwaltung enthalten. Sie sind zum Teil schon deutlich weiter in der Formulierung und konkreten Umsetzung. Wenn die Schweiz nicht intensiv und koordiniert die Chancen und Risiken der KI an die Hand nimmt, wird der Abstand weiter zunehmen.<sup>115</sup>

<sup>113</sup> vgl. ETH Zürich (2018): «Algorithmus verbessert Erwerbschancen von Flüchtlingen», <https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2018/01/algorithmus-verbessert-erwerbschancen-von-fluechtligen.html> sowie Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence (2018): «Improving refugee integration through data-driven algorithmic assignment», *Science*, Vol. 359, Issue 6373, S. 325-329, <https://science.sciencemag.org/content/359/6373/325>

<sup>114</sup> vgl. Ergebnisse der Umfrage zum Einsatz von künstlicher Intelligenz (KI) bei den Kantonen vom 15. Juli 2019, Kanton Luzern, unveröffentlichtes Manuskript.

<sup>115</sup> Auf internationaler Ebene haben Länder wie China, Finnland oder Singapur KI zur Schlüsseltechnologie der Digitalisierung erklärt und Pläne angekündigt, um die Forschung und Entwicklung im Bereich KI massiv und rasch voranzutreiben. Auch die europäische Kommission bringt KI voran und formuliert drei Stossrichtungen:



#### 6.14.4 Bewertung und Handlungsbedarf

Konkrete Überlegungen und daraus resultierend Empfehlungen betreffend algorithmische bzw. automatische Entscheidungsprozesse sind noch verfrüht. Es geht im Gegenteil darum, KI basierten Systemen keine autonome Entscheidungsgewalt zu überantworten, sondern diese als Unterstützungsfunktion zu etablieren, die raschere und bessere Entscheidungen des Mitarbeiters / der Mitarbeiterin der Verwaltung ermöglichen.

Empfehlungen werden sich aus den Erfahrungen erster konkreter Projekte ergeben und allen künftig folgenden Projekten als Leitplanken dienen. Aus der Analyse ergeben sich folgende Aktionsfelder:

Aktionsfeld 1: Gemeinsame Datengrundlagen	
Durch eine gemeinsame Datengrundlage der Verwaltung ergeben sich mehr Anwendungsmöglichkeiten für KI, da mehr Daten zur Verfügung stehen und diese stärker verknüpft sind.	
Schaffung und Bereitstellung von Datenbeständen innerhalb der Bundesverwaltung	<p>Für die Entwicklung von KI müssen grosse und geschützte Datenbestände verfügbar sein. Die Bundesverwaltung muss Datenbestände schaffen und bereitstellen, welche einen produktiven Datenaustausch ermöglichen. Hierzu sind Regelungen für einen sicheren Umgang mit Daten zu definieren.</p> <p><b>Zuständigkeit:</b> Bundesämter mit grossen Datenmengen (u.a. EZV, BFS, ESTV, BLW)</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

Aktionsfeld 2: Kompetenznetzwerk KI in der Bundesverwaltung	
Die verwaltungsübergreifende Erfassung von Prozessen sowie ein übergreifender Datenzugriff sind Voraussetzung, um das Potenzial von KI in der Bundesverwaltung auszuschöpfen.	
Vertiefte Abklärung zur Schaffung eines Kompetenznetzwerkes KI mit spezifischem Fokus auf technischen Aspekten der Anwendung von KI in der Bundesverwaltung	<p>Die verwaltungsübergreifende Erfassung von Prozessen sowie ein übergreifender Datenzugriff sind eine der Voraussetzungen, um das Potenzial der KI in der Bundesverwaltung auszuschöpfen. Der überdepartementale Aufbau und Austausch von Wissen und Erfahrung sind für eine wirtschaftliche und koordinierte Entwicklung von KI-Lösungen in der Bundesverwaltung wesentlich. Fragmentierte Lösungen sind dagegen nicht zielführend. Eine gemeinsame Anlaufstelle bzw. ein Kompetenznetzwerk mit speziellem Fokus auf technische Aspekte für die konkrete Anwendung von KI in der Bundesverwaltung könnte eine Lösung darstellen. Dabei sollte die erwähnte Anlaufstelle bzw. das Netzwerk insbesondere beratende Funktionen wahrnehmen.</p> <p>Das EFD (ISB) prüft in Zusammenarbeit mit dem EDI (BFS) und unter Einbezug der übrigen Departemente und der BK den Mehrwert (u.a. Bedürfnisanalyse) und die Machbarkeit einer gemeinsamen Anlaufstelle bzw. eines Kompetenznetzwerkes mit speziellem Fokus auf technische Aspekte der Anwendung von KI in der Bundesverwaltung. Diese Stelle soll für die Anwendung in der Bundesverwaltung beratende Funktion haben. Mit dem Ziel einer gesamtheitlichen Vernetzung und Technologiebetrachtung sollen auch andere Technologien der digitalen Transformation (bspw. Blockchain, IoT, Big-Data etc.) und die Herausforderungen von deren Anwendung berücksichtigt werden.</p> <p><b>Zuständigkeit:</b> ISB, BFS      <b>Status:</b> Prüfung zuhanden Bundesrat</p>
<b>Zusätzlicher Handlungsbedarf:</b> Ja	

[1] Förderung der Investitionen, [2] Vorbereitung auf sozioökonomische Veränderungen und [3] Gewährleistung eines geeigneten ethischen und rechtlichen Rahmens.

<p><b>Aktionsfeld 3: Chancen von KI aufzeigen (Kommunikation)</b>                  Aktive Information und das Aufzeigen von Chancen der KI können Ängste der Mitarbeitenden abbauen.</p>	
<p>Verstärkte Kommunikation über KI relevante Themen innerhalb der Bundesverwaltung</p>	<p>KI wird auch in der Bundesverwaltung Jobprofile ändern und den Adaptionsdruck deutlich erhöhen. Diesen Änderungsprozess gilt es zu begleiten, damit die Mitarbeiter den Prozess positiv erleben und aktiv unterstützen.</p> <p><b>Zuständigkeit:</b> ISB                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

<p><b>Aktionsfeld 4: Gesetzliche Grundlagen, Datenhoheit und Datensicherheit</b>                  KI muss in der Bundesverwaltung ermöglicht werden. Unnötige Hürden und Hindernisse sind abzubauen. Es muss ein Pioniergeist-Umfeld geschaffen werden.</p>	
<p>Überprüfung gesetzliche Grundlagen für Einsatz KI in der Bundesverwaltung</p>	<p>Mittel- bis langfristig sind für den Einsatz von KI gesetzliche Anpassungen nötig; u.a. müssen die Rahmenbedingungen im Bereich der Datenhoheit geklärt werden (Cyberkriminalität, Gewährleistung von Datensicherheit). Dabei wird die anstehende Revision des DSGVO mitberücksichtigt.</p> <p><b>Zuständigkeit:</b> ISB in enger Zusammenarbeit mit dem BFS und den betroffenen Bundesämtern  <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<p><b>Zusätzlicher Handlungsbedarf:</b> Nein</p>	

## 6.15 Weiterentwicklung des allgemeinen Rechtsrahmens im Hinblick auf künstliche Intelligenz

### 6.15.1 Übersicht

Wie in Kapitel 4 gezeigt wurde, ist der bestehende allgemeine Rechtsrahmen (1) vollumfänglich für die Anwendung von KI anwendbar, (2) auf Basis der Einschätzung des heutigen technologischen Standes im vorliegenden Bericht grundsätzlich geeignet, mit neuen Anwendungen und Geschäftsmodellen umzugehen und (3) eine allfällige Anpassung von Rechtsnormen derzeit nur anwendungsspezifisch angezeigt.

Gemäss den Ausführungen in Kapitel 3 ist zudem der Fokus heutiger und realistischerweise absehbarer KI-Anwendungen nach wie vor eng begrenzt. Die Entwicklungen schreiten jedoch rasch voran. In den kommenden Jahren dürften im Hinblick auf Anwendungsmöglichkeiten von KI neben rechtlichen insbesondere auch ethische und gesellschaftliche Fragen weiter an Bedeutung gewinnen.

### 6.15.2 Herausforderungen

#### Weiterentwicklung des Rechtsrahmens

Im Hinblick auf die Weiterentwicklung neuartiger Anwendungsmöglichkeiten können sich jedoch auch für den allgemeinen Rechtsrahmen zukünftig einige grundsätzliche Herausforderungen aus rechtlicher Sicht stellen. Namentlich im Hinblick auf die in Kapitel 4 diskutierten Probleme der Nachvollziehbarkeit, Transparenz und Diskriminierung. Und auch die Fähigkeit von KI-Systemen, zunehmend autonom agieren zu können, stellt den bestehenden rechtlichen Rahmen auf die Probe. Künstliche Intelligenz hat damit das Potenzial, zentrale Prämissen des Rechtsstaats, der auf der Grundlage basiert, dass der Mensch entscheidet und handelt, und durch den Menschen getragene Entscheidungsverfahren (z.B. Rechtsetzung, Rechtsprechung, Wahl- und Abstimmungen etc.) vorsieht, in Frage zu stellen.

Die Nutzung von KI im öffentlichen Sektor, sei es in der Verwaltung oder im Justizwesen, muss systematisch auf Grundlage einer Liste von Kriterien bewertet werden, die die Zuverlässigkeit des Prozesses, seine Transparenz, seine Neutralität und seine Integrität, die Einhaltung der Menschenrechte, insbesondere der Grundrechte, und die Vermeidung von Diskriminierung sowie die Vermeidung eines präskriptiven Charakters sicherstellen. Allerdings ist die Gefahr, dass diese Rechte verletzt werden, offensichtlich nicht auf den öffentlichen Sektor begrenzt. Der Bundesrat beobachtet diese Entwicklung sehr aufmerksam. Sollte er zum Schluss kommen, dass die gegenwärtigen rechtlichen Grundlagen nicht ausreichen, um den genannten Risiken zufriedenstellend zu begegnen, wird er die erforderlichen Massnahmen ergreifen und sich dabei insbesondere an Lösungen wie die bereits in anderen Bereichen bestehenden anlehnen (z. B. Humanforschung oder Gesetz über genetische Untersuchungen beim Menschen).

In einer solchen Situation wäre insbesondere zu prüfen, ob aus rechtsstaatlicher Sicht Vorkehrungen zu treffen sind, wie die Transparenz solcher Systeme gefördert und wie sie in einer für den Menschen verständlichen Form über die ihren Entscheidungen zugrundeliegende Logik Auskunft geben können, so dass deren Nachvollziehbarkeit und gegebenenfalls Überprüfbarkeit gewährleistet wird. Dabei sind Geschäfts- und Fabrikationsgeheimnissen Rechnung zu tragen (siehe auch die im Rahmen der DSGVO-Revision vorgesehenen Massnahmen, Kapitel 4.3).

Transparenz kann auch den Einsatz von KI in der Kundeninteraktion betreffen. Es gibt erste Systeme, die mit einem Menschen Telefongespräche führen, ohne dass es für den Menschen erkennbar ist, dass der Gesprächspartner eine Maschine ist (vgl. Kapitel 3.3 und 4.3). Es stellt sich dabei die Frage, ob Firmen die Kunden darüber vorab systematisch zu informieren haben. KI-Systeme sind vom Vorhandensein grosser Datenmengen abhängig. Insbesondere wo die öffentliche Hand KI-Applikationen einsetzt, ist sicherzustellen, dass es nicht infolge mangelhafter Datenqualität zu Verzerrungen, unzulässigen Ungleichbehandlungen, Diskriminierungen oder sonstigen Nachteilen für die betroffenen Personen führt.

Der Einsatz von KI-Systemen kann unter Umständen auch grenzüberschreitende Fragen aufwerfen. Aus der Perspektive des internationalen Privatrechts (IPR) scheinen sich jedoch im Bereich der KI trotz der zahlenmässig zunehmenden internationalen Sachverhalte *prima facie* keine neuen Probleme

zu ergeben, die sich nicht bereits durch das geltende IPR lösen liessen.<sup>116</sup> Auch im Strafrecht schafft der internationale Kontext a priori keine neuen Probleme, die nicht mit den vorhandenen Instrumenten gelöst werden könnten (insbesondere Art. 3 ff StGB sowie die internationalen Übereinkommen über die Rechtshilfe).

Es muss allerdings darauf hingewiesen werden, dass die Strafrechtshilfe im Bereich der «cyber-enabled criminality» oft mit zusätzlichen Schwierigkeiten verbunden ist. Diese Schwierigkeiten können für die Rechtshilfe im Bereich von Delikten, die mittels KI-Systemen begangen werden oder KI-Systeme betreffen, nicht ausgeschlossen werden.

### **Zukünftige Entwicklungen künstlicher Intelligenz und ethische Herausforderungen**

Während der Fokus des vorliegenden Berichts auf aktuellen Anwendungsmöglichkeiten von KI liegt, ist nicht auszuschliessen, dass sich mit der Weiterentwicklung der Technologien neue rechtliche, ethische und zivilgesellschaftliche Herausforderungen ergeben. Es bedarf daher der Weiterführung eines Dialogs zu übergeordneten Fragen der KI, welche unser geltendes Wertesystem betreffen und für dieses eine Herausforderung, oder möglicherweise gar eine Bedrohung darstellen.

Wo hingegen in spezifischen Bereichen neue Anwendungen aufgrund von bestehenden Regulierungen nicht möglich sind, anbietet sich das sog. «Sandboxing». In einer solchen sicheren und kontrollierten Testumgebung, die möglichst flexibel und daher administrativ schlank sein sollte, könnte durch die Lockerung gewisser Parameter (z.B. erleichterte Erteilung von Versuchsbewilligungen und eine Vereinfachung der hierfür vorgeschriebenen Verfahren) die Entwicklung neuer Technologien sowie innovativer Geschäftsmodelle vorangetrieben und zugleich getestet werden, welche regulatorischen Mindestanforderungen dafür notwendig sind bzw. wo das Recht eine gewisse Flexibilität aufweisen muss. Das Wissen und die Transparenz, die eine Sandbox liefern, ermöglichen zeitgerechte Anpassungen des Regulierungsrahmens, so dass der Nutzen von Innovationen volkswirtschaftlich besser beurteilt werden kann.

Schliesslich sind die Aktivitäten auf internationaler Ebene zu berücksichtigen: Der von mächtigen, transnational operierenden Technologieunternehmen geführte Prozess der Normensetzung sowie die von diversen internationalen Gremien erarbeiteten ethischen Prinzipien leiden an einem Demokratiedefizit. Für die Schweiz stellt sich daher die Frage, wie sie sich sinnvoll an diesen Prozessen beteiligen und sich darin positionieren kann, so dass sie ihre Werte und Vorstellungen darüber einbringen und die möglichen Auswirkungen auf die nationale Rechtsordnung abschätzen kann.

### **6.15.3 Bestehende Aktivitäten**

#### **Geltender Rechtsrahmen**

Verschiedene Abklärungen haben gezeigt, dass im Hinblick auf die Frage der Haftung bei autonom agierenden Maschinen die bestehenden Regelungen im Bereich der zivil- und strafrechtlichen Verantwortlichkeit und des internationalen Privatrechts derzeit genügen. Dass deren Anwendung auf Roboter Verantwortlichkeitslücken ergäben, hat sich noch nicht gezeigt. Insofern wurde bisher kein gesetzgeberischer Handlungsbedarf geortet.<sup>117</sup> Die Frage der Transparenz von KI-Anwendungen, soweit sie automatisierte Einzelentscheidungen betrifft, wurde im Rahmen der Revision des Datenschutzgesetzes entsprechend aufgenommen.

#### **Zukünftige Entwicklungen künstlicher Intelligenz und globale ethische Herausforderungen**

Angesichts der globalen Dimension der Entwicklung von KI ist es für die Schweiz wichtig, die internationalen Regulierungsbestrebungen eng mitzuverfolgen und insbesondere zu verstehen, wie in diesem Bereich neues Völkerrecht geschaffen wird und welche Auswirkungen dies auf die Schweiz haben kann. Der Bund bringt sich bereits aktiv in diese internationalen Diskussionen ein und engagiert sich dafür, dass beim Einsatz von KI etablierte Werte und Normen respektiert und alle relevanten Anspruchsgruppen einbezogen werden.

<sup>116</sup> In Fällen ausservertraglicher Haftungsgrundlagen besteht beispielsweise grundsätzlich eine Zuständigkeit in der Schweiz als Erfolgsort (Art. 129 IPRG; Art. 5 Ziff. 3 LugÜ), und regelmässig käme auch Schweizer Recht zur Anwendung (vgl. für die ausservertragliche Haftung z.B. Art. 133 IPRG, für Produktmängel Art. 135 IPRG, oder für Strassenverkehrsunfälle Art. 3 des Haager Übereinkommens von 1971 [SR 0.741.31]).

<sup>117</sup> vgl. z.B. 18.3445 Ip. Marchand-Balet mit Bezug auf selbstfahrende Fahrzeuge; 17.3040 Po. Reynard «Die Schaffung einer Rechtspersönlichkeit für Roboter prüfen» und Bericht des Bundesrates in Erfüllung des Postulats Leutenegger Oberholzer 14.4169 «Auto-Mobilität» bzgl. Sachverhalte mit Auslandsbezug.

#### 6.15.4 Bewertung und Handlungsbedarf

Der vorliegende Bericht ortet in einigen Themenfeldern rechtliche Herausforderungen von KI, die einer Klärung oder zumindest weiterhin einer engen Beobachtung bedürfen. Bezüglich des nationalen Rechtsrahmens ergibt sich derzeit kein zusätzlicher Handlungsbedarf. Handlungsbedarf stellt sich dagegen im internationalen Kontext. Dabei stehen besonders die Entstehung und Durchsetzung des internationalen Rechts im Vordergrund.

##### Aktionsfeld 1: Entstehung von KI-spezifischem Völkerrecht

Es ist vertieft zu prüfen, wie internationale Regeln zu KI entstehen, wie sie zu qualifizieren sind, inwiefern dadurch Völkerrecht geschaffen wird und welche Auswirkungen diese auf die Schweiz haben könnten.

Erarbeitung eines Berichts über die Entwicklung des Völkerrechts im KI-Bereich

Dem Bundesrat soll bis Ende 2020 ein Bericht über die Entwicklung des Völkerrechts im KI-Bereich vorgelegt und ggf. Massnahmen vorgeschlagen werden, wie sich die Schweiz dazu positionieren soll.

**Zuständigkeit:** EDA                      **Status:** Prüfung zuhänden Bundesrat

**Zusätzlicher Handlungsbedarf:** Ja

##### Aktionsfeld 2: Erkennbarkeit von KI-Systemen in der Interaktion mit Konsumenten

Verfolgen der Entwicklungen der Interaktion von KI-Systemen

Bezüglich der Erkennbarkeit von KI in der Interaktion mit Kunden (Chatbots, Telefonanrufe von KI) stellt das Bundesgesetz gegen den unlauteren Wettbewerb (UWG) prinzipiell eine geeignete Grundlage dar. Zu prüfen wäre die Frage jedoch erst bei Vorliegen einer konkreten Beschwerde.

**Zuständigkeit:** WBF (SECO)  
**Status:** Beobachtung im Rahmen bestehender Kompetenzen

**Zusätzlicher Handlungsbedarf:** Nein

## 6.16 Einsatz von künstlicher Intelligenz in der Justiz

### 6.16.1 Übersicht

KI hat auch als Hilfsmittel im Justizsystem ein erhebliches Potenzial. Gewisse heute entwickelte Tools zielen darauf ab, den Angehörigen der Rechtsberufe bei juristischen Nachforschungen zu helfen oder die Ergebnisse gerichtlicher Auseinandersetzungen vorherzusagen (sogenannte «prädiktive Justiz-Tools»). Andere Tools können eingesetzt werden, um die Gerichte bei der Verwaltung von Fällen zu unterstützen (zum Beispiel durch Analyse und Zuweisung von Anträgen an die zuständigen Gerichtsabteilungen) oder um die Leistung von Gerichten zu analysieren. Ausserdem können diese Tools eingesetzt werden, um Rechtsstreitigkeiten einfacher online beizulegen. KI wird in der Schweiz aktuell vereinzelt, in absehbarer Zukunft aber wohl häufiger zur Vorhersage von Prozessentscheiden eingesetzt werden. Im Vordergrund steht dabei die Simulation von Entscheiden aufgrund von Vergangenheitsdaten.

Entscheidungen im Rechtssystem sind jedoch mit Blick auf die Rechte der betroffenen Personen besonders sensibel, weshalb den möglichen, mit KI verbundenen Problemen von Transparenz und Diskriminierung besondere Beachtung zukommt. Insofern bedingt die Einbindung von KI-Systemen in das rechtstaatliche Gefüge, dass Entscheidungen (mit oder ohne menschliche Hilfe) in einer für den Menschen verständliche Form über die für ihre Entscheidungen massgebenden Gründe Auskunft geben können.

### 6.16.2 Herausforderungen

In bestimmten Bereichen kann es sinnvoll sein, im staatlichen Bereich (administrative und juristische Entscheidungen) auf automatisierte Entscheide zurückzugreifen. Das Recht, gehört zu werden, das Recht auf eine Entscheidungsbegründung oder die Beschwerdefähigkeit dürfen dadurch jedoch nicht eingeschränkt werden. In allen Fällen, in denen der Einsatz von KI zu einem erhöhten Risiko bzgl. der Rechte der betroffenen Person führt, ist auf die Einhaltung dieser Vorgaben zu achten. Es gilt zu vermeiden, dass Entscheidungsfindungen auf irgendeine Art an eine wenig zuverlässige Maschine delegiert werden.<sup>118</sup> Der Europarat hat einen Ethikkodex für den Einsatz von KI im Justizsystem und dessen Umfeld entwickelt<sup>119</sup>, der einzuhaltende Grundsätze definiert und die einzelnen Einsatzmöglichkeiten unterschiedlichen Risikokategorien zuordnet.

«Predictive analytics»<sup>120</sup> birgt aufgrund der Verarbeitung vergangenheitsorientierter Daten die Gefahr, den «status quo» in der Rechtsprechung einzufrieren und wenig Raum für Rechtsfortentwicklung zuzulassen. Ebenso wenig werden äussere Umstände bzw. irrationale/emotionale Vorgänge und Ausnahmeerscheinungen berücksichtigt. Die Nutzung von «predictive analytics» zur Kontrolle von Richterinnen und Richtern ist insoweit heikel, als diese für ihre Urteile gegenüber einer nicht exakten Technologie rechenschaftspflichtig würden. Dieses Risiko birgt aus grund- und menschenrechtlicher Sicht die Gefahr einer Beeinträchtigung des Rechts auf ein unabhängiges und unparteiisches Gericht (Art. 30 BV), da das Gericht durch den Einsatz von KI unter äusseren Druck geraten könnte.

Maschinell berechnete Gewinnchancen werden als objektiv wahrgenommen und haben dadurch einen grossen Einfluss auf die Prozessentscheidungen der Parteien, unter Umständen sogar des Gerichts. Der Einsatz von Vorhersagealgorithmen könnte zu vermehrten Ablehnungen führen (Beispiel: Ablehnung, falls geringere Erfolgschancen als 50 Prozent ggü. Ablehnung, falls die Erfolgschancen erheb-

---

<sup>118</sup> Der Bericht des Europarats (Referenz s. unten) teilt die Risiken und Chancen in vier Kategorien ein: zu befürwortende Anwendungen; mögliche Anwendungen, für die starke methodische Vorsichtsmassnahmen getroffen werden müssen; mögliche Anwendungen unter dem Vorbehalt zusätzlicher wissenschaftlicher Untersuchungen; nur unter höchsten Vorbehalten mögliche Anwendungen. «Predictive Analytics» fällt unter die zweite Kategorie, die Nutzung zur Kontrolle von Richtern unter die dritte.

<sup>119</sup> Ref. Europäische Kommission für die Effizienz der Justiz (2018): «Europäische Ethik-Charta für den Einsatz künstlicher Intelligenz in Gerichtssystemen und ihrem Umfeld», <https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b>

<sup>120</sup> «Predictive analytics» hat das Potenzial, die Rechtsberatung und Justiz zu verändern, weil bspw. Anwaltsansätze anhand der Schwierigkeit des Falls (d.h. wenn geringe Gewinnchancen vorhersagt werden) bemessen und die «Performance» der Anwältinnen und Anwälte verglichen werden können; Versicherungsprämien an das Prozessrisiko angepasst werden können; Prozessfinanzierer ihre Fälle effizienter auswählen können; Prädispositionen von RichterInnen durch KI erkannt werden können.

lich geringer scheinen als die Gefahr, zu unterliegen). Dieses Risiko kann aus grund- und menschenrechtlicher Sicht zu einer Beeinträchtigung des Rechts auf Zugang zu einem Gericht führen (Art. 29a BV, Art. 6 Ziff. 1 EMRK und Art. 14 Abs. 1 Uno-Pakt II).

Der Staat muss demnach dafür sorgen, dass der Zugang zum Recht nicht aus KI-bezogenen und wirtschaftlichen Gründen faktisch verwehrt wird. Davon besonders betroffen sind nicht-mittellose Rechtssuchende. Aber auch für mittellose können geringe, durch KI berechnete Gewinnchancen insoweit grund- und menschenrechtlich problematisch sein, als ihnen gestützt darauf die unentgeltliche Rechtspflege wegen Aussichtslosigkeit verweigert würde.

Künstliche Intelligenz kann auch eingesetzt werden, um Entscheidungen zu fällen, Inkonsistenzen zu erkennen oder Entscheidungen zu unterstützen (z. B. Berechnung von Schadenersatzansprüchen). Es ist wichtig, dass das Gericht von den Schlussfolgerungen der Maschinen abweichen kann und/oder dass die Entscheidung durch eine natürliche Person überprüft werden kann.

### 6.16.3 Bestehende Aktivitäten

Der DSGVO-Revisionsentwurf betrifft auch den Einsatz von KI in der Justiz und schreibt eine formelle gesetzliche Grundlage vor, wenn es sich bei der Bearbeitung von Daten um ein Profiling handelt oder wenn der Zweck oder die Art und Weise der Verarbeitung zu einem Eingriff in die Grundrechte der betroffenen Person führen können.

Die Strafverfolgungsbehörden und die Gerichte des Bundes und der Kantone führen gemeinsam und schweizweit in den kommenden Jahren den elektronischen Rechtsverkehr und die rechtsgültige elektronische Akten- resp. Dossierführung ein. Mit dem Projekt «Justitia 4.0» sollen alle Dokumente, Vermerke und Vorgänge eines Gerichtsverfahrens elektronisch gespeichert und ausgetauscht werden können. Über ein zentrales Eingangsportale haben die Prozessbeteiligten (Richter, Anwälte, Behörden usw.) Einblick in die für sie relevanten Einträge. Ein konkreter Einsatz von KI ist jedoch zum jetzigen Zeitpunkt nicht geplant.

### 6.16.4 Bewertung und Handlungsbedarf

Aktionsfeld 1: KI-basierte Entscheid-Unterstützung in Verwaltung und Justiz (Predictive Analytics)	
Beobachtung der Entwicklungen bei KI-basierter Entscheid-Unterstützung in Verwaltung und Justiz (Predictive Analytics)	Der Einsatz von KI-basierten Applikationen zur Erarbeitung von Entscheidungsgrundlagen, zur Vorbereitung von Entscheiden oder zur Unterstützung von Entscheidungsträgerinnen und -trägern kann sowohl in der Verwaltung als auch in der Justiz sinnvoll sein. Wie die Erfahrung aus anderen Ländern und auf internationaler Ebene zeigen, ist dies aber auch mit Risiken verbunden. Beim Einsatz solcher Technologien im Verwaltungs- bzw. Justizbetrieb muss die verantwortliche Verwaltungs- bzw. Justizbehörde daher im Einzelfall vertieft prüfen, ob diese Technologien besondere Auswirkungen auf die Grund-, Menschen- und Verfahrensrechte der Parteien haben und ob diese Auswirkungen besondere gesetzliche Grundlagen erfordern. Dabei kann sich die verantwortliche Verwaltungs- bzw. Justizbehörde an den Arbeiten zwischenstaatlicher Organisationen orientieren (bspw. an den «Ethics guidelines for trustworthy AI» oder am Kapitel IV der Empfehlungen der OECD).  <b>Zuständigkeit:</b> BJ, EDA <b>Status:</b> Beobachtung im Rahmen bestehender Kompetenzen
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## 6.17 Künstliche Intelligenz, Daten und Immaterialgüterrecht

### 6.17.1 Übersicht

Für die Anwendung aktueller KI-Methoden ist die Verfügbarkeit von Daten zentral. Für den Bund stellen sich diesbezüglich wichtige Fragen bezüglich der Weiterentwicklung der **Datenpolitik**. Diese will in erster Linie den Zugang zu Daten, und namentlich zu frei verfügbaren Daten (Open Data) als Rohstoff einer digitalen Wirtschaft und Gesellschaft fördern, um zeitgemässe und kohärente Rechtsgrundlagen und Rahmenbedingungen zu schaffen, mit denen sich die Schweiz als attraktiver Standort für eine Wertschöpfung durch Daten positioniert. Die Datenpolitik definiert aber auch den Rechtsrahmen, innerhalb dessen Daten zulässig gesammelt, verknüpft und ausgewertet werden können. Zielkonflikte bestehen dabei besonders mit den Ansprüchen des **Datenschutzes**, welcher sich an neuen Möglichkeiten von KI zu Auswertung enormer Datenmengen orientieren muss. Geistiges Eigentum ist sowohl ein Anreiz für schöpferische Tätigkeit und Innovationen als auch ein Hilfsmittel für die Weitergabe von Wissen. Die Produktion immaterieller Güter könnte durch den Einsatz von KI im Rahmen schöpferischer Tätigkeiten sowie der F&E vereinfacht und deutlich gesteigert werden. KI könnte im Lauf der Zeit eine immer wichtigere Rolle für schöpferische und erfinderische Arbeitsabläufe spielen. Dies dürfte zu einer Weiterentwicklung des Immaterialgüterrechts führen. Darüber hinaus stützt sich KI auch auf Daten, **die Schutzrechten unterliegen**. Dabei kann es sich beispielsweise um das Urheberrecht an einem Text, einer Zeichnung oder einer Fotografie handeln.

### 6.17.2 Herausforderungen

KI-Anwendungen brauchen meist grosse Datenmengen in adäquater Form. Für den Bund stellt sich in erster Linie die Frage, wie die Verfügbarkeit von Daten gefördert werden kann. Im Bereich der **Datenpolitik** ist zudem das Fehlen eines gemeinsamen Standards eines der am häufigsten hervorgehobenen Hindernisse. Darüber hinaus müssen Daten auffindbar sein, d.h. sie müssen katalogisiert und/oder recherchierbar und wiederverwendbar sein.

KI-Auswertungen erlauben auch neue Formen der Auswertung, welche den **Datenschutz** vor neuartige Herausforderungen stellen. Die zunehmende Fähigkeit, verschiedene Datensätze zu verknüpfen und verschiedene Arten von Informationen abzugleichen, macht eine Unterscheidung zwischen persönlichen und unpersönlichen Daten immer schwieriger. Zugleich verfügen KI-Systeme über die Möglichkeit, aus (der Kombination mehrerer) unpersönlicher Datenelemente persönliche Informationen abzuleiten. Solche Daten, die ursprünglich nicht personenbezogen waren, aber durch die KI identifiziert werden können, werfen Fragen im Hinblick auf Zustimmung, Zweckbestimmung und Nutzung auf. Darüber hinaus sind Datenverarbeitungen immer weniger transparent und es wird daher auch schwieriger, die angemessene Information der betroffenen Personen zu gewährleisten und die Rolle zu verstehen, die die betreffenden Daten bei Entscheidungsfindungen spielen. Zusätzlich besteht die Gefahr, dass Algorithmen (unbewusste) Diskriminierungen verstärken, wenn die zugrundeliegenden Daten von geringer Qualität sind. Die Schweiz hat im Bereich Datenschutz nur wenig Handlungsspielraum. Eine erhebliche Abweichung der Regulierung in der Schweiz von dem von der EU festgelegten Standard könnte den freien Datenfluss behindern und dadurch negative Konsequenzen mit sich bringen.

Bei der Datenverarbeitung mithilfe von KI wird eine digitale Kopie der betreffenden Daten angelegt. Sind die verarbeiteten Daten jedoch geschützt, beispielsweise durch ein **Urheberrecht**, verletzt diese digitale Kopie die Schutzrechte der Autorin bzw. des Autors. Da es um die verfassungsmässige Garantie der Eigentumsrechte der entsprechenden Inhaber oder Berechtigten geht, stellt sich die berechtigte Frage, unter welchen Umständen Daten durch Dritte verwendet werden können. Sollten dank der KI in Zukunft immer mehr immaterielle Güter zu immer geringeren Kosten produziert werden können, dürfte sich eine Weiterentwicklung der Rolle und des Stellenwerts des Immaterialgüterrechts aufdrängen. Sobald KI-Systeme in der Lage sein werden, unabhängige Entscheidungen zu treffen, wenn sie also zu Erfindungen oder schöpferischer Tätigkeit in der Lage sind, wird sich die Frage stellen, welchen Schutz das Immaterialgüterrecht den betreffenden Werken gewährt. Wenn es nicht mehr möglich ist, einen schöpferischen oder erfinderischen Akt mit Sicherheit einem Menschen oder einem KI-System zuzuordnen, wird es nicht mehr wirklich möglich sein, die Kriterien «geistige Schöpfung» und



«erfinderische Tätigkeit», die intrinsisch mit der menschlichen Natur verknüpft sind, anzuwenden. Daher könnte ein Systemwechsel erforderlich sein.

### 6.17.3 Bestehende Aktivitäten

Beim freien Zugang zu Datenbeständen (Open Data bzw. Open Government Data OGD) sowohl in der Bundesverwaltung als auch bei den bundesnahen Unternehmen wurden in den letzten Jahren Fortschritte erzielt. Mit der **OGD-Strategie** fördert der Bundesrat Verfügbarkeit, Transparenz und Effizienz der Verwendung von Behördendaten.

Der derzeit in der parlamentarischen Prüfung befindende **DSG-Revisionsentwurf** berücksichtigt hinsichtlich des Schutzes personenbezogener Daten bereits gewisse Herausforderungen, die sich mit KI ergeben. Für den Schutz der Privatsphäre in Zusammenhang mit KI sind mehrere vorgesehene Massnahmen von besonderer Bedeutung (vgl. Kapitel 4).<sup>121</sup> Darüber hinaus wurde die Frage nach den rechtlichen Grundlagen und ihrer Aktualität in Zusammenhang mit personenbezogenen und nicht personenbezogenen Daten im Jahr 2017 im Rahmen der Strategie Digitale Schweiz untersucht. Aus dieser Untersuchung hat sich kein grundlegender Bedarf ergeben, die derzeitigen rechtlichen Grundlagen über den DSG-Revisionsentwurf und punktuellen Handlungsbedarf hinaus zu überarbeiten. Auch die für die Datenverarbeitung und -sicherheit zuständige Expertengruppe<sup>122</sup> hat sich intensiv mit den Fragen der Verarbeitung personenbezogener und nicht personenbezogener Daten und des Zugriffs auf diese Daten in Zusammenhang mit der Digitalisierung, einschliesslich des Einsatzes von KI, befasst. Die aus dieser Arbeit hervorgegangenen Empfehlungen werden derzeit geprüft.

Der **Revisionsentwurf zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte** enthält eine Bestimmung, die das Urheberrecht einschränkt und ausdrücklich «Text und Data Mining» im Rahmen der Grundlagenforschung erlaubt. Dies wird die Forschung erheblich vereinfachen und zur Stärkung des Forschungsstandorts Schweiz beitragen.

### 6.17.4 Bewertung und Handlungsbedarf

Aktionsfeld 1: Datenpolitik	
Weiterführung bestehende Arbeiten zur Datenpolitik des Bundes	<p>Die Verfügbarkeit und der Zugang zu Daten ist zentral. Der Bund verfügt zu diesem Zweck über eine breit aufgestellte Datenpolitik. Mit der OGD-Strategie hat der Bundesrat die Verfügbarkeit und den Zugang zu Verwaltungsdaten weiter gestärkt. Die Arbeiten sind bereits eingeleitet worden und können im Rahmen der laufenden Aktivitäten der Verwaltung erfolgen.</p> <p><b>Zuständigkeit:</b> BAKOM, GS EDI/BFS</p> <p><b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

Dies gilt insbesondere für die Verpflichtung, eine Folgenabschätzung durchzuführen, Transparenzpflichten der Datenverantwortlichen und die auf automatisierte Entscheidungen anzuwendenden Regeln, den Grundsatz des Datenschutzes durch Technik und datenschutzfreundliche Voreinstellungen, die Meldepflicht für Sicherheitsverstösse sowie die Aufforderung, Verhaltenskodizes einzuführen.<sup>122</sup> Bericht der Expertengruppe zur Zukunft der Datenbearbeitung und Datensicherheit, 2018, <https://www.news.admin.ch/newsd/message/attachments/53591.pdf>

<sup>122</sup> Bericht der Expertengruppe zur Zukunft der Datenbearbeitung und Datensicherheit, 2018, <https://www.news.admin.ch/newsd/message/attachments/53591.pdf>

Aktionsfeld 2: Datenschutz	
Weiterführung bestehende Arbeiten beim Datenschutz	<p>Hinsichtlich des Datenschutzes stellt sich die Frage, ob es nicht erforderlich ist, sektorspezifisch anstatt mittels einer Überarbeitung der allgemeinen gesetzlichen Grundlagen einzugreifen. Darüber hinaus stellt sich die Frage, ob die Lösung für die festgestellten Probleme nicht wenigstens teilweise eher im technischen als alleine im rechtlichen Bereich zu suchen ist. Derzeit werden zahlreiche Massnahmen geprüft (DSG-Revision in parlamentarischer Beratung, Prüfungsauftrag des Bundesrats betreffend die Übertragbarkeit, Empfehlungen der Expertengruppe hinsichtlich Datenverarbeitung und -sicherheit), wobei die Schweiz in diesem Bereich nur schwerlich isoliert agieren kann. Daher müssen zunächst die laufenden Arbeiten zu Ende geführt werden. Gleichzeitig muss beobachtet werden, wie sich die Situation sowohl in technischer Hinsicht als auch international (insbesondere auf Ebene der EU und des Europarats) weiterentwickelt.</p> <p><b>Zuständigkeit:</b> BJ                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	
Aktionsfeld 3: Geistiges Eigentum	
Weiterführung bestehende Arbeiten beim geistigen Eigentum	<p>Es besteht kein dringender Bedarf, das Immaterialgüterrecht zu ändern. Das derzeit geltende Recht bietet für die unmittelbare Zukunft genügend Flexibilität. Die Verwaltungs- und Justizbehörden besitzen ausreichend Handlungsspielraum, um juristische Fiktionen / unbestimmte Rechtsbegriffe weiterzuentwickeln, mit denen sich das Aufkommen künstlicher Intelligenz begleiten lässt. Unabhängig davon sind Diskussionen und Überlegungen geboten, um zu gewährleisten, dass das Immaterialgüterrecht die Entwicklung der künstlichen Intelligenz mittel- und langfristig weiterhin sinnvoll unterstützt.</p> <p><b>Zuständigkeit:</b> IGE                      <b>Status:</b> Umsetzung im Rahmen bestehender Kompetenzen</p>
<b>Zusätzlicher Handlungsbedarf:</b> Nein	

## Anhang 1: Gesamtübersicht Aktionsfelder

In den untersuchten Themenbereichen stellt der Bericht eine Vielzahl an bereits ergriffenen Massnahmen, Initiativen und aufgegleisten Abklärungen dar. Diese werden in erster Linie innerhalb bestehender Aktivitäten, Kompetenzen und etablierter Verfahren adressiert. Im Zentrum der Weiterführung der vielfältigen Arbeiten des Bundes im Hinblick auf die Herausforderungen der künstlichen Intelligenz stehen die folgenden Aktivitäten im Mittelpunkt:

<b>Internationale Gremien und KI (BAKOM, EDA)</b>
<p><b>Aktionsfeld 1: Informations- und Wissensaustausch und Koordination der Positionen des Bundes in internationalen Gremien</b></p> <ul style="list-style-type: none"> <li>Nutzung der «Plateforme Tripartite» als interdisziplinäres nationales Kompetenznetzwerk zu KI-Themen und entsprechenden internationalen Gremien und Prozessen sowie für die Koordination der Positionen des Bundes in internationalen Gremien im Bereich KI</li> </ul> <p><b>Aktionsfeld 2: Globale Gouvernanz</b></p> <ul style="list-style-type: none"> <li>Stärkung der globalen Gouvernanz</li> <li>Integration KI in Aussenpolitische Strategie 2020 - 2023</li> </ul> <p><b>Aktionsfeld 3: Internationales Genf</b></p> <ul style="list-style-type: none"> <li>Bessere Vernetzung und Zusammenarbeit der KI-relevanten Akteure</li> <li>Prüfung zur verstärkten Zusammenarbeit für «AI for Good Summit»</li> <li>Stärkung der Geneva Internet Platform</li> <li>Stärkung Internationales Genf als Zentrum für digitale Gouvernanz in Aussenpolitische Strategie 2020 - 2023</li> </ul>
<b>Digital Europe Programme (SBFI u.a.)</b>
<p><b>Aktionsfeld 1: Schweizer Beteiligung an Horizon Europe und am Digital Europe Programme</b></p> <ul style="list-style-type: none"> <li>Prüfung Beteiligung «Horizon Europe» und «Digital Europe Programme»</li> </ul>
<b>Veränderungen der Arbeitswelt (SECO)</b>
<p><b>Aktionsfeld 1: Auswirkungen der KI auf den Arbeitsmarkt</b></p> <ul style="list-style-type: none"> <li>Weiterverfolgung der Entwicklungen des Schweizer Arbeitsmarkts</li> </ul>
<b>KI in Industrie und Dienstleistungen (SECO)</b>
<p><b>Aktionsfeld 1: KI in der Wirtschaft</b></p> <ul style="list-style-type: none"> <li>Weiterverfolgung der Entwicklungen von KI in Industrie und Dienstleistungen</li> </ul>
<b>KI in der Bildung (SBFI, Kantone und andere relevante Akteure)</b>
<p><b>Aktionsfeld 1: Gewährleisten angemessener Kompetenzen</b></p> <ul style="list-style-type: none"> <li>Vermittlung der für den Einsatz künstlicher Intelligenz erforderlichen Kompetenzen auf allen Bildungsstufen</li> </ul> <p><b>Aktionsfeld 2: Gewährleisten einer transparenten und verantwortungsvollen Anwendung von KI im Bildungsbereich</b></p> <ul style="list-style-type: none"> <li>Gewährleisten einer transparenten und verantwortungsvollen Anwendung von KI im Bildungsbereich</li> </ul>
<b>Anwendung von KI in Wissenschaft und Forschung (SBFI)</b>
<p><b>Aktionsfeld 1: Kompetenzen in Forschung und Innovation</b></p> <ul style="list-style-type: none"> <li>Gewährleistung der Forschungs- und WTT-Kompetenzen im Rahmen der BFI-Politik</li> </ul>

**KI in der Cybersicherheit und Sicherheitspolitik** (EDA, armasuisse (VBS), NDB, Kompetenzzentrum Cyber (EFD), BK, Armee (GS-VBS), BABS, WBF, VBS)

**Aktionsfeld 1: Aussensicherheitspolitische Auswirkungen**

- Prüfung von aussenpolitischen Implikationen bei Einsatz von KI-gestützten Systemen

**Aktionsfeld 2: Bedrohungsformen und Doktrin**

- Prüfung Cybersicherheit im Rahmen neuer Bedrohungsformen angesichts der Nutzung von KI
- Prüfung Propaganda und Informationsoperationen angesichts der Nutzung von KI
- Prüfung der Kriegsführung und der Fähigkeiten unterhalb der Kriegsschwelle angesichts Nutzung von KI

**Aktionsfeld 3: Fähigkeiten und Kapazitäten**

- Verstärkte Integration und Nutzung von KI-Lösungen in Streitkräften und im nachrichtendienstlichen Bereich
- Prüfung Upgrade-Möglichkeiten beim Akquisitionsprozess von kritischen Systemen
- Vermehrte Berücksichtigung von KI-Komponenten bei Lieferanten- und Subakkordanten von kritischen Systemen
- Regelmässige Überprüfung des Technologiestands bei Betreibern von kritischen Infrastrukturen

**Aktionsfeld 4: Antizipation durch Zusammenarbeit, Forschung und Prüfstände**

- Verstärkte Zusammenarbeit mit leitenden Bildungs- und Forschungsinstituten
- Berücksichtigung der Entwicklungen im KI Bereich beim Cyberlagebild
- Verstärkte Teilnahme an internationalen Gremien und Forschungsinitiativen im KI Bereich
- Überprüfung der Notwendigkeit, der Bedeutung und des Potentials einer schweizerischen KI-Prüfstelle

**KI, Medien & Öffentlichkeit** (BAKOM, BK, EDA)

**Aktionsfeld 1: Schweizerische Governance im Bereich Intermediäre**

- Erarbeitung Governance-Bericht im Bereich Intermediäre

**Aktionsfeld 2: Beobachtung der Entwicklung im Bereich Medien**

- Verfolgen der Entwicklungen zum Einsatz von KI im Medienbereich

**Automatisierte Mobilität und KI** (ASTRA, BAZL, BAV, swisstopo, VBS (BABS), UVEK)

**Aktionsfeld 1: Einsatz von KI in automatisierten Fahrzeugen**

- Koordination bestehende Arbeiten bezüglich automatisierten Fahrzeugen
- Erarbeitung Luftverkehrsmanagement für Unmanned Aircraft Systems (UAS)
- Abklärungen für Pilotprojekt zur Unterstützung der Steuerung des Bahn- respektive des Strassenverkehrs

**Aktionsfeld 2: Zwingender Datenaustausch für KI in Automatisierter Mobilität**

- Umsetzung bestehender Massnahmenpläne
- Aufbau «Verkehrsnetz Schweiz» zur Geolokalisierung
- Abklärungen für Pilotprojekt mobile breitbandige Sicherheitskommunikation (MSK)

**Aktionsfeld 3: Datenschutz in Automatisierter Mobilität**

- Koordinationsnetzwerk mit dem EDÖB

**Aktionsfeld 4: Rechtssetzung und gesellschaftliche Akzeptanz von KI in Automatisierter Mobilität**

- Bewilligungen für automatisiertes Fahren
- Koordination Zulassungsverfahren für automatisierte Fahrzeuge
- Abklärung Fehlertoleranzen von KI
- Koordinationsnetzwerk im Bereich Recht und Internationales

<b>KI im Gesundheitswesen</b> (BAG, Swissmedic)
<b>Aktionsfeld 1: Humanforschung</b> <ul style="list-style-type: none"><li>• Prüfung gesetzliche Grundlagen im Bereich Weiterverwendung von Proben und Daten sowie Biobanken</li><li>• Einheitliche Positionierung im Bereich Datennutzung und Big Data</li></ul> <b>Aktionsfeld 2: Heilmittelgesetz HMG</b> <ul style="list-style-type: none"><li>• Prüfung von zukünftigen Lösungsansätzen in der Arzneimittelentwicklung</li></ul>
<b>KI in der Finanzwirtschaft</b> (EFD)
<b>Aktionsfeld 1: Verhaltenspflichten</b> <ul style="list-style-type: none"><li>• Weiterverfolgung der Entwicklungen bei Verhaltenspflichten</li></ul> <b>Aktionsfeld 2: Operationelle Risiken</b> <ul style="list-style-type: none"><li>• Weiterverfolgung der Entwicklungen operationelle Risiken bei Finanzmarktinstituten</li></ul> <b>Aktionsfeld 3: Bestimmung der Versicherungsprämie in der Privatversicherung</b> <ul style="list-style-type: none"><li>• Weiterverfolgung der Entwicklungen bei Versicherungsprämien in der Privatversicherung</li></ul>
<b>KI in der Landwirtschaft</b> (BLW)
<b>Aktionsfeld 1: Auswirkungen der KI in der Landwirtschaft</b> <ul style="list-style-type: none"><li>• Weiterverfolgung der Entwicklungen der KI in der Landwirtschaft</li></ul>
<b>Energie, Klima, Umwelt und KI</b> (BFE, BAFU)
<b>Aktionsfeld 1: KI im Energiesektor</b> <ul style="list-style-type: none"><li>• Weiterverfolgung der Entwicklungen im Bereich «Energie»</li></ul> <b>Aktionsfeld 2: KI in Umwelt und Klima</b> <ul style="list-style-type: none"><li>• Weiterverfolgung der Entwicklungen im Bereich «Umwelt und Klima»</li></ul>
<b>KI in der Verwaltung</b> (ISB, BFS, Bundesämter mit grossen Datenmengen (u.a. EZV, BFS, ESTV, BLW))
<b>Aktionsfeld 1: Gemeinsame Datengrundlagen</b> <ul style="list-style-type: none"><li>• Schaffung und Bereitstellung von Datenbeständen innerhalb der Bundesverwaltung</li></ul> <b>Aktionsfeld 2: Kompetenznetzwerk KI in der Bundesverwaltung</b> <ul style="list-style-type: none"><li>• Vertiefte Abklärung zur Schaffung eines Kompetenznetzwerkes KI mit spezifischem Fokus auf technischen Aspekten der Anwendung von KI in der Bundesverwaltung</li></ul> <b>Aktionsfeld 3: Chancen von KI aufzeigen (Kommunikation)</b> <ul style="list-style-type: none"><li>• Verstärkte Kommunikation über KI relevante Themen innerhalb der Bundesverwaltung</li></ul> <b>Aktionsfeld 4: Gesetzliche Grundlagen, Datenhoheit und Datensicherheit</b> <ul style="list-style-type: none"><li>• Überprüfung gesetzliche Grundlagen für Einsatz KI in der Bundesverwaltung</li></ul>
<b>Weiterentwicklung des allgemeinen Rechtsrahmens im Hinblick auf künstliche Intelligenz</b> (EDA)
<b>Aktionsfeld 1: Entstehung von KI-spezifischem Völkerrecht</b> <ul style="list-style-type: none"><li>• Erarbeitung eines Berichts über die Entwicklung des Völkerrechts im KI-Bereich</li></ul> <b>Aktionsfeld 2: Erkennbarkeit von KI-Systemen in der Interaktion mit Konsumenten</b> <ul style="list-style-type: none"><li>• Verfolgen der Entwicklungen der Interaktion von KI-Systemen</li></ul>
<b>Einsatz von KI in der Justiz</b> (BJ, EDA)
<b>Aktionsfeld 1: KI-basierte Entscheid-Unterstützung in Verwaltung und Justiz (Predictive Analytics)</b> <ul style="list-style-type: none"><li>• Beobachtung der Entwicklungen bei KI-basierter Entscheid-Unterstützung in Verwaltung und Justiz (Predictive Analytics)</li></ul>

**KI, Daten und Immaterialgüterrecht (BAKOM, BJ, IGE, GS EDI/BFS)**

**Aktionsfeld 1: Datenpolitik**

- Weiterführung bestehende Arbeiten zur Datenpolitik des Bundes

**Aktionsfeld 2: Datenschutz**

- Weiterführung bestehende Arbeiten beim Datenschutz

**Aktionsfeld 3: Geistiges Eigentum**

- Weiterführung bestehende Arbeiten beim geistigen Eigentum

## Anhang 2: Maschinelles Lernen

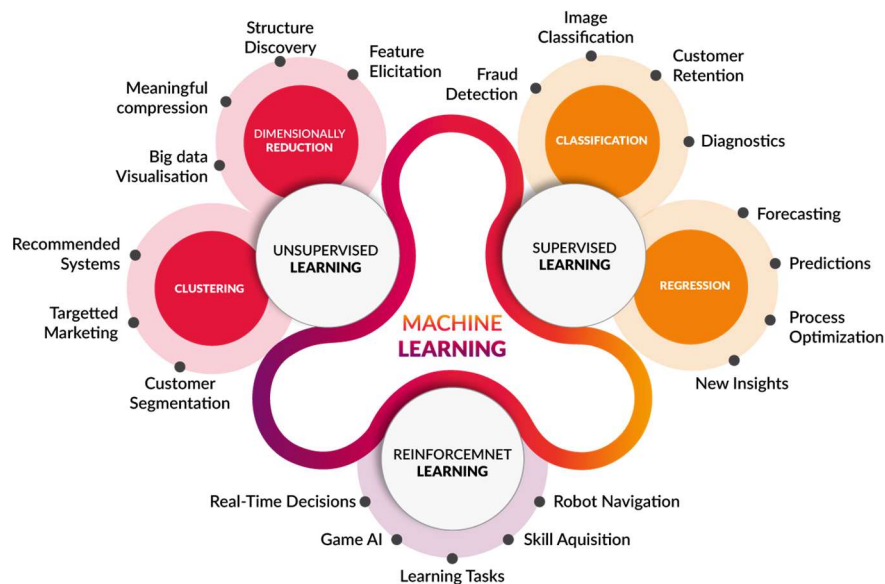
In diesem Anhang werden neben dem maschinellen Lernen selbst die folgenden, zentralen Begriffe erläutert: überwachtes Lernen, unüberwachtes Lernen, bestärkendes Lernen, Deep Learning, Künstliche Neuronale Netzwerke (KNN), Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN).

**Maschinelles Lernen (ML)** bezeichnet Ansätze, bei welchen Maschinen die Fähigkeit besitzen, sich ihr eigenes Wissen anzueignen, indem sie Muster aus Rohdaten extrahieren und auf dieser Basis automatisiert robuste Prognosen in komplexen Daten erstellen.<sup>123</sup>

### Kategorien von Lernansätzen

Derzeit lassen sich drei hauptsächliche Kategorien unterscheiden, wie solche Systeme lernen: überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen. Diese Ansätze haben spezifische Vor- und Nachteile und sind daher für unterschiedliche Anwendungen geeignet (Abbildung 14).

Abbildung 14: Lernansätze, Funktionen und Anwendungsgebiete des maschinellen Lernens



Quelle: "Towards Data Science: Coding Deep Learning For Beginners", verfügbar unter

<https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

**Überwachtes Lernen (supervised learning):** Überwachtes Lernen hat das Ziel, ein durch Beispieldaten vorgegebenes Verhalten zu imitieren. Basierend auf einem Trainingsdatensatz von bekannten Eingabe-Ausgabe Paaren (z.B. Bildern mit Annotationen, welche Objekte sich darauf befinden) werden Abhängigkeiten identifiziert, die Vorhersagen auf neuen (nicht annotierten) Daten ermöglichen (diese Objekte z.B. also auf neuen Bildern erkennen).

*Typische Anwendungen:* Bildklassifikation, Spamfilter, medizinische Diagnose.

**Unüberwachtes Lernen (unsupervised learning):** Ist kein explizites Lernziel in den Daten vorhanden, kann mit unüberwachtem Lernen eigenständig nach Mustern (z.B. verschiedene Gruppen) in Da-

<sup>123</sup> Die Definitionen und Beschreibungen in diesem Anhang basieren auf folgenden Arbeiten:

- Matt Taddy (2018): «The Technological Elements of Artificial Intelligence».
- Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016): "Deep Learning".
- IEC White Paper (2018): "Artificial intelligence across industries", International Electrotechnical Commission, <https://basecamp.iec.ch/download/iec-white-paper-artificial-intelligence-across-industries-en/>
- Schweizerische Akademie der Technischen Wissenschaften SATW (2019): «Technology Outlook 2019».
- Yann LeCun, Yoshua Bengio & Geoffrey Hinton (2015): "Deep Learning".

ten gesucht werden. Die Algorithmen bestimmen die zugrundeliegende Struktur des Datensatzes ohne Informationen über Zielkriterien. Bei einer Reihe von Bildern könnte ein solcher Algorithmus beispielsweise erkennen, dass die Objekte in den verschiedenen Bildern nicht identisch sind. Ohne die Objekte zu kennen, können dann verschiedene Kategorien gebildet werden.

*Typische Anwendungen:* Kundensegmentierung oder Produktempfehlungen.

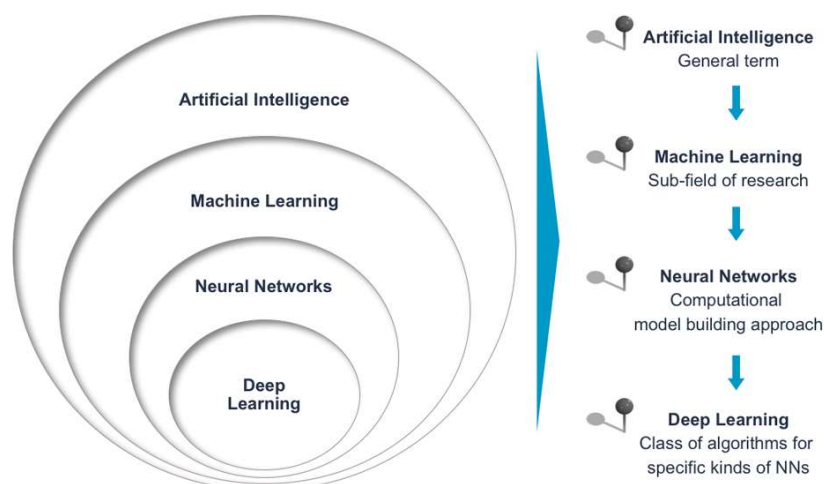
**Bestärkendes Lernen (reinforcement learning):** Beim bestärkenden Lernen lernt das System statt aus Daten durch Interaktion mit seiner Umgebung (typischerweise in einer Computer-Simulation). Das System probiert eigenständig unterschiedliche Lösungswege und erhält aus seiner Umwelt *Rückmeldungen* (Feedbacks), die ein bestimmtes Vorgehen belohnen oder bestrafen. Dabei wird nicht vorgezeigt, welche Aktion in welcher Situation die beste ist. Vielmehr werden durch wiederholte Handlungen und Rückmeldungen die Auswirkungen von Handlungen in bestimmten Situationen erlernt.

*Typische Anwendungen:* KI in Spielen, robotische Steuerung

## Deep Learning

Innerhalb des maschinellen Lernens gibt es eine Vielzahl an Modellen und Algorithmen, die für unterschiedliche Anwendungen geeignet sind. Die aktuell erfolgreichsten Anwendungen im Bereich künstlicher Intelligenz sind jedoch wesentlich auf eine bestimmte Klasse von Methoden im Bereich maschinelles Lernen, dem sog. **Deep Learning** auf Basis von **künstlichen neuronalen Netzwerken**, zurückzuführen (Abbildung 15).

Abbildung 15: Verschiedene Abstraktionslevel von künstlicher Intelligenz



Quelle: Capgemini: Artificial Intelligence, Machine Learning und Data Science: Same same but different?!, verfügbar unter <https://www.capgemini.com/de-de/2017/09/artificial-intelligence-machine-learning-und-data-science-same-same-but-different/>

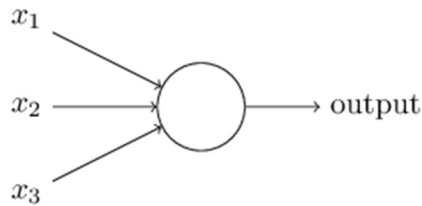
Siehe auch: Ian Goodfellow and Yoshua Bengio and Aaron Courville (2016), "Deep Learning", MIT Press

**Künstliche Neuronale Netzwerke (KNN)** sind die zentrale Grundlage der Algorithmen die heute im Bereich des Deep Learning Verwendung finden. KNN sind vage nach dem Vorbild der Neuronen in einem Gehirn modelliert. Sie bestehen typischerweise aus Schichten von Knoten (künstliche Neuronen), die durch verstellbare Gewichtsverknüpfungen (Synapsen) miteinander verbunden sind.

Die Knoten sind zentrale Elemente von KNN. In ihrer einfachsten Form nehmen sie Informationen entgegen und entscheiden mit einer Gewichtung dieser Informationen, ob als Ergebnis ein Ereignis eintritt (vgl. Abbildung 16). Z.B. kann die Entscheidung mit dem Fahrrad zur Arbeit zu Fahren von zwei Faktoren abhängen: (a) ob der Tag ein Arbeitstag ist, und (b) ob das Wetter gut ist. Mit der Zeit *lernt* das System, dass die Bedeutung von Arbeitstagen viel wichtiger ist als das Wetter und passt die Gewichte entsprechend an.

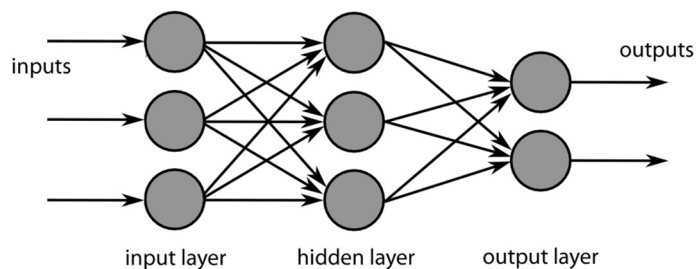


Abbildung 16: Vereinfachte Darstellung eines Knotens eines künstlichen neuronalen Netzes



Ein KNN besteht aber aus zahlreichen, oft Millionen von Knoten, die in Schichten angeordnet sind. Jeder Knoten in einer Schicht ist verbunden mit allen (oder mehreren) Knoten in der nächsten Schicht und kann entsprechend seine Informationen weitergeben (Abbildung 17). So kann die Information, ob mit dem Fahrrad zur Arbeit gefahren wurde, für zahllose weitere Entscheidungen wichtig sein, z.B. über die Entscheidung, ob das Mittagessen im Park oder in einem Restaurant eingenommen wird.

Abbildung 17: Vereinfachte Darstellung eines künstlichen neuronalen Netzes

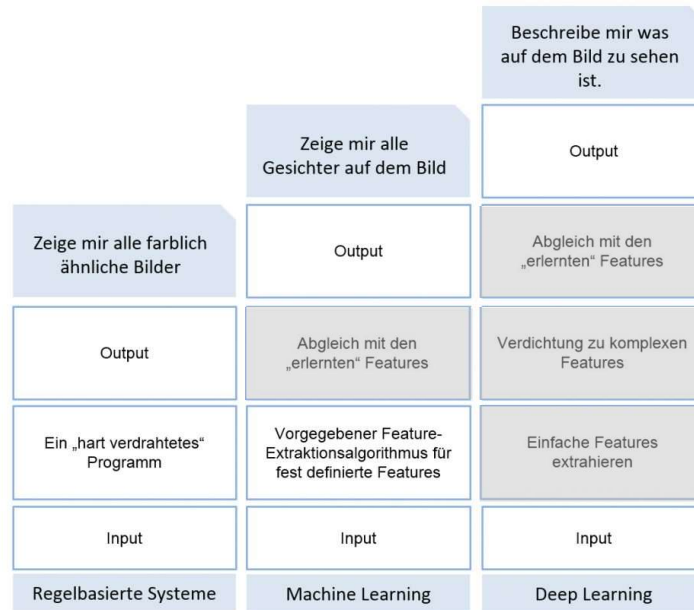


Die Qualität von ML-Anwendungen hängt wesentlich davon ab, wie gut der Output anhand der vorhandenen Merkmale (Features) charakterisiert werden kann (Repräsentation). Die Auswahl der richtigen Merkmale ist eine komplexe und zeitaufwändige Aufgabe, die hochspezialisiertes Fachwissen erfordert. Oft ist dies auch fast unmöglich. So lässt sich ein Auto zwar einfach anhand der Merkmale Reifen, Scheiben, Spiegel etc. beschreiben, solch abstrakte Konzepte aber auf Basis von Pixeln eines Bildes zu beschreiben (die der Computer benötigt), ist für einen Menschen kaum allgemein möglich.

Ein Teilgebiet des maschinellen Lernens (**Repräsentationslernen**) löst dieses Problem, indem nicht nur der Zusammenhang zwischen Merkmalen und Ergebnis (z.B. zwischen Vorhandensein eines Reifens und einem Auto), sondern auch die geeigneten Merkmale selbst, die zur Klassifizierung notwendig sind, automatisiert erlernt werden.

**Deep Learning** geht dabei noch einen Schritt weiter, indem nicht nur die Merkmale, sondern eine eigene *Hierarchie* von Repräsentationen eingeführt und erlernt wird: komplexe Repräsentationen werden durch immer einfachere Repräsentationen aufgebaut und definiert. Damit kann ein solches System eigene Repräsentationen aufbauen und selbst bestimmen, welche Merkmale zur Repräsentation auf der jeweiligen Stufe und für den Output jeweils notwendig und relevant sind (Abbildung 18).

Abbildung 18: Machine Learning/Deep Learning versus regelbasierte Systeme



Anmerkung: grau schattierte Kästchen kennzeichnen Komponenten, die aus Daten lernen können.

Quelle: Jaxenter: Maschinelle Bilderkennung mit Big Data und Deep Learning, verfügbar unter <https://jaxenter.de/big-data-bildanalyse-50313>, nach Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016), “Deep Learning”, MIT Press.

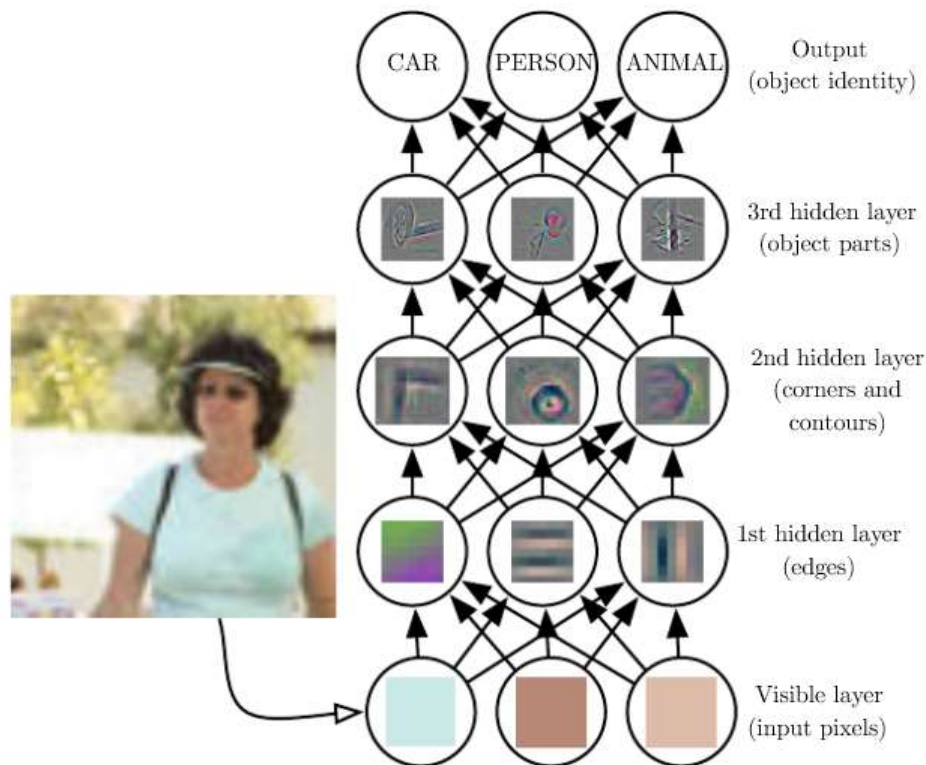
**Deep Neural Networks (DNN)** sind die zentralen Modelle innerhalb des DL. Diese stellen im Grunde genommen nichts anderes als KNN dar, wobei im Vergleich zu einem einfachen KNN zwischen der Input- und der Outputschicht mehrere versteckte Ebenen liegen (d.h. mehr als ein «hidden layer» in Abbildung 17). Solche Systeme können heute Milliarden von Knoten in dutzenden von Schichten aufweisen.<sup>124</sup> Diese Schichten werden als «versteckt» bezeichnet, weil ihre Werte nicht in den Daten angegeben sind und nicht vorgegeben werden müssen. Stattdessen bestimmt das Modell selbstständig, welche Konzepte zur Erklärung der Zusammenhänge in den beobachteten Daten nützlich sind.

Abbildung 19 zeigt, wie ein Deep-Learning-System das Konzept eines Personenbildes erfassen kann, indem die komplizierte Zuordnung in eine Reihe einfacherer, verschachtelter Zusammenhänge (d.h. eine eigene Hierarchie von Repräsentationen) unterteilt wird.

Die Eingabe eines Bildes auf Pixel-Ebene wird zunächst mit sehr einfachen Konzepten erfasst. So werden z.B. auf dieser Ebene nur Kanten identifiziert, indem die Helligkeit benachbarter Pixel verglichen wird. Angesichts der Beschreibung der Kanten durch die erste Ebene sucht die zweite Ebene nach weiter abstrakteren Konzepten, z.B. Ecken und Konturen. Damit kann die dritte Ebene bereits ganze Teile von Objekten erkennen, indem verschiedene Konturen und Ecken zusammengefasst werden. Schliesslich kann diese Information zur Identifikation auf die im Bild enthaltenen Objektteile verwendet werden. Am Ende des Prozesses klassifiziert ein solches System die im Bild enthaltenen Objekte.

<sup>124</sup> Google benutzt bspw. für die Bilderkennung in «Google Photo» ein KNN mit 30 Schichten. Androidcentral (2015): “Google is building deep neural networks to help improve its search engine”, <https://www.androidcentral.com/google-building-deep-neural-networks-help-improve-their-search-engine>

Abbildung 19: Illustration eines Deep Learning Modells



Quelle: Ian Goodfellow and Yoshua Bengio and Aaron Courville (2016): "Deep Learning", MIT Press.

Das System erhält mit jedem neuen Trainingsbeispiel Feedback über die Erkennungsleistung. Die unzähligen Parameter werden daraufhin jeweils angepasst. Auf diese Weise lernt das System nach und nach, was bspw. am Gesicht eines Menschen essenziell ist. Das Training wird so lange wiederholt, bis die Fehlerrate zufriedenstellend ausfällt. Nach dieser Trainingsphase sollte das System in der Lage sein, mit den vorhandenen Parametern (welche nun nicht mehr verändert werden), die Erkennung auch in neuen, unbekanntem Daten vorzunehmen. Dazu ist dann eine wesentlich geringere Rechenleistung notwendig.

Unter den unzähligen Varianten von DNN sind schliesslich drei hervorzuheben:

**Convolutional Neural Networks (CNN):** Diese sind spezifisch daraus ausgerichtet, mit (hauptsächlich) zweidimensionalen Daten umzugehen, weshalb sie insbesondere für die Verarbeitung von Bildern geeignet sind. In den Bereichen Bilderkennung, Personenidentifikation, autonomer Robotik und selbstfahrenden Autos, aber auch in der Spracherkennung ist diese, sehr erfolgreiche Variante der KNN heute stark dominierend und erzielt spektakuläre Resultate. Neuere CNN-Architekturen verfügen über 10 bis 20 Schichten, Hunderte von Millionen Gewichten und Milliarden von Verbindungen zwischen den Knoten. Während das Training solch grosser Netzwerke noch vor zwei Jahren Wochen dauern konnte, haben die Fortschritte bei der Parallelisierung von Hardware, Software und Algorithmen die Trainingszeiten auf wenige Stunden reduziert.

**Recurrent Neural Networks (RNN):** Eine weitere bedeutsame Erweiterung sind RNN. Während CNN besonders gut mit Daten umgehen können, die in (zwei) Dimensionen angeordnet sind, sind RNN auf sequentielle Daten spezialisiert. Daher sind diese Modelle besonders erfolgreich bei der Textverarbeitung. Eine der erfolgreichsten Implementierungen von RNN ist das sog. long short-term memory (LSTM), welches am **Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA) in Lugano** entwickelt wurde. Die darauf basierenden Algorithmen werden heute in rund 3 Milliarden Smartphones täglich benutzt und betreiben u.a. die Spracherkennung und Übersetzung von Google sowie die Übersetzung bei Facebook.

**Generative Adversarial Networks (GAN):** Schliesslich stellt die Entwicklung von GAN ein weiterer wichtiger Meilenstein in der Entwicklung von ML dar. Die relativ simple Idee hinter GAN ist, dass zwei KNN miteinander im Wettbewerb stehen: eines versucht (auf Basis eines vorgegebenen Trainingsatzes) neue Daten zu generieren, die von vorgegebenen Daten nicht unterscheidbar sind, während das andere die Daten bewertet und versucht, die neuen Daten zu erkennen und von den echten, vorgegebenen Daten zu unterscheiden. Diese Technik lernt somit, neue Daten mit denselben Statistiken wie der Trainingsatz zu generieren. Beispielsweise kann ein auf Fotografien trainiertes GAN neue Fotografien erzeugen, die für menschliche Beobachter authentisch aussehen und viele realistische Eigenschaften aufweisen. Während die Anwendungsmöglichkeiten relativ breit sind, wurden GAN einer breiten Öffentlichkeit v.a. durch gefälschte Fotos und Videos («Deep Fakes») bekannt.

## Anhang 3: Bibliographie

### Kapitel 1 – 5

- Ackerman E. (2017): "Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms". IEEE Spectrum: <https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms>
- Agrawal, Ajay, Gans, Joshua and Goldfarb, Avi, (2018): "Prediction, Judgment, and Complexity: A Theory of Decision Making and Artificial Intelligence" in "The Economics of Artificial Intelligence: An Agenda", National Bureau of Economic Research, Inc.
- Gopala K. Anumanchipalli, Josh Chartier & Edward F. Chang (2019): "Speech synthesis from neural decoding of spoken sentences"; Nature, Volume 568, pages 493–498.  
<https://techcrunch.com/2019/04/24/scientists-pull-speech-directly-from-the-brain/>
- ASGARD (2017): «The European Artificial Intelligence Landscape», <https://asgard.vc/the-european-artificial-intelligence-landscape-more-than-400-ai-companies-made-in-europe/>
- Berkeley Haas (2018): "Minority homebuyers face widespread statistical lending discrimination, study finds", <https://newsroom.haas.berkeley.edu/minority-homebuyers-face-widespread-statistical-lending-discrimination-study-finds/>
- Bundesrat (2014): «Eine Industriepolitik für die Schweiz», Bericht des Bundesrats vom 16.04.2014 in Erfüllung des Postulats Bischof.
- Bundesrat (2017): «Bericht über die zentralen Rahmenbedingungen für die digitale Wirtschaft», verfügbar unter <https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirtschaftspolitik/digitalisierung.html>
- Bundesrat (2017): Bericht in Erfüllung der Postulate 15.3854 Reynard «Automatisierung. Chancen und Risiken» vom 16. September 2015 und 17.3222 Derder «Digitale Wirtschaft. Die Arbeitsplätze der Zukunft und Massnahmen für ihre Förderung in der Schweiz identifizieren» vom 17. März 2017, verfügbar unter <https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirtschaftspolitik/digitalisierung.html>
- Bundesrat (2018) «Gesamtschau der Innovationspolitik», Bericht in Beantwortung des Postulats Derder 13.3073 vom 13. März 2013.
- Bundesrat (2018): Bericht des Bundesrates «Rechtliche Grundlagen für Distributed Ledger Technologie und Blockchain in der Schweiz».
- Carlini und Wagner (2018): "Audio Adversarial Examples: Targeted Attacks on Speech-to-Text", arXiv:1801.01944 [cs.LG] , <https://arxiv.org/abs/1801.01944>
- CNN Business (2019): "IBM's fast-talking AI machine just lost to a human champion in a live debate", verfügbar unter <https://edition.cnn.com/2019/02/11/tech/ai-versus-human-ibm-debate/index.html>
- Iain M. Cockburn, Rebecca Henderson, Scott Stern (2018): "The Impact of Artificial Intelligence on Innovation - An Exploratory Analysis", (bibliographic info) (download) version of January 10, 2018 (Working Paper version).
- Conseil de l'Europe DGI (2017)12 : Etude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires, p. 29 ss.

## Herausforderungen der künstlichen Intelligenz

Conseil de l'Europe (2018) : « Discrimination, artificial intelligence and algorithmic decision-making ».

EconSight (2019): «Künstliche Intelligenz, Globale Entwicklungen, Anwendungsgebiete, Innovationstreiber und Weltklasseforschung», <https://www.econsight.ch/artificial-intelligence/>

Elsevier (2018): «AI Report»,  
<https://public.tableau.com/profile/isabella.cingolani1149#!/vizhome/ElseviersAIprogramme/Dashboard?publish=yes>

Scott Fortmann-Roe (2012): «Understanding the Bias-Variance Tradeoff»; <http://scott.fortmann-roe.com/docs/BiasVariance.html>

Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016): “Deep Learning”, MIT Press,  
<http://www.deeplearningbook.org>

The Guardian (2018): “Women must act now, or male-designed robots will take over our lives”,  
<https://www.theguardian.com/commentisfree/2018/mar/13/women-robots-ai-male-artificial-intelligence-automation>

Jovanovic/Rousseau (2005): «General Purpose Technologies, Handbook of Economic Growth», in: Aghion/Durlauf (Hrsg.), Handbook of Economic Growth, S. 1181–1224, Elsevier.

Kaplan Andreas; Michael Haenlein (2018): “Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence”, Business Horizons, 62(1).

Regina Kiener/Walter Kälin/Judith Wyttenbach, Grundrechte, 3. Auflage, 2018.

Sebastian Lapuschkin, Stephan Wäldchen, Alexander Binder, Grégoire Montavon, Wojciech Samek & Klaus-Robert Müller (2019): «Unmasking Clever Hans predictors and assessing what machines really learn”, Nature Communications, volume 10, Article number: 1096 (2019).

Yann LeCun, Yoshua Bengio & Geoffrey Hinton (2015): “Deep Learning”, Nature Vol. 521:436-44, May 2015.

J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon (1955). “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

MIT Technology Review (2017): “The Dark Secret at the Heart of AI”,  
<https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>

Sendhil Mullainathan and Jann Spiess (2017): “Machine Learning - An Applied Econometric Approach”; Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 87–106.

OECD (2017): Science, Technology and Industry Scoreboard 2017 – The Digital Transformation, verfügbar unter <https://www.oecd.org/sti/oecd-science-technology-and-industry-scoreboard-20725345.htm>

OECD (2019): «Artificial Intelligence in Society», <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>

Roland Berger (2018): “Artificial Intelligence - A strategy for European startups”,  
[https://www.rolandberger.com/publications/publication\\_pdf/roland\\_berger\\_ai\\_strategy\\_for\\_european\\_startups.pdf](https://www.rolandberger.com/publications/publication_pdf/roland_berger_ai_strategy_for_european_startups.pdf)

## Herausforderungen der künstlichen Intelligenz

- Schweizerische Akademie der Technischen Wissenschaften SATW (2019): «Technology Outlook 2019», verfügbar unter <https://www.satw.ch/de/frueherkennung/technologies/>
- SBFI (2017), «Herausforderungen der Digitalisierung für Bildung und Forschung in der Schweiz»; [https://www.sbf.admin.ch/dam/sbf/de/dokumente/webshop/2017/bericht-digitalisierung.pdf.download.pdf/bericht\\_digitalisierung\\_d.pdf](https://www.sbf.admin.ch/dam/sbf/de/dokumente/webshop/2017/bericht-digitalisierung.pdf.download.pdf/bericht_digitalisierung_d.pdf)
- E. Strickland (2019): «IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care», in IEEE Spectrum, vol. 56, no. 4, pp. 24-31, April 2019.
- Su, Vargas und Kouichi (2017): “One pixel attack for fooling deep neural networks”, arXiv:1710.08864 [cs.LG], <https://arxiv.org/abs/1710.08864>
- Szegedy et al. (2014): “Intriguing properties of neural networks”, arXiv:1312.6199v4 [cs.CV], <https://arxiv.org/abs/1312.6199v4>
- Matt Taddy (2018): «The Technological Elements of Artificial Intelligence», Chapter in forthcoming NBER book The Economics of Artificial Intelligence: An Agenda, Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb, editors.
- TA-SWISS (Hrsg.): «Wenn Algorithmen an unserer Stelle entscheiden: die Herausforderungen der künstlichen Intelligenz»; unveröffentlichtes Manuskript.
- Rodney Brooks (2017). "The Seven Deadly Sins of AI Predictions", MIT Technology Review, October 2017, verfügbar unter <https://www.heise.de/tr/artikel/Essay-Die-sieben-Todsunden-der-KI-Vorhersagen-4003150.html>
- Manuel Trajtenberg (2018): “AI as the Next GPT - A Political-Economy Perspective”, version of January 23, 2018 (Working Paper version).
- A. M. Turing (1950): «Computing Machinery and Intelligence» Mind 49: 433-460.
- VALUER (2018): The best AI startups in Europe, <https://valuer.ai/blog/the-best-ai-startups-in-europe/>
- Weizenbaum, Joseph (1966): “ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine”, verfügbar unter <http://www.cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1966.pdf>
- Yuan et al. (2018): “CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition”, arXiv:1801.08535 [cs.CR], <https://arxiv.org/abs/1801.08535>; [https://www.ics.uci.edu/~alfchen/yulong\\_ccs19.pdf](https://www.ics.uci.edu/~alfchen/yulong_ccs19.pdf)

## Internationale Gremien und KI

- Bericht der Projektgruppe «Internationale Gremien und künstliche Intelligenz», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

## Digital Europe Programme

Europäische Kommission (2018): «Künstliche Intelligenz für Europa»,  
<https://ec.europa.eu/transparency/regdoc/rep/1/2018/DE/COM-2018-237-F1-DE-MAIN-PART-1.PDF>

SBFI (2019): «Ergebnisse der Umfrage zum Digital Europe Programme (DEP)»,  
[https://www.sbf.admin.ch/dam/sbf/de/dokumente/2019/07/ergebnisse-dep.pdf.download.pdf/bericht\\_dep\\_d.pdf](https://www.sbf.admin.ch/dam/sbf/de/dokumente/2019/07/ergebnisse-dep.pdf.download.pdf/bericht_dep_d.pdf)

## Veränderungen der Arbeitswelt

Bundesrat (2017): «Bericht über die zentralen Rahmenbedingungen für die digitale Wirtschaft», verfügbar unter <https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirtschaftspolitik/digitalisierung.html>

Bundesrat (2017): Bericht «Auswirkungen der Digitalisierung auf Beschäftigung und Arbeitsbedingungen – Chancen und Risiken», verfügbar unter <https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirtschaftspolitik/digitalisierung.html>

Bundesrat (2018): Ergebnisbericht der Umfrage «Digitaler Test», Überprüfung regulatorischer Hindernisse für die Digitalisierung, verfügbar unter <https://www.seco.admin.ch/seco/de/home/wirtschaftslage---wirtschaftspolitik/wirtschaftspolitik/digitalisierung.html>

OECD (2019): «The Future of Work - Employment Outlook 2019», verfügbar unter [https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019\\_9ee00155-en](https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2019_9ee00155-en)

## KI in Industrie und Dienstleistungen

SATW (2019): «Künstliche Intelligenz in Industrie und Dienstleistungen», Bericht im Auftrag des Staatssekretariats für Bildung, Forschung und Innovation, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

## KI in der Bildung

educa.ch (2019): «Daten in der Bildung – Daten für die Bildung. Grundlagen und Ansätze zur Entwicklung einer Datennutzungspolitik für den Bildungsraum Schweiz». Bern.

Staatssekretariat für Bildung, Forschung und Innovation (2019): «Künstliche Intelligenz in der Bildung», verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

Tuomi, I. (2018): «The Impact of Artificial Intelligence on Learning, Teaching, and Education. Policies for the future», Eds. Cabrera, M., Vuorikari, R & Punie, Y., Publications Office of the European Union, Luxembourg, 2018.



## Anwendung von KI in Wissenschaft und Forschung

SATW (2019): «Künstliche Intelligenz in Wissenschaft und Forschung», Bericht im Auftrag des Staatssekretariats für Bildung, Forschung und Innovation, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

## KI in der Cybersicherheit und Sicherheitspolitik

Bericht der Projektgruppe «Künstliche Intelligenz in der Cybersicherheit und Sicherheitspolitik», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

ETH Zürich / CSS (2019): «Studie KI und Sicherheitspolitik - Künstliche Intelligenz, technologischer Wandel und nationale und internationale Sicherheitspolitik».

ETH Zürich / CSS (2019): «Policy Perspectives - Ein neutraler Hub für KI-Forschung».

EPFL (2019) : « Studie Cybersécurité et Politique de sécurité ».

OECD (2019): «Artificial Intelligence in Society», <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>

## KI, Medien & Öffentlichkeit

Bericht der Projektgruppe «Künstliche Intelligenz, Medien & Öffentlichkeit», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

Dreyer, Stephan; Schulz, Wolfgang (2019): Künstliche Intelligenz, Intermediäre und Öffentlichkeit. Bericht an das BAKOM erstattet durch das Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG) und das Leibnitz-Institut für Medienforschung | Hans-Bredow-Institut (HBI).

Eidgenössische Medienkommission EMEK (2018): Besonderheiten von Medien im digitalen Zeitalter: Gestaltungsoptionen für eine leistungsfähige Medienlandschaft auf ökonomischer und gesellschaftlicher Perspektive. Arbeitspapier der EMEK. Biel. Online verfügbar unter [https://www.emek.admin.ch/inhalte/dokumentation/22.01.2018\\_Besonderheiten\\_von\\_Medien\\_im\\_digitalen\\_Zeitalter/D\\_Medien\\_im\\_digitalen\\_Zeitalter\\_22.01.18.pdf](https://www.emek.admin.ch/inhalte/dokumentation/22.01.2018_Besonderheiten_von_Medien_im_digitalen_Zeitalter/D_Medien_im_digitalen_Zeitalter_22.01.18.pdf), zuletzt geprüft am 13.08.2019.

Fichter, Adrienne (2018): Die Schweiz wappnet sich für den Angriff aus dem Silicon Valley. In: Republik, 16.05.2018. Online verfügbar unter <https://www.republik.ch/2018/05/16/die-schweiz-wappnet-sich-fuer-den-angriff-aus-dem-silicon-valley>, zuletzt geprüft am 13.08.2019.

Gillespie, Tarleton (2017): #trendingistrending. Wenn Algorithmen zur Kultur werden. In: Robert Seyfert und Jonathan Roberge (Hg.): Algorithmenkulturen. Über die rechnerische Konstruktion der Wirklichkeit. Bielefeld: transcript Verlag, S. 75–106.

Goldhammer, Klaus; Dietrich, Kevin; Prien, Tim (2019): Künstliche Intelligenz, Medien und Öffentlichkeit. Wissenschaftlicher Bericht. Berlin.

Jarren, Otfried (2018b): Normbildende Macht. In: epd medien (24), S. 35–39.

Livingstone, Sonia (2019): Audiences in the Age of Datafication: Critical Questions for Media Research. In: Television & New Media 20 (2), S. 170–183.

Lobigs, Frank; Neuberger, Christoph (2018): Meinungsmacht im Internet und die Digitalstrategien von Medienunternehmen. Neue Machtverhältnisse trotz expandierender Internet-Geschäfte der

## Herausforderungen der künstlichen Intelligenz

traditionellen Massenmedien-Konzerne. Gutachten für die Kommission zur Ermittlung der Konzentration im Medienbereich (KEK). Leipzig.

Saurwein, Florian; Just, Natascha; Latzer, Michael (2017): Algorithmische Selektion im Internet: Risiken und Governance automatisierter Auswahlprozesse. In: kommunikation@gesellschaft, 22 Seiten.

Schweizerischer Bundesrat (2017): Rechtliche Basis für Social Media: Erneute Standortbestimmung. Nachfolgebericht des Bundesrates zum Postulatsbericht Amherd 11.3912 „Rechtliche Basis für Social Media“. Bern. Online verfügbar unter <https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/digitale-kommunikation/soziale-medien.html>, zuletzt geprüft am 13.08.2019.

Schweizer Presserat (Hrg.): Journalistenkodex. Online verfügbar unter <https://presserat.ch/journalistenkodex/erklaerung/>, zuletzt geprüft am 13.08.2019.

## Automatisierte Mobilität und KI

Bericht der Projektgruppe «Automatisierte Mobilität und künstliche Intelligenz», August 2019, verfügbar unter [www.sbf.admin.ch/ai-d](http://www.sbf.admin.ch/ai-d)

Bundesrat (2016): Automatisiertes Fahren – Folgen und verkehrspolitische Auswirkungen, Bericht des Bundesrates in Erfüllung des Postulats Leutenegger Oberholzer 14.4169 «Auto-Mobilität. Fahren ohne FahrerIn oder Fahrer. Verkehrspolitische Auswirkungen».

## KI in der Finanzwirtschaft

Su, Vargas und Kouichi (2017): “One pixel attack for fooling deep neural networks”, arXiv:1710.08864 [cs.LG], <https://arxiv.org/abs/1710.08864>

## Energie, Klima, Umwelt und KI

Bericht des Bundesrates (2018): «Umwelt Schweiz 2018», verfügbar unter <https://www.bafu.admin.ch/bafu/de/home/dokumentation/umweltbericht/umweltbericht-2018.html>

Bundesamt für Energie (2019): «Digitalisierung im Energiesektor - Dialogpapier zum Transformationsprozess», verfügbar unter <https://www.bfe.admin.ch/bfe/de/home/versorgung/digitalisierung.html>

Bundesamt für Energie (2019): Pilot-, Demonstrations- und Leuchtturmprogramm, <https://www.bfe.admin.ch/bfe/de/home/forschung-und-cleantech/pilot-demonstrations-und-leuchtturmprogramm.html>

### **KI in der Verwaltung**

Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence (2018): «Improving refugee integration through data-driven algorithmic assignment», *Science*, Vol. 359, Issue 6373, S. 325-329, <https://science.sciencemag.org/content/359/6373/325>

Ergebnisse der Umfrage zum Einsatz von künstlicher Intelligenz (KI) bei den Kantonen vom 15. Juli 2019, Kanton Luzern, unveröffentlichtes Manuskript.

ETH Zürich (2018): "Algorithmus verbessert Erwerbschancen von Flüchtlingen", <https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2018/01/algorithmus-verbessert-erwerbschancen-von-fluechtlingen.html>

### **Einsatz von KI in der Justiz**

Commission européenne pour l'efficacité de la justice (2018) : « Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement », <https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b>