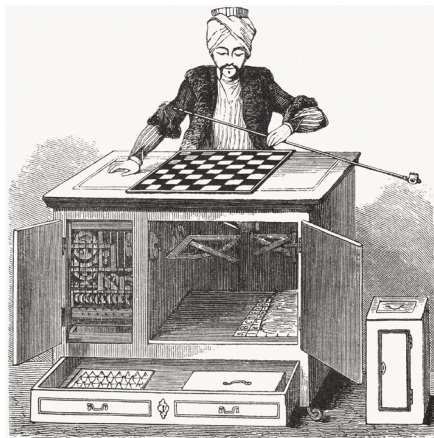


Hartmut
Hirsch-Kreinsen
*Das Versprechen
der Künstlichen
Intelligenz*
Gesellschaftliche
Dynamik einer
Schlüssel-
technologie



Das Versprechen der Künstlichen Intelligenz

Hartmut Hirsch-Kreinsen, Prof. i.R. Dr., ist Research Fellow an der Sozialforschungsstelle Dortmund.

Hartmut Hirsch-Kreinsen

Das Versprechen der Künstlichen Intelligenz

Gesellschaftliche Dynamik einer Schlüsseltechnologie

Campus Verlag
Frankfurt/New York

ISBN 978-3-593-51668-4 Print

ISBN 978-3-593-45311-8 E-Book (PDF)

ISBN 978-3-593-45312-5 E-Book (EPUB)

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlags unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Trotz sorgfältiger inhaltlicher Kontrolle übernehmen wir keine Haftung für die Inhalte externer Links.

Für den Inhalt der verlinkten Seiten sind ausschließlich deren Betreiber verantwortlich.

Copyright © 2023. Alle Rechte bei Campus Verlag GmbH, Frankfurt am Main.

Umschlaggestaltung: Campus Verlag GmbH, Frankfurt am Main.

Umschlagmotiv: Automatonschachspieler von Wolfgang von Kempelen, Holzschnitt, erschienen 1893

© ZU_09, 1224156169, istockphoto.com

Satz: le-tex xerif

Gesetzt aus der Alegreya

Druck und Bindung: Beltz Grafische Betriebe GmbH, Bad Langensalza

Beltz Grafische Betriebe ist ein klimaneutrales Unternehmen (ID 15985–2104-1001).

Printed in Germany

www.campus.de

Inhalt

| | |
|----------------------------------|----|
| Vorwort | 9 |
| Einleitung | 13 |
| Erwartungen und KI-Dynamik | 15 |
| Technologievorsprechen | 17 |
| Fragestellung | 19 |
| Methodische Basis | 20 |
| Zum Gang der Argumentation | 21 |

Teil I: Konzept und Begriffe

| | |
|---|----|
| 1. Promising Technology | 25 |
| 1.1 Leitmotiv und Koordinationsfunktion | 25 |
| 1.2 Wechselwirkungen mit gesellschaftlichen Bedingungen | 30 |
| 2. Entwicklungslinien der KI | 35 |
| 2.1 KI, ein Sammelbegriff | 35 |
| 2.2 »Paradigmen« der KI | 38 |
| 2.3 Maschinelles Lernen | 43 |
| 2.4 Weitere Teilbereiche der KI | 45 |
| 2.5 Koexistenz verschiedenster Konzepte | 46 |

Teil II: Dynamik der KI

| | | |
|-----|--|-----|
| 3. | Frühe Versprechungen in den 1970ern | 51 |
| 3.1 | Beginn in den USA | 51 |
| 3.2 | Bundesrepublik: KI in der wissenschaftlichen Nische | 57 |
| 4. | Jenseits der Nische – Kommerzieller Aufbruch in den 1980ern | 67 |
| 4.1 | Erweiterung des Technologieversprechens | 67 |
| 4.2 | Zusammenspiel von Wissenschaft, Politik und Unternehmen | 69 |
| 4.3 | Weitere Institutionalisierung der KI | 75 |
| 5. | Krise und Konsolidierung ab Ende der 1980er | 79 |
| 5.1 | Krisenursachen | 80 |
| 5.2 | Anschließend: Konsolidierung | 85 |
| 6. | Dynamik ab den 2010ern | 93 |
| 6.1 | »Big Bang« der KI | 93 |
| 6.2 | Ein gesellschaftlich orientiertes Technologieversprechen | 96 |
| 6.3 | KI-Community als »Driving Force« | 102 |
| 6.4 | Politik: Agenda Setting, Koordination und Förderung | 105 |
| 6.5 | Wissenschaft und Unternehmen | 116 |
| 6.6 | Unterstützung durch den öffentlichen Diskurs | 122 |
| 6.7 | Widersprüchliche Diffusion | 126 |
| 7. | Zwischenresümee: Entwicklungsphasen, Akteure und Interessen . | 131 |

Teil III: Perspektiven – Dynamik in Grenzen

| | | |
|-----|---|-----|
| 8. | Grenzen und Herausforderungen der KI | 141 |
| 8.1 | Anwendungsprobleme | 142 |
| 8.2 | Ungewisse ökonomische Effekte | 145 |
| 8.3 | Bremsende ökonomische Rahmenbedingungen | 148 |
| 8.4 | Herausforderung Normierung und Regulation | 152 |
| 8.5 | Hohe technologische Hürden | 155 |

| | | |
|------|--|-----|
| 9. | Eine europäische und deutsche KI-Variante? | 161 |
| 9.1 | Technologische Souveränität | 162 |
| 9.2 | Industrieorientierung | 164 |
| 9.3 | »Trustworthy AI« | 166 |
| 9.4 | Impact vertrauenswürdiger KI | 170 |
| 9.5 | Vertrauenswürdige KI als Alleinstellungsmerkmal? | 175 |
| 9.6 | Deutsche und Europäische KI im internationalen Vergleich . | 178 |
| 10. | Längerfristige Perspektiven | 181 |
| 10.1 | Ein neuer AI-Winter? | 181 |
| 10.2 | Weitreichende Versprechungen: Artificial General Intelligence | 183 |
| 10.3 | »Veralltäglichung« der KI | 187 |

Teil IV: KI – ein besonderer Innovationsmodus

| | | |
|------|---|-----|
| 11. | Institutioneller Rahmen | 193 |
| 11.1 | Nationales Innovationssystem | 193 |
| 11.2 | Industrieorientierung des deutschen Systems | 194 |
| 12. | Besonderheiten des KI-Innovationsmodus | 199 |
| 12.1 | Merkmale des Innovationsmodus | 199 |
| 12.2 | Innovationssystem unter Druck | 210 |
| 12.3 | Segmentationstendenzen | 216 |
| 13. | Innovationspolitik für KI | 221 |
| 13.1 | KI-orientierte Maßnahmen | 221 |
| 13.2 | Missionsorientierung der Innovationspolitik | 232 |

Teil V: Logik des Technologieversprechens

| | |
|--|-----|
| 14. Anschlussfähige Rhetorik | 243 |
| 15. Vereinfachungen und Kurzschlüsse | 247 |
| 15.1 Mehrdeutigkeit und unklare Metaphorik | 247 |
| 15.2 Schlüsseltechnologie: ein vager Begriff | 249 |
| 15.3 Verkürztes Innovationsverständnis | 252 |
| 15.4 Übersteigerte ökonomische Erwartungen | 255 |
| 15.5 Generalisierung von Einzelfällen | 258 |
| 15.6 Legitimationsbeschaffung | 260 |
| 16. Weiterungen: Technikutopie KI | 263 |
| Resümee: Der Mythos der intelligenten Maschine | 267 |
| Literatur | 275 |

Vorwort

Der Begriff Künstliche Intelligenz (KI) bezeichnet eine Technologie, mit der seit Jahrzehnten faszinierende gesellschaftliche Zukunftsaussichten verbunden werden und die völlig neue, bislang nicht gekannte ökonomische und soziale Anwendungspotenziale in Aussicht stellt. Auch gegenwärtig stellt die KI immer wieder ihre besondere Leistungsfähigkeit unter Beweis, seien es die immer besser funktionierenden Programme der Spracherkennung, die oftmals sehr hilfreichen Spamfilter für E-Mails, autonom fahrende Autos oder auch Kunst produzierende Neuronale Netze. Spätestens seit den 2010er-Jahren gilt KI daher als eine der Schlüsseltechnologien des 21. Jahrhunderts, mit deren Hilfe nicht nur weitreichende Produktivitätsfortschritte erreicht, sondern auch die vor uns liegenden ökonomischen, sozialen sowie ökologischen Herausforderungen bewältigt werden sollen. Das Wissen über diese Technologie und die Fähigkeiten, sie anzuwenden, wird als Voraussetzung dafür angesehen, dass nicht nur Deutschland, sondern die gesamte EU in der internationalen Technologiekonkurrenz nicht abgehängt und das erreichte Wohlstandsniveau nicht gefährdet werden. Daher versuchen Deutschland und die EU seit Längerem, durch hohe Investitionen und gezielte innovationspolitische Maßnahmen mit den das Feld der KI dominierenden USA und China gleichzuziehen. Dabei gewinnt der KI-Diskurs zeitweise den Charakter eines Hypes und die damit verknüpften Erwartungen haben geradezu technikutopischen Charakter. Obgleich sich mit der KI zugleich auch vielfältige Befürchtungen, Ängste und geradezu dystopische Perspektiven in Hinblick auf mögliche soziale und ethische Folgen verbinden, ist bis heute die öffentliche Faszination, die von dieser Technologie ausgeht, ungebrochen.

Resümiert man die circa 70-jährige Entwicklungsgeschichte der KI, so kann allerdings keineswegs von einer ungebrochenen Dynamik oder gar ei-

nem kontinuierlichen Aufschwung dieser Technologie gesprochen werden. Vielmehr ist der Entwicklungsprozess der KI von Aufschwung- und Krisenphasen gekennzeichnet. Dies betrifft insbesondere die Entwicklung in den USA, teilweise – mit dem bekannten amerikanisch-europäischen Timelag – aber auch die deutsche Situation. Denn die mit der KI verknüpften Visionen erwiesen sich teilweise als überzogen und die mit ihr verknüpften Erwartungen als völlig unrealistisch. Die Enttäuschungen der hoch gespannten Erwartungen hatten daher die Abkehr ursprünglich stark engagierter Akteure, den Verzicht auf weitere Investitionen in die KI-Entwicklung und ein schnelles Nachlassen des politischen und öffentlichen Interesses zur Folge. Insbesondere im letzten Jahrzehnt ist jedoch eine erneute nachhaltige Beschleunigung der KI-Dynamik und ein rasant steigendes gesellschaftliches Interesse an der KI-Thematik unübersehbar.

An dieser Entwicklungsgeschichte setzt die vorliegende Studie an und es werden aus einer technik- und innovationssoziologischen Sicht die gesellschaftlichen Bestimmungsfaktoren dieser Dynamik in den Blick genommen. Die weitverbreitete Auffassung ist, dass die KI-Dynamik hauptsächlich von der jeweils verfügbaren Leistungsfähigkeit der Rechnersysteme und den Grundlagen und Methoden der KI bestimmt wird. Aus einer sozialwissenschaftlichen Perspektive greift eine solche ausschließlich informatik- und technikzentrierte Begründung jedoch zu kurz. Vielmehr zeigt die Forschung instruktiv, dass Verlauf und Ergebnisse technologischer Innovationen maßgeblich als Resultate sozialer, ökonomischer und politischer Entscheidungen und Einflussfaktoren verstanden werden müssen. Dies gilt ohne Frage auch für die Entwicklung der KI, insbesondere für ihren gegenwärtigen Boom.

Die These der vorliegenden Studie ist daher, dass mit KI ein »Technologieversprechen« verbunden ist, das bei vielen Akteuren Interesse weckt, in der Öffentlichkeit weitreichende Zukunftserwartungen hervorruft und letztlich zu hohen Investitionen und vielfältigen innovationspolitischen Fördermaßnahmen führt. Diese These begründet den »roten Faden« der folgenden Argumentation. Davon ausgehend richtet sich die Analyse zum einen auf das konkrete Verlaufsmuster der Dynamik und die Zukunftsperspektiven der KI, zum Zweiten auf die damit zusammenhängenden Herausforderungen für Innovationsprozesse und Innovationspolitik sowie zum Dritten auf die Wirkungsmechanismen und die Rhetorik des Technologieversprechens selbst.

Diesen Fragen wird in der vorliegenden Studie mit dem Hauptfokus auf die deutsche KI-Entwicklung empirisch nachgegangen. Lediglich am Rande

und nur soweit für die vorliegende Analyse erforderlich, werden die »großen« und grundlegenden sowie heftig umstrittenen Fragen des KI-Diskurses angesprochen. Nämlich die weitreichende und bis heute kaum hinreichend zu beantwortende Frage, welche generellen gesellschaftlichen und politischen Folgen die KI nach sich zieht und wie diese zu bewerten sind, dann die Frage nach den bislang kaum geklärten vielfältigen rechtlichen und ethischen Herausforderungen »autonomer« KI-Systeme und schließlich die Frage, inwieweit der Begriff der KI selbst ein Mythos ist. Denn er vermittelt eine Vision, die letztlich daran scheitern muss, dass menschliche Intelligenz und menschliches Bewusstsein grundsätzlich nicht mit formalen Modellen und programmierten Algorithmen erfassbar sind.

Die vorliegende Studie entstand in den Jahren 2021 und 2022 als Expertise zum Thema »Dynamik der Künstlichen Intelligenz«. Sie stand im Kontext des vom BMBF geförderten Projektes »KI – Mensch – Gesellschaft« (KI.Me.Ge.), das bis Ende 2022 am Institut für Sozialwissenschaftliche Forschung e. V. (ISF München) durchgeführt wurde.¹ Besonders gedankt sei an dieser Stelle den Kollegen*innen des ISF München, namentlich Michael Heinlein, Norbert Huchler und Regina Wittal, für die Kooperation im Rahmen des Projektes KI.Me.Ge. Darüber hinaus sei Thorben Krokowski von der Sozialforschungsstelle der TU Dortmund für seine engagierte Mitarbeit an den durchgeführten Expertengesprächen und seine kenntnisreiche und kritische Begleitung der Expertise besonders gedankt. Ein herzlicher Dank geht auch an Edgar Illert, der das Manuskript und vor allem das Literaturverzeichnis professionell redigiert und schließlich auf einen lesbaren Stand gebracht hat. Weiterhin möchte ich Judith Wilke-Primavesi, der Programmleiterin des Campus Verlags, ausdrücklich für ihr Interesse an der Buchpublikation danken. Schließlich sei Eva Janetzko, wissenschaftliche Lektorin bei Campus, für ihre sehr hilfreiche Beratung und ihr Lektorat gedankt.

Berlin, im Herbst 2022

Hartmut Hirsch-Kreinsen

¹ Vgl. die Projekthomepage <https://www.kimege.de/> sowie die Zusammenfassung der Projektergebnisse, herausgegeben von Michael Heinlein und Norbert Huchler (2023).

Einleitung

Dass der Begriff der Digitalisierung nicht nur einen technologischen, sondern auch einen gesellschaftlichen Megatrend bezeichnet, ist weithin unstrittig. Erwartet werden völlig neue und ständig erweiterte Anwendungsmöglichkeiten digitaler Technologien und als Folge wird mit einem geradezu disruptiven gesellschaftlichen Wandel gerechnet. Unstrittig ist, dass in den letzten Jahren ein tiefgreifender Entwicklungsschub digitaler Technologien Platz gegriffen hat. Gründe hierfür sind die dramatische Steigerung der Leistungsfähigkeit der Computersysteme und die gleichzeitige massive Senkung ihrer Kosten. Darüber hinaus eröffnen das Internet und die weltweite datentechnische Vernetzung von Systemen und Orten sowie die rasant zunehmende global verfügbare Daten bis vor Kurzem kaum vorstellbare Anwendungsmöglichkeiten der digitalen Technologien.

Als ein treibendes Moment dieses Technologieschubs gilt Künstliche Intelligenz (KI) bzw. im internationalen Sprachgebrauch Artificial Intelligence (AI). Sie ist seit Jahren Gegenstand eines Booms, ja eines Hypes nicht nur in fachlichen Diskussionen, sondern auch in Politik und Gesellschaft insgesamt. KI kann ohne Übertreibung als »most promising technology of our era« angesehen werden (Brynjolfsson et al. 2017: 7) und aufgrund der ihr zugeschriebenen Fähigkeit, komplexe Probleme »autonom« und »smart« zu bearbeiten, erzeugt sie eine große Faszination in der Öffentlichkeit. Medial gefeierte Highlights wie der Sieg des IBM-Schachcomputers Deep Blue gegen den damaligen Schachweltmeister Garry Kasparov im Jahr 1997 oder der Sieg des KI-Systems Alpha Go im Jahr 2016 gegen den Weltchampion des Go-Spiels verstärken diese Sicht noch nachhaltig. Auch belegen vielfältige alltägliche Anwendungen immer wieder die besondere Leistungsfähigkeit von KI-Systemen und finden sich inzwischen in vielen alltäglich genutzten Systemen. Typisch ist hier das Smartphone, dessen

Funktionen ohne KI-Komponenten nicht denkbar sind. Andere bekannte Anwendungen sind Spamfilter, Suchmaschinen, maschinelle Übersetzungen, medizinische Diagnostik und Systeme der Spracherkennung. Freilich ist der Begriff KI auch ein Etikett, das Waren und Produkte interessant macht, das, um den Absatz zu steigern, inflationär auf viele Geräte, ob Autos, Onlineshops, Suchmaschinen oder Kühlschränke und Waschmaschinen, geklebt wird (Lenzen 2018: 21). Verschiedentlich wird deshalb auch sehr kritisch von »Fake AI« gesprochen (Kalthheuner 2021). Innovationspolitisch wird in Deutschland und der EU KI als »Schlüsseltechnologie« angesehen, die massiv gefördert werden muss, um im globalen Technologiewettbewerb insbesondere mit den USA und mit China nicht ins Hintertreffen zu geraten. Denn nur dann, so unisono die Ansicht vieler Expertinnen und Experten sowie Politiker*innen, könne das deutsche und das europäische Wirtschaftsmodell erfolgreich weiterentwickelt und der erreichte Wohlstand gesichert und ausgebaut werden.

Befördert wird die Faszination allerdings auch von einem bis heute weitverbreiteten unklaren Verständnis, was unter KI eigentlich zu verstehen ist. Dies ist nicht zuletzt Resultat des Umstandes, dass KI ein weites, unübersichtliches Feld unterschiedlichster Konzepte und Methoden umfasst und es bis heute offensichtlich keine allgemein akzeptierte Begriffsdefinition gibt. Daher werden teilweise auch für ein und dieselbe KI-Anwendung verschiedene, unscharfe Bezeichnungen verwendet wie zum Beispiel die oft undifferenziert synonym verwendeten Begriffe Maschinelles Lernen, Künstliche Neuronale Netze und Deep Learning. Auch dient der Terminus KI häufig lediglich als Schlagwort für die generelle Anzeige eines Trends aktueller hochtechnologischer Entwicklungen (Bauberger et al. 2021: 908). Und bis heute denken viele beim Stichwort KI – so etwas spöttisch eine im Rahmen der vorliegenden Studie interviewte KI-Expertin – an den Terminator aus dem bekannten Science-Fiction-Film mit Arnold Schwarzenegger aus den 1980er-Jahren.

Dabei ist KI keinesfalls ein neues und wirklich unbekanntes Technologiefeld. Vielmehr kann der Beginn ihrer Entwicklung grob auf die Mitte der 1950er-Jahre datiert werden, wo in den USA der Begriff Artificial Intelligence erstmals öffentlich gebraucht wird. Seitdem ist allerdings die KI-Entwicklung von einer Folge von Aufschwung- und Abschwungphasen gekennzeichnet, letztere als »AI-Winter« bezeichnet. So stellt schon vor Jahren ein kritischer Beobachter der KI-Entwicklung fest, dass diese nach einem immer wiederkehrenden Muster verlaufen: »Jedesmal wenn neue Ideen oder Kon-

zepte ins Spiel gebracht werden, tritt nach verblüffenden Anfangserfolgen die Entwicklung schon bald auf der Stelle, um sich dann zwischen unüberwindlichen Hindernissen zu verstricken.« (Brödner 1997: 189) Spätestens seit den 2010er-Jahren lässt sich nun ein kontinuierlicher Aufschwung, geradezu ein Boom der KI-Entwicklung identifizieren, einhergehend mit der Diffusion dieser Technologie in verschiedenste gesellschaftliche Anwendungsfelder sowie einem dazugehörigen Diskurs, der zeitweise den Charakter eines Hypes annimmt. Indes sind Prognosen über die zukünftige Entwicklung widersprüchlich. Sie reichen von der Erwartung eines weiterhin ungebrochenen Booms der KI bis hin zu einem erneuten Niedergang.

Erwartungen und KI-Dynamik

Mit der vorliegenden Studie wird an diese Debatten angeknüpft.¹ Eine allgemein akzeptierte Erklärung für die verschiedenen KI-Phasen ist, dass diese maßgeblich von der Entwicklung der Rechnertechnologie sowie den jeweils verfügbaren Konzepten und Methoden der KI bestimmt werden. Insbesondere wird die KI-Dynamik der letzten Jahre als Folge des schnell steigenden informationstechnologischen Fortschritts sowie der sich dadurch ständig erweiternden Nutzungsmöglichkeiten für die KI gesehen. Als zentrale Faktoren werden hierbei die massiv gestiegene Leistungsfähigkeit und Rechenkapazitäten von Computern, ihre dramatische Verbilligung, die Entwicklung nutzbarer komplexer KI-Methoden wie Neuronale Netze sowie schließlich Big-Data-Methoden und die Verfügung über große Datenmengen, besonders über das Internet, angeführt.

Die schnell fortschreitende informations- und datentechnologische Entwicklung ist ohne Frage eine notwendige Voraussetzung für diese Dynamik der KI, denn dadurch werden technologische Restriktionen für die

¹ Die KI-Entwicklung ist seit ihrem Anbeginn begleitet von einer breiten, heftigen und teilweise fundamentalen Kritik an ihren grundlegenden Annahmen und Konzepten. Im gegebenen Rahmen und der Fragestellung kann darauf allerdings nur selektiv eingegangen werden; vgl. zum Beispiel die Klassiker dieser Debatte Hubert Dreyfus (1972; 1992) und Joseph Weizenbaum (1978), die grundlegende Kritik an der KI von Peter Brödner (1997; 2019) sowie die aktuelle und grundsätzliche Auseinandersetzung von Erik J. Larsson (2021) mit dem gegenwärtigen KI-Boom und der von vielen als realistisch angesehenen Perspektive einer menschenähnlichen Maschinenintelligenz, der sog. Artificial General Intelligence.

Realisation von KI-Systemen kontinuierlich überwunden und die Grenzen ihrer Nutzung hinausgeschoben. Freilich handelt es sich dabei um lediglich technologische Potenziale, die aus soziologischer Sicht allein noch keine hinreichende Erklärung für die wechselnden Verlaufsmuster der KI-Entwicklung und den gegenwärtigen Boom sind. Die Frage ist damit nicht beantwortet, ob überhaupt und in welcher Weise die in bestimmten Entwicklungsphasen jeweils gegebenen technologischen Potenziale zu der tatsächlichen Nutzung, Diffusion und konkreten Anwendung dieser Technologie führen. Denn folgt man dem Common Wisdom der sozialwissenschaftlichen Technik- und Innovationsforschung, so besteht zwischen der Entwicklung neuer Technologien und ihren möglichen Anwendungspotenzialen einerseits, ihrer Verbreitung und Nutzung und schließlich ihren sozialen Konsequenzen etwa in Arbeitsprozessen andererseits kein deterministischer Zusammenhang. Vielmehr handelt es sich dabei um eine komplexe und wechselseitige Beziehung, die von einer Vielzahl nicht-technischer, ökonomischer, sozialer und politischer Faktoren geprägt wird. Deren Einfluss entscheidet darüber, in welcher Weise die Nutzungspotenziale der Technologien tatsächlich ausgeschöpft und realisiert werden und welche sozialen Konsequenzen letztlich damit auch verknüpft sind.²

Diese generellen techniksoziologischen und innovationstheoretischen Argumente können mit grundlegenden Überlegungen aus der *Sociology of Expectations* verknüpft werden. Danach ist wirtschaftliches Handeln im Kapitalismus von einer temporalen Struktur gekennzeichnet, die durch »Zukunftsimaginationen« überwunden und bewältigt wird. Konsequenz ist, dass Akteure, wollen sie Einkommen und Gewinne erzielen oder ihren Status verbessern, »Imaginationen einer wirtschaftlichen Zukunft« erzeugen und ihre Entscheidungen danach ausrichten, ob sie diese Zukunft verwirklichen oder vermeiden wollen (Beckert 2018: 13). Dies betrifft etwa Investitionsentscheidungen, die Vergabe von Krediten, aber eben auch die Dynamik und unterschiedlichen Verlaufsmuster technologischer Innovationen.

Wie die Innovationsforschung überzeugend zeigt, kommt »Erwartungen« über die zukünftigen Chancen neuer Technologien ein überaus hoher Einfluss auf Verlauf und Richtung von technologischen Innovatio-

2 Vgl. hierzu grundlegend zum Beispiel aus der Sicht der Innovationsforschung Rogers (2003) und Fagerberg (2005) sowie aus der Sicht der soziologischen Technikforschung Bijker et al. (1987) oder Pfeiffer (2018).

nen zu. Denn Innovationen zeichnen sich in der Regel durch Risiken und Ungewissheiten in Hinblick auf ihren Verlauf, ihre Realisierbarkeit sowie die angestrebten, zumeist ökonomischen Effekte aus. Dies trifft sowohl für den Beginn einer neuen, möglicherweise radikalen Entwicklung, aber auch für eine Weiterentwicklung schon existierender Technologien und die Entscheidungen über konkrete Nutzungsformen zu (van Lente/Rip 1998; Bender 2005; Borup et al. 2006; Konrad 2006; Rip 2018). Denn Erwartungen und damit verknüpfte Versprechungen, Visionen und Szenarien über mögliche und wünschenswerte Entwicklungsperspektiven reduzieren die Ungewissheiten und die Komplexität von Innovationen, eröffnen konkrete Handlungsperspektiven für die Entwickler, interessieren weitere beteiligte Akteure und koordinieren ihr Handeln. Oder in den Worten von Harro van Lente und Arie Rip: »Erwartungen und Geschichten über die Zukunft im Allgemeinen reduzieren die grundlegende Kontingenz in einem nichtdeterministischen Sinn, indem sie Entwürfe zur Verfügung stellen, die zum Handeln verwendet werden können.« (1998a: 217)

Technologieversprechen

Auf diesen Überlegungen basiert die These der folgenden Analyse: Mit der KI ist ein Technologieversprechen verknüpft, das die KI-Dynamik in ihren verschiedenen Phasen maßgeblich bestimmt und vorantreibt. Dieses Technologieversprechen ist als Narrativ zu verstehen, das Entwicklungspotenziale der KI, denkbare Möglichkeiten ihrer Anwendung und Nutzung sowie die damit zukünftig realisierbaren Gewinne in Aussicht stellt. Dabei kann es sich um die Lösung vorhandener technologischer Probleme und Engpässe, um die Steigerung ökonomischer Effizienz und Effektivität, wie aber auch um einen als wünschenswert angesehenen sozialen und gesellschaftlichen Wandel handeln. Solche von einflussreichen und an der Technologie und ihrer Entwicklung interessierten Akteuren vorgebrachten Technologieversprechen haben in ihren verschiedenen Phasen die KI-Entwicklung stets erneut und maßgeblich angestoßen. Dabei muss allerdings das Technologieversprechen für interessierte Adressaten in mehr oder weniger einsichtiger Weise mit dem gegebenen Stand der Forschung und einer erwartbaren Technologieentwicklung begründet sein. Damit sollen bei potenziell interessierten Akteuren Erwartungen in Hinblick auf den

technischen, ökonomischen oder auch sozialen Nutzen der Innovation geweckt werden und sie für die Beteiligung an entsprechenden Innovationsmaßnahmen, etwa durch Investitionen und Förderung, Forschungs- und Entwicklungsaktivitäten oder auch Anwendungen, gewonnen werden.

Dass Versprechungen, Erwartungen und Enttäuschungen die KI-Entwicklung prägen, zeigen immer wieder zumeist kritische Studien über die Entwicklungsgeschichte der KI (Dreyfus 1965; Cyranek/Coy 1994; Breiter 1995; Brödner 1997; Buchanan 2006; Larson 2021). Einerseits handelt es sich um Versprechungen, die sich angesichts nur beschränkter Leistungsfähigkeit und hoher Anwendungsprobleme der jeweils verfügbaren KI-Systeme und Methoden als überzogen erweisen. Resultat sind oftmals »Erwartungs-enttäuschungen« wichtiger beteiligter Akteure (Ahrweiler 1995: 22) und ihr Rückzug aus dem Technologiefeld. Andererseits aber bietet die schnelle und ganz offensichtlich dynamische Entwicklung der KI stets wieder Anlass für neuerliche Versprechungen, die mit großer Überzeugungskraft die Erwartungen und das Interesse vieler Akteure und der interessierten Öffentlichkeit an dieser Technologie wecken. Diese Versprechungen schöpfen ihre Kraft aus »transhistorischen« und »transkulturellen« fantasievollen Geschichten über die intelligente Maschine, die die Menschheitsgeschichte durchziehen und lange vor dem Aufkommen moderner Wissenschaften existierten (Cave et al. 2020).

Auf den Punkt bringt diese von Versprechen getriebene KI-Dynamik ein früher Protagonist der amerikanischen KI und Gründungsmitglied der American Association for Artificial Intelligence Bruce G. Buchanan: »The history of AI is a history of fantasies, possibilities, demonstrations, and promises.« (Buchanan 2006: 53) Ähnlich lautet die Feststellung eines deutschen KI-Wissenschaftlers aus dem Jahre 1994, der nicht erreichte Ziele und ein ständiges Hinausschieben und Erneuern der Perspektiven der KI-Entwicklung betont: »Das ist ein altes Spiel in der AI. Man sagt: In 10 Jahren werden wir das und das haben, und in 10 Jahren stellt man dann die gleiche Frage und sagt: In 20 Jahren werden wir das erreicht haben. In 20 Jahren stellt man wieder die gleiche Frage, und auch in 50 Jahren werden wir die gleiche Frage stellen.« (Cyranek/Coy 1994: 259)

Fragestellung

Auf diese Überlegungen bezieht sich die vorliegende Studie. Aus technik- und innovationssoziologischer Sicht werden die sozialen Bestimmungsfaktoren wie auch die gesellschaftlichen Folgen dieser widersprüchlichen Dynamik mit insgesamt drei Fragen in den Blick genommen.

Zum einen wird untersucht, inwieweit die KI-Dynamik von Technologieversprechen und daran orientierten kollektiven Erwartungen über ihre zukünftigen Potenziale getrieben wird. Der Fokus richtet sich dabei sowohl auf die vergangene Entwicklung als auch mögliche Zukunftsperspektiven.

Zum Zweiten wird gefragt, welchen neuen Herausforderungen mit dem Technologieversprechen KI für Innovationsprozesse und die institutionellen und politischen Rahmenbedingungen des Innovationssystems einhergehen. Denn die KI-Entwicklung konstituiert einen spezifischen Innovationsmodus, der mit den bisherigen Innovationspraktiken und -routinen und vor allem mit den tradierten Grundsätzen von Innovationspolitik nur mehr bedingt kompatibel ist.

Zum Dritten wird der Frage nachgegangen, auf welchen inhaltlichen und rhetorischen Voraussetzungen das Technologieversprechen KI und sein Einfluss auf den Verlauf der KI-Dynamik beruht. Deutlich wird, dass das Technologieversprechen KI letztlich seine Überzeugungskraft daraus zieht, dass es an den uralten Mythos der intelligenten Maschine anknüpft, die der menschlichen Intelligenz ebenbürtig, wenn nicht gar überlegen ist.

Mit dem Begriff der Dynamik werden dabei mehrere miteinander verschränkte Dimensionen der Genese einer neuen Technologie bezeichnet: die Entwicklung ihrer wissenschaftlichen und technologischen Grundlagen, die beteiligten Akteure und die von ihnen vorgetragenen Technologieversprechen, die innovationspolitischen Fördermaßnahmen, der Technologiediskurs und die Erwartungen über die Nutzungspotenziale und schließlich die Institutionalisierung eines soziotechnischen Feldes der KI.

Der Begriff Technologie wird dabei im Folgenden weit gefasst. Bezeichnet werden damit Wissensbestände und Rationalitätskriterien, die bestimmte Methoden und Verfahren mit Plausibilitätskriterien versehen, die die Angemessenheit der gewählten Mittel im Hinblick auf die gewünschten Zwecke feststellen und schließlich daraus sich ergebende Nutzungsmöglichkeiten und Systemausprägungen bestimmen (Lösch 2012). Eingeschlossen sind in dieser Definition konkrete Techniken und technische Systeme, die Realphänomene bezeichnen (Mayntz 2001). KI kann

daher als Technologie bzw. auch als technologisches Feld sehr verschiedener Methoden und Konzepte verstanden werden, mit denen kognitive Leistungen computertechisch reproduziert werden sollen (Ahrweiler 1995a: 43). Sie wird als Sammelbegriff für die Entwicklung, Konzeptualisierung und Anwendungen von »Algorithmen« bzw. »algorithmischen Systemen« verwendet, die in konkrete technische Artefakte und Programme sowie technische Prozeduren umgesetzt werden (Bauberger et al. 2021: 908).

Methodische Basis

Empirische Basis der Studie ist die Analyse der Entwicklungsdynamik der KI in der Bundesrepublik. Dabei werden allerdings internationale Bezüge keineswegs ausgeklammert, denn die KI-Entwicklung verläuft in hohem Maße in internationalen Kontexten.³

Methodisch hat die Studie eine breite qualitative Basis: Zum einen basiert sie auf den Ergebnissen der Analyse der laufenden wissenschaftlichen, politischen und öffentlichen Diskurse über KI, insbesondere auf der Durchsicht und Auswertung einer großen Zahl von »grauen« Dokumenten, Preprints, politischen Verlautbarungen, Websites, Studien und einschlägiger Fachpublikationen aus nationalen und internationalen Kontexten. Als besonders informativ erwiesen sich oftmals Interviews und Statements von KI-Expertinnen und Experten, die im Internet zugänglich waren.

Zum Zweiten beruht die Argumentation auf den Ergebnissen eigener ausführlicher Expertengespräche. Zwischen Oktober 2021 und März 2022 wurden im Rahmen des Projektes 19 Interviews mit 16 KI-Expertinnen und Experten aus Deutschland durchgeführt. Befragt wurden zehn grundlagen- und anwendungsorientierte Wissenschaftler*innen sowie sechs KI-Expertinnen und Experten aus KI-nahen Wissenschaftsdisziplinen und Anwendungsfeldern. Hervorzuheben ist, dass sich darunter vier Emeriti

³ Zur Historie der KI in der Bundesrepublik liegen Studien unterschiedlicher Provenienz vor. Hervorzuheben ist vor allem die wissenssoziologische Studie von Petra Ahrweiler, die Mitte der 1990er-Jahre vorgelegt wurde (Ahrweiler 1995/1995a). Zu nennen sind weiterhin die technikhistorische Studie von Seising & Dittmann (2018) sowie Rückblicke früher beteiligter Informatiker, zum Beispiel Konrad (1998), Siekmann (2009) und Bibel (2020). Die KI-Entwicklung in der DDR sei im Folgenden ausgeklammert (dazu Ahrweiler 1995/1995a).

aus der ersten Generation der KI-ler der Bundesrepublik befanden.⁴ Die Interviews wurden online durchgeführt und dauerten mindestens eine Stunde, oft waren sie deutlich länger. Die Expertengespräche umfassten teilweise mehrere Feedbackrunden, zum Beispiel per E-Mail zur Klärung noch offener Fragen oder zur Diskussion erster Thesen. Mit drei Experten konnten in einer zweiten ausführlichen Interviewrunde einzelne Fragestellungen vertieft werden. Die Interviews wurde aufgezeichnet und die zentralen Thesen der Gespräche transkribiert.

Zum Dritten bedient sich die Analyse einer Reinterpretation vorliegender eigener Forschungsergebnisse über den gesellschaftlichen Digitalisierungsprozess der letzten Jahre. Schließlich flossen in die Argumentation Ergebnisse und Erkenntnisse aus der laufenden Beteiligung des Autors an einschlägigen Workshops, Tagungen, wie aber auch informellen Gesprächen im Feld der KI ein.

Zum Gang der Argumentation

Die Fragestellungen und die leitende These, wonach die KI-Dynamik und ihre verschiedenen Entwicklungsphasen von Versprechungen getrieben werden, soll mit den folgenden Argumentationsschritten entfaltet werden:

Im *Teil I* werden die analytischen Grundlagen der Untersuchung ausgeführt; begründet wird das Konzept des Technologieversprechens und es wird der Begriff der KI genauer geklärt. Im *Teil II* wird ausgehend vom Konzept des Technologieversprechens die Dynamik der KI in der Bundesrepublik in Hinblick auf ihre verschiedenen Phasen und ihre wissenschaftlich-technologischen, akteursbezogenen, politischen und institutionellen Dimensionen nachgezeichnet. Es handelt sich dabei um den empirischen Hauptteil der Untersuchung. Daran wird im *Teil III* angeknüpft, in dem Thesen zur Perspektive und den zukünftig denkbaren Entwicklungspfaden der KI formuliert werden. Dabei wird vor allem auch auf das vielfältige Bündel von Grenzen und Herausforderungen einer schnellen weiteren KI-Entwicklung und Diffusion eingegangen und die Relevanz und Chancen

⁴ Sofern sie zitiert werden, sind die KI-Wissenschaftler*innen als W1 bis W11 und die KI-nahen Expertinnen und Experten als E1 bis E5 gekennzeichnet.

einer europäischen bzw. deutschen KI-Entwicklung erörtert. Im *Teil IV* wird nach den gesellschaftlichen Voraussetzungen und Folgen der Dynamik der KI gefragt. Konkret wird dabei der wechselseitige Zusammenhang zwischen den institutionellen Bedingungen des Nationalen Innovationssystems und der KI-Dynamik in den Fokus genommen. Davon ausgehend werden zentrale Anforderungen an eine KI-orientierte Innovationspolitik formuliert. Mit dem *Teil V* wird die Kategorie des Technologieversprechens aufgegriffen und es wird diskutiert, an welche besondere inhaltliche und rhetorische Architektur seine Wirksamkeit gebunden ist. Im *Resümee* wird abschließend das Argument ausgeführt, dass die anhaltende Überzeugungskraft des Technologieversprechens KI letztlich auf dem uralten Mythos der intelligenten Maschine beruht, die der menschlichen Intelligenz ebenbürtig, wenn nicht gar überlegen ist.

Teil I: Konzept und Begriffe

1. Promising Technology

1.1 Leitmotiv und Koordinationsfunktion

Die These vom Technologieversprechen KI kann präzisiert werden mit dem Konzept der Promising Technology (van Lente/Rip 1998; Bender 2005). Das Kernargument lautet, dass Akteure bei ihren Entscheidungen, an einer Technologieentwicklung teilzunehmen, sich zunächst noch an einem noch sehr allgemeinen Versprechen über diese erst noch zu entwickelnde Technologie orientieren. Das Versprechen offeriert im Prinzip interessierten Akteuren eine Entwicklungsperspektive, stellt zukünftige Nutzenpotenziale in Aussicht, eröffnet einen einzuschlagenden Weg für die Innovation und begründet auch die Möglichkeit für die Durchsetzung eigener Interessen. Es bietet die Voraussetzung dafür, weitere Akteure anzusprechen, sie in den Innovationsprozess einzubinden, ihr Handeln zielgerichtet zu koordinieren sowie Wissens- und Innovationsressourcen zu mobilisieren und Investitionen in Forschung und Entwicklung zu initiieren. Anders formuliert, das Technologieversprechen hat die Funktion eines Leitmotivs für heterogene Akteure mit ihren im Grunde sehr unterschiedlichen Interessen, die in einen Innovationsprozess eingebunden werden. In diesem Sinn lässt sich festhalten, dass Versprechen »[...] do not strive for truth or accuracy, but are meant to influence specific social processes in technological developments« (Geels/Smit 2000: 880).

Freilich ist dies sehr voraussetzungsvoll. Wie noch genauer zu begründen ist, ist hierfür eine rhetorische Architektur des Technologieversprechens erforderlich, die durch Generalisierung und Pauschalisierung ein hohes Maß Anschlussfähigkeit ermöglicht. Mehr noch, mit dem Verweis auf technologisch lösbare gesellschaftliche Herausforderungen wird zugleich an eine gesellschaftspolitische Debatte über wünschenswerte gesellschaftliche

Weiterentwicklung angeknüpft und es wird damit Interesse von Akteuren jenseits der zunächst angesprochenen Fachöffentlichkeit geweckt (Hirsch-Kreinsen 2016: 12 f.). Soziologisch gesprochen, fungiert ein Technologieversprechen als *Boundary Object* mit einem sehr allgemeinen Kern von Ideen, der mit verschiedensten Kontexten verknüpft und von unterschiedlichsten Akteuren aufgegriffen werden kann. Das heißt, die Formbarkeit eines solchen Grenzobjekts erlaubt eine Vielzahl von lokalen Interpretationen und Praktiken, die sich locker auf einen Gemeinsamkeiten stiftenden Kern beziehen lassen (Faust 2021: 72). Dabei kann ein Technologieversprechen entweder den Charakter einer explizit artikulierten und rhetorisch vorgebrachten Meinung, eines Vortrages oder eines Diskussionsbeitrages haben oder als Dokument, Objekt oder generell als materialisierte Darstellung präsentiert werden (Brown et al. 2003: 6).

Innovationsdynamik

Ein Technologieversprechen bezeichnet mithin einen Startpunkt, von dem aus ein Innovationsprozess angestoßen, koordiniert und vorangetrieben wird. Es entsteht eine Entwicklungsdynamik, die mit van Lente und Rip (1998; 1998a) analytisch mit mehreren Schritten grob präzisiert werden kann:

Im *ersten Schritt* wird ein Versprechen formuliert, das Zukunftsvisionen postuliert und damit verschiedenste Akteure interessiert und eine kollektive Orientierung herstellt. Entworfen wird ein solches Technologieversprechen in der Regel von einer kleinen Gruppe in der Sache engagierter und von den formulierten Visionen überzeugter Protagonisten etwa aus dem Wissenschaftsbereich, der Politik oder auch aus Unternehmen. Dabei weicht diese Gruppe oftmals vom Stand der Forschung und bisherigen technologischen Entwicklungspfaden ab und entwirft eine neue technologische Vision. Folgt man der Innovationsforschung, so ist die Voraussetzung hierfür ein »protected space« bzw. eine *Nische*, in der diese Gruppe institutionell geschützt und relativ autonom ihre Ideen entwickeln kann. Eine solche Nische kann geschaffen werden durch strategische Investitionen von Unternehmen, durch eine gezielte politische Förderung, oder aber auch durch die gezielte Nutzung vorhandener akademischer Freiräume. Nischen werden auch als »incubation rooms« für radikale Innovationen angesehen, da sie insbesondere vor dem Druck existierender Technologiepfade und der damit ver-

wobenen Interessen schützen (Geels 2004; Markard/Truffer 2008). Zudem müssen diese Akteure über hinreichende kommunikative und interaktive Ressourcen verfügen, um im späteren Verlauf ihrer Anstrengungen das Technologieversprechen zunächst einer Fachöffentlichkeit präsentieren zu können und das Interesse ihrer Mitglieder an dem Versprechen zu wecken. Angestoßen wird auf diese Weise ein Diskurs, in dem denkbare Optionen und Perspektiven einer neuen Technologie zwischen einer zunehmend größeren Zahl von interessierten Akteuren ausgetauscht werden.¹ Dieser Austausch- und Kommunikationsprozess ist die Voraussetzung dafür, dass das ursprünglich nur von einer kleinen Gruppe Interessierter vorgetragene Technologieversprechen verallgemeinert wird und die daran geknüpften Erwartungen kollektiven Charakter gewinnen (Konrad 2006: 431).

Der *zweite Schritt* wird als *Agenda Setting* gefasst. In dessen Verlauf wird die Vision der neuen Technologie fortgeschrieben und präzisiert, Anforderungen an die Entwicklung und die erforderlichen Innovationsschritte definiert, die Rollen verschiedener zu beteiligender Akteure beschrieben und damit insgesamt der Innovationsprozess strukturiert. Es findet eine Transformation der schriftlich niedergelegten, in Vorträgen präsentierten oder sonst wie fixierten Perspektiven, Visionen und Ziele in Handlungsbedarf und in kollektiv orientierte Handlungsanweisungen statt. Generiert wird damit eine Agenda, die wie ein Drehbuch das weitere Handeln der beteiligten Akteure in einer Weise strukturiert, die der beabsichtigten Innovation förderlich ist (Bender 2005: 174). Die Akteure mit ursprünglich divergierenden Interessen entwickeln ein gemeinsames Interesse an der Innovation, arbeiten zunehmend in einer *koordinierten Weise* zusammen und sie versuchen, Lösungen zu erarbeiten, die die in der Agenda spezifizierten Anforderungen erfüllen. Es entsteht so eine vorher so nicht existierende *Community of Practice*, das heißt eine vernetzte Akteurskonstellation, die, geteilten Entwicklungsperspektiven folgend, einen Diskurs über die zu verfolgenden Ziele führt und damit eine gemeinsame Wissensbasis über Potenziale und Entwicklungserfordernisse der neuen Technologie generiert (Wenger 1998). Dabei wird die Agenda fortlaufend spezifiziert, leitet

¹ Diskurs wird im Folgenden verstanden als ein kommunikativer Raum oder auch Diskursraum mit einer gemeinsamen Kultur mit Wir-Bewusstsein und einem gemeinsamen Orientierungszentrum für die Pflege einschlägiger Aktivitäten, ihrer Beobachtung, ihrer Konkurrenz und ihrer Stilisierung (Zifonun 2012: 237).

weitere Innovationsschritte an und formt damit einen sich abzeichnenden technologischen Entwicklungspfad.

Ein *dritter Schritt* stellt die zunehmende Verfestigung und Strukturierung dieses Interaktionszusammenhangs dar und er gewinnt den Charakter eines zunehmend *stabilen Handlungskontextes* mit spezifischer Logik und normativer Verbindlichkeit. Insofern kann dieser Prozess auch als *Institutionalisierung* von Handlungsverbindlichkeiten begriffen werden. Es entstehen soziale Strukturen, die durch vernetzte Akteurskonstellationen, arbeitsteilige Routinen und gefestigte Wahrnehmungsmuster bestimmten Handlungsweisen Optionen für weitere konkrete Innovationsverläufe eröffnen und andere mehr oder weniger ausschließen. Voraussetzung hierfür ist, dass sich die am Innovationsprozess beteiligten Akteure bei ihren Entscheidungen, Handlungen und Interaktionen an diesen Bedingungen orientieren – wodurch sie durchaus auch zu ihrer Reproduktion und Reputation beitragen.

Ergebnis ist die Emergenz eines neuen *soziotechnischen Feldes*. Mit diesem werden in institutionentheoretischer Perspektive jene Organisationen bezeichnet, »die gemeinsam einen abgegrenzten Bereich des institutionellen Lebens konstituieren: die wichtigsten Zulieferfirmen, Konsumenten von Ressourcen und Produkten, Regulierungsbehörden sowie andere Organisationen, die ähnliche Produkte oder Dienstleistungen herstellen bzw. anbieten« (DiMaggio/Powell 2000: 149). Mit dem Feldkonzept wird die Komplementarität im Prinzip unterschiedlicher Funktionen und Interessenlagen betont.² Sofern sich ein solches Feld durch Wettbewerb, Staat oder Professionen konstituiert, ist von einer stabilen Interaktion zwischen den beteiligten Akteuren auf dem Wege von Koalitions- und Machtprozessen, Informationsaustausch oder gegenseitiger Beobachtung auszugehen. Konstitutiv für ein soziotechnisches Feld ist der Fokus der beteiligten Akteure auf eine bestimmte Technologie und ihre Entwicklungsperspektiven.

Expectation-Requirement-Cycle

Diese von einem Technologieversprechen getriebene Innovationsdynamik kann zusammenfassend als Expectation-Requirement-Cycle mit verschiedenen Stufen gefasst werden (Geels/Smit 2000; Brown et al. 2003): Ein mit

² Vgl. hierzu die Diskussion des Begriffs der institutionellen Komplementarität bei Mayntz (2006).

dem Stand der Forschung und Technologieentwicklung mehr oder weniger konkret begründetes Technologieversprechen stößt einen Diskurs über Entwicklungsziele an, diese begründen Erwartungen und werden dann in Form einer Agenda in Innovationsanforderungen übersetzt. Diese generieren Handlungsverbindlichkeiten und koordinieren das Innovationshandeln heterogener Akteure. Im Erfolgsfall wird ein neues soziotechnisches Feld institutionalisiert und kontinuierlich erweitert.³ Dieser Prozess lässt sich grafisch wie folgt darstellen (Abbildung 1):

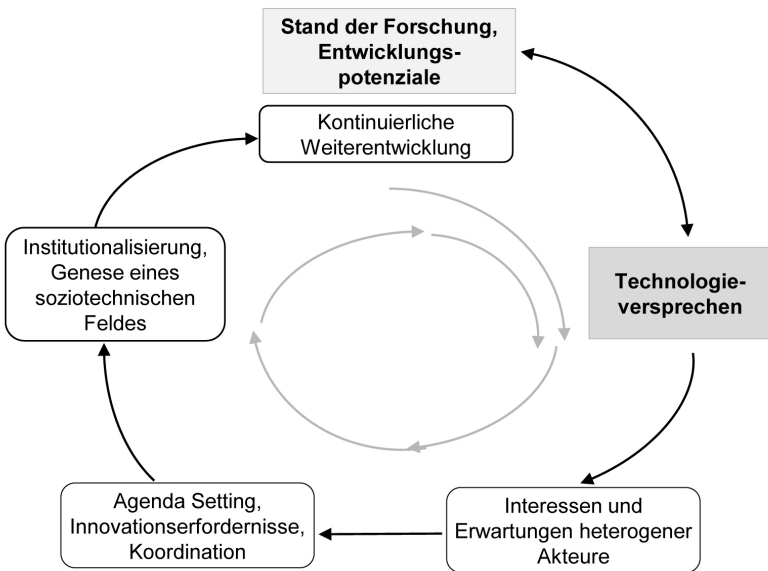


Abb. 1: Expectation-Requirement-Cycle
 Quelle: Eigene Darstellung in Anlehnung an Geels/Smit 2000

Empirisch kann die Dynamik des Kreislaufprozesses allerdings in verschiedener Weise verlaufen. Divergenzen können beispielsweise in Hinblick

³ Vgl. hierzu auch Christoph Deuschmann (1997), der diesen dynamischen Prozess technologischer und ökonomischer Entwicklung instruktiv als »Mythenspirale« gefasst hat. Zum Begriff des Mythos im Kontext der KI-Entwicklung vgl. das Resümee.