Zeitschrift für Medienwissenschaft, Jg. 15, Heft 29 (2/2023), https://doi.org/10.14361/zfmw-2023-150209. Published by transcript Verlag.

Ein Gespräch zwischen NOORTJE MARRES und PHILIPPE SORMANI

KI TESTEN

«Do we have a situation?»

Dieses Gespräch basiert auf einem Austausch zwischen Noortje Marres (University of Warwick) und Philippe Sormani (Universität Lausanne), der sich aktuellen Fällen von realweltlichen KI-Tests und den Situationen widmet, die sich daraus ergeben haben – oder auch nicht. Es fand am 25. Mai 2022 online im Rahmen der Siegener Ringvorlesung «Testing Infrastructures» statt. Die *ZfM* publiziert hier eine gekürzte Fassung, die den in Soziologie und Science and Technology Studies (STS) geführten Diskurs erstmals auf Deutsch dokumentiert. Eine englische Langfassung ist als Working Paper des Sonderforschungsbereichs «Medien der Kooperation» erschienen.

Einleitung

Befürworter*innen der ‹neuen› KI, sowohl in der Informatik als auch in den Sozial- und Geisteswissenschaften, haben behauptet, dass die heutigen, sehr großen Deep-Learning-Modelle radikal neue Fähigkeiten zum kontextbezogenen Beurteilen und Treffen von Entscheidungen entwickelt haben sowie ein Situationsbewusstsein aufweisen (vgl. Abb. 1 für eine spielerische Reflexion dieser Behauptung). Dieses Argument wird in hochkarätigen Veröffentlichungen, Konferenzberichten und *arXiv*-Papers angeführt,[1] aber auch nahegelegt durch Tests und Demonstrationen wie der Präsentation von DeepMind im Radcliffe Observatory in Oxford,[2] AlphaGos Sieg im Four Seasons Hotel in Seoul (Südkorea)[3] und Testversuchen mit selbstfahrenden Fahrzeugen in städtischen Zentren wie Phoenix (Arizona) und Coventry im Vereinigten Königreich.[4] Solche öffentlichen Demonstrationen haben nicht nur eine bemerkenswerte Rolle bei der Verbreitung der Behauptung gespielt, dass die ‹neue› KI über situative Intelligenz verfügt, sondern auch bei der Problematisierung solcher Behauptungen. Im Folgenden zeigen wir auf, wie sozial- und medientheoretische Studien zu KI-Tests sich mit Annahmen hinsichtlich situativer

1 Vgl. Yann LeCun, Yoshua Bengio, Geoffrey Hinton: Deep learning, in: *Nature*, Bd. 521, 2015, 436–444, doi.org/10.1038/nature14539; Rishi Bommasani u. a.: On the opportunities and risks of foundation models (Report), *arXiv.org*, 16.8.2021, doi.org/10.48550/arXiv.2108.07258; Jonathan Roberge, Michael Castelle: Toward an End-to-End Sociology of 21st-Century Machine Learning, in: dies.: *The Cultural Life of Machine Learning. An Incursion into Critical AI Studies*, Cham 2021, 1–29.

2 Vgl. *Artificial Intelligence and the Future*, Podcast, 55 Min., mit Demis Hassabis, Strachey Lecture 26.2.2016, Computer Science Series, Oxford University, podcasts.ox.ac.uk/artificial-intelligence-and-future (30.11.2022).

3 Vgl. Philippe Sormani: Logic-in-Action? AlphaGo, Surprise Move 37 and Interaction Analysis, in: Jean-Yves Beziau, Arthur Buchsbaum, Christophe Rey (Hg.): *Handbook of the 6th World Congress and School on Universal Logic*, Cham 2016, 355–357, 355; Michael Mair u. a.: Just what are we doing when we're describing AI? Harvey Sacks, the commentator machine, and the descriptive politics of the new artificial intelligence, in: *Qualitative Research*, Bd. 21, Nr. 3, 2021, 341–359.

4 Vgl. Noortje Marres: Co–existence or displacement. Do street trials of intelligent vehicles test society?, in: *The British Journal of Sociology*, Bd. 71, Nr. 3, 2020, 537–555.

Intelligenz der ‹neuen› KI auseinandersetzen können und sollten.

Unsere Diskussion ist entlang folgender Fragen gegliedert: Zunächst kehren wir zu einer klassischen Kritik zurück, die Soziolog*innen und Anthropolog*innen an KI geäußert haben, nämlich zu der Behauptung, dass die der KI-Entwicklung zugrunde liegende Ontologie und Epistemologie rationalistisch und individualistisch ist – und als solche durch blinde Flecken bezüglich der sozialen, situierten oder situativen Einbettung von KI gekennzeichnet ist.[5]

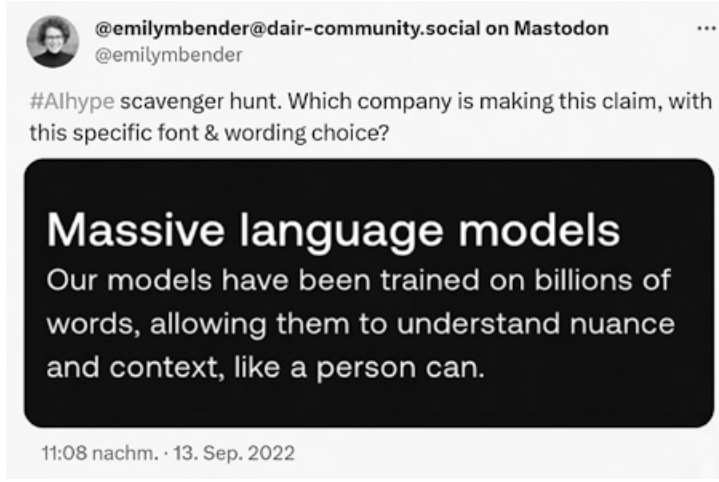

Abb. 1 **Tweet von Emily Bender: #AIhype scavenger hunt, 13.9.2022, Screenshot**

Wir fragen: Ist die Leistung und Bewertung maschineller Intelligenz in den gegenwärtigen Fällen von KI-Tests in sogenannten realen Environments weiterhin durch ein solches ‹soziales Defizit› gekennzeichnet? Als Nächstes befassen wir uns mit der Frage, ob und wie Techniksoziologie und Medientheorie KI-Tests in realen Settings situativ erklären können. Hier knüpfen wir an die Arbeiten des französischen Soziologen Louis Quéré an, indem wir uns mit der Frage beschäftigen: Was können wir aus den heutigen realen Tests von KI hinsichtlich der Verteilung von Kapazitäten (im Sinne handlungsermöglichender Ressourcen und Affordanzen) zwischen Artefakten, Umgebungen und Kontext in rechenintensiven Praktiken lernen?[6] Abschließend erörtern wir die Auswirkungen auf unser methodologisches Engagement für die Situation in der soziotechnischen KI-Forschung: Ist es sinnvoll, sich bei der Respezifikation der maschinellen Intelligenz weiterhin auf die Beschreibung von Situationen zu verlassen?[7]

Frage 1

Setzt die ‹neue› KI weiterhin auf die Ausklammerung von Situationen? Erfordert die Leistung und Bewertung der maschinellen Intelligenz weiterhin das Übersehen von Situationen und die Ausklammerung von sozialem Leben?

Noortje Marres Um diese erste Frage zu erörtern, möchte ich mit einer besonderen Herausforderung beginnen, die sich durch den Anstieg von lernbasierter, datenintensiver KI stellt. Eine besondere Herausforderung aus der Perspektive der Wissenschafts- und Techniksoziologie sind meiner Meinung nach die reißerischen Behauptungen, die in den letzten Jahren über die Fähigkeiten dieser Systeme zu situativer Intelligenz und kontextuellem Lernen aufgestellt wurden. Hier ein Zitat des Informatikers Percy Liang aus seiner Einführung zu einem Workshop der Stanford University über sogenannte große *foundation models*:

5 Lucy A. Suchman: *Plans and Situated Actions. The Problem of Human-Machine Communication*, Cambridge (UK), 1987; Lucy A. Suchman: *Human-Machine Reconfigurations. Plans and Situated Actions*, 2. Aufl., Cambridge u. a. (UK) 2007, doi.org/10.1017/CBO9780511808418; Susan Leigh Star: The Structure of Ill-Structured Problems. Boundary Objects and Heterogeneous Distributed Problem Solving, in: Les Gasser, Michael H. Huhns (Hg.): *Distributed Artificial Intelligence*, London 1998, 37–54, doi.org/10.1016/B978-1-55860-092-8.50006-X.

6 Vgl. Louis Quéré: The still – neglected situation?, in: *Réseaux. Communication – Technologie – Société*, Bd. 6, Nr. 2, 1998, 223–253.

7 In dieses anfängliche Gespräch wurden Fragen und Kommentare aus dem Publikum eingestreut, die für die vorliegenden Zwecke transkribiert wurden, einschließlich einer abschließenden Reflexion über das Turn-Taking in (Online-) Konversationen, und die sich in Gänze im vollständigen englischen Working Paper zu dieser Konversation finden.

> Foundation models […] are based on a decades old idea, self-supervised learning, meaning that, based on lots of raw data, you make up predictive exercises […] like weight training to develop the muscle for pattern recognition […]. Doing this at scale, results in the emergence of new capabilities, and one thing GPT can do is context learning, generalization to new tasks.[8]

In den 1980er und 1990er Jahren behaupteten die Science and Technology Studies (STS) sowie Forschungen über KI in diesem Bereich, dass Kontext und Situation genau das sind, was automatisierte Systeme *nicht* berücksichtigen können: Dies war die Kritik, die Lucy Suchman[9] und andere an älteren «expert systems» übten und die Suchman in Bezug auf die Robotik als «unreconstructed form of realism in roboticists' constitution of the ‹situation›» bezeichnete, selbst wenn «references to the situated nature of cognition and action have become ‹business as usual› within AI research».[10] Heute hat es den Anschein, dass dieses Argument für KI aus mehreren Gründen nicht mehr ganz zutrifft. So wird die Idee, die ‹neue› KI sei zu situativer Intelligenz oder kontextuellem Lernen fähig, auch von STS-Wissenschaftler*innen aufgegriffen. Nehmen wir Harry Collins, der in seinem Buch *Artifictional Intelligence* argumentiert:

> The problem for AI is how it can develop social abilities, because this would require the full embedding of AI in language speaking social communities in society. The problem of AI is the problem of engagement with social context. AI engagement with the Internet has resolved this to some extent.[11]

Collins scheint wie Liang der Meinung zu sein, dass das Trainieren von Computermodellen anhand großer Mengen von Daten aus dem Internet das Problem des sozialen Kontexts in der KI (zumindest teilweise) gelöst hat.

Was ich jedoch in Bezug auf diese Art von Behauptungen hervorheben möchte, ist, wie unglaublich selektiv sowohl ein Soziologe wie Collins als auch ein Informatiker wie Liang in ihren Definitionen dessen sind, was als relevanter Kontext oder als relevante Situation für die KI gilt. Kontext scheint als das vorherige Auftreten einer bestimmten Äußerung oder Interaktion in textlichen oder visuellen Daten definiert zu sein, wodurch die meisten der Merkmale, die Soziolog*innen als entscheidende Attribute von Situationen ansehen (Verkörperung, Materialität, Kopräsenz), ausgeschlossen werden. Darüber hinaus scheint ihr Begriff des Kontexts diejenigen Arten von Situationen auszugrenzen, die durch die Einführung von KI in die Gesellschaft selbst hervorgerufen werden. Dafür gibt es einige Beispiele, ebenso wie für das Fehlschlagen von Prozessen des kontextuellen Lernens. Viele werden den Fall des rassistischen Online-Chatbots Tay kennen, der sich radikalisierte, nachdem er von der *4chan*-Community trainiert wurde und uns das Schauspiel eines rassistischen Online-Chatkurses bescherte, woraufhin er umgelernt und schließlich aufgelöst wurde.[12]

Dieser Fall zeigt uns, dass es eine Menge Kontext gibt, der von der KI nicht berücksichtigt wird. Mehr noch, er deutet darauf hin, dass in der kontextuellen

8 Videoaufzeichnung: Workshop on Foundation Models: Welcome and Introduction, mit Percy Liang, am Centre for Research on Foundation Models (CRFM) der Stanford University, 23/24.08.2021, crfm.stanford.edu/workshop.html (30.11.2022).

9 Vgl. Suchman: *Plans and Situated Actions*.

10 Lucy A. Suchman: Feminist STS and the Sciences of the Artificial, in: Edward J. Hackett u. a. (Hg.): *The Handbook of Science and Technology Studies*, 3. Aufl., Cambridge (MA) 2007, 139–164, hier 148f.

11 Harry Collins: *Artifictional Intelligence. Against Humanity's Surrender to Computers*, Cambridge 2018, 162.

12 Vgl. Sanjay Sharma, Phillip Brooker: #notracist: Exploring racism denial talk on Twitter, in: Jessie Daniels, Karen Gregory, Tressie McMillan Cottom (Hg.): *Digital Sociologies*, Bristol 2016, 463–485.

Auseinandersetzung mit der KI eine Menge *perverser Sozialisation* stattfindet bzw. Sozialisation scheitert. Meines Erachtens ist die feministische Kritik an der Blindheit der Maschinen gegenüber der Welt so aktuell wie eh und je. In der Tat haben Soziolog*innen und die Sozialforschung im weiteren Sinne die Aufmerksamkeit auf diese Art von problematischen Wechselwirkungen zwischen Artefakt, Environment und Kontext gelenkt, wie etwa bei der perversen Sozialisation von Tay. Sie haben gezeigt, wie toxische Online-Umgebungen, die das akzeptierte soziale Umfeld für das Training großer Sprachmodelle zu sein scheinen,[13] monströse Formen der KI mitproduzieren. Sie haben auch gezeigt, wie die Inszenierung einer öffentlichen Situation mit einem rassistischen Chatbot durch Microsoft zur Verwirklichung dieser Perversion beigetragen hat,[14] eine ‹schlechte› Situation, die aus der kontextuellen Blindheit der KI-Entwickler*innen resultierte.

Was sich aus diesen Studien jedoch leider noch nicht als Erkenntnis ergeben hat: Diese Fälle von perverser Sozialisation stellen eine Herausforderung für die akzeptierten Definitionen dessen dar, was in der Informatik als kontextuelles Lernen von KI gilt. Statt eines kritischen Verständnisses der methodischen und konzeptionellen Herausforderungen, die sich ergeben, wenn Computersysteme *im* sozialen Leben und *als* soziales Leben operieren, sehen wir oft, dass diese Art von Fällen wie des rassistischen Chatbots als *ethische Probleme* gerahmt werden. Dies hat zur Folge, dass die gesamte situative Logik, wie ein Bot rassistisch wird, außerhalb des erkenntnistheoretischen Rahmens der KI-Entwicklung und -Forschung liegt. Situative KI, wie sie ein*e Soziolog*in verstehen würde, d. h. KI-Systeme, die in sozialen Situationen agieren, findet daher in der KI-Entwicklung und -Forschung immer noch relativ wenig Beachtung. Und deshalb halte ich es für sehr wichtig, dass die Sozial- und Kulturwissenschaften weiterhin darauf bestehen, dass *situative Handlungen* von KI methodische und konzeptionelle Probleme mit KI aufzeigen können.

Ich möchte noch ein weiteres Zitat anführen, um zu zeigen, wie leicht die ‹Löschung› der situativen Logik in der KI-Entwicklung und -Forschung geschehen kann. Es stammt aus einem Experteninterview, das ich vor Kurzem mit einem Ingenieur für vernetzte autonome Fahrzeuge geführt habe. Ich fragte ihn nach der Komplexität der Situationen, denen automatisierte Fahrzeuge auf der Straße begegnen. Seine Antwort war:

> We believe quite strongly that the complexity in driving on the roads, is not in observing where the road ends and the pedestrian crossing starts and where the traffic lights are, these static tasks of identification have been solved for a long, long time actually. The real challenge is modeling the behavior of other so called agents, because they're not necessarily totally rational or perfect or identical.[15]

Es mag also den Anschein haben, dass die Art von Interaktionen, die in situ auftreten, bei der Entwicklung von automatisierten Fahrzeugsystemen berücksichtigt werden. Der Ingenieur erklärte dann aber, dass diese Situationen durch

13 Vgl. Emily M. Bender u. a.: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in: FAccT '21: *Proceedings of the 2021 ACM conference on Fairness, Accountability, and Transparency*, 1.3.2021, 610–623, doi.org/10.1145/3442188.3445922.

14 Vgl. Gina Neff, Peter Nagy: Talking to bots: Symbiotic agency and the case of Tay, in: *International Journal of Communication*, Bd. 10, 2016, 4915–4931.

15 Dieses Zitat stammt aus einem Expert*inneninterview vom 28. Mai 2022, einem von zwölf Interviews mit britischen Expert*innen für vernetzte und automatisierte Fahrzeuge (Connected and Automated Vehicles, CAV), die ich [Noortje Marres] zwischen 2022 und 2023 geführt habe.

die Festlegung der statistischen Eigenschaften der Regelbefolgung vollzogen werden können (*can be dealt with*). Er fuhr fort:

> [T]hey still do follow rules, rules of the road, but probably more fundamentally statistical properties, based on experience. And the way we learn these rules is through ... as children or as young adults … is observing the patterns of how vehicles move. And what we are implicitly learning is physics.[16]

Seiner Ansicht nach sind die statistischen Modelle in den Gehirnen von Individuen der eigentliche Gegenstand dessen, was ein intelligentes System simulieren muss, um eine Situation auf der Straße erfolgreich zu navigieren. Dass eine Situation erst durch die *Interaktion* zwischen Akteur*innen vor Ort zustande kommt – diese Tatsache wird ausgeklammert. Ich denke also, dass wir weiter darauf bestehen müssen und in Erinnerung rufen sollten, dass Situationen interaktiv und kontextabhängig vollzogen werden; wofür meiner Meinung nach Kontingenzen da sind. Philippe wird dazu noch mehr sagen.

Philippe Sormani Genau, aber um zunächst auf die erste Frage einzugehen, die wir uns für heute gestellt haben, möchte ich mit zwei konzeptionellen Bemerkungen beginnen und dann einige empirische Beispiele vorstellen, an denen ich in den letzten Jahren gearbeitet habe.

Zunächst einmal finde ich den Begriff maschinelle Intelligenz nach wie vor interessant, denn er erinnert einerseits an die gemeinsame Grundlage von dem, was oft als KI oder ‹gute altmodische KI› (*good old-fashioned AI*) diskutiert wird – also Top-down-Programmierung, regelbasiert –, und andererseits an aktuelle Formen des maschinellen Lernens und insbesondere des Deep Learnings, bei dem Muster in großen Datensätzen erkannt und auf dieser Grundlage Vorhersagen und Wahrscheinlichkeiten berechnet werden. Aber es gibt eine Gemeinsamkeit, und der Begriff maschinelle Intelligenz fasst dies recht gut zusammen – ein wirklicher Eisberg verborgener Grundannahmen, sodass man regelbasiertes Verhalten beispielsweise auf ‹Code› reduzieren kann (was auf Alan Turing zurückgeht, ebenso wie der Begriff der maschinellen Intelligenz). Dominique Cardon u. a. weisen darauf hin, dass aktuelle Formen des maschinellen Lernens oft in Begriffen der KI angepriesen und beworben werden, ungeachtet ihres forschungspolitischen Zwecks Mitte der 1950er Jahre, und zwar um maschinelles Lernen als brauchbaren Forschungszweig in diesem Bereich zurückzustufen.[17] Die sich daraus ergebende Kontroverse wirft jedoch die Frage nach den Gemeinsamkeiten – den Gemeinsamkeiten zwischen sogenannter ‹symbolischer KI› und maschinellem Lernen – auf und danach, wie diese kunstvoll eingesetzt werden – z. B. in und als Teil einer Demonstration von Technologie.[18]

Die Frage «Erfordert die Leistung und Bewertung der maschinellen Intelligenz weiterhin die ‹Löschung› von Situationen und die Ausklammerung von sozialem Leben?» würde ich dahingehend näher bestimmen, dass ich sie nicht nur als Ja/Nein-Frage formuliere, sondern so umformuliere, dass sie die Frage

16 Ebd.

17 Vgl. Dominique Cardon, Jean-Philippe Cointet, Antoine Mazières: La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle, in: *Réseaux. Communication, Technologie, Société*, Bd. 5, Nr. 211, 2018, 173–220, doi.org/10.3917/res.211.0173.

18 Vgl. z. B. Philippe Sormani: Remaking Intelligence? Of Machines, Media, and Montage, in: *Tecnoscienza. Italian Journal of Science and Technology Studies*, Bd. 13, Nr. 2, 2022, 57–85.

nach dem Wie und Warum der Demonstration und Bewertung von maschineller Intelligenz einschließt. *Wie* und *warum* klammern Technologiedemonstrationen und -bewertungen oftmals das, was in der Frage als Situationen und soziales Leben identifiziert wird, aus?[19]

Abb. 2 **AlphaGo-Exhibition-Match, Aja Huang (links), gegenüber Lee Sedol (rechts), aus dem Dokumentarfilm *AlphaGo* (Regie: Greg Kohs, USA 2017), Orig. i. Farbe**

Was die empirischen Beispiele angeht, so konzentriere ich mich derzeit auf *edtech in interaction* und darauf, wie seit langem bestehende Vorstellungen von maschineller Intelligenz in die Auseinandersetzung der Teilnehmer*innen mit *edtech* im Klassenzimmer einfließen (*edtech* steht für Bildungstechnologie, in der Regel digital). Abgesehen davon war ich vor fünf Jahren nicht der Einzige, der die Aufregung um die ‹neue› KI bemerkte, unter anderem weil sie *deep (machine) learning* als KI anpries. Die Fälle, an denen ich zu arbeiten begann, waren zunächst öffentliche Demonstrationen von Technologien mit dem Etikett KI, wie z. B. das AlphaGo-Exhibition-Match im Jahr 2016, bevor ich Videomaterial von Testfahrten (mit SmartShuttles) und nun *edtech* in pädagogischen Experimenten (mit verschiedenen Bildungsrobotern) untersuchte. Bevor ich in jedem Fall auf die Frage nach dem Warum zurückkomme, möchte ich kurz auf die Frage nach dem *Wie* eingehen:

Als öffentliche Demonstration eines hochentwickelten KI-Systems wurde im März 2016 das AlphaGo-Exhibition-Match aus Seoul übertragen, bei dem das System gegen Lee Sedol, den damaligen südkoreanischen Go-Champion, antrat. *Wie* wurde das AlphaGo-Exhibition-Match zunächst als ‹Symmetrie-Spektakel› zwischen AlphaGo und Lee Sedol inszeniert (vgl. Abb. 2)? Zweitens interessierte ich mich für SmartShuttle-Testfahrten, ein Interesse, das ich zusammen mit Jakub Mlynář weiterverfolgt habe. Auf der PostBus-Website wurde der SmartShuttle als «first intelligent bus in the world» vorgestellt.[20] Daher wieder die Frage: *Wie* wurden die Straßen – der Bus, andere Fahrzeuge, Fußgänger*innen usw. – so installiert und inszeniert, dass der Bus als intelligent erscheinen konnte? Und für die ‹Mars-Mission› als Klassenexperiment, mit der ich mich kürzlich befasst habe: *Wie* wurde sie inszeniert, sodass Schüler*innen im Klassenzimmer die Möglichkeit hatten, jeden Roboter von der Erde aus zu steuern? Ich werde später auf einige der technischen Einzelheiten zurückkommen. In der Zwischenzeit wollen wir uns damit befassen, inwiefern die erwähnten Demonstrationen von Technologie Situationen des sozialen Lebens ausklammern oder sogar ‹löschen›.

Eine Möglichkeit, dies zu tun, besteht darin, genauer zu betrachten, wie Kontingenzen gehandhabt werden und wie, als Konsequenz dieser Handhabung, die

19 Natürlich könnte unsere zweite Frage auch ähnlich umformuliert werden: Wie und warum ist die reale Welt heute so beschaffen, dass KI-Tests in ihrem Rahmen stattfinden können, ohne dass eine ‹problematische› Situation – ein Notfall, ein Zwischenfall oder ein Unfall – entsteht, sondern dass stattdessen ein Gefühl der Normalität, eine alltägliche Szene, erhalten bleibt?

20 Artikel auf dem Unternehmensblog der PostBus Ltd.: Autonomous driving, ohne Datum, *postauto.ch/en/about-us-and-news/innovation/autonomous-driving* (16.5.2023).

eine oder andere Version von maschineller Intelligenz in Erscheinung tritt, ähnlich wie bei der Spezifizierung von reflexiv-konstitutivem Experimentieren im Labor – also wie bei der «work of making an experiment work».[21] Einen Ausgangspunkt bietet eine Differenzierung, die Harold Garfinkel getroffen hat: die vorläufige Unterscheidung dessen, was er einerseits als «standing contingencies» und andererseits als «local contingencies» benennt.[22] Wir könnten erstere als «manifestly standing contingencies» bezeichnen, die an eine bestimmte Praxis gebunden sind – sagen wir, ein Go-Exhibition-Match –, im Unterschied zu den lokal produzierten Kontingenzen,[23] also Kontingenzen, die sich im Verlauf der Handlung ergeben, wo es immer diesen unerwarteten oder schwer zu erwartenden Verlauf gibt, zu dem die Teilnehmer*innen beitragen und mit dem sie konfrontiert werden (auch bekannt als Situationen des sozialen Lebens).

In Bezug auf das AlphaGo-Exhibition-Match gibt es mehrere Merkmale, die als *standing contingencies* bezeichnet werden können und die mit dem Go-Spiel und seiner Inszenierung als ein solches Exhibition Match verbunden sind. Erstens werden die Spiel- und Kommentator*innenräume gezeigt, nicht aber der Kontrollraum, geschweige denn die Recheninfrastruktur, die für den Ablauf des Matchs erforderlich ist. Ein zweites Beispiel stammt aus dem Dokumentarfilm *AlphaGo*[24] und ist der Moment, in dem die PR-Verantwortliche sagt: «We need somewhere to put AlphaGo under here» (TC 00:26:23). Gemeint ist der AlphaGo-Laptop, den sie dann unter dem Spieltisch platziert, an dem der Profispieler Lee Sedol gegen das Programm antritt, wobei die Züge von Aja Huang, einem AlphaGo-Teammitglied, ausgeführt werden. Und ein drittes Beispiel wäre die Situation, als Demis Hassabis, der CEO des Unternehmens hinter dem AlphaGo-Programm (DeepMind), Sedol anruft, um ihn zum Exhibition-Match gegen das Programm einzuladen, was ebenfalls im Dokumentarfilm gezeigt wird. Während dieses Anrufs erscheint hinter Hassabis das Whiteboard, auf dem notiert ist, wie die Veranstaltung ablaufen soll. Wie bei den ersten beiden Beobachtungen wird dies im Film jedoch auch nicht weiter ausgeführt. Zusammenfassend können diese Aspekte als ein Ensemble von (*manifestly*) *standing contingencies* in Bezug auf diese Demonstration oder dieses Exhibition-Match gesehen werden, und ihre lokale Handhabung lässt die soziale Situation, auf die sich die Demonstration stützt, teilweise aus dem Blickfeld verschwinden.

Ich möchte nun noch kurz auf die Frage des *Warums* eingehen. Und das ist auch eine Frage, die Phil Agre, ein kritischer Computerwissenschaftler, Garfinkel im Hinblick auf sein Interesse an lokalem Kontingenzmanagement und experimenteller wissenschaftlicher Praxis stellte. «What are the contingencies for?»[25] Wozu soll man sie auflisten? Und Garfinkel würde diese Frage als weitere Kontingenz in die Liste aufnehmen, denn die Warum-Frage wird mitunter auch für die Beteiligten, und *last but not least* auch für Soziolog*innen, relevant. Und in dieser Hinsicht gibt es meines Erachtens drei Punkte, die zumindest angeführt werden sollten. Erstens die Frage der *Zurechenbarkeit* (*accountability*): Wie treten

21 Vgl. Harold Garfinkel: *Studies of Work in the Sciences*, London u. a. 2022.

22 Ebd.

23 Vgl. Ebd., 90f., Anm. 34.

24 *AlphaGo – The Movie*, Regie: Greg Kohs, Moxie Pictures, USA 2017. Der komplette Film wurde am 13.3.2020 auf den YouTube-Account *Google DeepMind* hochgeladen: youtu.be/WXuK6gekU1Y (1.11.2022).

25 Garfinkel: *Studies of Work in the Sciences*, 24, vgl. weiter ebd., 39–55.

die Dinge in Erscheinung? Wie werden sie gezeigt? Was sind die Konsequenzen? Zweitens die Frage, wie der Kontext angesichts der Eigendynamik einer Situation gehandhabt wird. Dies steht in Zusammenhang mit Technologiedemonstrationen und dem Punkt, den Joseph Lampel vor 20 Jahren in einem Aufsatz mit dem Titel «Show-and-Tell: Product Demonstrations and Path Creation of Technological Change» dargelegt hat.[26] Der Punkt ist folgender: Damit eine Technologie vertrauenswürdig erscheint, darf sie nicht zu detailliert dargestellt werden. Und natürlich ist das, was präsentiert wird, sorgfältig ausgearbeitet. Dies ist ein weiterer Aspekt – wie wird kritisches Nachforschen behindert oder abgelehnt, während «commitment evaluation routines»[27] für die vorgestellte Technologie im Vordergrund stehen und gefördert werden. Und drittens – das ist ein weiteres klassisches Problem – die Gefahr der Verdinglichung: Wenn wir das lokale Kontingenzmanagement aus der Betrachtung ausklammern, einschließlich der dramaturgischen Verwendung der Unterscheidung zwischen Vorder- und Hinterbühne, dann laufen wir Gefahr, die maschinelle Intelligenz *ex nihilo* zu verdinglichen. Umgekehrt erhält man einen klareren Sinn für die soziale Situation und ihren lebendigen Verlauf, die typischerweise bei Demonstrationen von Technologie vorausgesetzt und weitgehend ausgelassen werden. Zurück zu Noortje.

Frage 2

Was lässt sich aus der Verteilung von Kapazitäten zwischen Artefakten, Environment und Kontext in gegenwärtigen realweltlichen KI-Testsituationen lernen?[28]

N.M. Lassen Sie uns eine sehr allgemeine Frage aufgreifen, nämlich die, wie Forscher*innen zwischen den Rollen von Artefakten, Environment und Kontext bei der Untersuchung von computerbasierten Praktiken unterscheiden sollten, zu denen wir die rechenintensiven Arrangements von KI zählen können. Mit dieser Frage berufen wir uns auf die Arbeit des französischen Soziologen Louis Quéré, der in einem Artikel mit dem Titel «The still – neglected situation?» die Bedeutung dieser Unterscheidung – zwischen Artefakten, Environment und Kontext – hervorhob und sich gegen die Vermischung dieser Begriffe aussprach.[29]

Ich möchte zwei Gründe hervorheben, warum die Fragen «Was gehört zum Artefakt?», «Was gehört zum Environment?» und «Was gehört zum Kontext oder zur Situation?» von besonderer Relevanz für die Untersuchung der zeitgenössischen KI und des KI-Testens sind.

Zum ersten ist oft darauf hingewiesen worden, dass bei Demonstrationen von KI spektakuläre Fähigkeiten – Fähigkeiten, die auf Intelligenz hindeuten – der Maschine selbst zugeschrieben werden, dass diese bei näherer Betrachtung aber abhängig von aktiven Beiträgen aus der Umgebung der Maschine sind – einschließlich der Menschen, die ihr ordnungsgemäßes Funktionieren

26 Joseph Lampel: Show-and-Tell: Product Demonstrations and Path Creation of Technological Change, in: Raghu Garud, Peter Karnøe (Hg.): *Path Dependence and Creation*, Mahwah 2001, 303–327, 304, doi.org/10.4324/9781410600370.

27 Ebd.

28 Vgl. Quéré: The still – neglected situation?

29 Ebd.

Abb. 3/4 **Coventry Autodrive – Versuch mit autonomen Fahrzeugen, Screenshots aus Live-Videos des *Coventry Telegraph*, 15.11.2017 (29.5.2023)**

gewährleisten, wie Philippe gerade dargelegt hat.[30] Die Untersuchung von KI-Demonstrationen und -Tests ist von dieser analytischen Verpflichtung geprägt: Durch die Untersuchung von KI-Tests in sozialen Umgebungen, wie z. B. bei den Tests von selbstfahrenden Autos auf der Straße (vgl. Abb. 3 und Abb. 4), können wir feststellen, wie die der KI zugeschriebenen Urteils- und Entscheidungsfähigkeiten in situ zustande kommen. In diesem Zusammenhang ist vor allem die obige von Quéré aufgeworfene Frage von Bedeutung. Wenn wir Demonstrationen von KI untersuchen, können wir fragen: Welchen Beitrag leisten das Artefakt, das Environment und der Kontext jeweils in Bezug auf die Ausführungen der KI?

Im Fall einer Testfahrt mit autonomen Fahrzeugen im Stadtzentrum von Coventry, die im November 2017 von einer Reporterin mit ihrer Handykamera aufgezeichnet wurde (vgl. Abb. 3 und Abb. 4), können wir also fragen: Welchen Beitrag leisten die Sicherheitspylonen, die wir neben dem Fahrzeug sehen? Welchen Beitrag leisten die Zäune? Was ist mit der Beschriftung des Fahrzeugs? Und was ist mit dem Sicherheitspersonal, das im Hintergrund steht? Bei der Untersuchung von KI-Tests auf der Straße – und es gibt viele andere Fälle, wie etwa die Gesichtserkennungstechnologien, die in den letzten Jahren auf Bahnhöfen und von der Polizei im Vereinigten Königreich erprobt wurden – können und sollten wir uns somit mehr auf teilnehmende Beobachtung verlassen und weniger auf die formalen Beschreibungen von KI-Technologien, die von der Informatik, dem Technologiesektor und der Industrie für den öffentlichen Konsum produziert werden. Nehmen wir z. B. die öffentliche Ankündigung der Coventry-Testfahrt:

> The UK's largest trial to date of connected and autonomous vehicles technology on public roads explor[es] the benefits of having cars that can ‹talk› to each other and their surroundings – with connected traffic lights, emergency vehicle warnings and emergency braking alerts. The vehicles rely on sensors to detect traffic, pedestrians and signals but have a human on board to react to emergencies. *The trials are testing a number of features and most importantly seeking to investigate how self-driving vehicles interact with other road users.*[31]

30 Vgl. Bruno Latour: Social Theory and the Study of Computerized Work Sites, in: Wanda J. Orlikowski u. a.: *Information Technology and Changes in Organizational Work*, Boston 1996, 295–307, doi.org/10.1007/978-0-387-34872-8_18.

31 Ryan Tute: Driverless vehicle testing on public roads hailed as landmark moment, in: *Infrastructure Intelligence*, 24.11.2017, infrastructure-intelligence.com/article/nov-2017/driverless-vehicle-testing-public-roads-hailed-landmark-moment (22.4.2023), Herv. NM.

Die anderen Verkehrsteilnehmer*innen werden hier als Fußgänger*innen dargestellt, ohne dass Zäune, Pylonen oder Sicherheitspersonal im Bild zu sehen sind.

Es überrascht vielleicht nicht, dass ein intelligentes Navigationssystem von Grund auf nicht die Erwartungen erfüllt, die derartige Werbebeschreibungen wecken – etwa dass es in der Lage sei, sein Verhalten mit anderen Verkehrsteilnehmer*innen in situ zu koordinieren. Stattdessen ist dieser KI-Versuch von Verwirrung (*confusion*) geprägt und stützt sich auf alle möglichen Requisiten und Elemente im Raum, einen Zaun mitten auf der Straße, Passant*innen, die nicht ganz verstehen, was vor sich geht, und Wachpersonal, das den Verkehr regelt. Was sagt uns diese etwas konfuse Situation? Ich denke, sie sagt uns, dass wir bei der Betrachtung von KI-Versuchen, die sich als Teil des gesellschaftlichen Lebens entfalten, auf eine ganz andere Art von Situation stoßen, eine, die sich deutlich von stereotypischen Situationen unterscheidet, die in KI-Versuchsdemonstrationen inszeniert werden.

In den Straßen von Coventry konnte ich keine Systeme ausfindig machen, die versuchten, als Menschen oder soziale Akteur*innen eingestuft zu werden. Stattdessen fand ich eine hochgradig künstliche Situation vor, eine, in der nicht klar ist, ob und wie die Technologie funktioniert, in der die Akteur*innen ziemlich desorientiert wirken (und dies auch werden) und in der sehr stark auf Requisiten zurückgegriffen wird. Ich spreche hier also den oben erörterten Punkt der STS an, dass die Intelligenz, die der Maschine zugeschrieben wird, in Wirklichkeit durch ein ganzes Kollektiv von Akteur*innen während der Situation erreicht wird (Zaun, Sicherheitspersonal, Pylonen, Schilder usw.).

Aber eine Testsituation wie diese wirft auch ein Licht auf einen vielleicht weniger offensichtlichen Punkt: Die Einführung von KI in die Gesellschaft bringt Modifikationen des gesellschaftlichen Environments mit sich, die meiner Meinung nach die Unterscheidung zwischen Artefakt und Environment, wie sie Quéré vornimmt, stören und bis zu einem gewissen Grad untergraben. Die Performance jenes selbstfahrenden Fahrzeugs als Artefakt wird durch Eingriffe und Veränderungen des Settings erreicht. In Abbildung 3 gibt es auch andere, weniger sichtbare Veränderungen des Environments, die im Rahmen der Versuche in Coventry vorgenommen wurden: die Installation von Straßenrand-Einheiten – *roadside units* –, die Sensoren enthielten und Kommunikation zwischen Fahrzeugen ermöglichten, sowie die Verbesserung der Beschilderung auf der Straße, damit diese maschinell erkannt werden konnte.

Wenn wir also erstens über die Urteils- und Entscheidungsfähigkeit von Maschinen nachgedacht haben, dann ist das hier nun mein zweiter Punkt: Die Untersuchung von KI-Tests in situ macht für mich deutlich, dass es zwischen diesen verschiedenen konstitutiven Elementen des sozialen Lebens – Artefakt, Environment und Kontext – zu *leakage* kommen kann, mit anderen Worten: dass deren Beziehung von Durchlässigkeit gekennzeichnet ist. Quéré grenzt diese Elemente voneinander ab, indem er argumentiert, dass ein Environment scharf von dem Kontext unterschieden werden muss, in dem sich Alltagserfahrungen entfalten:

> [A]n environment in itself has neither axes nor directions since we are the ones who set them in different ways; these settings give rise to an ‹environing experienced world›. […] *It is the orientation of experience that gets one from the environment to the situation*, because situations come under the register of the organization of experience, which is not the case of environments. Someone who is disoriented is still in an environment.[32]

Meine These ist, dass diese Unterscheidung zwischen Environment und Kontext, und vielleicht auch Artefakt, im Rahmen von realem Testen der heutigen KI sozio-materiell gesehen eine Rekonfiguration erfährt: Da die Einführung von KI in Gesellschaften die Einfügung von rechen- und datenintensiven Technologien (*devices*) in den Hintergrund des sozialen Lebens rückt und in der Tat die Modifikation infrastruktureller Environments in der Gesellschaft mit einbezieht, wird künstliche Intelligenz buchstäblich zur *Verwirklichung dieser Modifikation von Environments*. Das Artefakt kann ohne dieses modifizierte Environment nicht funktionieren, und es sind diese rechnerisch ausgestatteten Umgebungen, die die Navigation und die Kommunikation von Fahrzeug zu Fahrzeug ermöglichen, die eine entscheidende Rolle bei der Orientierung spielen.

Ich glaube, dass die Unterscheidung zwischen Artefakt und Environment bzw. Kontext in Gesellschaften mit KI zunehmend verwischt wird. Der Fall von KI-Tests in der Gesellschaft trägt dazu bei, dies zu verdeutlichen. Die Ausrichtung der Erfahrung, die Quéré unter dem Begriff der Situation zusammenfasst, ist genau das, worum es bei der Gestaltung des Environments geht: Sensoren im Setting leiten die Fahrzeuge, damit sie navigieren können; Zäune lenken die Wahrnehmung des Tests. Die Situation, so könnte man sagen, ist das, was durch die Einbettung rechenintensiver Systeme in das sozio-materielle Environment entsteht. Diese Systeme verändern die Bedingungen für soziale Routinen in den Settings. Sie verändern die Art und Weise, wie sich soziales Leben in ihnen entfalten kann.

Wir sollten jedoch beachten, dass die soziale Situation auf der Straße gestört, ja sogar desorientiert wird, wenn das Environment so modifiziert wird, dass es den Maschinen Orientierung bietet. Um auf die oben erwähnte Verwirrung zurückzukommen und dies etwas dramatischer auszudrücken: Solange der analytische Fokus weiter darauf liegt, wie das KI-System die Situation in realen Tests wie dem in Coventry bewältigt – wie es sich orientiert und mit einem Zusammenbruch (bzw. einer ‹Störung›) umgeht –, sehen wir nicht wirklich, wie soziales Leben durch die Einführung ebendieses Systems aktiv desorientiert wird und in manchen Fällen vielleicht sogar zusammenbricht. In solchen Testsituationen kann die Koordinierung der Interaktionen zwischen Fahrzeugen und Fußgänger*innen nicht mehr wie gewohnt ablaufen. Die restriktiven Maßnahmen, die die KI zum Funktionieren benötigt, machen eine normale Interaktion auf der Straße unmöglich. Und dieses Schema, bei dem die Erleichterung der maschinellen Orientierung zu Desorientierung einer breiteren sozialen Situation führt, wiederholt sich im größeren Maßstab. Ich denke dabei an die

32 Quéré: The still – neglected situation?, 288, Herv. NM.

zutiefst störenden und schädlichen Auswirkungen auf die Welt, die sich aus den fortgesetzten Investitionen in die Automobilität ergeben.

Dies ist ein weiterer Grund, warum wir unsere Darstellung von KI entnaturalisieren sollten, indem wir uns auf Situationen des Testens bzw. der Demonstration konzentrieren. Dieser empirische Fokus ermöglicht es uns insbesondere, die naturalistische Fiktion und die daraus resultierende Täuschung abzulehnen, die mit der Frage einhergeht, wie KI mit Pannen umgeht. Methodisch wird bei Collins suggeriert, KI gelte als intelligent, wenn ihr der Umgang mit Zusammenbrüchen gelingt.[33] Aber was ist eigentlich die Beziehung zwischen KI und Zusammenbruch? Ein naturalistischer Ansatz führt dazu, dass er die Orientierungslosigkeit, die Verwirrung, die Störung und den Zusammenbruch verschleiert, die als Folgen der Einführung von KI in das gesellschaftliche Leben entstehen. Die Sozialwissenschaft, die das Artefakt so behandelt, als ob es ein*e soziale*r Akteur*in wäre, nimmt die obigen Werbebeschreibungen als gegeben hin, anstatt die Situation zu analysieren und zu beobachten, was tatsächlich passiert, wenn KI in das soziale Leben eintritt.

Zusammenfassend lässt sich sagen: Das Testen von KI lädt dazu ein, die prägenden Unterscheidungen zwischen Artefakt, Environment und Kontext neu zu untersuchen. In der Tat wird es bei der Analyse von Gesellschaften mit KI zu unserer Aufgabe, zu untersuchen, wie die Kapazitäten zwischen Artefakt, Environment und Kontext bei der Implementierung von KI und der daraus resultierenden situativen Desorientierung umverteilt werden.

P.S. Vielen Dank, Noortje. Aus ethnomethodologischer Sicht interessiere ich mich erst seit (relativ) kurzer Zeit für KI-Technologien oder Technologien mit KI-Etikett, einschließlich ihrer Nutzung und Entwicklung sowie der Forschung darüber – z. B. für KI im Bereich der Bildungstechnologie (*edtech*), wenn auch nicht ausschließlich. Und natürlich bin ich damit nicht allein.[34]

In Bezug auf unseren Diskussionspunkt, die Frage 2, schließe ich mich den von dir [Noortje Marres] und David Stark vorgebrachten Argumenten für Kontinuität an. In dem Aufsatz mit dem Titel «Put to the test» verbindet ihr «expert-led testing *and* social experimentation».[35] Eine Trennung der beiden hingegen, so schreibt ihr es auch, «risks rendering invisible the testing situations that the sociology of testing should elucidate»,[36] was aber auch jede zeitgenössische Ethnomethodologie des Experimentierens betrifft.

Allerdings habe ich auch hier zwei konzeptionelle Vorbehalte gegenüber der Formulierung dieser Frage aus ethnomethodologischer Sicht. Und ich möchte einen dritten hinzufügen.

Erstens glaube ich nicht, dass die Verteilung von Kapazitäten und die Art und Weise, wie sie zugewiesen werden, ein guter Ausgangspunkt für eine Situationsbeschreibung sind. Wie bereits betont wäre es besser, damit zu beginnen, wie *Kontingenzen* in situ gehandhabt werden und wie sie offenkundig gehandhabt werden, damit sie beschrieben werden können – und sei es nur, um einen

33 Vgl. Collins: *Artifictional Intelligence*.

34 In Bezug auf *edtech* in Interaktion kann ich [Philippe Sormani] auf lehrreiche Gespräche mit meinem Kollegium (Lausanne) verweisen, insbesondere mit Marc Audétat, Julien Bugmann, Farinaz Fassa, Guillaume Guenat und Audrey Hostettler, da wir uns eingehend mit der *Relokalisierung maschineller Intelligenz* befasst haben.

35 Noortje Marres, David Stark: Put to the Test: For a New Sociology of Testing, in: *The British Journal of Sociology*, Bd. 71, Nr. 3, 2020, 423–443, hier 428, Herv. PS.

36 Ebd.

«distributed essentialism»[37] im Sinne der Akteur*innenidentifikation zu vermeiden. Das heißt, eine ethnomethodologische Beschreibung – eine Beschreibung, die sich auf die alltäglichen Methoden praktischer Aktivitäten, ihre besondere Verständlichkeit und ihren situierten Vollzug konzentriert – endet dort, wo ein soziologisches Modell der Erklärung von Handlungen beginnt oder beginnen könnte. Ein solches Modell setzt insofern das voraus, was die Beschreibung liefert – eine erkennbare Situation, eine wahrnehmbare Konfiguration, eine sich entfaltende Interaktion (in Bezug auf Akteur*innen, die identifiziert werden, Kapazitäten, die verteilt werden, Hindernisse, die ausgemacht werden usw.).

Mein *zweiter* Vorbehalt bezieht sich auf die Vorstellung von rechenintensiven Praktiken, die als konstitutiv für Computer-basierte Artefakte aufgefasst werden – seien es Programme, Programme mit Sensoren oder Programme mit Sensoren und Aktoren, um Johnsons und Verdicchios Unterscheidung zu folgen.[38] Auch hier erscheinen weder rechenintensive Praktiken noch computerbasierte Artefakte als guter Ausgangspunkt für die Beschreibung, da sie (als Konzepte) zu eng gefasst sind. Man läuft Gefahr, die beteiligten *kulturellen* Artefakte zu übersehen, die Art und Weise, wie sie in situ produziert werden und welche *Praktiken* sie konstituieren, seien es nun ‹rechenintensive› oder andere Arten von Praktiken. In dem Bild, das Noortje vorhin gezeigt hat (vgl. Abb. 3 und Abb. 4), sehen wir zuerst *ein* Auto oder Fußgänger*innen, d. h., wir sehen sie als kulturelle Artefakte oder verkörperte Akteur*innen und nicht als Computer-basierte oder rechenintensive Akteur*innen. Wenn diese Artefakte und Akteur*innen umgekehrt als «sociotechnical assemblages»[39] zu betrachten sind – *embodied* und *entangled*, *never pure*, *never alone* –, welche Art von Assemblagen sind sie dann, wie werden sie eingesetzt und wie erfüllen sie ihre Funktionen in situ? Und wie sieht dann die situierte Praxeologie eines (wenn nicht des) «cultural life of machine learning» aus?[40]

Ein möglicher *dritter* Vorbehalt ergibt sich aus Quérés wichtiger Unterscheidung zwischen «environment, context *and* situation».[41] Ich stimme dir zu, Noortje, dass KI-Systeme, damit sie als Teil einer Testfahrt funktionieren können, erfordern, dass das Environment modifiziert wird und dass diese instrumentelle Modifikation sich als problematisch, wenn nicht sogar störend für den regulären Verkehr erweisen kann. In diesem Sinne wird die soziale Situation auf der Straße gestört. Aber auch eine gestörte Situation bleibt eine soziale Situation. Quérés Unterscheidung ist also nützlich, um auf den Unterschied zwischen einem spezifischen *Kontext* oder einer selektiven Kontextualisierung eines Straßenenvironments einerseits und der Art und Weise, wie sich eine alltägliche Verkehrs*situation* (unabhängig von ihren Teilnehmer*innen, ihrem Kontext oder ihrer Umgestaltung) in ihren vielfältigen Einzelheiten tatsächlich entfaltet, andererseits hinzuweisen. Eine Situation kann sich per se als irreduzibel auf ein Environment oder einen Kontext erweisen, so wie es David Sudnow beschreibt: «[A] lost newcomer *finds* himself suddenly in the midst of a Mexico City traffic circle.»[42]

37 Steve Woolgar: What Happened to Provocation in Science and Technology Studies?, in: *History and Technology*, Bd. 20, Nr. 4, 2004, 339–349, hier 344, doi.org/10.1080/07341510420003043 21.

38 Vgl. Deborah G. Johnson, Mario Verdicchio: Reframing AI Discourse, in: *Minds and Machines*, Bd. 27, 2017, 575–590.

39 Göde Both: *Keeping Autonomous Driving Alive: An Ethnography of Visions, Masculinity and Fragility*, Opladen 2020.

40 Roberge, Castelle: Toward an End-to-End Sociology of 21st-Century Machine Learning.

41 Quéré: The still – neglected situation?, 243, Herv. PS.

42 David Sudnow: *Ways of the Hand. The Organization of Improvised Conduct*, London u. a. 1978, 30, Herv. PS.

Lassen Sie mich nun auf meine drei Beispiele zurückkommen: *edtech* in Interaktion bei der ‹Mars-Mission› im Klassenzimmer, ein SmartShuttle, der einen spontanen Zwischenstopp einlegt, und AlphaGo auf der Bühne. Wenn dies empirische Beispiele sind, wofür sind sie dann Beispiele?

Ich habe zuvor das lokale Management praktischer Kontingenzen als ein Phänomen von ethnomethodologischem Interesse erwähnt. Wie richten die Teilnehmer*innen bestimmte Geräte, Systeme oder Infrastrukturen ein, nutzen sie und interagieren mit ihnen, sodass man sagen kann, ihr Betrieb weise KI-Fähigkeiten auf? Wie tun sie das, und zwar auf erkennbare Weise? Und was sind die Kontingenzen – die «locally lived constraints»[43] –, auf die sie dabei stoßen, mit denen sie zu kämpfen haben und/oder die sie unterlaufen? Jeder der drei genannten Fälle bietet eine empirische Antwort auf die aufgeworfenen Fragen, eine Antwort, die jedes Mal durch die angetroffenen Kontingenzen Quérés Unterscheidung zwischen Kontext und Situation zum Ausdruck bringt.

Im zweiten Spiel des AlphaGo-Exhibition-Matchs im März 2016 kam es zu jenem 37. Zug, der als ein ganz besonderer Zug des KI-Systems bezeichnet wurde: «AlphaGo somehow taught the world completely new knowledge»[44]. Diese an die journalistischen Medien gerichtete Mitteilung, wie sie ursprünglich auf der DeepMind-Website veröffentlicht wurde, fasst auch den Überraschungsmoment der beiden englischsprachigen Spielkommentatoren zusammen, als sie die Besonderheit von AlphaGos Zug 37 während des Spiels zuerst bemerkten.[45] Die Diskrepanz zwischen ihrem Kommentar, der bestimmte Spielzüge antizipiert hatte, und dem tatsächlichen Spielzug zeigt, dass die Spielkommentatoren Mühe hatten, sich einen Reim darauf zu machen. Die Diskrepanz wurde also zu ihren *locally lived constraints* für ihre anschließende Analyse, ganz zu schweigen von Lee Sedols Antwortzug (wie er nach der Partie bestätigte). Im Sinne von Quéré markiert die angetroffene Kontingenz den Unterschied zwischen einem erwarteten, wenn auch projizierten Kontext und der sich tatsächlich entfaltenden Situation.

Ein ähnlicher Fall konnte bei einer Testfahrt mit dem SmartShuttle beobachtet werden, bei dem der Ausgangspunkt ebenfalls eine lokal angetroffene Diskrepanz war, bei der etwas geschah, das die Teilnehmer*innen in diesem Setting nicht erwarten sollten oder konnten. In diesem Fall wurden die Kapazitäten des Shuttles von einem Fahrgast kommentiert, der sagte: «It is very good how it does, going around things. I'm amazed that it goes through the narrow places so easily – yeah.»[46] In diesem Moment hielt der Bus abrupt an, und der Betreiber kommentierte diesen Halt als ein immer wiederkehrendes Problem: «[A]nd here she [the van, *la navette* auf Französisch] does each time the same [thing] to us.» Das heißt, der Shuttle blieb einfach ohne erkennbaren Grund stehen, zumindest für den Betreiber. Auch hier scheint der Kontext, dieses Mal eine Testfahrt, anders zu sein als die Situation, die die Teilnehmer*innen vorfinden und mit der sie umgehen müssen.

43 Garfinkel: *Studies of Work in the Sciences*, 23.

44 Zitat ursprünglich veröffentlicht auf dem DeepMind-Blog der Entwickler*innen als Eintrag «The story of AlphaGo so far». Link nicht mehr gültig, Zitat aber noch einsehbar in diversen Newsartikeln, z. B. Cynthia Harvey: Deep Learning and Artificial Intelligence, in: *Datamotion*, 14.6.2018, *datamation.com/applications/deep-learning-and-artificial-intelligence* (28.6.2013).

45 Die Analyse des Spielzugs Nr. 37 von den beiden englischsprachigen Kommentatoren Michael Redmond und Chris Garlock kann eingesehen werden im YouTube-Video: Move 37!! Lee Sedol vs. AlphaGo Match 2, hochgeladen von Account Daniel Estrada am 12.3.2016, *youtu.be/JNrXgpSEEIE* (28.6.2023). Für eine detaillierte Videoanalyse vgl. Philippe Sormani: Interfacing AlphaGo, in: *Social Studies of Science* (im Erscheinen).

46 Dieses und das folgende Zitat stammen aus der Videoaufzeichnung der besprochenen Testfahrt.

Und ein dritter Fall, mein derzeitiger Schwerpunkt: ‹Mars-Missionen›. In meinem Wahlkanton findet dieses Projekt in der Schule statt, wo Schüler*innen eingeladen werden, kleine mobile Roboter zu programmieren, mit denen sie sich virtuell auf dem Mars bewegen und eine Mission erfüllen können, wobei die Marsoberfläche an einer technischen Hochschule inszeniert und den Schüler*innen über YouTube gezeigt wird. Im Klassenzimmer durchbrechen dann *einige* der Schüler*innen mit ihren Kommentaren die allgemeine Begeisterung: «Aber Herr Lehrer, wir sind hier nicht auf dem Mars» (in Bezug auf den Videostream), «Wir werden den Wettbewerb verpassen» (in Bezug auf ihren Sporttag) und «Ich würde gerne filmen» (in Bezug auf die Kameraausrüstung).[47] Die Liste der Kontingenzen in situ ließe sich noch erweitern. Auch hier wird deutlich, dass sich die jeweilige Situation nicht auf einen bestimmten Kontext oder ein vorläufig kontrolliertes Environment reduzieren lässt (z. B. eine im Klassenzimmer inszenierte ‹Mars-Mission›).[48]

Es gibt also verschiedene Arten von Kontingenzen, die bewältigt werden müssen, wenn verschiedene Arten von maschineller Intelligenz bei verschiedenen Demonstrationen von Technologie, Testfahrten und/oder pädagogischen Experimenten inszeniert werden. Natürlich sind diese Demonstrationen, Versuche und Experimente auch so konzipiert, dass die Kontingenzen als Teil ihres praktischen Managements verschwinden, eines Managements, das sie funktionieren lässt, damit maschinelle Intelligenz sich manifestiert, sei es als rhetorischer Effekt, als Navigationsanforderung oder als pädagogische Aufgabe. Diesem beabsichtigten Verschwinden der alltäglichen Praktikabilität sind jedoch Grenzen gesetzt. In einem meiner Lieblingszitate formuliert Lucy Suchman dies wie folgt: «Lived practice inevitably exceeds the enframing moves of its own procedures of order production».[49] Okay, ich glaube, das war's von mir zu dieser zweiten Frage, und ich denke, wir sind wieder bei Noortje.

N.M. Das ist großartig, Philippe, ich werde zwei kurze Punkte ansprechen. Erstens den Umgang mit Kontingenz und vielleicht auch die Unmöglichkeit, kontingente Situationen zu bewältigen. Ich interessiere mich sehr für dieses Thema, zum Teil auch, weil ich im Moment Agnes Hellers sozialtheoretisches Werk *Can Modernity Survive?* lese.[50] Diese Sozialtheoretikerin, die sich an Georg Lukács orientiert, legt großen Wert auf die Kontingenz des Alltagslebens als das, was die Moderne in gewisser Weise auszeichnet: In der Moderne wird das Alltagsleben als kontingent erlebt, es ist eine Form des Lebens, in der Routinen und Praktiken untersucht und in Frage gestellt, tatsächlich getestet und modifiziert werden können, weil sie als kontingent erkannt werden, und das ist ein Schlüsselaspekt von Hellers Verständnis der Moderne, und in der Tat auch von ihrem Verständnis, warum wir wirklich daran arbeiten müssen, damit diese überleben kann. Also, *ja*.

Aber ich denke auch, und das ist mein zweiter Punkt, dass diese Frage der Kontingenz der Situation und die Verteilung der Kapazitäten innerhalb der

47 Zitate von Teilnehmer*innen der Mars-Mission, Feldnotizen PS.

48 Diese lokalen Gegebenheiten bieten wiederum unzählige pädagogische Möglichkeiten, wie ein Teammitglied an der örtlichen Universität für Lehrerbildung feststellte. Für das programmatische Argument vgl. Michael E. Lynch: Garfinkel's Studies of Work, in: Douglas W. Maynard, John Heritage (Hg.): *The Ethnomethodology Program. Legacies and Prospects*, Oxford, New York 2022, 114–138, doi.org/10.1093/oso/9780190854409.003.0004.

49 Suchman: *Feminist STS and the Sciences of the Artificial*, 193.

50 Agnes Heller: *Can Modernity Survive?*, Berkeley 1990.

Situation wirklich eng miteinander verbunden sind. Einer der Gründe, warum ich immer auf der Frage nach der Verteilung oder Umverteilung von Kapazitäten innerhalb einer Situation bestehe, ist, dass die Art und Weise, in der Kapazitäten innerhalb des Artefakts konzentriert oder konsolidiert werden – neben der Aufwertung der Maschine durch die Spezifizierung ihrer Kapazitäten als unglaublich, hochentwickelt, außergewöhnlich … –, einer der Effekte davon ist, dass die Kontingenz verschwindet. Es lässt die Maschine als unentbehrlich erscheinen, als diejenige, die die anstehende Aufgabe notwendigerweise ausführen muss – außergewöhnlich und daher unersetzlich zu sein, dies ist eines der Risiken. Die Feststellung, dass die Kapazitäten verteilt sind, bedeutet hingegen: Jede Verteilung der Kapazitäten ist in der gegebenen Situation kontingent und kann sich ändern. Ich würde also sagen, dass das Beharren auf der Kontingenz und das Beharren auf der Verteilung der Kapazitäten vielleicht gar nicht so sehr im Widerspruch zueinander stehen, wie du, Philippe, vielleicht meinst.

P.S. Vielen Dank, Noortje. Lass mich kurz antworten. Mir ist aufgefallen, dass du in der Tat damit beginnst, Quéré methodologisch zu lesen, und zwar in dem Sinne, dass du seine Frage dahingehend verstehst oder auffasst, wie Forscher*innen zwischen Artefakten, Environment und Kontext (wenn nicht sogar zwischen Kontext und Situation) unterscheiden *sollten*.[51] Und das war auch sein Plädoyer, sein Argument, da er der Meinung war, dass sie zu sehr in einen Topf geworfen werden. Dies galt zumindest zu der Zeit, als er schrieb, und hinsichtlich der Forschungssituation, die Quéré in den späten 1990er Jahren kommentierte, insbesondere in Bezug auf Arbeitsplatzstudien (in *Human-computer interaction* und *Computer-supported cooperative work*) und die objektorientierte Soziologie – die Akteur-Netzwerk-Theorie. In diesem Zusammenhang würde ich auch zustimmen, dass wir dies vielleicht nicht als methodologische Frage behandeln, sondern als Phänomen betrachten können: Wie wurde die obige Unterscheidung von Forscher*innen – Soziolog*innen, KI-Forscher*innen oder auch Teilnehmer*innen – in bestimmten Situationen getroffen? Wir sollten es als empirisches Phänomen betrachten, anstatt über Methodologie zu streiten, ganz zu schweigen von Ontologie.

Damit sind wir wieder bei der Frage nach der Verteilung der Kapazitäten, und ich frage mich: Geht es in deiner Argumentation um multiple Kausalitäten, also darum, wie verschiedene Fähigkeiten zu, sagen wir, einer laufenden Handlung beitragen? Vielleicht ist das eine zu starke Formulierung, aber sie erlaubt mir, einen Kontrast zu dem zu setzen, was Quéré meiner Meinung nach anstrebte. Zumindest auf seiner phänomenologischen Seite scheint sein Hauptinteresse nicht darin gelegen zu haben, wie Kapazitäten verteilt und zugeschrieben werden können, sondern vielmehr darin, wie eine Situation als verständlich hergestellt wird – als ein Ganzes, als eine Gestalt einer bestimmten Art, die es nur unter dieser Voraussetzung erlaubt, bestimmte Akteur*innen zu identifizieren, und zwar im Hinblick auf eine bestimmte Kontextualisierung. Es

51 Vgl. Quéré: The still – neglected situation?

ist also etwas, das vor der Verteilung von Kapazitäten und deren Zuschreibung an verschiedene Akteur*innen kommt. Aber das eine schließt das andere nicht aus – oder bedingt es typischerweise. Wie du sagtest, besteht die politische Gefahr darin, dass die Art und Weise, in der Kapazitäten verteilt werden, auf heikle Weise Kontingenzen verschwinden lässt, obwohl diese Gefahr vielleicht gerade der Zweck eines erfolgreichen Engineerings ist, zumindest in den typischen Begriffen der Ingenieur*innen!
[…]

Offene Fragen und Forschungsperspektiven

So, do we have a situation?
Nein, insofern immer noch ein Situationsdefizit in der Art und Weise besteht, wie KI konzipiert, implementiert und diskutiert wird – entgegen der Behauptung, dass maschinengestützte Systeme zu kontextuellem Lernen fähig sind.

Ja, insofern die Einführung von KI in das gesellschaftliche Leben kritische Momente, öffentliche und politische Situationen hervorruft, die im öffentlichen Diskurs unterbestimmt bleiben.

Wie geht es also weiter?
Ein Ansatz besteht darin, die Testsituationen der KI wiederherzustellen und schließlich neu zu definieren, wobei die Begriffe Experiment und Experimentieren in den Prozess eingebettet und erweitert werden. Dieses Gespräch hat erste Ansätze in diese Richtung gebracht, andere sind schon seit einiger Zeit im Gange[52] oder müssen noch artikuliert werden, insbesondere in Bezug auf KI und die zeitgenössischen Varianten des maschinellen Lernens. In diesem Sinne sind unser Gespräch und das längere Working Paper eine Einladung zu weiterer Ausarbeitung und kritischem Austausch, sowohl on- als auch offline.

52 Vgl. z. B. Tanja Bogusz: *Experimentalism and Sociology: From Crisis to Experience*, Cham 2022; Georgina Born, Andrew Barry: Art-Science: From Public Understanding to Public Experiment, in: dies. (Hg.): *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences*, London u. a. 2013, 247–272; Noortje Marres, Michael Guggenheim, Alex Wilkie (Hg.): *Inventing the Social*, Manchester 2018.

Die englische Langfassung des Gesprächs ist im Mai 2023 erschienen: dx.doi.org/10.25819/ubsi/10332. Wir [Noortje Marres und Philippe Sormani] danken Johannes Schick für die Organisation und Carolin Gerlitz für die Moderation des Gesprächs sowie den Teilnehmer*innen für ihre Teilnahme und Beiträge. Das vorliegende Gespräch wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 262.513.311 – SFB 1187 «Medien der Kooperation». Philippe Sormanis Überlegungen wurden zudem vom Projekt «Relocating Machine Intelligence» (SNF ProjektNr. 407740_1187541) inspiriert und werden dieses weiter inspirieren. Übersetzung und Redaktion: DeepL, Inga Schuppener, Sebastian Gießmann.