Lasse Ferdinand Quarck: Die künstliche Intelligenz in der Strafrechtsdogmatik. Zur Verantwortung beim Einsatz von intelligenten Agenten und zur (Be-)Strafbarkeit von „e-Personen“

*von Gunnar Spilgies**

2025, Springer, ISBN 978-3-658-50503-5, S. 261, € 84,99

Das Thema der strafrechtlichen Verantwortlichkeit für Rechtsgutsverletzungen beim Einsatz von KI ist mittlerweile monographisch intensiv bearbeitet. Die Mehrzahl der Autoren lehnt hierbei eine eigenständige strafrechtliche Verantwortlichkeit der KI sowohl de lege lata als auch de lege ferenda in Übereinstimmung mit der herrschenden Meinung ab.[1] Zuletzt hat jedoch *Schiemann* in dieser Zeitschrift eine Arbeit kritisch gewürdigt, in der sich die Autorin für die Ausgestaltung eines „Maschinenstrafrechts“ ausspricht.[2] Nunmehr unternimmt auch *Lasse Ferdinand Quarck*, der schon vor einiger Zeit die Einführung einer KI-Strafbarkeit als „langfristig unumgänglich“ bezeichnet hat,[3] in seiner von der Rechtswissenschaftlichen Fakultät der Christian-Albrechts-Universität zu Kiel als Dissertation angenommenen Arbeit den Versuch, die Notwendigkeit einer strafrechtlichen Verantwortlichkeit sog. starker KI als E-Person darzulegen und strafrechtsdogmatisch zu begründen. Die ausgezeichnete Bewertung der Arbeit („summa cum laude“), die Verleihung des Promotionspreises der Schleswig-Holsteinischen-Universitäts-Gesellschaft sowie der vom Autor selbst formulierte Anspruch, ein „besonderes Merkmal“ seiner Arbeit sei „die tiefgehende Analyse der Handlungs- und Schuldfähigkeit von Künstlicher Intelligenz, auch im Vergleich zu Unternehmen“ (S. 18), lassen eine gewinnbringende Lektüre erwarten. Ob die Arbeit dieser Erwartung tatsächlich gerecht wird, soll die nachfolgende Besprechung zeigen.

I. Aufbau der Arbeit

Die Arbeit ist in sieben Abschnitte gegliedert: Nach der „Einleitung“ (S. 8–18) und dem theoretischen Teil „Begriffliche und technische Grundlagen“ (S. 19–68) folgen die Hauptabschnitte „Verantwortlichkeit für die KI und Verantwortungsdiffusion“ (S. 69–121) und „Verantwortlichkeit der KI“ (S. 122–211). Daran schließen sich noch die beiden kurzen Abschnitte „Täterschaftliche Begehung mit oder mittels KI“ (S. 212–235) sowie „Folgeüberlegungen“ (S. 236–241) an. Auf S. 242 werden „Abschließende Thesen“ präsentiert. Mit dem Literaturverzeichnis (S. 243–261) endet die Arbeit.

II. Kritische Auseinandersetzung

1. „Einleitung“ (S. 8 ff.)

Einleitend betont *Quarck*, angesichts des Voranschreitens des flächendeckenden Einsatzes von KI im Zeitalter der vierten industriellen Revolution (Industrie 4.0, Internet of Things) müssten auch im Strafrecht „Standards und Regeln für die aufkommenden digitalen Probleme“ entwickelt werden. Vor diesem Hintergrund formuliert *Quarck* insbesondere folgende Kernfragestellungen:

„Kann ein Verantwortungsvakuum aus strafrechtlicher Sicht hingenommen werden?“ und „Kommt eine Strafbarkeit der KI selbst infrage und unter welchen Voraussetzungen? Kann eine KI insbesondere handlungs- und schuldfähig sein und Rechtssubjektivität besitzen?“ (S. 17).

2. „Begriffliche und technische Grundlagen“ (S. 19 ff.)

a) *Quarck* nähert sich diesen Fragen, indem er in einem theoretischen Grundlagenteil zunächst den Begriff der künstlichen Intelligenz erläutert, wobei er den Begriff analytisch zerlegt und sich als Erstes dem Begriff der Intelligenz widmet. Nachdem er vorweg erklärt, es gebe viele Intelligenzen und eine klare Definition fehle – womit er ein Vorurteil weiter verbreitet, das die Intelligenzforschung längst entkräftet hat[4] –, bestimmt er Intelligenz einerseits *ontologisch-deskriptiv* „als Kumulation verschiedener Fähigkeiten“, wie z.B. logisches Denken, Gedächtnisleistungen, Sprachgebrauch (S. 23), resümiert dann aber andererseits einem *soziologisch-askriptiven* Ansatz folgend, „dass Intelligenz keine binäre Eigenschaft darstellt, die entweder vorliegt oder nicht. Sie manifestiert sich vielmehr in der flexiblen Interaktion der intelligenten Entität mit ihrer Umgebung – sie entsteht also aus dem ‚sozialen Handlungszusammenhang‘“ (S. 24). In Fn. 74 verweist *Quarck* auf „*Görz/Nebel*, Künstliche Intelligenz, S. 11“, die ausdrücklich von der „Intelligenz für uns“ sprechen und keine Notwendigkeit sehen, eine „Intelligenz per se“ zuzuschreiben.

---

* *Gunnar Spilgies* ist freier Autor.

1 Vgl. *Lohmann*, Strafrecht im Zeitalter von Künstlicher Intelligenz, 2021, S. 105 ff., 142 f. (Bespr. von *Schiemann*, KriPoZ 2022, 478 f.); *Woltmann*, Der soziale Roboter, 2022, S. 166, 189 f. (vgl. aber S. 191 ff.: partielle Rechtssubjektivität); *Ibold*, Künstliche Intelligenz und Strafrecht, 2024, S. 260 ff.; *Schäfer*, Artificial Intelligence und Strafrecht, 2024, S. 517 ff. (Bespr. von *Schiemann*, KriPoZ 2025, 431 ff.); *Preetz*, Rechtsgutsverletzungen durch KI-Systeme im Spiegel der Funktion des Strafrechts, 2025, § 10 (S. 462); *Simmler*, Strafrechtliche Verantwortung beim Zusammenwirken von Mensch und Maschine, 2025, S. 296, 455.

2 Vgl. *Möllenbeck*, Maschinenstrafrecht, 2025 (Bespr. von *Schiemann*, KriPoZ 2026, 63 f.); vgl. auch noch *Fincan*, Artificial Intelligence and Legal Issues, 2023, S. 172 ff., der eine Neuprogrammierung, Wiedergutmachung und temporäres Abschalten als KI-Sanktionen befürwortet.

3 Siehe *Quarck*, ZIS 2020, 65 ff. (69).

4 Vgl. *Stern/Neubauer*, Psychologische Rundschau 2016, 15 (17 ff.).

Im Folgenden oszilliert *Quarcks* Intelligenzbegriff zwischen diesen beiden begriffsmethodischen Ansätzen. So wird auf S. 26 wieder einerseits die soziale Konstruktion von Intelligenz herausgestellt:

„Das bedeutet, dass Intelligenz, egal ob natürlich oder künstlich, jedenfalls durch unsere gesellschaftliche und soziale Wirklichkeit zugeschrieben wird. […] Diese Zuschreibung erfolgt anhand bestimmter objektiver Anhaltspunkte, anhand derer es für den Betrachter logisch ist, dass das beobachtete Individuum Intelligenz besitzt."

Die Belege in den Fn. 78, 79 beziehen sich auf „*Krämer*, Synästhetische Soziologie, S. 55", der beschreibt, wie im Alltag anhand von „Zeichen der Intelligenz" versucht wird, „echte" Intelligenz vorzuspielen, sowie auf „*Burchard*, Künstliche Intelligenz als Ende des Strafrechts?, S. 7", der „KI als non-essentialistisches Konstrukt" betrachtet. Dann aber heißt es andererseits:

„Dies [objektive Anhaltspunkte der Zuschreibung] können beispielsweise Intelligenz-Tests sein […]. Intelligenz bemisst sich also aus der Interaktion einer Entität mit ihrer Umwelt, bzw. von Entitäten untereinander und der Beobachtung dieser Interaktion durch eine oder mehrere andere Entitäten. Diesbezüglich können auch durchaus Trugschlüsse vorliegen."

Wenn Quarck „Intelligenz-Tests" als objektive Anhaltspunkte für die Zuschreibung in Betracht zieht und „Trugschlüsse" im Rahmen der Zuschreibung für möglich hält, so setzt er implizit voraus, dass es eine „echte" Intelligenz gibt. Intelligenz kann aber nicht zugleich ontologisch („echt") in der Person vorhanden sein und sozial konstruiert („gemacht") werden. Sie kann sich nicht sowohl als „soziales Faktum" aus der konkreten Interaktion ergeben als auch als ein „innerer Zustand" darstellen (vgl. S. 28). Da *Quarck* nicht sauber zwischen der ontologischen und soziologischen Bestimmung der Intelligenz trennt, gerät er in einen logischen Zirkel: Er behauptet zwar, Intelligenz sei eine Zuschreibung, behandelt sie aber wie die Feststellung einer ontologischen Tatsache. Die explizit askriptive Begriffsbestimmung wird durch implizit ontologische Rechtfertigungskriterien unterlaufen. Die Argumentation beruht damit auf einem methodischen Kategorienfehler.

Die psychologische Ursache hinter diesem Kategorienfehler ist nicht schwer auszumachen: Dem Bedürfnis, askriptiv zu verfahren, weil Intelligenz ontologisch nicht direkt wahrnehmbar ist, steht das Bedürfnis nach Objektivität gegenüber, das darauf gerichtet ist, den deskriptiven Anspruch weiterzuverfolgen, und zu dem Versuch führt, durch „Ontologisierung" Objektivität zu sichern, um die soziale Zuschreibung nicht als „willkürlich" oder „bloße Fiktion" erscheinen zu lassen. Dies alles wird hier so detailliert ausgeführt, weil diese kategoriale Vermischung der ontologischen und soziologischen Begriffsbestimmung, wie sich zeigen wird, die gesamte Arbeit durch zieht. *Quarck* selbst verweist darauf, dass eine „derartige Zuschreibung innerer Zustände" auch in der Rechtswissenschaft ein bekanntes Vorgehen sei: „wird doch im Ergebnis auch hinsichtlich etwa dem Vorsatz oder der Schuld genauso verfahren" (S. 28). An späterer Stelle bezeichnet *Quarck* diesen logischen blinden Fleck der Kategorienvermischung als ein „Muster" im Rahmen seiner Arbeit, das er wie folgt beschreibt: „Die Zuschreibung anhand von Plausibilitätsgesichtspunkten unter Berücksichtigung der rechtlichen Notwendigkeit für ein gerechtes und verbindliches Zusammenleben" (S. 204).

b) Diesem „Muster" gemäß bestimmt der Autor denn auch den Begriff der „künstlichen" Intelligenz (S. 29 ff.). Diese werde nämlich genauso wie menschliche Intelligenz einerseits als ein „soziales Faktum" kontextabhängig durch unser soziales System zugeschrieben und sei kein „Dauerzustand" (S. 33), andererseits sei „es durchaus denkbar, künstlich intelligente Entitäten einem IQ-Test zu unterziehen" (S. 34). Letztlich meine der Begriff künstliche Intelligenz, „dass nicht die Intelligenz als solche künstlich ist, sondern der Agent" (S. 35). Es stelle sich die Frage, „ob und wenn ja wie, jedenfalls einigermaßen, allgemeingültig festgestellt werden kann, ob eine künstliche Entität intelligent ist oder nicht" (S. 36). Eine Möglichkeit, die Zuschreibung von Intelligenz in der Interaktion Mensch-Maschine zu verobjektivieren, sei, „neben den oben angesprochenen IQ-Tests, der […] Turing-Test" (S. 36). Der Turing-Test ist nun tatsächlich geeignet, den Prozess der sozialen Zuschreibung von Intelligenz durch standardisierte Bedingungen zu „objektivieren", aber gerade deshalb ist er nicht mit einem IQ-Test als objektivem Maß „wirklicher" Intelligenz vergleichbar. *Quarck* verfolgt die Suche nach allgemeingültigen Maßstäben nicht weiter, da der interdisziplinäre Rahmen der Arbeit dadurch überschritten werde (S. 37). Auch die Frage, wann eine artifizielle Entität Fähigkeiten besitze, wodurch die Schwelle zur menschengleichen, sog. starken KI überschritten werde, die Grundvoraussetzung für den E-Personenstatus sei, könne derzeit nicht beantwortet werden, da die Entwicklung derartiger KI-Systeme „noch einige Jahrzehnte dauern" werde. Eine Antwort hierauf müsse in der Zukunft vielmehr von einem „multidisziplinären Gremium" gefunden werden (S. 38). Den Widerspruch, dass ein Gremium eine solche ontologische „Schwelle" nicht feststellen kann, wenn die KI-Eigenschaft durch soziale Zuschreibung entsteht und kein Dauerzustand ist, erkennt *Quarck* nicht.

c) Anschließend gibt *Quarck* einen Überblick über die Art und Weise, wie KI lernt und welche Fähigkeiten sie bereits heute besitzt (S. 39 ff.). Er kommt zu dem Schluss, die Mehrheit der KI-Forscher sei sich einig, „dass es nur noch eine Frage der Zeit ist, bis es sog. starke, also gemessen an ihren Fertigkeiten menschengleiche KI geben wird" (S. 52 f.). Die in Fn. 206 angeführten Belege auf *Hilgendorf*, *Russell* und *Miebach* stützen diese „Konsensthese" jedoch gerade nicht. *Hilgendorf* äußert sich nur zu rechts- und technikphilosophischen Fragen bzgl. auto-

nomer Roboter und lässt den Robotikforscher *Hans Moravec* zu Wort kommen.[5] *Russell* hält eine Superintelligenz zwar für möglich, verteidigt diese Position aber gegen Skeptiker – von einer mehrheitlichen Einigkeit der KI-Forscher geht *Russell* also nicht aus.[6] *Miebach* resümiert sogar explizit: „Die KI-Experten sind sich nicht einig in der Einschätzung, ob Superintelligenz von KI -Systemen überhaupt erreichbar ist."[7] Obwohl namentlich der von *Quarck* zitierte KI-Experte *Russell* in seinem Aufsatz die Gefahr sieht, „dass schlecht konstruierte superintelligente Maschinen eine große Bedrohung für die Menschheit darstellen würden"[8], hält *Quarck* die Sorge vor dystopischen Zukunftsszenarien dagegen für unbegründet, denn er geht technokratisch optimistisch davon aus, dass „Mechanismen in den Algorithmus implementiert werden können, die die Vornahme überraschender Handlungen verhindern bzw. minimieren" (S. 53).

d) Im Folgenden ordnet der Autor den Roboter begrifflich und klassifikatorisch ein, da dieser „am wahrscheinlichsten im Mittelpunkt der zum jetzigen Zeitpunkt vorstellbaren Fälle von Rechtsgutsverletzungen stehen wird" (S. 53 ff.). Besonders bedeutsam für das Strafrecht seien autonome Roboter, weil ihr erweiterter Aktionsspielraum das Risiko von Rechtsgutsverletzungen durch „überraschende" Aktionen steigere (S. 58 ff.). Die Autonomie des Roboters bestimmt *Quarck* wiederum gemischt ontologisch-soziologisch: Neben einem bestimmten Grad der selbständigen Entscheidung in technischer Hinsicht verlangt *Quarck* als ontologische Voraussetzung, der autonome Agent müsse „Handlungskontexte differenzieren können sowie Folgenbewusstsein in der unbekannten Situation haben" (S. 60). Der Nachweis in Fn. 234 geht zurück auf „*Loh*, InTer 2017, 220 (225)", die u.a. das Folgenbewusstsein zur psychischen Zuschreibungsbedingung für Verantwortung erklärt. Der Zweitnachweis auf „*Sorge*, Softwareagenten, S. 8" geht fehl, da der Autor am angegebenen Ort für die Annahme von Autonomie voraussetzt, dass der Agent sein Verhalten kontrollieren und ohne Fremdeingriff handeln könne. Den Einwand, es „handele sich um eine reine Datenverarbeitung und nicht um einen in diesem Sinne intentionalen Vorgang" (S. 62), kontert *Quarck* dann seinem „Muster" gemäß mit folgender Erwägung:

„Die Zielbestimmung des Systems und die daraus folgenden Aktionen reichen […] für eine Intentionalität aus. Entscheidend für die Annahme von Autonomie ist also die externe Perspektive, unabhängig von – sowohl bei Mensch als auch Maschine ja ggf. nur fiktiven – inneren Zuständen wie Folgenbewusstsein und Intentionalität im humanistischen Sinne. Das alles gilt dann, wenn ein Rückschluss von den äußeren Handlungen auf die inneren Zustände möglich ist (anders wird etwa bei der Annahme von Vorsatz auch nicht verfahren)" (S. 62 f.).

Zunächst argumentiert *Quarck* hier explizit für eine konstruktivistische Begriffsbestimmung. Dies belegt auch der Verweis in Fn. 241 auf den Aufsatz von „*Kemper*, cognitio 2018, 1 (12)", die Intentionalität als sozial konstruiert betrachtet und sich u.a. auf die „intentional stance" des Philosophen *Daniel Dennett* stützt. Dieser verbale Externalismus, der sogar durch den pejorativen Hinweis auf „nur fiktive" innere Zustände bestärkt wird, wird dann aber durch den Rekurs auf „innere Zustände" wieder aufgehoben. So wird letztlich der zunächst ausgeschlossene Ontologismus doch wieder zur Bedingung der sozialen Zuschreibung gemacht. Logisch vertretbar ist aber nur entweder ein Normativismus oder ein Ontologismus. Daher ist auch der erneute Vergleich mit der Annahme von Vorsatz, scil. dolus eventualis (vgl. schon S. 28 sowie oben a), verfehlt, sofern *Quarck* damit bezweckt, sein methodisches Vorgehen durch eine Praxis des Vorsatznachweises zu legitimieren, die selbst unter dem Verdacht steht, denselben methodischen Kategorienfehler zu begehen.[9] Der Vorsatz-Vergleich suggeriert zudem einen Konsens, der zumindest verbal nicht existiert: Vielmehr streitet die h.M. für eine wortgetreue ontologisch-deskriptive Bestimmung des Vorsatzes als „psychischer Tatsache", während die Gegenansicht den Versuch einer „*normativierenden Korrektur* des üblichen Psychologismus"[10] unternimmt und den Vorsatz soziologisch-askriptiv als „normative Konstruktion" bestimmt, mit anderen Worten also bewusst fingiert.[11]

e) Zum Abschluss des Grundlagenteils hebt *Quarck* sodann hervor, dass sich aus dem erhöhten Schadensrisiko beim Einsatz autonomer Roboter ein „Bedürfnis der dogmatischen Einordnung autonomer KI" ergebe (S. 64 ff.). Dieser Abschnitt weist Redundanzen zur Einleitung der Arbeit auf und liest sich streckenweise wie ein zweites Einleitungskapitel. Im Vorgriff auf die weitere Untersuchung kündigt *Quarck* an, im Folgenden darzulegen, „warum es aus strafzwecktheoretischen Erwägungen einer KI-Strafbarkeit bedarf" (S. 66).

3. „Verantwortlichkeit für die KI und Verantwortungsdiffusion" (S. 69 ff.)

a) Es folgt der erste Hauptabschnitt, in dem die menschliche Verantwortlichkeit für Rechtsgutsverletzungen beim Einsatz von KI und das Problem der Verantwortungsdiffusion untersucht werden. Bei strafrechtlichen Erfolgen aufgrund algorithmischer Fehlentscheidungen komme gegenwärtig als Erstes eine Fahrlässigkeitsstrafbarkeit der involvierten menschlichen Akteure in Betracht. Ausgehend von drei Fallbeispielen beschäftigt sich *Quarck* daher mit der grundlegenden Fahrlässigkeitsdogmatik (S. 72 ff.), den Fahrlässigkeitselementen (S. 75 ff.) sowie exkursartig mit Problemen des autonomen Fahrens

5 Siehe *Hilgendorf*, in: Hilgendorf/Hötitzsch (Hrsg.), Das Recht vor den Herausforderungen der modernen Technik, 2015, S. 32 f.
6 Siehe *Russell*, Digitale Welt, 3/2021, 34 ff.
7 *Miebach*, Digitale Transformation von Wirtschaft und Gesellschaft, 2020, S. 320.
8 *Russell*, Digitale Welt, 3/2021, 34.
9 Vgl. hierzu *Momsen*, KriPoZ 2018, 76 ff., der am Schluss seines Aufsatzes die Gerichte in die Pflicht nimmt, „sich in die Psychologie des Angeklagten einzufühlen" (100).
10 *Jakobs*, Kritik des Vorsatzbegriffs, 2020, S. 39 Fn. 120 (Hervorhebung im Original).
11 Vgl. exemplarisch die Kontroverse zwischen *Puppe*, ZIS 2014, 66 (68 ff.), und *Fischer*, ZIS 2014, 97 ff., sowie *Leitmeier*, HRRS 2016, 243 ff., der *Puppe* sekundiert.

(S. 105 ff.). Als Ergebnis hält er fest, dass eine Fahrlässigkeitsstrafbarkeit für Hersteller, Programmierer und Benutzer regelmäßig mit Verweis auf das erlaubte Risiko und die Sozialadäquanz ausscheide, wenn die Sorgfaltspflichten beim Einsatz von KI eingehalten worden seien (S. 113). Insbesondere für Fälle mit selbstlernender KI konstatiert *Quarck* damit eine „Verantwortungsdiffusion". Die vielfach geäußerten Zweifel am Vorliegen einer solchen Verantwortungsdiffusion[12] werden nicht diskutiert. Auch die Frage der Praxisrelevanz wird nicht aufgeworfen, obwohl *Quarck* zuvor im Grundlagenteil das Risiko von Fehlentscheidungen durch starke KI noch als gering eingeschätzt hat (vgl. S. 53 sowie oben 2 c).

b) Das Phänomen der Verantwortungsdiffusion führt den Autor zu der Frage, „ob dies ein Zustand ist, der aus strafrechtlicher Perspektive hingenommen werden kann" (S. 114). Der anschließende kursorische Blick auf die Straftheorien („Die Strafzwecke") bleibt ohne tieferen Bezug zum weiteren Argumentationsgang. Vielmehr erklärt *Quarck* unter pauschalem Verweis auf die Strafzwecke in apodiktischer Weise:

„Der Zustand eines solchen rechtlichen Vakuums kann nicht hingenommen werden. Denn insbesondere unter der Berücksichtigung der generalpräventiven Strafzwecke besteht die Gefahr, dass es zu einer Erosion des Rechts kommt. Bliebe delinquentes Verhalten ohne Reaktion, […] würde das Normvertrauen erschüttert" (S. 118).

Beim Lesen dieser Zeilen kommt dem Rezensenten unmittelbar in den Sinn: Es gab in Kiel einmal eine von *Joachim Hellmer* begründete strafrechtskritische Tradition, die alternative Lösungsansätze jenseits der Strafe in den Mittelpunkt stellte. Insoweit kämen bei Vorliegen einer Verantwortungsdiffusion in KI-Fällen z.B. restorative Modelle oder technische Compliance statt Strafen in Betracht. *Quarck* bricht fundamental mit dieser Tradition. Theoretisch beschwört er, das „Strafrecht muss […] ultima ratio bleiben" (S. 119), um dann aber einem Punitivismus das Wort zu reden und zu postulieren:

„Es kann aus Perspektive derjenigen, die von den technischen Neuerungen betroffen sind (also im Prinzip alle Menschen) in der Tat nicht angehen, dass KI in den Rechtsverkehr entlassen wird, die sanktionslos machen darf, was sie will" (S. 120).

Auf diese Weise könnte man aber ebenso gut ein generalpräventives Strafbedürfnis gegenüber Rechtsgutsverletzungen von Kindern „begründen". In Wahrheit handelt es sich um die bloße Behauptung eines solchen Strafbedürfnisses, also um die Erschleichung des Beweisgrundes (Petitio Principii). Die Begründung unterstellt, was erst zu beweisen wäre: Die Frage ist ja gerade, ob ein solches Strafbedürfnis wegen eines Normgeltungsschadens in KI-Fällen überhaupt entsteht, was voraussetzt, dass der agierende Roboter in den Augen der Rechtsgemeinschaft ein taugliches Zuschreibungssubjekt ist „mit der Kompetenz, die Normgeltung zu desavouieren"[13]. *Quarcks* Behauptung eines Strafbedürfnisses liegt damit dieselbe Grund-Folge-Verwechselung zugrunde wie der Argumentation der Vertreter funktionaler Schuldlehren.[14] In diesem Zusammenhang muss erstaunen, dass *Quarck* dieser Verwechselung nicht gewahr wird, obwohl er seine generalpräventive Begründung des Strafbedürfnisses in KI-Fällen gegen den funktionalen Schuldbegriff an späterer Stelle mit den Worten verteidigt:

„Ein gesellschaftliches, generalpräventives Bedürfnis nach Sanktion kann überhaupt nur dann entstehen, wenn Schuld zugerechnet werden kann. Andernfalls würde, im Falle der KI, nur eine Maschine und somit schon kein taugliches Zurechnungssubjekt vorliegen" (S. 187).

Der erste Hauptabschnitt endet mit dem Vorschlag des Autors zur „Auflösung der Verantwortungsdiffusion": der Strafbarkeit des intelligenten Agenten selbst (S. 121).

4. „Verantwortlichkeit der KI" (S. 122 ff.)

a) Im folgenden zweiten Hauptabschnitt der Arbeit widmet sich *Quarck* nun der eigentlichen Kernfrage, ob die Verantwortungsdiffusion durch eine „Verantwortlichkeit der KI" de lege lata oder de lege ferenda „aufgelöst" werden könne, wobei eine materielle Strafbarkeit nur für starke, autonome KI in Betracht komme. Zu Beginn stellt *Quarck* hierzu „Vorüberlegungen" an, in denen er den Begriff der Verantwortlichkeit seinem gemischt ontologisch-soziologischen „Muster" gemäß sowohl auf kognitive Fähigkeiten (Kommunikationsfähigkeit, Handlungsfähigkeit, Urteilskraft) zurückführt als auch soziologisch zuschreibend versteht (S. 124 f.). Da nur starke KI verantwortlich sein könne, zeige sich erneut die „Korrelation von Intelligenz, Autonomie und Verantwortung" (S. 125). Ein Bezug zur weiteren Argumentation lässt sich wiederum nicht ausmachen.

b) Mit Blick auf die materielle Strafbarkeit von KI erkennt *Quarck* dann bzgl. der Handlungs- und Schuldfähigkeit das Problem, „dass irgendwie geartete innere Zustände Strafbarkeitsvoraussetzung sind und diese inneren Vorgänge eines künstlichen neuronalen Netzwerks möglicherweise anders ablaufen als die in einem menschlichen Gehirn" (S. 126). Diese ontologische Bestimmung des Handlungs- und Schuldbegriffs wird nicht problematisiert, vielmehr stelle sich die Frage, „ob sich Begriffsbedeutungen und die Dogmatik von Handlung und Schuld, die originär für den Menschen erdacht worden sind, auf KI übertragen und anwenden lassen" (S. 126).

c) Die Frage der Handlungsfähigkeit von KI reduziert *Quarck* nach thematisch weitschweifigen allgemeinen Ausführungen zur dogmatischen Bedeutung der Handlung schließlich auf die Frage, ob KI-Verhalten das Kriterium der „Willensgetragenheit" nach den verschiedenen

[12] Vgl. z.B. *Ziemann*, in: Hilgendorf/Günther (Hrsg.), Robotik und Gesetzgebung, 2013, S. 190 f.; *Lohmann* (Fn. 1), S. 202; *Schäfer* (Fn. 1), S. 21 ff., 508 ff.; *Wohlers*, in: Wohlers/Seelmann (Hrsg.), Schuldgrundsatz, 2024, S. 263.

[13] *Jakobs*, Strafrecht Allgemeiner Teil, 2. Aufl. (1991), 17/2.

[14] Vgl. zu diesem Standardargument gegen die funktionale Schuldlehre nur *Eisele*, in: TK-StGB, 31. Aufl. (2025), Vor §§ 13 ff. Rn. 117 m.w.N.

Handlungsbegriffen erfülle (S. 133 ff.). Nach einer breiten Diskussion der Handlungsbegriffe kommt *Quarck* zu dem Ergebnis, dass starke KI aufgrund ihrer Autonomie (Folgenbewusstsein, Intentionalität) und Lernfähigkeit nach den als vorzugswürdig erachteten natürlichen, intentionalen und personalen Handlungsbegriffen unter dem Aspekt der „Willensgetragenheit" als handlungsfähig anzusehen sei (S. 138, 146 f., 149). Die Handlungsfähigkeit könne aber dennoch angezweifelt werden, da die Handlungsbegriffe „menschliches" Verhalten voraussetzten.

Rückschlüsse für die Handlungsfähigkeit intelligenter Agenten erhofft sich *Quarck* daher von der parallelen Diskussion über die Handlungsfähigkeit einer anderen nichtmenschlichen Entität: die des Verbands (S. 150 ff.). In einer ausführlichen Stellungnahme zu den einzelnen Modellen der Verbandshandlung gelangt *Quarck* zu dem Schluss, dass die Handlungsfähigkeit von Verbänden zu verneinen sei (S. 159 ff.). Ohne die nötige Tiefe kritisiert *Quarck* dabei die systemtheoretische Begründung der Verbandshandlung, wenn er ihr pauschal vorwirft, sie laufe Gefahr, „ins Uferlose abzudriften" (S. 162), und führe „zu einer völligen handlungstheoretischen Loslösung der Individuen und des Verbands voneinander" (S. 163). Daher erkennt *Quarck* auch das Potenzial des systemtheoretischen Ansatzes nicht, der KI qua Teilnahme an gesellschaftlichen Kommunikationsprozessen Handlungsfähigkeit zuzuschreiben (vgl. S. 166), wiewohl er den einschlägigen Aufsatz von „*Teubner*, AcP 218 (2018), 155 (164 ff.)" zitiert. Vielmehr lehnt er eine Konstruktion der Handlungsfähigkeit von KI „in Anlehnung an die holistischen Modelle" ab und relativiert den Nutzen der systemtheoretischen Ansätze mit der lapidaren Bemerkung, diese zeigten nur „die Bedeutung, die soziale Systeme bei der Herausarbeitung von Handlungsfähigkeit haben" (S. 167).

Im Anschluss legt der Autor seinen „Entwurf eines strafrechtlichen Handlungsbegriffs für intelligente Agenten" vor: Starke KI sei autonom, ihr Agieren damit „willensgetragen". Das „Menschsein" sei keine notwendige Voraussetzung für die Willensbildung und auch andere Entitäten könnten grundsätzlich hierzu fähig sein. Die „Willensgetragenheit" bzw. „Intentionalität" bestimmt *Quarck* wiederum genau wie die Autonomie (vgl. S. 62 f. sowie oben 2 d) seinem gemischt ontologisch-soziologischem „Muster" gemäß:

„Willensgetragenheit im Sinne eines intentionalen Handelns wird also wiederum nicht als psychischer Zustand einer bestimmten, eingegrenzten physischen Entität, sondern als Zuschreibung zielgerichteten Handelns durch einen Beobachter verstanden […]. Wenn das System einem Verhalten Intentionalität in diesem Sinne zuschreibt, ist es auch eine Handlung" (S. 168).

Der unbefangene Leser deutet diese Ausführungen als eine rein soziologisch-konstruktivistische Bestimmung der Willensgetragenheit. Der aufmerksame Leser erkennt dagegen, dass *Quarck* zuvor „innere Zustände" zur Strafbarkeitsvoraussetzung erklärt (vgl. S. 126 sowie oben b) und eine systemtheoretische Konstruktion der Handlungsfähigkeit der KI explizit abgelehnt hat (vgl. S. 167 sowie soeben) und daher die zugeschriebene Willensgetragenheit bzw. Intentionalität gerade nicht als reines „soziales Konstrukt" oder *Dennett*'sche „intentional stance" versteht, sondern letztlich doch ontologisch als einen „inneren Zustand". Im späteren Fazit drückt *Quarck* dies denn auch klar aus: „Für die Willensgetragenheit der Handlung reicht lediglich eine irgendwie geartete Zielgerichtetheit im Sinne einer Intentionalität, welche als innerer Zustand zwangsläufig zugeschrieben werden muss" (S. 210).

Am Ende sieht *Quarck* das „eigentliche Problem" darin, „Grenzen festzusetzen, wann die KI so stark ist, dass sie handlungsfähig wird", und meint: „Der zentrale Anknüpfungspunkt hierfür ist die Autonomie des intelligenten Agenten" (S. 170). Wenn aber starke KI handlungsfähig ist, kann nicht die Handlungsfähigkeit vom Grad der Stärke abhängen, und wenn starke KI autonom ist, kann die Autonomie nicht gleichzeitig das Kriterium sein, um festzustellen, ab wann KI stark genug ist. *Quarck* konstruiert hier ein Problem, das sich nach seinen eigenen Prämissen gar nicht stellt. Freilich lässt sich dieser Vorwurf der Probleminszenierung letztlich allgemein gegen die Untersuchung der Handlungsfähigkeit von KI erheben. Da nach *Quarck* starke, autonome KI per definitionem intentional agiert (vgl. S. 60, 62 f. sowie oben 2 d) und damit handlungsfähig ist, gerät die Frage, ob KI handlungsfähig ist, zu einer Scheindebatte ohne echten Erkenntniswert. Dementsprechend konstatiert *Quarck* auch schon vorab der Frage nach der Handlungsfähigkeit von KI, dass nur starke KI „über Fähigkeiten verfügt, die Autonomie dergestalt begründet, dass daraus Kommunikationsfähigkeit, Handlungsfähigkeit und Urteilskraft entstehen und aus diesen Kompetenzen dann genuine Handlungen hervorgehen" (S. 125).

d) Die Frage der Schuldfähigkeit von KI untersucht der Autor, indem er zunächst den psychologischen, normativen und funktionalen Schuldbegriff kritisch analysiert und im Anschluss prüft, ob das KI-Verhalten als Grundlage für eine normative Schuldzuschreibung in Betracht kommt (S. 173 ff.). Da der psychologische Schuldbegriff keine Bedeutung mehr habe, wendet sich *Quarck* sogleich dem normativen Schuldbegriff zu. Mit Verweis auf „BGHSt 2, 194 (200)" wird der Inhalt des Schuldvorwurfs dahingehend bestimmt, dass dem Täter mit dem Unwerturteil der Schuld vorgeworfen werde, sich für das Unrecht entschieden zu haben, obwohl er sich rechtmäßig hätte verhalten können. Notwendige Voraussetzung des Schuldvorwurfs sei daher „schon begrifflich die Willensfreiheit: ohne freien Willen gibt es auch keine freie Entscheidung, die man persönlich vorwerfen könnte" (S. 175). Zwar könne der freie Wille nach Erkenntnissen der Hirnforschung empirisch nicht nachgewiesen werden, der Gesetzgeber habe jedoch in § 20 StGB „eine klare Entscheidung für ein indeterministisches Menschenbild getroffen" (S. 177).

Begründet werde die Willensfreiheit unterschiedlich: mit Hilfe des „subjektiven Freiheitsempfindens" oder als „Teil der sog. gesellschaftlichen Rekonstruktion der

Wirklichkeit“ (*Schünemann*). Der von *Quarck* behauptete Zusammenhang zwischen der *Schünemann*’schen Willensfreiheitsbegründung und *Kohlrauschs* „staatsnotwendiger Fiktion“ der Willensfreiheit besteht indes nicht. Den Vertretern einer analogischen Schuldfeststellung nach Maßgabe des Verhaltens einer „fiktiven, sozialen Vergleichsperson“ hält *Quarck* einen „Trugschluss“ vor: „Bezweifelt man generell die Willensfreiheit für den Menschen, so muss man dies dann konsequenterweise auch für die Maßfigur tun, die dem Täter nebengestellt wird“ (S. 178). Dieser Vorhalt beruht aber auf einem doppelten Missverständnis der analogischen Schuldlehre: Erstens bezweifeln deren Vertreter die Willensfreiheit keineswegs generell, vielmehr soll nur das Problem der Nachweisbarkeit der Willensfreiheit „gelöst“ werden, und zweitens kommt es bzgl. der Maßfigur nicht auf ihre Willensfreiheit an, sondern darauf, ob sie anstelle des Täters rechtmäßig gehandelt hätte.

Im weiteren Verlauf wird die Argumentation zunehmend undurchsichtiger und widersprüchlicher. So bezweifelt *Quarck*, dass die Schuld eine solche Willensfreiheit voraussetze, wie sie die Hirnforschung verstehe, und meint: „Ein interdisziplinär identisches Verständnis von Willensfreiheit wäre somit ein Kategorienfehler“ (S. 179). Beide Aussagen werden nicht begründet. Nach *Quarck* folgt daraus, „dass die persönliche Erfahrung und eine daraus folgende Fiktion von Willensfreiheit im biologischen Sinne für das Vorliegen von Willensfreiheit im juristischen Sinne genügt“ (S. 179). In Fn. 773 verweist *Quarck* auf „*Burkhardt*, Das Magazin 2003, 21 (23)“ sowie „*Hirsch*, ZStW 106 (1994), 746 (763)“. Beide vertreten am angegebenen Ort aber keine fiktionalistische Position. Wenig einleuchtend ist auch, dass aus der „persönlichen Erfahrung“ von Willensfreiheit eine Willensfreiheits*fiktion* folgen soll. Dann erwägt *Quarck*, dass für die Schuldfähigkeit „eine Fiktion gar nicht erforderlich sei, sondern vielmehr die Willensfreiheit gar nicht im absoluten Indeterminismus zu suchen sei“ (S. 180). Der Zusammenhang erschließt sich hier nicht. Unklar ist auch, wie *Quarcks* Aussage, „dass selbst bei einer Beweisbarkeit des deterministischen Menschenbildes dieses keinerlei Auswirkungen auf das Schuldstrafrecht hätte“ (S. 180 f.), mit *Quarcks* ontologischer Willensfreiheitsprämisse des Schuldvorwurfs und seiner Feststellung zusammenpasst, der Gesetzgeber habe in § 20 StGB eine klare Entscheidung für ein indeterministisches Menschenbild getroffen (vgl. S. 177 sowie soeben). Der Nachweis in Fn. 779 auf „*Hochhuth*, JZ 2005, 745 (752)“ geht überdies fehl. Dann heißt es wieder: „Das gesamte Rechtssystem basiert auf der Voraussetzung, ‚dass es so etwas wie Willensfreiheit und persönliche Verantwortung gibt‘“ (S. 182), wobei das Zitat nachweislich der Fn. 789 von „*Pauen*, Allgemeine Zeitschrift für Philosophie 2001, 23“ stammt, der jedoch einen kompatibilistischen Freiheitsbegriff vertritt. Zum Abschluss der Betrachtung des normativen Schuldbegriffs vergleicht *Quarck* die Bestimmung der Strafbegründungsschuld – wie schon die Bestimmung der Autonomie (vgl. S. 62 f. sowie oben 2 d) und der Intelligenz (vgl. S. 28 sowie oben 2 a) – mit der Vorsatzbestimmung. Seinem gemischt ontologisch-soziologischen „Muster“ gemäß führt er aus, die Vorsatzbestimmung sei ein „normativer Zuschreibungsakt“ und kein „psychologischer Erkenntnisakt“:

„Stattdessen werden regelmäßig objektive Anhaltspunkte oder Plausibilitätsgesichtspunkte herangezogen, etwa, indem aufgrund der besonderen Gefährlichkeit des Verhaltens zumindest auf Eventualvorsatz geschlossen wird“ (S. 182 f.).

Die kategoriale Vermischung der soziologischen und ontologischen Begriffsbestimmung, durch einen „normativen Zuschreibungsakt“ auf Eventualvorsatz als „inneren Zustand“ zu „schließen“, spiegelt sich hier sogar in den Fußnotennachweisen wider: So zitiert *Quarck* in Fn. 795 einerseits „*Schünemann*, in: Schünemann, Grundfragen des modernen Strafrechtssystems, S. 183 f.“, der die Schuldfähigkeit ontologisch bestimmt, und andererseits „*Gómez-Jara Díez*, ZStW 119 (2007), 290 (307)“, der die Unternehmensschuldfähigkeit als normative Konstruktion versteht. Genauso in Fn. 796: Der erste Nachweis geht auf eine Entscheidung des *BGH* (NStZ 2003, 431 [431 f.]), die beansprucht, den Vorsatz ontologisch zu bestimmen. Dann wird vergleichend aber auch auf „*Jakobs*, Strafrecht AT, § 17 Rn. 23“ und „*Puppe*, Kleine Schule des juristischen Denkens, S. 48 ff.“ verwiesen, die beide einem funktional-askriptiven Ansatz folgen. Dieser positive Verweis hindert *Quarck* wiederum nicht, *Jakobs*’ normativen Ansatz in der nachfolgenden Diskussion des funktionalen Schuldbegriffs zu kritisieren, womit er ganz nebenbei seine frühere Ansicht aufgibt.[15]

Zu Beginn dieser Diskussion werden *Jakobs* und *Roxin* als „bedeutendste Vertreter“ eines „rein präventiv ausgerichteten Schutzstrafrechts“ bezeichnet, die für eine „Abkehr vom bisherigen Schuldstrafrecht“ plädierten (S. 183). Maßgeblich für die Verantwortlichkeit des Täters seien „Nützlichkeitsgesichtspunkte, vor allem die Generalprävention“. Namentlich nach *Jakobs* bedürfe es für die Verantwortungszuschreibung keiner Willensfreiheit. Die generalpräventive Schuldbegründung konfrontiert *Quarck* sodann mit Standardeinwänden gegen die *spezialpräventive Straf*begründung (S. 185). Diese Kritik irritiert nicht nur wegen dieser Schieflage, sondern auch deshalb, weil *Quarck* selbst – trotz der generalpräventiven Herleitung der Notwendigkeit einer KI-Strafbarkeit (vgl. S. 118 ff. sowie oben 3 b) – wegen der Möglichkeit der Umprogrammierung der KI die spezialpräventive Wirksamkeit einer KI-Strafe im späteren Verlauf der Arbeit hervorhebt (S. 237). Weiter führt er gegen den funktionalen Schuldbegriff an, dieser ignoriere im Rahmen der Verantwortungszuschreibung den Fokus auf die persönliche Schuld

[15] Siehe *Quarck*, ZIS 2020, 65 (68), wo es noch heißt: „Ein solches funktionales Verständnis von Schuld ermöglicht nun auch eine Zuweisung der Verantwortlichkeit an intelligente Agenten. […] Es kommt also weder bei Mensch noch Maschine darauf an, ob die getroffene unrechte Entscheidung auf determinierten biologischen bzw. algorithmischen Vorgängen basiert oder auf rechtlich fehlerhafter freier Willensbildung.“

und er impliziere selbst eine Schuldzuschreibung (S. 185 f.). Ein generalpräventives Strafbedürfnis könne „nur dann entstehen, wenn Schuld zugerechnet werden kann“ (S. 187). Abgesehen davon, dass sich *Quarcks* Kritik gegen seine eigene generalpräventive Herleitung eines Strafbedürfnisses in KI-Fällen richtet (vgl. S. 118 ff. sowie oben 3 b), offenbart diese Kritik die fehlende Unterscheidung zwischen der epistemischen Zuschreibung deskriptiver Begriffe und der normativen Zuschreibung askriptiver Begriffe, die dem methodischem Kategorienfehler zugrunde liegt.

Im „Zwischenfazit zur Schuld“ gerinnt dann dieser Kategorienfehler zu einem alles vereinenden ontologisch-soziologischen Schuldbegriff (S. 187 f.): Der Unterschied zwischen den Schuldbegriffen, so *Quarck*, bestehe „lediglich in der Perspektive: Der normative Schuldbegriff stellt auf das individuelle Selbstempfinden ab, während der funktionale das Kollektive in den Blick nimmt. Zugeschrieben wird gleichermaßen, aus welcher Perspektive ist vor allem eine Glaubensfrage […].“ Vorzugswürdig sei „der (individuell-)normative Schuldbegriff, allerdings ohne die Zugrundelegung eines absoluten Indeterminismus und in Bewusstsein der nur marginalen Unterschiede. Grundlage für die normative Zuschreibung ist der epistemisch logische Schluss, dass die subjektive Wahrnehmung von Willensfreiheit und die Möglichkeit deren Existenz für die Bejahung nach einem rechtswissenschaftlichen Verständnis ausreicht. Dieser Schluss findet sich auch im Rahmen des funktionalen Schuldbegriffes wieder und ist somit anerkanntermaßen das entscheidende Merkmal für die Zuschreibung“.

In Wahrheit existiert die suggerierte Einigkeit zwischen Vertretern einer traditionellen und einer funktionalen Schuldlehre nicht. *Quarck* bekommt die differenzierte Debatte um den Schuld(fähigkeits)begriff, in der sich ontologisch-deskriptive und soziologisch-askriptive Lehren mit jeweils unterschiedlichen Lesarten der „Zuschreibung“ gegenüberstehen, gar nicht in den Blick bzw. blendet sie aus.

Als Leser fragt man sich an dieser Stelle, wie es dem Autor auf der Grundlage seiner letztlich ontologisch-indeterministischen Bestimmung des Schuldbegriffs gelingen will, der KI normativ Schuldfähigkeit zuzuschreiben. Denn bis hierin wurde ja von der Programmierbarkeit und damit Determinierbarkeit von KI ausgegangen (vgl. S. 39 ff., 42, 53). Erstaunlicherweise heißt es nun aber:

„Eine nachweislich determinierte KI kann nicht schuldfähig sein, da dann bereits die bloße Möglichkeit der Willensfreiheit als eine normative Grundlage für eine Verantwortungszuschreibung entfällt und bei sicherem Wissen der vollständigen Determiniertheit die Zuschreibung von Schuld nicht plausibel wäre. Sie wäre dann schon nicht darauf angelegt im ‚Bewusstsein der Freiheit‘ zu handeln (und auch ansonsten kein Mensch […])“ (S. 188).

In Fn. 827 postuliert *Quarck* dann, in dieser Untersuchung werde „von der Indeterminiertheit autonomer Roboter ausgegangen“. Dieses Indeterminismus-Postulat überrascht auch deshalb, weil der Mensch damit etwas erschaffen können soll, das sich, wie *Quarck* konstatiert (vgl. S. 176 f. sowie soeben), nicht einmal beim Menschen selbst empirisch nachweisen lässt. Zudem übersieht *Quarck*: Wenn KI nun aber wirklich indeterminiert *ist*, bedarf es gar keiner normativen Zuschreibung mehr, weil die Indeterminiertheit sicher feststeht. Damit wäre eigentlich die Frage der Schuldfähigkeit der KI positiv beantwortet. Doch wird die Diskussion über „Anhaltspunkte für und gegen eine Schuldzuschreibung“ ohne erkennbaren roten Faden weitergeführt (S. 189 ff.). Erwogen wird z.B., die Steuerungsfähigkeit von KI in der Interaktion zwischen Mensch und Maschine wie bei einem „Turing-Test, nur wesentlich ausdifferenzierter“, festzustellen (S. 190). In dem Fortpflanzungsbedürfnis, das einer Determiniertheit nahekomme, zeige sich: „Die Annahme von Freiheit ist in unserer Gesellschaft conditio sine qua non für Verantwortung und Verantwortungssubjektivität“ (S. 190). Warum aber soll das Fortpflanzungsbedürfnis die Freiheitsannahme für Verantwortung belegen? Eine erklärende Herleitung fehlt. Weiter wird behauptet, sofern das Verhalten von KI „plausibel erklärt und prognostiziert werden kann, so kann ihr auch Bewusstsein und ein freier Wille zugeschrieben werden“ (S. 191). Auch dieser Zusammenhang harrt einer Begründung, denn eine Verhaltenserklärung und -prognose steht mit dem Indeterminismus-Postulat gerade im Widerspruch. Allerdings geht der Verweis in Fn. 840 auf „*Erhardt/Mona*, in: Gless/Seelmann, Intelligente Agenten und das Recht, S. 80“, die einen kompatibilistischen Willensfreiheitsbegriff vertreten. Wie das zu deuten ist, bleibt unklar.

Die anschließende Betrachtung der Verbandsschuld soll weitere Rückschlüsse für die Schuldfähigkeit intelligenter Agenten liefern (S. 192 ff.). Im Ergebnis lehnt *Quarck* zwar eine Verbandsschuld ab, denn Schuld dürfe „nicht nur als bloße Urheberschaft verstanden werden, da ihr, die Annahme des sittlich freiverantwortlich handelnden Individuums zugrundeliegt, und zwar auch dann, wenn sie zugeschrieben wird“ (S. 197). Der Gedanke, dass ein Verband ein „sozialer Akteur“ sei, lasse sich aber auf intelligente Agenten übertragen (S. 199). Erfahrungen aus Österreich mit der Verbandsstrafbarkeit zeigten keine negativen Rückwirkungen auf das allgemeine Strafrecht, so dass dies auch für die Schuldfähigkeit von KI nicht zu erwarten sei (S. 200).

Schließlich diskutiert der Autor die Einführung eines „e-Personenstatus als notwendige Bedingung der Schuldzuschreibung“ (S. 200 ff.). Denn strafrechtliche Verantwortlichkeit „gelingt nur bei Anerkennung als Person“. De lege lata werde Rechtssubjektivität für intelligente Agenten in allen Rechtsbereichen überwiegend abgelehnt. De lege ferenda bestimmt *Quarck* dann den Begriff der Rechtsperson – im Hinblick auf einen möglichen E-Personenstatus von KI – seinem „Muster“ gemäß sowohl als psychische Eigenschaft als auch als soziales Urteil. So bemerkt er dem ontologischen Ansatz folgend, „dass in den verschiedenen Rechtsgebieten für die Annahme von Rechtssubjektivität bestimmte Fähigkeiten oder Eigenschaften vorausgesetzt werden, so dass jedenfalls de lege

ferenda ein Personenstatus für KI denkbar ist“ (S. 203). Daraus folgert *Quarck*: „Der Begriff der Rechtsperson ist somit nicht statisch, sondern ein dogmatischer Begriff, dessen Bedeutung von der Rechtswissenschaft bestimmt werden muss und in der Geschichte auch schon immer auf verschiedene, epochen- und kontextabhängige Weise bestimmt worden ist“ (S. 203). Die begriffsmethodische Zweideutigkeit dieser Aussage spiegelt sich in den Belegen in Fn. 901: Verwiesen wird auf „*Erhardt/Mona*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, S. 83“ und „*Fincan*, Artificial Intelligence and Legal Issues, S. 65 f.“, die einen ontologischen Personenbegriff vertreten, sowie auf „*Gómez-Jara Díez*, ZStW 119 (2007), 290 (307 f.)“, der die Rechtsperson als normatives Konstrukt des Rechtssystems betrachtet. Im Anschluss an diesen Satz folgt als Ergänzung ein Zitat von „*Jakobs*, Das Schuldprinzip, S. 29“, in dem dieser dem soziologisch-askriptiven Ansatz folgend die Person im Recht, wie *Gómez-Jara Díez*, als soziales Konstrukt betont: Diese werde „im Recht generalisierend-normativ bestimmt“ und es sei „nicht einmal so, daß ihre Selbstbestimmung zur Rechtlichkeit in einem psychologisierend verstandenen Sinn erwartet würde […]“.

Im Geiste dieses ontologisch-soziologischen Ansatzes heißt es weiter, die Annahme von Rechtssubjektivität im Strafrecht setze daher voraus, „dass der Mensch normalerweise Fähigkeiten besitzt, die die Annahme von Schuld ermöglichen. Hierfür kommt es auch auf die soziale Wirklichkeit an sowie auf die Frage, ob die Anerkennung als Rechtsperson zweckmäßig ist“ (S. 204). Dementsprechend müsse die KI als taugliche Normadressatin „Folgenbewusstsein sowie die Fähigkeit zur Abwägung“ besitzen (S. 205) – Voraussetzungen, die starke KI nach *Quarck* per definitionem erfüllt. Dabei beeilt sich *Quarck* – gegen die askriptive Sichtweise von *Teubner*[16] gerichtet und mit Verweis auf das Objektivitätsbedürfnis (siehe oben 2 a) – festzustellen:

„Es kann nach hier vertretener Ansicht aber nicht gänzlich auf die kognitiven Fähigkeiten verzichtet werden. […] Würde man indes zulassen, dass allein die soziale Wirklichkeit ausschlaggebend sein kann, könnte ein bloßer Eliza-Effekt ohne tatsächliche Fähigkeiten ausreichen und wäre missbrauchsanfällig und unbestimmt. Nicht jede beliebige Entität sollte durch bloße Übereinkunft Rechtssubjektivität erhalten können“ (S. 208).

Um seine ontologisch-soziologische Bestimmung der Rechtsperson zu verteidigen, bedient sich *Quarck* an dieser Stelle eines Strohmann-Arguments: Der abermalige Willkür-Vorwurf gegen eine systemtheoretische Begriffsbestimmung beruht auf einer verkürzten Darstellung und zeichnet ein Zerrbild, denn *Teubner* lässt keineswegs allein die soziale Wirklichkeit als konstitutiv für Rechtssubjektivität genügen, sondern knüpft die Zuschreibung von Rechtssubjektivität für KI an rechtliche Autonomiekriterien, die er unter dem Schlagwort „Entscheidung unter Ungewissheit“ zusammenfasst.[17] *Teubners* Kriterien sind sogar konkreter als *Quarcks* schlichter Verweis auf „Folgenbewusstsein sowie die Fähigkeit zur Abwägung“ (S. 205), wobei *Teubner* im Übrigen auch Intentionalität im *Dennett*’schen Sinne als Autonomiekriterium ansieht.

Zwar gibt *Quarck* zu Beginn seiner Betrachtung der Verbandsschuld zu bedenken, wegen „der fehlenden Eigenschaft als Mensch könnte sich die Schuldfähigkeit von KI Kritik ausgesetzt sehen“ (S. 192), jedoch geht er diesem Einwand des Anthropozentrismus als Grundlage der Schuldzuschreibung im Weiteren nicht nach, auch nicht bei der Frage eines möglichen E-Personenstatus von KI. Er blendet insbesondere die Tatsache aus, dass gegenwärtig in der sozialen Wirklichkeit nur ein Mensch schuldfähig sein kann, d.h. „eine Person, die als Gleicher definiert wird“ und die „zur Zeit der Tat ein Subjekt mit der Kompetenz ist, die Normgeltung in Abrede zu stellen“[18]. Notwendige Bedingung der Schuldzuschreibung ist also derzeit, dass der Akteur überhaupt zur Gattung „Mensch“ gehört, er also kein Roboter ist. Da ein Roboter keine den anderen gleiche Person ist, fehlt ihm die Kompetenz, die Normgeltung zu desavouieren.[19] Das mag *Quarck* bedauern, ändert aber solange nichts an der Schuldzuschreibung, bis „die Menschen humanoide Roboter als alter ego begreifen, d.h. dem Verhalten künstlicher Intelligenz den gleichen Orientierungswert für ihr eigenes soziales Erleben und Handeln wie dem Verhalten anderer Menschen zuerkennen.“[20] Stattdessen greift *Quarck* zu einem rhetorischen Kniff, wenn er die Verleihung der Rechtssubjektivität an KI mit einer sozialen Wirklichkeit begründet, die er selbst erst für die Zukunft vorhersagt:

„Sieht die soziale Wirklichkeit entsprechend der Prognosen so aus, dass intelligenten Agenten eine Personalität durch die Gesellschaft zugeschrieben wird, so ist dies ein starkes Indiz dafür, dass auch das Recht dieser Wirklichkeit folgen sollte. Das Recht […] hat auch die Aufgabe gesellschaftliche Entwicklungen aufzunehmen und abzubilden“ (S. 209).

Schon vor diesem Hintergrund ist unverständlich, dass *Quarck* am Schluss der Arbeit eine Kriminalstrafe als KI-Sanktion in der Sache mit der Erwägung in Zweifel zieht, dass in der sozialen Wirklichkeit einem Roboter die Schuldfähigkeit mangels „Subjektqualität“ gerade nicht sozial zugeschrieben wird (vgl. S. 240 f. sowie unten 6 b).

Am Ende der Untersuchung einer KI-Schuld plädiert *Quarck* für die Schaffung eines „e-Personenstatus“, und zwar entweder „durch eine Reformierung des StGB“ oder „durch spezialgesetzliche Regelungen außerhalb des StGB“ in einem „KI-StGB“, nicht aber durch eine vollständige Gleichstellung mit dem Menschen (S. 209 ff.).

16 *Teubner*, AcP 218 (2018), 155 (171).
17 Vgl. *Teubner*, AcP 218 (2018), 155 (174 ff.).
18 *Jakobs* (Fn. 13), 17/48.
19 Vgl. *Seher*, in: Gless/Seelmann (Hrsg.), Intelligente Agenten und das Recht, 2016, S. 58 f.; *Fateh-Moghadam*, ZStW 131 (2019), 863 (877 f.); *Wohlers*, in: Wohlers/Seelmann (Fn. 12), S. 274 ff.
20 *Frister*, Strafrecht Allgemeiner Teil, 10. Aufl. (2023), § 3 Rn. 11.

5. „Täterschaftliche Begehung mit oder mittels KI" (S. 212 ff.)

Es folgt ein Abschnitt, in dem *Quarck* untersucht, welche Auswirkungen eine Rechtssubjektivität von KI auf die Strafbarkeit menschlicher Tatbeteiligter hätte. Auf eine nähere Betrachtung wird hier verzichtet.

6. „Folgeüberlegungen" (S. 236 ff.)

Zum Abschluss der Arbeit stellt der Autor „Folgeüberlegungen" zum Wesen der KI-Sanktion und zum Umfang der Rechtssubjektivität an, die „nicht mehr Teil der eigentlichen Arbeit" seien. Dass darin der bisherigen Argumentation nachträglich das Fundament entzogen wird, ahnt man als Leser daher nicht.

a) Zunächst aber offenbart *Quarck* nochmals seine punitive Einstellung, wenn er die general- und spezialpräventive Wirksamkeit einer KI-Strafe betrachtet. Obwohl *Quarck* in seiner Arbeit stets die „soziale Wirklichkeit" beschwört, erscheint ihm die in der Gesellschaft herrschende Ansicht der Strafunfähigkeit von Robotern nur als unerwünschte Strafeinschränkung:

„Wenn nun aber die breite Masse der Bevölkerung einen intelligenten Agenten als ‚unbestrafbar' ansieht, weil er ja ‚nur eine Maschine' und daher unter anderem unempfänglich für ein Strafübel ist […], so können generalpräventive Erwägungen nur eingeschränkt zur Geltung kommen" (S. 237).

Den umgekehrten Schluss, dass die Gesellschaft intelligenten Agenten als Maschinen keine Schuldfähigkeit zuschreibt und ein Strafbedürfnis mangels Normgeltungsschadens daher gar nicht entsteht, vermag *Quarck* an dieser Stelle offensichtlich nicht zu erwägen (paradoxerweise drei Seiten später aber schon [vgl. sogleich unter b]). Vielmehr hofft er, „dass bei flächendeckendem Einsatz und entsprechend hochentwickelter KI der Eindruck, es handele sich ‚nur um eine Maschine' zusehends aufgeweicht wird" (S. 237), und sehnt die generalpräventiven Strafbedürfnisse gleichsam herbei. Das grundlegende Problem, dass eine Bestrafung von KI einerseits die dahinterstehenden Menschen von der Verantwortung entlastet, andererseits jedoch wesentlich betrifft,[21] wird nicht thematisiert.

Die Straflust des Autors zeigt sich auch in seiner Euphorie für eine Umprogrammierung der KI als Strafe, durch die sich spezialpräventiv „maximale Wirkung" erzielen ließe, die beim Menschen durch Resozialisierung nicht erreichbar sei (S. 237). Schon *Ziemann* wies seinerzeit gegen dieselbe Argumentation des Philosophen *Andreas Matthias* in seiner 2008 erschienenen Pionierarbeit mit dem Titel „Automaten als Träger von Rechten" nicht nur auf die verfassungsrechtlichen Bedenken eines solchen Vorschlags einer „strafweisen Umprogrammierung der Maschine" hin und sah darin „eine Maßnahme überwunden geglaubten Strafrechtsdenkens", sondern zeigte sich außerdem beunruhigt angesichts dieser „Unbekümmertheit im Umgang mit Strafe".[22] Gleichzeitig offenbart *Quarcks* Glaube an eine mögliche Umprogrammierung der KI eine rein ergebnisorientierte Argumentation: Während er nämlich zur Abwehr dystopischer Szenarien die Programmierbarkeit von KI-Aktionen betont (vgl. S. 53), verweist er zur Begründung der KI-Strafbarkeit auf deren Unvorhersehbarkeit (vgl. S. 64 ff., 123) und postuliert zur Begründung der KI-Schuldfähigkeit gar deren Indeterminiertheit (vgl. S. 188), um bei der Frage der KI-Bestrafung unversehens wieder alle Hoffnung in deren Programmierbarkeit zu setzen. Je nach gewünschtem Ergebnis stützt sich *Quarck* also mal technokratisch optimistisch auf die Kontrollierbarkeit von KI-Aktionen, mal auf deren Unvorhersehbarkeit, ja Indeterminiertheit. Dass eine wesensmäßig unvorhersehbar agierende KI eben gerade nicht umprogrammierbar ist, wird nicht gesehen. Schließlich ist auch unklar, welchen Bezug diese spezialpräventive Strafbegründung eigentlich zur ursprünglich ausschließlich *generalpräventiven* Herleitung einer notwendigen KI-Strafbarkeit (S. 118 ff., siehe oben 3 b) hat.

b) Im Anschluss nun prüft *Quarck*, ob es sich bei der KI-Sanktion überhaupt um eine Kriminalstrafe handelt, und meint, man könne „durchaus zu dem Schluss kommen, dass […] eine KI-Sanktion keine Kriminalstrafe sein kann" (S. 240 f.). Als Leser fragt man sich irritiert: Hatte *Quarck* nicht gerade die general- und spezialpräventive Wirkung einer KI-*Strafe* ausgelotet? Und hatte er nicht zuvor in seiner „eigentlichen Arbeit" entschieden und mit beachtlichem Aufwand für eine *strafrechtliche* Verantwortlichkeit der KI gestritten, weil „es aus strafzwecktheoretischer Sicht nicht hinnehmbar ist, eine Verantwortungslücke bestehen zu lassen" (S. 126), und im Ergebnis explizit eine entsprechende „Reformierung des StGB" oder ein „eigenes KI-StGB" vorgeschlagen (S. 209 f.)? Überraschenderweise gibt *Quarck* an dieser Stelle seine bisherige Argumentation implizit auf und entzieht damit seiner „eigentlichen Arbeit" nachträglich das argumentative Fundament. Zur Begründung führt er u.a. an, der Missbilligungscharakter einer KI-Sanktion könne mit der Erwägung angezweifelt werden, dass die KI

„zwar ein, aber kein gleichwertiges Mitglied der sozialen Gemeinschaft ist. Dann nämlich wäre der durch die Missbilligung betroffene soziale Geltungsanspruch der KI weniger stark ausgeprägt als beim Menschen und der Vorwurf entsprechend weniger schwer. Eine KI könnte also per se nicht so sehr gegen die gesellschaftlichen Regeln verstoßen wie ein Mensch, weil sie, trotz Ähnlichkeit, keiner ist" (S. 240).

[21] Vgl. hierzu z.B. *Ziemann*, in: Hilgendorf/Günther (Fn. 12), S. 191; *Fateh-Moghadam*, ZStW 131 (2019), 863 (877 f.); *Schäfer* (Fn. 1), S. 518 ff.; *Ibold* (Fn. 1), S. 262 ff.

[22] *Ziemann*, in: Hilgendorf/Günther (Fn. 12), S. 189 f.

Dass dieser Zweifel am Missbilligungscharakter einer KI-Sanktion auf die fehlende Schuldfähigkeit der KI rekurriert, scheint *Quarck* nicht (mehr[23]) zu sehen. Anders als noch drei Seiten zuvor (vgl. S. 237, oben a) dient ihm die soziale Wirklichkeit plötzlich als Beleg dafür, dass der KI in der Sache keine Schuldfähigkeit zugeschrieben wird, weil sie kein Mensch ist. Da *Quarck* aber offensichtlich die KI unbedingt sanktionieren will, wechselt er unvermittelt die Sanktionsart: „Denkbar wäre alternativ eine Form des Maßregelrechts oder Sanktionen sui generis" (S. 241). Das Sanktionsinstrument scheint also gleichgültig, nach dem Motto: Wenn Strafe nicht funktioniert, dann eben eine Maßregel! Dieser Wechsel zu einem der Strafe widersprechenden Sanktionsmodell ist ein weiterer pronoucierter Ausdruck für den punitiven Ansatz des Autors.

III. Literaturauswertung, Sprache

Nach dieser inhaltlichen Kritik noch ein kurzes Wort zur wissenschaftlichen Arbeitsweise des Autors. Zu bemängeln ist die Literaturbasis der Arbeit, die ein auffallendes Ungleichgewicht aufweist. Während *Quarck* populärwissenschaftliche Publikationen und tagesaktuelle Medien breit rezipiert, ist einschlägige Fachliteratur nur selektiv ausgewertet. Grundlegende Quellen bleiben unberücksichtigt, wie z.B. die vorliegend zitierten Beiträge von *Sascha Ziemann* und *Bijan Fateh-Moghadam* sowie selbst die Pionierarbeit von *Andreas Matthias*, in der dieser eine rechtliche Verantwortung autonomer Maschinen und entsprechende Gesetzesänderungen fordert. Insgesamt erscheint die Quellendichte der Arbeit daher zu dünn. Im theoretischen Grundlagenteil wird durch die zahlreichen fehlerhaften Seitenzahlangaben in den Fußnoten zusätzlich der Zugang zu den Quellen behindert.

Zudem wird die Lektüre durch das sichtliche Bemühen des Autors um wissenschaftliche Diktion sprachlich erschwert. Zum Beispiel: „Der Herausarbeitung eines Begriffsverständnisses von Intelligenz, das die Rechtsfindung beim Einsatz von KI fördert, ist die Annahme einer solchen ‚extended intelligence' allerdings nicht dienlich" (S. 25). Auch grammatikalisch fehlerhafte Satzkonstruktionen stören den Lesefluss. Zum Beispiel: „Hierfür wurden insbesondere der (hinsichtlich des Menschen vorzugswürdigere) natürliche bzw. intentionale Handlungsbegriff weiterentwickelt um Erwägungen, die aus Ideen der Verbandshandlung entstammen, ergänzt" (S. 170).

IV. Fazit

Nach dem Gesagten ergibt sich, dass die Arbeit die hohen Erwartungen nicht zu erfüllen vermag, die sowohl durch die formalen Auszeichnungen als auch durch den vom Autor selbst proklamierten Anspruch einer tiefgehenden Analyse geweckt werden. Zum einen leidet die Argumentation unter einem methodischen Kategorienfehler, weil die ontologisch-deskriptive und soziologisch-askriptive Begriffsbestimmung vermischt werden. Daher gelingt es *Quarck* insbesondere nicht, die Handlungs- und Schuldfähigkeit sowie die Rechtssubjektivität starker KI widerspruchsfrei zu begründen. Zum anderen drückt sich in *Quarcks* Forderung einer strafrechtlichen Verantwortlichkeit starker KI ein Punitivismus aus, der eine Verantwortungslücke und ein Strafbedürfnis bei Rechtsgutsverletzungen durch starke KI schlicht behauptet. Dass *Quarck* diese Forderung im Schlussteil implizit aufgibt, weil er den Strafcharakter der KI-Sanktion bezweifelt, ist der zentrale Widerspruch der Arbeit. Eine wissenschaftliche Arbeit, die ihre Kernthese letztlich selbst in Frage stellt, bleibt eine Antwort schuldig.

[23] An früherer Stelle hat *Quarck* den Zusammenhang zwischen Missbilligungscharakter der Strafe und Schuldfähigkeit durchaus gesehen, allerdings den umgekehrten Schluss gezogen: „Wenn nun künstliche Intelligenzen […] schuldfähig im Sinne einer (aus unserer sozialen Wirklichkeit folgenden) Verantwortungszuschreibung sind, dann kann die Strafe ihren missbilligenden Charakter auch gegenüber KI entfalten" (ZIS 2020, 65 [69]).