



Teststation in Rüttenscheid (Essen), Oliver Heise: *symptom*, Folkwang Universität der Künste, September 2022

TEST

Einleitung in den Schwerpunkt

Als OpenAI am 30. November 2022 die Version 3.5 seines Sprachmodells ChatGPT freischaltet, beginnt ein nahezu beispielloser Test der neuen Künstlichen Intelligenz (KI) durch die Öffentlichkeit. Innerhalb von fünf Tagen gewinnt ChatGPT eine Million neue Nutzer*innen. Im Januar 2023 greifen bereits über 100 Millionen User*innen auf die KI-Anwendung zurück. Flankiert wird diese rasante Entwicklung von zahlreichen journalistischen Artikeln und Social-Media-Postings, die den generativen Charakter des Sprachmodells und seine repräsentationalen Möglichkeiten ausgiebig testen. Vergleichbar dynamische Nutzungspraktiken waren zwar schon typisch für Bildgeneratoren wie Dall-E und Stable Diffusion. Dennoch hat der öffentliche Test von ChatGPT eine neue Qualität. Zum einen kommt es zu Beschleunigungseffekten bei der Implementierung maschinellen Lernens, deren öffentliche und kritische Diskussion noch aussteht. Zum anderen bestätigen diese Beschleunigungseffekte soziologische Diagnosen zum Aufstieg der Testgesellschaft, in der datenintensive Praktiken und Infrastrukturen des Testens zeitlich und räumlich ubiquitär werden.¹

Im selben Zug führt die wirtschaftliche Konkurrenz zwischen Open AI, einem maßgeblich von Microsoft finanzierten Unternehmen, und Alphabet/Google zu einer Dynamik, in der ständig neue KIs veröffentlicht werden. Dabei werden neue Interaktionsmöglichkeiten und das Management von Bias im maschinellen Lernen nicht mehr vorsichtig ausgetestet. Wurde ChatGPT noch vorab durch menschliche Arbeit in Kenia gegen Diskriminierungseffekte trainiert,² gibt es mittlerweile eine Fülle neuer, auch Open-Source-basierter Sprachmodelle, bei denen die kommerziell motivierte Delegation von Arbeit vorab gleich durch öffentliches, verteiltes Testen ersetzt wird.³ Teils reicht für deren Nutzung wie bei Metas KI-Modell namens LLaMA ein lokaler Rechner aus.

¹ Vgl. 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, doi.org/10.1111/1468-4446.12746.

² Vgl. Billy Perrigo: The \$2 Per Hour Workers Who Made ChatGPT Less Toxic, in: *Time*, 18.1.2023, time.com/6247678/openai-chatgpt-kenya-workers (18.5.2023).

³ Vgl. Sung Kim: List of Open Sourced Fine-Tuned Large Language Models (LLM), Beitrag im Blog *Geek Culture*, 30.3.2023, medium.com/geekculture/list-of-open-sourced-fine-tuned-large-language-models-llm-8d95a2e0dc76 (18.5.2023).

Digitale Medientechnologien wie *large language models* stellen damit den vorläufigen Höhepunkt einer Entwicklung des Testens dar, auf die die Medienwissenschaft bislang nur punktuell geantwortet hat.⁴ Zwar hat Walter Benjamin den Zusammenhang von industrialisierten Testregimen, Sport und filmischer Medienkultur bereits 1936 in einer wenig rezipierten Passage seines «Kunstwerk»-Aufsatzes brillant beschrieben. Doch seine These – «*Der Film macht die Testleistung ausstellbar, indem er aus der Ausstellbarkeit der Leistung selbst einen Test macht*» – ist kaum erschlossen worden.⁵ Vor der filmischen Apparatur, so Benjamin, wird die Praxis des Schauspielens zu einer Testleistung ersten Ranges. Sie testiert und ist mit der sportlichen Leistung verwandt. Die arbeitenden städtischen Massen suchen abends die Kinos auf und würdigen dort die Leistung der Filmdarsteller*innen,⁶ die für sie Revanche am täglichen bürokratisierten Test ihrer Körper und Sinne nehmen. Benjamins Sensibilität für die industrielle, sportliche und administrative Formierung menschlicher Wahrnehmungsweisen diagnostizierte bereits eine Alltäglichkeit des Testens, für die die Medienästhetik des Films eine verwandte, aber nicht identische Respondenz bedeutete.

Für die Wahlverwandtschaft von körpertechnischen Medien, medizinischen, gesellschaftlichen und industriellen Testregimen interessiert sich die rezente Medienforschung erneut. Sie schließt hierfür an Medienanthropologie, Wissenschaftsgeschichte⁷ und Science and Technology Studies an. So sind insbesondere in den Sound Studies wegweisende Fallstudien entstanden, darunter Jonathan Sterne's Rekonstruktion der laboratorischen, teilöffentlichen Entwicklung des MP3-Formats. Bei Sterne wird das Testen von Sound-codierenden Algorithmen durch die Expertise geschulter Hörer*innen zu einer technologischen Performance, die in einer Testumgebung zu präzisen, theatralen Repräsentationen führt.⁸ Ausgehend von der wissenshistorischen Genealogie von Hör- und Klangtests in der Moderne halten Alexandra Hui, Mara Mills und Viktoria Tkaczyk fest, dass es sich bei modernen Tests um materialisierte Netzwerke handelt, die Praktiken, Ideen, Werte und Institutionen umfassen.⁹ Das Testen hat sich auf diese Art und Weise zu einer Kulturtechnik entwickelt, die in Herstellung, Marketing, Erhaltung, Reparatur, Sicherheit, Qualitätskontrolle und forensischer Analyse zur Geltung kommt.¹⁰

Mit dem vorliegenden Heft der *Zeitschrift für Medienwissenschaft* folgen wir dieser Diagnose und fragen, wie – und durch welche Mediatoren – sich Medien und Tests wechselseitig konstituieren. Besondere Aufmerksamkeit sollen dabei Politiken des Testens erfahren. Wir schlagen vor, Tests als offene Situationen zu verstehen,¹¹ in denen mit teils etablierten, teils sich erst während des Testens etablierenden Maßstäben soziotechnisch Entscheidungen getroffen werden. Mittels Tests wird das Neue und Unerwartbare adressiert. Manche Tests wollen es identifizieren, klassifizieren und kontrollieren. Andere wollen Möglichkeitsräume schaffen, in denen sich das Neue entfalten kann. Vorher nicht Vergleichbares oder Inkommensurables wird im Test vergleichbar oder kommensurabel gemacht. Für einen medienkulturwissenschaftlichen Begriff des Tests gilt: In

⁴ Vgl. den grundlegenden Entwurf von Avital Ronell: *The Test Drive*, Chicago 2005.

⁵ Walter Benjamin: Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit (zweite Fassung), in: ders.: *Gesammelte Schriften*, Bd. VII.1, hg. v. Rolf Tiedemann, Hermann Schweppenhäuser, Frankfurt/M. 1991, 350–384, hier 365. Herv. i. Orig.

⁶ Benjamin schreibt «Filmdarsteller» im Maskulinum.

⁷ Vgl. Ehler Voss (Hg.): *Mediality on Trial. Testing and Contesting Trance and other Media Techniques*, Berlin, Boston 2020; David Keller: *Person and Form. Eine Medien- und Wissenschaftsgeschichte der Persönlichkeitsdiagnostik*, Tübingen 2021.

⁸ Vgl. Jonathan Sterne: *MP3. The Meaning of a Format*, Durham, London 2012, 150.

⁹ Vgl. Alexandra Hui, Mara Mills, Viktoria Tkaczyk: *Testing Hearing. An Introduction*, in: Viktoria Tkaczyk, Mara Mills, Alexandra Hui (Hg.): *Testing Hearing. The Making of Modern Aurality*, Oxford, New York 2020, 1–19, hier 5.

¹⁰ Vgl. ebd., 3 f.

¹¹ Vgl. Ulrike Bergermann: *Test und Testosteron. Medien des Selbstversuchs*, in: Julia Bee, Nicole Kandier (Hg.): *Differenzen und Affirmationen. queer/feministische Positionen zur Medialität*, Berlin 2020, 63–89, hier 63.

den Mikroentscheidungen des verteilten und verteilenden Testens steht das Soziale selbst auf der Probe. Die wissensgenerierende Kraft des Tests beruht auf heterogenen Assemblagen von prüfenden und bewertenden Medien,¹² deren Eigensinn produktiv gemacht wird. Die in diesem Heft versammelten Beiträge verdeutlichen: Kein Test ohne Medien – kein Medium ohne Test.

Wer verfügt einen Test, wer hat Mitsprache beim Formulieren von Kriterien und Bedingungen? Ist eine Testsituation für alle Beteiligten überhaupt als solche erkennbar? Gerade im Kontext digitaler Plattformmedien ist dies meist nicht der Fall. Verfahren des datenbasierten Testens kennzeichnen technisierte und digitalisierte Lebenswelten – von spielerischen und situierten Praktiken, mit denen opake Medientechnologien angeeignet werden (*unboxing*, *You Tube as Test Society*¹³), bis hin zu großflächigen Tests, die ihrerseits vom Stresstest des Finanzsystems über die datenintensiven Sozialforschungen großer Plattformen, agile Entwicklungsstrategien (ehemals Perpetual Beta) bis zur Universalisierung von Technologien maschinellen Lernens wie ChatGPT reichen.

David Stark und Noortje Marres haben bereits vor dem Ausbruch der Covid-19-Pandemie die fortwährende Ausweitung von Testverfahren als eine Signatur von rechen- und datenintensiven *test societies* ins Auge gefasst und sich dabei aus der Perspektive der Science and Technology Studies auf das wechselseitige Verhältnis von Test und Gesellschaft(en) konzentriert. Denn Tests finden im 21. Jahrhundert nicht mehr allein in Labor und Werkstatt, Büro oder Studio statt, sondern in sämtlichen Lebens- und Arbeitsbereichen.¹⁴ Sie sind generativ, datenintensiv und verteilt. Sie schaffen neue Umgebungen, in denen sich Gesellschaften durch Test- und Sensorpraktiken konstituieren und auf permanentes Testen ausgerichtet werden. Testgesellschaften können kontrollgesellschaftliche Elemente enthalten, zeichnen sich aber gerade in der digitalen Gegenwart durch einen offeneren Umgang mit Testsituationen aus. Anstelle direkter Überwachung dominiert in ihnen das permanente datenintensive Monitoring. Neben der Frage, wie Tests Entscheidungen treffen und Zukünfte möglich oder unmöglich machen, problematisieren die Beiträge dieses Heftes auch die beteiligten Akteur*innen und ihre Handlungsmacht.

Ausweitung der Testsituation

Der Test stellt eine Situation der Überprüfung, Bewertung oder Entscheidung dar und fordert zugleich das Verständnis von Situationen heraus.¹⁵ Die hier versammelten Beiträge zum Testen adressieren heterogene Testsituationen, in denen Medien, Akteur*innen, Umwelten, Praktiken, aber auch unterschiedliche Temporalitäten zusammenwirken.¹⁶ In ihrer Zusammenschau zeigen sie ein breites Spektrum, in dem der Test mal mehr, mal weniger klar konturiert als Situation erkennbar wird – und dadurch seine politische Brisanz entfaltet.

In der Objektgeschichte des Rorschachtests scheint die Testsituation zunächst begrenzt und begrenzbare zu sein: DAVID BUCHELIS Beitrag zeichnet nach,

¹² Vgl. zur Relation von Rechtfertigung, Prüfen und Testen Jörg Potthast: *The Sociology of Conventions and Testing*, in: Claudio E. Benzecry, Monika Krause, Isaac Ariail Reed (Hg.): *Social Theory Now*, Chicago 2017, 337–360.

¹³ Vgl. die beiden Workshops zu *You Tube as Test Society* an der Universität Siegen: www.mediacoop.uni-siegen.de/de/veranstaltungen/workshop-youtube-as-test-society, 11.–12.12.2017 und www.mediacoop.uni-siegen.de/en/events/workshop-youtube-as-a-test-society-ii, 19.–20.4.2018 (19.5.2023).

¹⁴ Vgl. Marres, Stark: *Put to the Test*. Vgl. zur historischen Genealogie und zur Rolle des Testens für die Entstehung des modernen (Hör-) Subjekts: Hui u. a. (Hg.): *Testing Hearing*, 4 f.

¹⁵ Vgl. zum Situationsbegriff Sebastian Gießmann u. a. (Hg.): *Materiality of Cooperation*, Wiesbaden 2023.

¹⁶ Vgl. auch die klassische Unterscheidung zwischen «prospective, current, and retrospective testing of technologies» bei Trevor Pinch: «Testing – One, Two, Three ... Testing!»: *Toward a Sociology of Testing*, in: *Science, Technology and Human Values*, Bd. 18, Nr. 1, 1993, 25–41, hier 27 f., doi.org/10.1177/01622439930180010.



wie Rorschachtests als Mittel der Psychodiagnostik entwickelt und als ein etabliertes Testverfahren stabilisiert wurden, das sich durch Standards und rigorose Testbedingungen auszeichnet. Ihr Ziel ist es, ein <verräterisches Sprechen> zu generieren, durch das sich das zu testende Subjekt offenbart und klassifizierbar wird. Bucheli schildert den langen Prozess des Experimentierens und Stabilisierens ebenjener Standards und erläutert, dass der Rorschachtest in seinen prototypischen Stadien selbst Gegenstand zahlreicher Tests war. Zugleich erkundet er die massenmedialen und paratextuellen Nutzungen des Rorschachtests, in denen nicht nur die Bedingungen, sondern auch die Grenzen der Testsituation neu verhandelt werden.

Die Bilder von OLIVER HEISE, die die rezente Geschichte der pandemischen Testinfrastruktur dokumentieren, zeigen ein spannungsvolles Wechselspiel zwischen diskreten Testsituationen, ihren Infrastrukturen und ihrer raumzeitlichen Ausweitung. Offizielle Covid-19-Tests haben einerseits streng standardisiert abzulaufen. Andererseits fügen sie sich in Orte des Alltags wie Tankstellen, Baumärkte, Bahnhöfe oder Shoppingcenter ein (Abb. 1). Diese infrastrukturelle

Abb. 1 Testzentrum am Barmer Bahnhof (Wuppertal), Oliver Heise: *symptom*, Folkwang Universität der Künste, September 2022

Verschränkung zeigt, wie das Fortbestehen des Alltags in der Pandemie an ein Fortbestehen dauerhafter Testsituationen gekoppelt wurde, die möglichst niedrigschwellig gestaltet sein sollten. So ist die Testsituation zwar einerseits klar begrenzt, kann jedoch niemals abgeschlossen werden, da das Testergebnis lediglich als Momentaufnahme stabilisiert ist und zur weiteren Teilnahme am Alltag regelmäßig wiederholt und zertifiziert werden muss.¹⁷

In seinem Beitrag zum Radar erzählt CHRISTOPH BORBACH eine andere Geschichte der Ausweitung der Testsituation. Er zeigt, wie die Entwicklung und Nutzung dieser Technologie durch Kaskaden ineinandergreifender industrieller Tests gekennzeichnet ist. Diese zentrale medienhistorische Rolle des Tests ergibt sich, so Borbach, aus der Tatsache, dass menschliche Sinne die Funktionalität der Technologie nicht direkt überprüfen können. Da Störungen der Technologie sich anders als bei anderen Medien nicht direkt zeigen, wurden heterogene Tests und Simulationen notwendig, um das Funktionieren des Radars zu überprüfen und seine Nutzbarkeit zu stabilisieren. Zudem entstanden im Kontext der Radarproduktion Hochsee-Testsimulatoren, die realweltliche Bedingungen nachbauten und als Vorläufer virtueller Tests zu verstehen sind.

Die von Marres und Stark diagnostizierte Migration der Tests aus dem Labor in den Alltag demonstriert der Beitrag von GABRIELE SCHABACHER und SOPHIE SPALLINGER zu den sogenannten Zukunftsbahnhöfen der Deutschen Bahn, in denen in ausgewählten realen Bahnhöfen neue Technologien, Materialitäten und Praktiken öffentlich getestet werden. Angesiedelt an der Grenze zwischen Marketingmaßnahme und Innovationslabor *in the wild* dienen diese Testräume dazu, materielle und datenbasierte Gestaltungsoptionen von Mobilität zu explorieren, die sonst rechtlich nicht möglich sind, wie z. B. automatisierte Gesichtserkennung oder KI-basierte Gefahrenwarnung in Echtzeit. Sie tragen so dazu bei, diese Technologien zu normalisieren. Die Testsituation wird dabei nicht auf den ersten Blick als solche erkennbar. Was eigentlich getestet wird, steht im Vorfeld nicht immer fest – sind es Technologien der Gesichtserkennung, ihre soziale Akzeptanz, ihre eigene Testbarkeit? Im Zukunftsbahnhof treten zudem die umgebenden und umweltlichen Qualitäten¹⁸ von Tests sowie ihre körperlich-affektiven Dimensionen zu Tage, die zugleich Grundlage des permanenten Testens sind.

Eine andere Entgrenzung des Tests erörtert DANIELA HOLZER in ihrem Beitrag zu den Survival-Tests der YouTube-Realityshow *7 vs. Wild*. Auch hier hat der Test das Labor verlassen. Die Serie testet die Fähigkeit der Kandidat*innen, in der Wildnis zurechtzukommen und ohne die Infrastrukturen der Zivilisation zu überleben. Zugleich versteht Holzer die Serie als soziales Experiment neoliberaler Zukunftsszenarien, das die gesellschaftlichen Fähigkeiten zur Selbstoptimierung und Selbstkontrolle auf die Probe stellt. Im Kontext der Plattform YouTube ist die Serie zudem selbst der permanenten Testung ihrer Popularität durch plattformisierte Interaktionsformen wie Likes, Kommentare und Views ausgesetzt.

¹⁷ Vgl. zur Zertifizierung Lawrence Busch: *Standards. Recipes for Reality*, Cambridge (MA), London 2011, doi.org/10.7551/mitpress/8962.001.0001.

¹⁸ Zu umgebenden Medien vgl. Florian Sprenger: *Epistemologien des Umgebens. Zur Geschichte, Ökologie und Biopolitik künstlicher environments*, Bielefeld 2019, doi.org/10.1515/9783839448397.

Dieselbe Entgrenzung entfaltet mit datenintensiven Tests in virtuellen und auf KI basierenden Umgebungen eine weitere Dimension, die es zu erkunden gilt. Durch Virtualisierung können Objekte und Umgebungen an immer neuen Testsituationen teilhaben. Zugleich sinken die Testbedingungen und damit ihre Aushandlung zunehmend in Infrastrukturen hinab, die von wenigen Akteur*innen kontrolliert werden. STEFAN RIEGER legt dar, wie virtuelle und auf Simulationen basierende Tests als Öffnung von Möglichkeitsräumen, z. B. für die ressourcenschonende Entwicklung neuer Technologien, genutzt werden können. Virtuelles Testen dient dadurch weniger einer Vergewisserung, sondern erprobt noch nicht stabilisierte Technologien und Szenarien. Derart offene virtuelle Tests stellen jedoch stets ihre eigene Virtualität zur Disposition, wenn die Leistungen des virtuellen *testbed* mit realen Umgebungen verglichen werden.

In ihrem Gespräch zu KI-Tests widmen sich NOORTJE MARRES und PHILIPPE SORMANI der feministischen Frage der «Blindheit der Maschinen gegenüber der Welt»¹⁹ und erörtern, inwiefern Künstliche Intelligenz gegenwärtig zunehmend in der Lage ist, durch konstante Datenerfassung, u. a. mit Sensoren, kontext- und situationssensibel zu operieren. Sie problematisieren, dass fehlgeschlagene KI-Anwendungen – wie z. B. Microsofts Twitter-Chatbot Tay, dem von Nutzer*innen binnen 24 Stunden durch Interaktion eine rassistisch-sexistische Sprache antrainiert wurde – als ethisches Problem mangelnder Werte gerahmt werden statt als Ergebnis fehlender Kontextsensibilität. Ihrer These nach muss KI permanent beweisen, inwiefern der Kontextbezug glückt oder nicht. Gelingt der Situationsbezug, so verbleibt die Arbeit der KI unsichtbar in der computerisierten Infrastruktur – gelingt er nicht, wird die KI Gegenstand öffentlicher Kontroversen.

So verschiebt sich der Fokus innerhalb der Beiträge von der Analyse der Medien des Testens hin zum Testen der Medien selbst, insofern Medien samt ihrer Vermittlungsleistungen kritisch zur Disposition gestellt werden. Tests erscheinen so als grundlegender Bestandteil der kooperativen Verfasstheit digitaler, lernender Medien.

Datengesellschaft: eine potenziell permanente Testsituation

Quer durch alle Beiträge tritt die zentrale Rolle von Daten in Testpraktiken und -situationen hervor, die zugleich eine politische Dimension des Testens markiert. Tests basieren auf Daten und stellen selbst Daten her, die Teil weiterer Bewertungs- und Entscheidungspraktiken werden. Dies können Daten sein, die gezielt für die Testsituation erhoben werden, wie z. B. die Kameradaten zur Gesichts- und Gefahrenerkennung in Testbahnhöfen, die verschiedenen Messungen bei der Testung von Radartechnologien oder die sensorische Erfassung von Umgebungen durch autonom fahrende Autos. Doch es können auch solche Daten sein, auf die digitale Medien zugreifen können und mittels derer sie trainiert werden.²⁰

¹⁹ Vgl. den Beitrag von Marres und Sormani in dieser Ausgabe.

²⁰ Vgl. Kate Crawford: *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, London 2021, doi.org/10.2307/j.ctv1ghu45t; Jenna Burrell: How the Machine «Thinks»: Understanding Opacity in Machine Learning Algorithms, in: *Big Data & Society*, Bd. 3, Nr. 1, 2016, 1–12. doi.org/10.1177/2053951715622512; Massimo Airoldi: *Machine Habitus. Toward a Sociology of Algorithms*, Cambridge 2022.

21 Vgl. zur Rolle der Hochgeschwindigkeitskamera und elektrischer Sensoren im Falle des Crashtests Greg Siegel: *Forensic Media. Reconstructing Accidents in Accelerated Modernity*, Durham, London 2014, 143 f.

22 Vgl. David Gelperin, Bill Hetzel: The Growth of Software Testing, in: *Communications of the ACM*, Bd. 31, Nr. 6, 1988, 687–695, doi.org/10.1145/62959.62965; Glenford J. Myers, Corey Sandler, Tom Badgett: *The Art of Software Testing*, Hoboken, New Jersey 2012 [1979].

23 Vgl. David Stark: Testing and Being Tested in Pandemic Times, in: *Sociologica*, Bd. 14, Nr. 1, 2020 = Sonderband: *Society after COVID-19*, hg. von dems., Elena Esposito, Flaminio Squazzoni, 67–94, doi.org/10.6092/issn.1971-8853/10931.

24 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, doi.org/10.1111/1468-4446.12730.

25 Vgl. Digital Culture & Society, Bd. 9, Nr. 1, 2023: *Taming Digital Practices: On the Domestication of Data-Driven Technologies*, hg. v. Tim Hector, David Waldecker, Niklas Strüver, Tanja Aal.

26 Vgl. Jonathan Kropf, Stefan Laser (Hg.): *Digitale Bewertungspraktiken. Für eine Bewertungssoziologie des Digitalen*, Wiesbaden 2019.

27 Vgl. Carolin Gerlitz: What Counts? Reflections on the Multivalence of Social Media Data, in: *Digital Culture & Society*, Bd. 2, Nr. 2, 2016, S. 19–38, doi.org/10.25969/mediarep/941.

28 Vgl. zur multiplen Situiertheit Erhard Schüttelpelz: Vom Werkzeug zum Behälter, vom Behälter zum Medium: Die Ausweitungen des Körpers, in: Nina Engelhardt, Johannes F. M. Schick (Hg.): *Erfinden, Schöpfen, Machen. Körper- und Imaginationstechniken*, Bielefeld 2021, 25–72, doi.org/10.14361/9783839448373-002; Michael Dieter u. a.: *Multi-Situated App Studies: Methods and Propositions*, in: *Social Media + Society*, Bd. 5, Nr. 2, 2019, 1–15, dx.doi.org/10.1177/2056305119846486.

Die Daten- und Rechenintensität digitaler Medien schafft auf der einen Seite die Grundlage für permanentes Testen, auf der anderen Seite macht sie dieses aber auch unabdingbar, muss die Art und Weise, wie Daten verarbeitet werden, doch selbst immer wieder überprüft und angepasst werden. Daher ist das Testen und Kalibrieren der Medien des Testens ein wesentliches Element ihrer Geschichte.²¹ Tests bilden ein Grundprinzip der Gestaltung analoger wie digitaler Medien, wurden als Methode zur Softwareentwicklung zentral²² und prägten die Idee des Perpetual Beta im frühen Web 2.0. Sie entfalten ihre sozio-technischen Implikationen heute in verbreiteten Praktiken des A/B-Testing durch und auf digitalen Plattformen und sind charakteristisch für agile Entwicklungspraktiken.²³ Digital vernetzte Medien sind nicht darauf ausgerichtet, abgeschlossen zu werden, sondern werden anhand zunehmend entgrenzter Testreihen weiterentwickelt, ob als öffentliche Demonstration autonomen Fahrens,²⁴ als Tests generativer Sprachmodelle, in der Entwicklung neuer Apps oder dem häuslichen Testen neuer Technologien.²⁵ Dies beinhaltet Praktiken des öffentlichen Testens in und mit heterogenen Datenöffentlichkeiten, die sich bis zur kontroversen Aushandlung neuer Technologien maschinellen Lernens und der sensorischen Aufrüstung medialer Räume erstrecken.

Während die Bewertungs-, Valorisierungs- und Entscheidungspraktiken datenbasierter Medien zunehmend in den Fokus kritischer Medien- und Sozialforschung gerückt sind,²⁶ ist ihre Verschränkung mit testenden Praktiken noch nicht umfassend diskutiert. Tests, so unsere These, sind zentraler Bestandteil unserer Datengesellschaft. Sie multiplizieren die Situationen, an denen Daten teilnehmen,²⁷ und verlangen daher nach trans- und postsituativen Zugängen.²⁸

Die Analyse von Testgesellschaften fordert das Feld der Datenkritik²⁹ heraus, sich noch stärker der Verschränkung von online und offline, digitalen und physisch-materiellen Umgebungen, menschlichen und technischen Sensoren zuzuwenden. Je mehr Tests in die computerisierte Infrastruktur absinken, umso mehr operieren sie dabei zwischen den Modi *in the wild* und *in the lab*. Im Einsatz digitaler Zwillinge und Simulationen in der Echtzeittestung, z. B. bei der Optimierung von Fabriken, in der autonomen Mobilität oder in generativen KI-Modellen verschränkt sich das Labor in neuer Form mit der Gesellschaft. Damit bringen daten- und rechenintensive Testsituationen sowohl menschliche als auch technische Sensorien zusammen und verknüpfen sie auf neue Weise. In der Forschung hat dies zurück zu einer grundlegenden Frage³⁰ der Medienwissenschaft geführt: Wie erweitern, verschränken und ko-konstituieren³¹ sich menschliche Sinne durch Medien und – wie wir ergänzen wollen – durch ihre Tests? Im Gegensatz zu Instrumenten tragen Medien zu einer Denaturierung der Sinne bei und ko-konstituieren in konkreten Situationen, was Sinneswahrnehmung bedeutet. Dementsprechend trägt die Einbettung von Sensoren in Medien und Umgebungen zu einer Rekonfiguration der menschlichen Sinne bei. Der Einsatz sensorischer Medien, die Daten zu Beweugung im Raum, Temperatur, Position, Vitalfunktionen, Puls, Hautwiderstand,

Mimik oder Gestik erfassen und z. B. in der YouTube-Serie *7 vs. Wild* oder in den <Proxys> der Testung von Radartechnologie zum Einsatz kommen, bringt menschliche und technische Sensoren in Austausch. Tests sind in einem solchen sozio-technischen Verständnis dieses neuen «sozialen Sensoriums»³² eingebettet und gestalten somit die Politiken des Sinnlichen.³³ Die Frage lautet also nicht nur, inwiefern ubiquitäre Tests als solche wahrnehmbar sind, sondern – im Sinne der bei Benjamin angelegten Frage – wie Tests und ihre Medien selbst unsere Wahrnehmung und Sinneskonstitution prägen.

Vermengt werden in diesem Kontext präzisions-geschlossene Testergebnisse (richtig/falsch, ja/nein) mit offenen Tests, deren Ziel die Erkundung von Möglickeitsräumen darstellt. In den Zukunftsbahnhöfen tritt diese Verschränkung besonders zu Tage – so zielen die Tests zur Gesichtserkennung einerseits auf die Präzisierung der Erkennung unter realweltlichen Bedingungen ab. Andererseits werden sie auch dann als erfolgreich angesehen, wenn sie scheitern, da sie die soziale Akzeptanz und Gewöhnung der Technologien selbst testen. Sie tragen damit im Sinne von David Stark nicht zu einer geschlossenen *predictability* zukünftiger Szenarien, sondern zu einem Klima der *preparedness* bei.³⁴ Oder sie stellen, wie im Beitrag von Stefan Rieger analysiert, das Verhältnis zwischen dem Realen und dem Imaginativ-Virtuellen per se in Frage.

Digitale Medien werden potenziell zu einer permanenten, multiplen Test-situation, die es im Sinne einer kritischen Medienforschung sichtbar zu machen gilt. Während Marres und Stark argumentieren, dass Tests die Labore verlassen haben und nun mitten in der Gesellschaft stattfinden und diese testen, ergibt sich aus Sicht der Medienforschung ein weiterer Befund: Tests sind fundamentaler Bestandteil digitaler, datenbasierter und lernender Medien. Lernende Medien sind im Dauertestbetrieb und verschränken dabei die involvierten Akteur*innen, Umgebungen und Artefakte auf neue Weise. Ihr Ubiquitär-Werden und ihre Entgrenzung erfolgen jedoch weder naht- noch bruchlos, sondern werden zunehmend kontrovers.

Politiken des Testens

So hat insbesondere der Verlauf der Covid-19-Pandemie gezeigt, wie ältere biopolitische Testregime wieder aufgenommen und aktualisiert werden können. Zugleich wurden Test- (Abb. 2) und Impfzentren, PCR- und Schnelltests von neuen Apps zum Tracing und Tracking, Registrieren und Identifizieren³⁵ flankiert, die wiederum alte biosoziale Kontrolltechniken remediatisiert haben. Nahezu jedes Detail pandemischer (Test-)Infrastrukturen unterlag in demokratischen Gesellschaften fortwährender öffentlicher Prüfung und kontroversen Aushandlungsprozessen. Diese wurden insbesondere durch divergierende Interpretationen der pandemischen Statistik noch befeuert. In diesem Sinne lässt sich die Pandemie als fortwährende Lektion hinsichtlich der Situationsbezogenheit und existenziell notwendigen Kollektivität des Testens

²⁹ Vgl. Andrew Iliadis, Federica Russo: Critical Data Studies: An Introduction, in: *Big Data & Society*, Bd. 3, Nr. 2, 2016, doi.org/10.1177/2053951716674238.

³⁰ Vgl. Chris Salter: *Sensing Machines: How Sensors Shape our Everyday Life*, Cambridge (MA) 2022.

³¹ Vgl. Marshall McLuhan: *Understanding Media: The Extensions of Man. Critical Edition*, hg. v. Terrence W. Gordon, Berkeley 2011 [1964].; Sebastian Scholz: Sensing the «Contemporary Condition»: The Chronopolitics of Sensor-Media, in: *Krisis | Journal for Contemporary Philosophy*, Bd. 41, Nr. 1, 2021, 135–156, doi.org/10.21827/krisis.41.1.36967; Joseph Vogl: Becoming-Media: Galileo's Telescope, in: *Grey Room*, Nr. 29, Herbst 2007, 14–25, doi.org/10.1162/grey.2007.1.29.14.

³² Adam Yuet Chau: The Sensorial Production of the Social, in: *Ethnos*, Bd. 73, Nr. 4, 2008, 485–504, doi.org/10.1080/00141840802563931.

³³ Vgl. Jacques Rancière: *Die Aufteilung des Sinnlichen. Die Politik der Kunst und ihre Paradoxien*, hg. v. Maria Muhle, Berlin 2008.

³⁴ Vgl. Stark: *Testing and Being Tested in Pandemic Times*.

³⁵ Vgl. Sebastian Gießmann: Identifizieren, in: Heiko Christians, Nikolaus Wegmann, Matthias Bickenbach (Hg.): *Historisches Wörterbuch des Mediengebrauchs*, Bd. 3, Köln 2022, 134–162.



Abb. 2 Testzentrum am Bahnhof Steinbeck (Wuppertal), Oliver Heise: *symptom*, Folkwang Universität der Künste, September 2022, Orig. i. Farbe

auffassen. Dies umfasste die Orientierung an grafischen Verläufen («flatten the curve») und numerischen Repräsentationen (Ansteckungszahlen, R-Wert, prozentualer Anteil von Geimpften an der Bevölkerung), aber auch den Übergang der Testpraxis vom offiziellen Getestet-Werden zum eigenverantwortlich durchzuführenden Selbsttest.

Keine Ausweitung der Testzone kann Absolutheit beanspruchen. Selbst wenn Tests ubiquitär und sozial verpflichtend oder zu einem Teil von daten- und rechenintensiver Hintergrundkooperation werden, bleiben auf Basis von Tests getroffene Entscheidungen situationsbezogene Mikroentscheidungen. Diese irreduzible Verteiltheit des Testens – das selbst verteilenden Charakter hat – macht es zu einem fundamentalen Bestandteil von Gesellschaften, die biosoziales Testen kaum ohne die kontroverse Kombination von Datenpraxis und Sozialstatistik vollziehen können. Kaskadierendes Testen schafft Kontroll- und Normalisierungsinstanzen, in denen gegen einen bestehenden oder sich gerade etablierenden Standard getestet wird. Auch in miteinander verbundenen Tests wird ein Ergebnis anhand von Skalen und/oder auf wenige Elemente reduzierten Displays bewertet, um daraufhin – mittels einer Klassifikation oder Selbstklassifikation – entscheiden zu können. Diese biopolitische Dimension des Testens krieert durch ihre verteilte Situiertheit regulative Umgebungen, mit denen die Offenheit von Testsituationen eingeengt wird.

Demgegenüber verfügen Testgesellschaften aber auch über die politische Möglichkeit, durch exploratives, erprobendes Testen neue Möglichkeitsräume entstehen zu lassen. Die Konjunktur von Simulationen, *living labs*, virtuellen und realen *testbeds* lässt sich als ein Indiz hierfür auffassen, ebenso wie die Modifikationen menschlicher Körper im Selbstversuch,³⁶ aber auch das spielerische Gegen-Testen und die Analyse opaker Medientechnologien.³⁷ Unser Umgang mit der Allgegenwart von Testsituationen, die Daten- und Rechenintensität, Sensorik und maschinelles Lernen kombinieren, erfordert nicht nur in den Wissenschaften soziale Imagination, neue Fertigkeiten und Fähigkeiten, kreative Skills und kritische *data literacies*. Mit einem kurzen Ausprobieren von ChatGPT ist es jedenfalls nicht getan. Dies ist kein Test.

Danksagung

Wir möchten der Redaktion der *ZfM* für die organisatorische Unterstützung und den engagierten Austausch über das Schwerpunkt-Thema sehr herzlich danken. Den Gutachter*innen des Peer-Review-Verfahrens verdankt dieses Heft eine Vielzahl produktiver Vorschläge zur Verbesserung, die wir besonders würdigen wollen. Ausdrücklicher Dank gilt Inga Schuppener für die interne organisatorische und redaktionelle Unterstützung. Die Arbeit an dieser Einleitung wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 262513311 – SFB 1187 «Medien der Kooperation».

³⁶ «Der Test ist Medium neuer Körper und Subjektivierungen», so Ulrike Bergermann (Test und Testosteron, hier 88).

³⁷ Vgl. Maria Eriksson u. a.: *Spotify Teardown. Inside the Black Box of Streaming Music*, Cambridge (MA), London 2019, insb. die Kapitel zu Interventionen, 19–20, 69–70, 105–106, 139–140, 173–174, 193–194.

SEBASTIAN GIEßMANN, CAROLIN GERLITZ