Der Elektronische Rechtsverkehr 44

Cornelius Kleiner

Die elektronische Person

Entwurf eines Zurechnungs- und Haftungssubjekts für den Einsatz autonomer Systeme im Rechtsverkehr

Nomos

Der Elektronische Rechtsverkehr

Herausgegeben von
Prof. Dr. Alexander Roßnagel und
Prof. Dr. Gerrit Hornung, LL.M.
in Zusammenarbeit mit
dem TeleTrusT Deutschland e.V.

Band 44

Cornelius Kleiner

Die elektronische Person

Entwurf eines Zurechnungs- und Haftungssubjekts für den Einsatz autonomer Systeme im Rechtsverkehr

Nomos

Onlineversion
Nomos eLibrary

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Zugl.: Freiburg im Br., Albert-Ludwigs-Univ., Diss., 2020

ISBN 978-3-8487-7787-7 (Print)
ISBN 978-3-7489-2192-9 (ePDF)

1. Auflage 2021
© Nomos Verlagsgesellschaft, Baden-Baden 2021. Gesamtverantwortung für Druck und Herstellung bei der Nomos Verlagsgesellschaft mbH & Co. KG. Alle Rechte, auch die des Nachdrucks von Auszügen, der fotomechanischen Wiedergabe und der Übersetzung, vorbehalten. Gedruckt auf alterungsbeständigem Papier.

Meinen Eltern

Vorwort

Die vorliegende Arbeit wurde im Sommersemester 2020 von der rechtswissenschaftlichen Fakultät der Albert-Ludwigs-Universität Freiburg im Breisgau als Dissertation angenommen. Sie entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Zentrum für Angewandte Rechtswissenschaft/ Institut für Informations- und Wirtschaftsrecht (ZAR/IIWR) am Karlsruher Institut für Technologie (KIT).

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Thomas Dreier, M.C.J. für die Unterstützung meines Promotionsvorhabens und seine hervorragende Betreuung. Herrn Prof. Dr. Boris Paal, M.Jur. (Oxford) danke ich für die zügige Erstellung des Zweitgutachtens.

Den Herausgebern und dem Verlag Nomos danke ich für die Aufnahme dieser Arbeit in die Schriftenreihe „Der Elektronische Rechtsverkehr".

Dir, liebe Marlene, danke ich für den Beistand und den Rückhalt, den du mir gibst – während der gesamten Promotionszeit und darüber hinaus.

Euch, meinen Eltern, danke ich für Eure bedingungslose Unterstützung, die Ermöglichung meiner gesamten Ausbildung, die Durchsicht dieser Arbeit sowie Eure hilfreichen Anmerkungen dazu – Euch ist diese Arbeit gewidmet.

Karlsruhe, im September 2020 *Cornelius Kleiner*

Inhaltsverzeichnis

Teil 1: Einleitung 19

Teil 2: Autonome Systeme 21

A. Begriff des autonomen Systems 21

B. Historischer Überblick 22

C. Agententypen und Grundlagen 23

I. Agententypen 23

II. Symbolverarbeitender und konnektionistischer Ansatz 25

III. Programmierung in Programmiersprachen 25

IV. Expertensysteme 26

V. Neuronale Netze 26

VI. Kombinatiossysteme/ hybride Architektur 29

VII. Lern- und Anpassungsfähigkeit künstlicher Intelligenz 30

1. Training neuronaler Netze 30

2. Genetische/ evolutionäre Algorithmen 30

VIII. Multiagentensysteme 31

Teil 3: Behandlung *de lege lata* 33

A. Agentenerklärungen 33

I. Willenserklärung 34

II. Übermittlung unter Verwendung elektronischer Systeme 36

III. Erklärungen deterministischer Systeme 37

1. Willenserklärung des deterministischen Systems selbst 37

2. Willenserklärung des Verwenders 38

IV. Erklärungen autonomer Systeme 39

1. Willenserklärung des Verwenders 39

2. Botenschaft 41

3. Eigene Willenserklärung des autonomen Systems 42
a) Objektiver Tatbestand 43
b) Subjektiver Tatbestand 44
aa) Selbstbewusstsein als Voraussetzung 44
bb) Funktionale Betrachtung 45
cc) Erklärungsbewusstsein und Geschäftswille 46
dd) Zwischenergebnis 47
c) Willenserklärung und Rechtsfolge 47
aa) Dualistische Ansichten 48
bb) Monistische Ansichten 49
cc) Zwischenergebnis 50
d) Zwischenergebnis 50
4. Zurechnung der Erklärung zum Verwender 51
a) Stellvertretung 51
b) Stellvertretung analog 51
c) Blankett 52
aa) Abgabe oder Zurechnung/ Botenrecht/ Ermächtigung 53
bb) Stellvertretung 54
cc) Stellvertretung analog bzgl. der offenen Blanketterklärung 55
dd) Stellvertretung analog bzgl. der verdeckten Blanketterklärung 56
(1) Schädlichkeit der Durchbrechung des Offenkundigkeitsprinzips 56
(2) Irrelevanz der Identität des Erklärungsurhebers 57
(3) Botenrecht analog bzgl. der verdeckten Blanketterklärung 60
ee) Zwischenergebnis 60
ff) Anwendbarkeit auf autonome Systeme 61
(1) Arbeitsteilige Erklärung 62
(2) Geschäfts- und Rechtsfähigkeit als Voraussetzung der Ausfüllungsbefugnis 63
(3) Die Haftung analog § 179 BGB 64
(4) Zwischenfazit 65

(5) Quantitative Grenzen/ qualifizierte elektronische Signatur iSd. § 126a BGB 66
gg) Fazit zur Blanketterklärung 67
d) Zurechnung aufgrund Vertrags-/ Bürgschaftslösung 68
e) Duldungsvollmacht 68
f) Zurechnung aufgrund Rechtsscheins 69
aa) Rechtsschein bzgl. des Vorliegens von Vertretungsmacht 71
bb) Rechtsschein bzgl. der Urheberschaft der Erklärung 73
cc) Anschein eines deterministischen Systems 74
dd) Zurechenbarkeit/Zurechnungsprinzip 74
(1) Willensprinzip 75
(2) Verschuldensprinzip 76
(3) Veranlassungsprinzip 78
(4) Risikoprinzip 80
ee) Zwischenergebnis 81
g) Elektronische Willenserklärung nach *Wiebe* 82
5. Zwischenergebnis 85

B. Haftung 86
I. Haftung bei einfachen elektronischen und deterministischen Systemen 86
1. Haftung des Systems selbst 87
2. Haftung des Verwenders 87
3. Haftung des Herstellers 88
II. Haftung bei autonomen Systemen 89
1. Haftung des Verwenders 89
a) Vertragliche Haftung 90
aa) Schuldverhältnis/ Pflichtverletzung 90
(1) Vertragliche Pflicht 91
(2) Culpa in contrahendo 92
(3) Culpa post contractum finitum 92
(4) Öffentlich-rechtliche Schuldverhältnisse 93
bb) Vertretenmüssen 93
(1) § 276 BGB 93
(2) § 278 BGB 95

(3) § 278 BGB analog 95
(a) Planwidrige Regelungslücke 95
(b) Vergleichbare Interessenlage 96
(c) Verschuldensmaßstab 98
(d) Verschuldenserfordernis des Gehilfen 98
(e) Verschuldensfähigkeitserfordernis des Gehilfen 99
(f) Menschliches Verhalten als Zurechnungsvoraussetzung 100
(g) Zwischenergebnis 101
(4) Vertragliche Modifikation des Vertretenmüssens 102
(5) Risikoprinzip 102
(6) § 242 BGB 103
b) Vertragliche Vereinbarung 103
c) Störerhaftung 103
d) Deliktische Haftung 105
aa) § 823 BGB 105
bb) § 831 BGB analog 106
(1) Planwidrige Regelungslücke 106
(2) Interessenlage 106
(3) Widerrechtlichkeit des Gehilfenverhaltens 108
(4) Zwischenergebnis 110
e) § 833 BGB analog 110
aa) Satz 1 analog 110
(1) Analogiefähigkeit von Gefährdungshaftungstatbeständen 110
(2) Fließender Übergang von Verschuldens- zu Gefährdungshaftung 113
(3) Analogiefähigkeit der Gefährdungshaftung nach Bauer 114
(4) Zwischenergebnis 115
bb) Satz 2 analog 115
f) Steuerung eines Fahrzeugs durch autonome Systeme 117
aa) Halterhaftung § 7 StVG 118
(1) Die Haltereigenschaft 118

(2) Die Entscheidung des autonomen Systems als höhere Gewalt 119
(3) Teleologische Reduktion des § 7 StVG 120
bb) Insassenhaftung § 18 StVG/ § 823 I BGB 120
cc) Herstellerhaftung 121
dd) Zwischenergebnis 122
g) Aufopferungshaftung 123
aa) § 904 BGB analog 123
bb) Allgemeiner Aufopferungsanspruch 124
h) Zwischenergebnis 125
2. Haftung des autonomen Systems selbst 125
3. Haftung des Herstellers/ Inverkehrbringers 126
a) Produkthaftung 126
b) Produzentenhaftung 129
4. Haftung des Verkäufers/ Vermieters/ Bereitstellers 129
a) Kaufvertrag 130
b) Sonstiges Vertragverhältnis 131
C. Anweisungslösung (de lege lata) 131
D. GmbH-Lösung (de lege lata) 132

Teil 4: Lösungen *de lege ferenda* 133

A. Verbotslösung 133
B. Gesetzliche Gefährdungshaftung/ Zurechnung der Erklärung zu natürlicher Person 133
I. Zur Haftung 134
II. Zur Zurechnung von Erklärungen 136
C. Elektronische Person 137
I. Überblick über die elektronische Person 137
II. Entschließung des europäischen Parlaments 140
III. Überblick über die elektronische Person nach dieser Arbeit 141

Teil 5: Die elektronische Person 145

A. Rechtsfähigkeit über die natürliche Person hinaus 145
I. Fiktionstheorien 146
II. Die Theorie des subjektlosen Zweckvermögens 147

III. Die Theorie des natürlichen Rechtsträgers (reale Verbandseinheit/ Zwecktheorien) 148
IV. Theorie der realen Verbandspersönlichkeit 149
V. Einordung der elektronischen Person 150
B. Moralischer oder funktionaler Ansatz 152
C. Anknüpfungspunkt für die elektronische Person 153
D. Entstehung der elektronischen Person 155
I. Technische Voraussetzungen 155
II. Anmeldung und Eintragung 155
1. Bedürfnis einer Kennzeichnung 156
2. Konstitutive oder deklaratorische Eintragung 157
3. Voraussetzungen der Anmeldung 158
E. Rechtsstellung der elektronischen Person 159
I. Die elektronische Person als Teilnehmerin am Rechtsverkehr 159
II. Elektronische Personen als Gegenstand des Rechtsverkehrs 159
III. Die Doppelstellung der elektronischen Person 161
F. Die Elektronische Person als Teilnehmerin am Rechtsverkehr 162
I. Die funktionale Willenserklärung 162
1. Rechts- und Geschäftsfähigkeit autonomer Systeme 163
2. Anwendung des Minderjährigenrechts 164
3. Teilgeschäftsfähigkeit 164
4. Anwendung der Vertretungsregeln §§ 164 ff. BGB 165
5. § 56 HGB 166
6. Anfechtung 166
a) Anwendbarkeit der Anfechtungsregeln 167
b) Abgabe der Anfechtungserklärung 167
II. Verschuldensfähigkeit autonomer Systeme 167
1. Strafrecht 168
a) Auswahl zwischen Handlungsalternativen 169
b) Auswirkungen der Nichtanwendbarkeit von Art. 1 I GG 170
2. Zivilrecht 171
a) Persönlichkeitsentfaltung 173

b) Anreizsetzung/ Verschuldensäquivalent autonomer Systeme 173
III. Haftung 175
1. Haftung für ein autonomes System vor Eintragung 175
2. Haftung der elektronischen Person 177
a) Haftung der elektronischen Person für das autonome System 177
aa) Deliktshaftung 177
bb) Vertragliche Haftung 178
cc) Zusammenwirken mehrerer elektronischer Personen/ elektronischer Personen und Menschen 178
dd) Verhältnis der Eigenhaftung der elektronischen Person zur Produkthaftung 179
ee) Innerbetrieblicher Schadensausgleich 180
ff) Sonderfall autonomes Fahren 181
b) Haftung der elektronischen Person für ihren organschaftlichen Vertreter und für Erfüllungsgehilfen 183
3. Haftung des Verwenders für die elektronische Person 183
a) Haftung für Schäden Dritter 183
b) Haftung für Schäden der elektronischen Person 184
IV. Haftungsmasse 184
1. Natürliche/juristische Person muss einstehen (vollständige Bürgschaftslösung) 185
2. Eigene Haftungsmasse der elektronischen Person 185
a) Einzahlung durch den Anmelder 186
b) Erfolgsabhängige Errichtung der Haftungsmasse 186
c) Vertragliche Bürgschaft einer natürlichen Person 187
d) Pflichtversicherung für gefahrgeneigte elektronische Personen 188
e) Höhe der Haftungsmasse 189

f) Nachschusspflicht/Schwankung der Haftungsmasse 190
g) Versicherung statt Haftungsmasse im Einzelfall 191
h) Zwischenergebnis 191
3. Vollständige Pflichtversicherungslösung 191
V. Wissenszurechnung beim Einsatz autonomer Systeme 192
VI. Umfang und Grenzen der organschaftlichen Vertretungsmacht 194
VII. Das Innenverhältnis zwischen Verwender und elektronischer Person 196
1. Vollmachterteilung 196
2. Das zugrundeliegende Rechtsverhältnis 197
3. Der Inhaber als Verwender 198
VIII. Das Verhältnis der elektronischen Person zu Gesellschaften 199
1. Die elektronische Person als Gesellschafterin 199
a) Personengesellschaften 199
aa) Gesellschaftszweck außerhalb des Anwendungsbereich der elektronischen Person 200
bb) Gesellschaftszweck im Anwendungsbereich der elektronischen Person 201
b) Kapitalgesellschaften 202
2. Unternehmerische Tätigkeit der elektronischen Person 202
IX. Sachenrecht 203
1. Eigentum 203
2. Besitz 204
a) Ausübung durch das Organ 205
b) Funktionaler Besitzwillen 206
c) Entbehrlichkeit eines Besitzwillens 206
d) Zwischenergebnis 207
3. Besitzdienerschaft 208
X. Anwendbarkeit des Strafrechts auf die elektronische Person 208
1. Strafbegründende Handlung des autonomen Systems 209

2. Parallele zur Unternehmensstrafbarkeit 211
a) Übertragbarkeit der Figur des korporativen Täters 211
b) Übertragbarkeit der Figur des singulären/kollektiven Täters 213
3. Zwischenergebnis 214
4. Persönliche Strafbarkeit des Organs und Straftaten gegen die elektronische Person 214
XI. Anwendbarkeit des Ordnungswidrigkeitenrechts auf elektronische Personen 214
XII. Grundrechtsfähigkeit elektronischer Personen, Art. 19 III GG 215
XIII. Allgemeines Persönlichkeitsrecht elektronischer Personen 217
XIV. Die elektronische Person als Urheberin i.S.d. UrhR 217
XV. Besteuerung der elektronischen Person 219
G. Die Elektronische Person als Gegenstand des Rechtsverkehrs 219
I. Das Recht an der elektronischen Person 220
II. Verkehrsfähigkeit der elektronischen Person 221
1. Übertragung der elektronischen Person 221
2. Übertragung des autonomen Systems 221
3. Vervielfältigung des autonomen Systems 222
III. Abwicklung der elektronischen Person 222
IV. Der Bereitstellungsvertrag zwischen dem Inhaber des Rechts an der elektronischen Person und dem Verwender 223
V. Urheberrechtlicher Schutz autonomer Systeme 223
1. Schutz als Computerprogramm 224
2. Schutz als Datenbankwerk 224
3. Schutz als Datenbank 226
H. RmbH 227
I. Vertretung /Geschäftsfähigkeit 228
II. Haftung 229
III. Fazit zum RmbH 230

Teil 6: Fazit 231
A. Einführung 231
B. Zurechnung von Willenserklärungen 232

C. Haftung 234
D. Die elektronische Person 236

Literaturverzeichnis 241

Teil 1: Einleitung

Ausgehend von der industriellen Revolution, im Zuge derer der Einsatz technischer Hilfsmittel menschliche Arbeit im großen Stil zu ersetzen begann, setzte sich diese Entwicklung im zwanzigsten Jahrhundert mit der digitalen Revolution und dem Einsatz von Computertechnologie fort. Durch deren Einsatz konnten immer komplexere mechanische Tätigkeiten, aber auch das Treffen von Entscheidungen, automatisiert werden. Dies geschah, indem diese Entscheidungen und deren Voraussetzungen in Computerprogrammen festgelegt wurden.

Durch die Fortschritte im Bereich der „künstlichen Intelligenz" ist es möglich, Software zu entwickeln, die nicht auf das Konkretisieren von vorher in Programmen exakt festgelegten Entscheidungen beschränkt ist, sondern aufgrund ihrer Lernfähigkeit imstande ist, Entscheidungen selbstständig zu treffen. Die Gesamtheit der Softwarekomponenten, die diese von einer Einzelfallprogrammierung durch einen Menschen unabhängigen Entscheidungen ermöglichen, wird in dieser Arbeit als „autonomes System" bezeichnet.

Da Entscheidungen eines solchen autonomen Systems gerade nicht anhand eines festgelegten Programms getroffen werden, sind sie nicht im Einzelnen vorhersagbar. Dies ist zugleich die Prämisse dieser Arbeit. Der Umstand, dass durch autonome Systeme Entscheidungen getroffen werden können, die in ihrem konkreten Inhalt nicht auf eine menschliche Willensbetätigung rückführbar sind, führt bei deren Einsatz zu Zurechnungsproblemen, wenn das autonome System mit der Umwelt interagiert. Eine solche Interaktion ist sowohl physisch denkbar, beispielsweise wenn ein autonomes System zur Steuerung eines Fahrzeugs eingesetzt wird oder Zugriff auf einen Roboterarm hat, aber auch rechtsgeschäftlich, wenn ein autonomes System als Agent zum Abschluss von Verträgen eingesetzt wird.

Am Beginn dieser Arbeit steht zunächst eine kurze Einführung zu dem in der Arbeit verwendeten Begriff des „autonomen Systems" und dessen technischer Grundlagen.[1] Anschließend wird in Teil drei[2] der Arbeit gezeigt, dass die Erklärungen und Aktionen autonomer Systeme mit dem Instrumentarium des geltenden Rechts nicht vollständig fassbar sind. In den

---

1 S. 21, Teil 2: Autonome Systeme.
2 S. 33, Teil 3: Behandlung de lege lata.

Teilen vier[3] und fünf[4] der Arbeit wird daher die Schaffung einer neuen juristischen Person, der elektronischen Person, als Zurechnungssubjekt für Erklärungen und Aktionen autonomer Systeme vorgeschlagen und diese entworfen. Ziel der Einführung einer elektronischen Person ist es, den Einsatz autonomer Systeme im Rechtsverkehr zu ermöglichen. Dieses Ziel ist dabei sowohl der Grund als auch die Grenze der Einführung der elektronischen Person.

Die Seitenzahlen bei Querverweisen innerhalb dieser Arbeit beziehen sich stets auf die Überschrift des Gliederungspunkts, auf den verwiesen wird.

3 S. 133, Teil 4: Lösungen de lege ferenda.
4 S. 145, Teil 5: Die elektronische Person.

Teil 2: Autonome Systeme

A. Begriff des autonomen Systems

Der Begriff „autonomes System“ bezieht sich in dieser Arbeit auf die Gesamtheit der Softwarekomponenten, die entscheidend für die Interaktion des Systems mit seiner Umwelt, natürlichen Personen oder anderen autonomen Systemen sind. In welcher Form diese Interaktion erfolgt, ist zunächst zweitrangig. Sie kann in der Ausgabe eines Texts über einen Bildschirm oder über eine Sprachausgabe, in der Ansteuerung von Stellmotoren oder in der Kommunikation mit einem Server bestehen. Entscheidend ist lediglich, dass die ausgeführte Aktion nicht direkt auf eine steuernde natürliche Person zurückgeführt werden kann.

Eine direkte Rückführbarkeit ist beispielsweise gegeben, wenn konkrete Steuerbefehle per Fernbedienung erfolgen oder ein Programm für vordefinierte Inputs bestimmte vordefinierte Outputs liefert. Dann kann von einem reinen Reflex gesprochen werden; die Aktion muss dann derjenigen natürlichen Person zugeordnet werden, die die jeweilige Entscheidung getroffen hat.

Moderne Computerprogramme entfernen sich jedoch immer weiter von reinen Reflexgebilden mit einem überschaubaren Repertoire hin zu nicht im Einzelnen nachvollziehbaren Blackboxes mit nur generell gehaltenen Zielanweisungen.

Autonome Systeme zeichnen sich dadurch aus, Aktionen unabhängig von unmittelbarer Steuerung durchzuführen. Das Programm oder die Entscheidungsinstanz in einem Programm, die ein solches Verhalten ermöglicht, wird als künstliche Intelligenz (KI) bezeichnet. Dabei gibt es verschiedene Ansätze, was genau Ziel der KI-Forschung ist. Bei *Russell/Norvig* wird zwischen vier Ansätzen unterschieden: Zu erreichen, dass Programme wie Menschen denken, dass sie wie Menschen handeln, dass sie rational denken oder dass sie rational handeln.[5] Ein System wird dann als rational begriffen, wenn es auf der Basis seiner Erkenntnisse „richtig“ handelt.[6]

Ein besonderes Problem der Einstufung von Programmen als intelligent stellen bereits die verschiedenen Ansätze dar, menschliche Intelligenz zu

---

5 *Russell/Norvig*, S. 23.
6 *Russell/Norvig*, S. 22.

beschreiben. Für die Frage nach der rechtlichen Einordnung eines autonomen Systems ist es jedoch von untergeordneter Bedeutung, ab wann ein Programm als künstlich Intelligent anzusehen ist, solange die Wahl der Aktion selbstständig und nicht bis ins Detail vorhersehbar ist.

Auf der Annahme, dass es bei entwickelten autonomen Systemen nicht immer möglich ist, die konkreten Aktionen vorauszusehen[7], fußt diese Arbeit. Im Folgenden soll kurz auf die technische Entwicklung autonomer Systeme eingegangen und erklärt werden, weshalb es bei komplexen autonomen Systemen Handlungserwartungen, aber keine Sicherheit geben kann.

Da dies keine technische Arbeit ist, beschränkt sich die Darstellung im Folgenden auf den Versuch, dem Leser einen groben Überblick zu verschaffen.[8] Hierzu wird zunächst ein Überblick über die Geschichte der künstlichen Intelligenz gegeben. Im Anschluss daran werden einige technische Grundkonzepte erläutert, aufgrund derer eine Vorhersehbarkeit der Aktionen autonomer Systeme scheitert.

B. Historischer Überblick[9]

Die Geschichte der KI begann bereits im Jahr 1943 mit einer Arbeit von *McCulloch* und *Pitts*, in der diese die Grundstruktur neuronaler Netze beschrieben.[10] 1949 demonstrierte *Hebb* ein Verfahren, nach dem solche neuronalen Netze anhand einer Aktualisierungsregel neuen Gegebenheiten angepasst werden können; die Lernfähigkeit neuronaler Netze nach dieser

---

7 So *Clemens*, NJW 1985, 1998, 2001 f.; *Kirn/Müller-Hengstenberg*, S. 1 f.; Spindler/Schuster-*Spindler*, § 164, Rn. 12; *Grützmacher*, CR 2016, 695, 696; *Denga*, CR 2018, 69, 70; *Müller-Hengstenberg/Kirn*, CR 2018, 682, 683; *Stiemerling*, CR 2015, 762, 764; *Teubner*, AcP 218, 155, 164, 174; *Mayinger*, S. 14; *Spindler*, CR 2015, 766, 766; siehe auch Fn. 111.

8 Für eine Einführung s. *Ertel*, Grundkurs künstliche Intelligenz, 4. Auflage, Springer Vieweg 2016; für eine umfassende Einführung s. *Russell/Norvig*, Künstliche Intelligenz, 3. Auflage, Pearson 2012; für eine Einführung, die sich an Laien richtet, s. Matthias, S. 21 ff.; für einen kurzen Überblick s. *Stiemerling*, CR 2015, 762 ff.

9 Nach *Russell/Norvig*, S. 39 ff.. Für eine umfangreiche historische Einführung siehe ebenda.

10 *McCulloch/Pitts*, Bulletin of Mathematical Biophysics, Vol. 5, 1943, p. 115–133; *Russell/Norvig*, S. 39.

Regel wird als Hebb´sches Lernen bezeichnet.[11] 1950 bauten *Minsky* und *Edmonds* den ersten Computer auf der Basis neuronaler Netze.[12]

1956 fand die als Geburtsstunde der künstlichen Intelligenz bezeichnete Dartmouth-Konferenz statt, ein Workshop, an dem *Mc Carthy*, *Minsky* und *Edmonds* teilnahmen, aber auch Vertreter von IBM, der Princeton University, der Carnegie Tech und des MIT.[13] Nachdem zu Beginn der Entwicklung der künstlichen Intelligenz die Nachbildung menschlicher Nervenverbindungen durch neuronale Netze standen, konzentrierte sich die Wissenschaft ab den späten sechziger Jahren neben Experimenten mit genetischen Algorithmen[14] vor allem auf wissensbasierte Systeme wie die sogenannten Expertensysteme.[15] Ziel von Expertensystemen ist es, ausgehend von definierten Regeln zu Schlüssen zu gelangen. *MYCIN* war beispielsweise ein Expertensystem zur Diagnose von Blutinfektionen; anhand von 450 Regeln konnte es Blutinfektionen ebenso zuverlässig wie ein menschlicher Experte diagnostizieren.[16]

Seit 1986 rückt der Fokus wieder verstärkt auf neuronale Netze und damit den konnektionistischen Ansatz.[17]

C. Agententypen und Grundlagen

I. Agententypen

Es wird zwischen mehreren Agententypen unterschieden, welche hier in aufsteigender Komplexität dargestellt werden. Allen Typen gemein ist, dass sie – in unterschiedlichen Abstraktionsstufen – einen Input in einen Output umsetzen.[18]

Der einfachste Agent ist der Reflex-Agent. Ein Reflex-Agent setzt einen Input anhand einer starren wenn-dann-Regel in einen Output um, ohne vorherige Inputs oder Folgen zu beachten.[19] Ein Beispiel für einen solchen einfachen Reflex-Agenten ist etwa ein Staubsaugerroboter, der bei jeder

11 *Russell/Norvig*, S. 39.
12 *Russell/Norvig*, S. 40.
13 *Russell/Norvig*, S. 40 f.
14 *Russell/Norvig*, S. 45.
15 *Russell/Norvig*, S. 46 ff.
16 *Russell/Norvig*, S. 47.
17 *Russell/Norvig*, S. 48 f.
18 *Ertel*, S. 18.
19 *Russell/Norvig*, S. 76; *Ertel*, S. 19.

Kollision mit einer Wand eine 90-Grad-Drehung nach links durchführt.[20] Ein solcher Staubsaugerroboter wird unter Umständen einen Raum vollständig saugen oder sich irgendwo festfahren.

Soll der Roboter ein Signal geben oder sich abschalten, sobald er sich festgefahren hat, benötigt er einen Zähler, wie oft er sich bereits um 90 Grad gedreht hat und angestoßen ist. Das Wissen darüber, wie die Welt aussieht, beziehungsweise der Teilausschnitt, der für den Agenten interessant ist, wird Modell genannt.[21] Im Falle des Saugroboters beinhaltet sein Modell, wie oft er bereits angestoßen ist. Agenten, die ein Modell verwenden, nennt man „modellbasierter Agent"[22]; der Roboter mit Zähler ist also ein modellbasierter Reflexagent.[23]

Möchte man nun, dass der Staubsaugerroboter den Raum vollständig und planmäßig saugt und anschließend zu seiner Ladestation zurückkehrt, genügt es nicht, dass er jeweils auf Grundlage seines aktuellen Zustandes entscheidet.[24] Der Agent muss zusätzlich die Fähigkeit haben, mehrere Schritte vorauszuplanen, um das Ziel zu erreichen.[25] Dazu benötigt auch dieser Agent ein Modell, in diesem Fall eine Karte des Raumes, die Information, wo er und die Station sich befinden sowie wo er bereits gesaugt hat. Einen solchen Agenten bezeichnet man daher als (modellbasierten) zielbasierten Agenten.[26]

Nun gibt es nicht nur ein erfolgreiches Vorgehen, den Raum zu saugen. Der Agent muss jedoch eine Route wählen. Dies kann entweder zufällig erfolgen oder aufgrund einer Nutzenfunktion. Eine solche erlaubt es dem Agenten, sein Vorgehen so zu wählen, dass sein Nutzen maximiert wird.[27] Dabei kann der Nutzen beispielsweise ein möglichst geringer Energieverbrauch, eine geringe Lautstärke, die besonders schnelle Reinigung der Wohnung oder die Fertigstellung um 17:00 Uhr sein. Ein solcher Agent wird als nutzenbasierter Agent bezeichnet.[28]

Besitzt ein Agent darüber hinaus die Fähigkeit zu lernen, kann er den Erfolg seiner Aktionen bewerten und auf dieser Grundlage neue Regeln er-

---

20 Vgl. *Russell/Norvig*, S. 76, 61 f.
21 *Russell/Norvig*, S. 78.
22 *Russell/Norvig*, S. 78.
23 Vgl. *Russell/Norvig*, S. 78.
24 *Russell/Norvig*, S. 80.
25 *Russell/Norvig*, S. 80.
26 Vgl. *Russell/Norvig*, S. 80 f.
27 *Russell/Norvig*, S. 81; *Ertel*, S. 20.
28 Vgl. *Russell/Norvig*, S. 81 f.

stellen und sein Verhalten anpassen.[29] Ein lernfähiger Agent besitzt die Fähigkeit, neue Ansätze zu erproben, auch wenn sie aus aktueller Sicht wenig erfolgsversprechend erscheinen.[30]

Damit ein Agent die oben beschriebenen Fähigkeiten und Charakteristika erhält, sind verschiedenen Komponenten nötig, welche im Zusammenspiel zum erwünschten Erfolg führen. So gibt es mehrere Möglichkeiten, Lernfähigkeit zu implementieren oder die Nutzenfunktion zu berechnen. Im Folgenden sollen daher einige Grundprinzipien vorgestellt werden, auf denen einzelne Komponenten fußen können.

II. Symbolverarbeitender und konnektionistischer Ansatz

Grundsätzlich wird im Forschungsfeld der künstlichen Intelligenz zwischen dem symbolverarbeitungs- und dem konnektionistischen Ansatz unterschieden.[31] Der symbolverarbeitende Ansatz basiert auf der Repräsentation von Wissen durch abstrakte Zeichen (Schriftzeichen, Zahlen, Symbole etc.).[32] Beim konnektionistischen Ansatz werden Wissen und Handlungsanweisungen nicht symbolisch dargestellt, sondern in Form von Gewichtungen in einem neuronalen Netz.[33] Dadurch wird versucht, die „Bauart biologischer intelligenter Systeme […] nachzubilden."[34]

III. Programmierung in Programmiersprachen

Die Programmierung konventioneller Programme erfolgt in Programmiersprachen. Durch sie werden symbolisch Handlungsanweisungen festgeschrieben, wie etwa die Anweisung an den Staubsaugerroboter, bei frontaler Kollision eine Drehung um 90 Grad durchzuführen. Programme sind (abhängig von der verwendeten Programmiersprache) menschenverständlich und können im Gegensatz zu neuronalen Netzen grundsätzlich ohne weiteres nachvollzogen werden.

---

29 Vgl. *Russell/Norvig*, S. 83 ff.; *Ertel*, S. 20.
30 *Russell/Norvig*, S. 84.
31 *Spreckelsen/Spitzer*, S. 28 ff.
32 *Spreckelsen/Spitzer*, S. 29.
33 *Spreckelsen/Spitzer*, S. 29 f.
34 *Spreckelsen/Spitzer*, S. 29.