

Von Gesichtswelblichkeit und Verweiblichung

Eine kritische Analyse der statistischen Physiognomik mit KI

Kris Vera Hartmann

Dieser Eindruck eines Weibes in männlicher Kleidung wird gesteigert durch den spärlichen Haarwuchs im Gesicht, der zudem bis auf ein Schnurrbärtchen rasiert ist, den tänzelnden Gang, das schüchterne, gezierte Wesen, die weiblichen Züge, den schwimmenden neuropathischen Ausdruck der Augen, die Spuren von Puder und Schminke, den stutzermäßigen Zuschnitt der Kleidung mit busenartig hervortretendem Oberkleid, die gefranste, damenartige Halsschleife und das von der Stirn abgescheitelte, glatt zu den Schläfen abgebürstete Haar. (Krafft-Ebing 1894: 278)

The gender atypicality of gay faces extended beyond morphology. Gay men had less facial hair, suggesting differences in androgenic hair growth, grooming style, or both. They also had lighter skin, suggesting potential differences in grooming, sun exposure, and/or testosterone levels. Lesbians tended to use less eye makeup, had darker hair, and wore less revealing clothes (note the higher neckline), indicating less gender-typical grooming and style. (Wang/Kosinski 2018: 251f.)

1. Einleitung

Diese beiden Zitate, die über 120 Jahre auseinanderliegen, sind auf den ersten Blick sehr unterschiedlich und doch sehr ähnlich. Das erste stammt aus der *Psychopathia Sexualis* von Krafft-Ebing (1894), einem Werk der frühen Sexualwissenschaft. Hier wurden Abweichungen von der heterosexuellen Norm als psychiatrische Pathologien betrachtet und zur Veranschaulichung dienten detaillierte, anekdotische Beschreibungen der Erscheinung (und des Verhaltens) von Patient*innen mit ›abweichendem Sexualverhalten‹. An dieser Stelle wird ein Mensch mit ›conträrer Sexualempfindung‹ beschrieben und die Abweichungen von dieser Norm werden vor allem als Weiblichkeit beschrieben. Obwohl sich die Sexualwissenschaft Ende des 19. Jahrhunderts als moderne Wissenschaft verstand, teilte sie wesentliche Methoden und Erkenntnisformen mit der vormodernen, teilweise metaphysischen Physiognomik.

Das zweite Zitat stammt aus einer Studie von Wang und Kosinski (2018), die belegen soll, dass künstliche Tiefe Neuronale Netze (Deep Neural Networks: DNN) aus Fotografien besser ablesen können, welche sexuelle Orientierung die abgebildete Person habe, als Menschen. Wang und Kosinski beschreiben hier Teile der Ergebnisse ihrer statistischen Analyse und kommen zu dem Schluss, homosexuelle Männer (und heterosexuelle Frauen) hätten eine stärkere ›Gesichtsweiblichkeit‹ als heterosexuelle Männer oder lesbische Frauen. Obwohl Kosinski und Wang völlig andere Methoden verwenden als Krafft-Ebing und sie auf eine Pathologisierung verzichten, werden in ihrer Studie ähnliche Vorstellungen über Geschlecht und sexuelle Orientierung sowie deren Erkennbarkeit aufgrund optischer Merkmale sichtbar.

Somit wird an dieser Stelle deutlich, wie sich mathematisch-exakte Methoden mit einer irrationalen Erkenntnisform wie der Physiognomik verbinden, so dass diese durch Verfahren des maschinellen Lernens wiederbelebt wird. Auch wenn die Sexualpathologie Krafft-Ebings sich selbst zur wissenschaftlichen Psychiatrie und nicht zur Physiognomik zählt, ähnelt ihre Erkenntnisform – insbesondere die detaillierte anekdotische Beschreibung von Personen – jenen Versuchen, die aus dem physiologischen Äußeren des Menschen auf deren Charakteristika schließen. Zurückgehend auf antike Bestrebungen, das Wesen, den Charakter oder andere Eigenschaften im Äußeren von Menschen zu erkennen, wurde die Physiognomik in der Neuzeit vor allem von Johann Caspar Lavater geprägt. Sein vierbändiges Werk *Physiognomische Fragmente* (1775) kann als Teil einer Anti-Aufklärung verstanden werden, die sich dem »Mystizismus und Wunderbaren« zugewandt hat (vgl. Gebarek 2003: 35). Lavaters Ansicht nach, lasse sich durch die Ebenbildlichkeit des Menschen zu Gott, das Göttliche im Äußeren des Menschen erkennen (vgl. ebd.: 40). Insbesondere die »Gesichtszüge des Menschen [ließen] auf dessen charakterliche Eigenschaften schließen [...], d.h. vor allem auf psychische Zustände, das Temperament sowie auf moralische und intellektuelle Fähigkeiten« (ebd.: 38). Die physiognomische Lehre wurde im 19. Jahrhundert weiterverfolgt, auch wenn der christliche Bezug in den Hintergrund trat und von eher naturreligiös-romantisierenden Vorstellungen abgelöst wurde. Sie stand damit weiterhin im Gegensatz zu der sich gleichzeitig etablierenden experimentellen Naturwissenschaft (vgl. Schmölders 2007: 31).¹

Während in der Physiognomik Lavaters vor allem mit anekdotischen Beschreibungen von Gemälden von Personen gearbeitet wurde (vgl. Schmölders 2007), vollzog sich mit der Fotografie eine Ver(natur)wissenschaftlichung und eine teilweise Integration in die Bereiche der medizinischen und psychiatrischen Diagnostik. Wichtige Vertreter dieser Strömung waren Jean-Martin Charcot, Francis Galton und Cesare Lambroso, die »einen szientifisch aufgeladenen Begriff von visueller Diagnose [forcierten], die den Außenseiter, den Verbrecher, den Kranken, aber auch den

1 Einen umfassenden Überblick, detaillierte Analysen und eine Quellensammlung zur Physiognomik bietet Schmölders (2007).

rassischen Mustertypus mit der nötigen eidetischen Evidenz vorstellen zu können versprach.« (ebd.: 32) Die physiognomische Anschauung mittels Fotografien wurde bis in die Psychiatrie des 20. Jahrhunderts weitergeführt, wie Regener (2010) in ihrer Studie zeigte.²

Wang und Kosinski (2018) versuchen mit Deep Learning Algorithmen menschliche Eigenschaften – hier die sexuelle Orientierung – aus Abbildungen von Gesichtern zu ›lesen‹ und schließen damit sowohl an die alte wie auch an die modernisierte Physiognomik an. Die Ergebnisse nehmen sie zum Anlass, naturalisierende Schlussfolgerungen über die Entstehung der sexuellen Orientierung zu ziehen. Zugespißt formuliert, versuchen Wang und Kosinski Homosexualität in den Gesichtern von Menschen zu verorten und erstellen sogar Abbildungen von ›archetypischen‹ homosexuellen und heterosexuellen Gesichtern (vgl. ebd.: 251). Sie wollen ihre Arbeit zwar explizit nicht in die Tradition physiognomischer Diskurse stellen, da diese zu Recht wissenschaftlich diskreditiert seien (vgl. ebd.: 246), sprechen im Weiteren jedoch davon, es gäbe vielleicht doch Verbindungen zwischen dem Äußeren und anderen Eigenschaften von Menschen. Menschen wären vielleicht nur nicht dazu in der Lage, diese zu erkennen:

Importantly, the low accuracy of humans when judging character from others' faces does not necessarily mean that relevant clues are not prominently displayed. Instead, people may lack the ability to detect or interpret them. (ebd.: 247)

Wo Menschen an ihrer begrenzten Auffassungsgabe und ihren Vorurteilen – wohl gemerkt nicht an den Prämissen – scheitern, scheinen sie dem Maschinalgorithmus zuzutrauen, erfolgreich zu sein. Dahinter lässt sich die Vorstellung vermuten, maschinelle Verfahren verfügen über bessere Methoden, größere Intelligenz oder schlicht eine größere Macht, die Wahrheit zu erkennen.

Aus der Perspektive der kritischen KI-Forschung³ wurde Wang und Konsinskis maschinelle Physiognomik bereits eingeordnet. Gemeinsam mit anderen Studien zur Erkennung und Prognose von bestimmten Eigenschaften aus Gesichtern⁴ werden sie z. B. von Campolo und Crawford als Teil einer ›Verzauberung‹ verstanden, die

2 Die photographische Physiognomik und mit ihr die Phrenologie, die Vermessung von Menschenschädeln und die damit zusammenhängenden Rückschlüsse auf entsprechende ›Rassen‹ wurde ebenfalls mit der Eugenik und der Rassenkunde im 20. Jahrhundert weiter betrieben, nicht zuletzt im Nationalsozialismus (vgl. Schmölders 2007: 31ff.). In den 1980er-Jahren analysierte z. B. Gould (1988) die – bis dahin anscheinend noch verbreitete – rassistische Schädellehre und mit ihr die falsche Zuschreibung von unterschiedlicher Intelligenz je nach ›Rasse‹ und kritisierte sie als wissenschaftlich nicht haltbar.

3 Für eine informations-ethische Perspektive auf die ›unheimliche‹ Rückkehr der Physiognomik vgl. Bendel (2018).

4 In anderen Studien wurde z. B. jüngst versucht, die politischen Orientierung aus Gesichtsabbildungen zu erkennen (Kosinski 2021; Rasmussen et al. 2022).

durch die Kombination von »Vorhersagegenauigkeit und mysteriösen oder unerklärlichen Eigenschaften« (Campolo/Crawford 2020; Übersetzung d. Vf.) entsteht. Das digitalisierte Verfahren erhält dabei die mythisch aufgeladene Konnotation der Übermenschlichkeit, die die Verantwortung der Wissenschaftler*innen und die deterministische Macht der sozialen Klassifikation und Kontrolle verdeckt (vgl. ebd.: 1). Keyes et al. (2021) sehen in der so verstärkten Bedeutung von KI als Quelle wissenschaftlichen Wissens eine Entwicklung, die dazu verwendet werden kann, gefährliche Ideen über die Identität von Menschen zu legitimieren (vgl. ebd.: 158). Diese Befürchtung lässt sich anhand der Studie von Wang und Kosinski leicht nachvollziehen: Auch wenn nicht-heteronormatives sexuelles Begehren zumindest in einigen Teilen der Welt akzeptiert oder toleriert wird, gibt es selbst in diesen Ländern reaktionäre Bewegungen und Bestrebungen, die durch eine Essentialisierung und vermeintliche Erkennbarkeit von Abweichungen gestärkt werden könnten.

Im Folgenden wird die Studie von Wang und Kosinski aus einer kritischen soziologischen Perspektive analysiert. Dabei wird aufgezeigt, dass die proklamierten Ergebnisse nicht haltbar sind, da das Verfahren stark von zuvor definierten Kategorisierungen abhängt, falsche Generalisierungen getroffen werden und (forschungs-)ethische Schwierigkeiten nicht ausreichend Beachtung finden. So kann gezeigt werden, dass die eingesetzten Algorithmen keineswegs exakte Instrumente zur ›Wahrheitsfindung‹ sind – wie unterstellt wird – und damit der pseudowissenschaftlichen Physiognomik näherstehen als beabsichtigt.

2. Kritik der maschinellen Erkennung der sexuellen Orientierung

Sozialwissenschaftliche Schwierigkeiten zeigen sich an verschiedenen Stellen der Studie von Wang und Kosinski. Zum besseren Verständnis soll diese, bzw. das zugehörige Paper, das im *Journal of Personality and Social Psychology* veröffentlicht wurde, an dieser Stelle kurz rekapituliert werden, bevor anschließend die kritischen Punkte anhand des Vorgehens der Autor*innen erläutert werden.

2.1. Aufbau der Studie von Wang und Kosinski

Nach der bereits erwähnten Bezugnahme auf die historische Physiognomik (vgl. Wang/Kosinski 2018: 246) wenden sich Wang und Kosinski verschiedenen psychologischen Studien zum Zusammenhang von äußerer Erscheinung und anderen Charaktereigenschaften zu und fragen, ob Menschen diese erkennen können (vgl. ebd.: 246f.). Auf diese Weise gelangen sie zur Pränatalen Hormontheorie (PHT), die als ihre Ausgangsprämisse verstanden werden kann. Die PHT führt die Entstehung von Homosexualität auf eine Unter- bzw. Überexposition von Androgenen im Mutterleib zurück (vgl. ebd.: 247). Diese biologische Erklärung resultiert in einem natura-

lisierten Determinismus: Die sexuelle Orientierung sei bereits vor der Geburt festgelegt. Das entspricht einer Festschreibung und biologischen Essentialisierung, die im Widerspruch zu z. B. sozialkonstruktivistischen Perspektiven, die die soziale und historische Gewordenheit und daher auch Wandelbarkeit solcher Identitätskonzeptionen annimmt, steht. Die PHT muss daher selbst als umstrittene Hypothese gelten. Neben dieser Einordnung der eigenen Forschung als einer technisch optimierten Physiognomik auf der einen Seite und dem naturwissenschaftlichen Rückgriff auf die Biologie auf der anderen Seite stehen Wang und Kosinski bereits hier im eingangs erwähnten Spannungsfeld von Irrationalismus und Rationalismus. Sie zitieren im Weiteren verschiedene (psychologische) Studien zur Messung der ›Gesichtswеiblichkeit‹ – der ›facial femininity‹ (ebd.) – von homosexuellen Menschen. Deutlich wird hier die bereits erwähnte Vorstellung von sexueller Orientierung als Verweiblichung oder Vermännlichung, auf die im späteren Verlauf noch weiter eingegangen wird.⁵

Die Studie ist in fünf Untereinheiten aufgeteilt:

- *Studie 1.a-c*: Erkennung und Extraktion der Gesichtselemente und -eigenheiten (facial elements and features) je nach sexueller Orientierung (Datenset mit Gesichtsbildungen von 35.326 Personen)
- *Studie 2 und 3*: Untersuchung der identifizierten Eigenheiten nach ›geschlechtsuntypischen‹ Merkmalen (›Gender-Atypicality‹)
- *Studie 4*: Vergleichsstudie mit menschlichen Vorhersagen
- *Studie 5*: Kontrollstudie mit unabhängigem Datenset (Facebook-Bilder)

2.2. Datenauswahl und -zusammenstellung

Bei der Datenauswahl entschieden sich Wang und Kosinski für Bilder, die auf einer – ungenannten – US-amerikanischen Datingplattform veröffentlicht wurden, da diese in großen Mengen gesammelt werden können, ökonomisch günstiger sowie repräsentativer seien als Bilder kleinerer Laborstudien. Da die Fotografien von den Nutzer*innen selbst aufgenommen wurden, wird ihnen außerdem ein großer ›ökologischer‹ Wert zugesprochen (vgl. ebd.: 248).

Dieses Verfahren wird in den Datenwissenschaften als Datensammlung ›in the wild‹ (Harvey/LaPlace 2021) bezeichnet und unterläuft einen Grundsatz der Forschung an bzw. mit Menschen. So ist es forschungsethisch nicht vertretbar, Menschen zu Forschungszwecken heranzuziehen, die nicht ihre informierte

5 Darüber hinaus begründen Wang und Kosinski ihre Studie auch damit, homosexuelle Menschen, politischen Entscheidungsträger*innen, Firmen und die Gesellschaft ›als ganze‹ vor den Gefahren der Akkuratheit der Gesichtserkennung zu warnen (vgl. Wang/Kosinski 2018: 248).

Zustimmung hierzu geben. Aus forschungsethischer Perspektive muss hier die Verletzung der Autonomie kritisiert werden (vgl. hierzu auch Bendel 2018). So ist die sexuelle Orientierung als Teil der – nicht unbedingt als fest gedachten – Identität eine höchst private Angelegenheit, versteht man Autonomie mit Rössler (vgl. Rössler 2001: 100ff.) als Möglichkeit, sich selbst die Frage zu stellen, wie man leben will und – soweit die äußeren, sozialen Bedingungen es zulassen – auch tatsächlich so zu leben. Rösslers Konzeption von Autonomie ist dabei entscheidend mit Privatheit verbunden, die sie in drei Dimensionen (lokal, dezisional und informationell) differenziert (vgl. ebd.: 138). Insbesondere die informationelle Privatheit ist die Dimension, die von Algorithmen wie den von Wang und Kosinski beschriebenen, verletzt werden kann, sollten sie ohne die Zustimmung und das Wissen der betreffenden Personen eingesetzt werden.⁶ Geben Menschen ihre sexuelle Orientierung zum Zwecke des Datings an, ist dies ihre – kontextgebundene – Entscheidung, die sich aber nicht verallgemeinern lässt. Auch wenn sie in Kauf nehmen, dass ihre Daten weiterverwendet werden, wie dies im Internetzeitalter meist der Fall ist (vgl. Seubert 2011: 220), haben sie diese Informationen nur für ihre potenziellen Dating-Partner*innen geteilt. Werden die Bilder, die sie mit dieser Information veröffentlichen, dazu genutzt, die Homosexualität ›als solche‹ bei Menschen zu erkennen, verletzt dies eindeutig die Privatsphäre. Die Anwendung von Algorithmen dieser Art auf weitere Personen, die die fraglichen Informationen gar nicht selbst im Internet veröffentlicht haben, muss selbstverständlich erst recht als illegitimer Eingriff in die Privatsphäre gewertet werden.

Auch aus sozialwissenschaftlicher Perspektive werden bei einem Blick auf die Auswahl der Daten die ersten Einschränkungen der Studie sichtbar. Es wird nicht reflektiert, dass es sich hier um Fotografien handelt, die als solche immer nur einen zeitlich und räumlich begrenzten Ausschnitt der Wirklichkeit abbilden. Sowohl die soziale Performanz und die Inszenierung für bestimmte Zwecke als auch die ungleiche Verwendung von Datingplattformen in der Bevölkerung bleiben unberücksichtigt, so dass trotz der hohen Zahl der analysierten Bilder, nicht von einer repräsentativen Grundgesamtheit⁷ gesprochen werden kann, die es erlauben würde allgemeine Schlüsse zu ziehen. Darüber hinaus zeigt sich in den Daten eine mangelnde Diversität, welche Wang und Kosinski am Ende des Artikels reflektieren. Sie erklären, dass sie keine ausreichende Anzahl an nicht-weißen Homosexuellen in den Daten finden konnten und vermuten, dies hänge mit einer unterschiedlichen Verteilung von Vorurteilen gegenüber Schwulen zusammen und der unterschiedlichen

6 Den Kern der informationellen Privatheit bildet nach Rössler die Kontrolle über den Zugang zum Wissen über die eigene Person, die dazu dient zu wissen, was andere über einen selbst wissen (vgl. Rössler 2001: 201), so dass abgeschätzt werden kann, »in welcher Beziehung sie aufgrund dieses Wissens zu ihr stehen« (ebd.: 205).

7 Zum Begriff der Grundgesamtheit: z.B. Häder 2010: 141.

Nutzung von Online-Dating-Plattformen je nach rassifizierter Gruppenzugehörigkeit (vgl. Wang/Kosinski 2018: 255). Trotz der starken Einschränkung der untersuchten Grundgesamtheit, halten sie an ihrer Generalisierung fest:

We believe, however, that our results will likely generalize beyond the population studied here. They are consistent with the PHT of sexual orientation, which was supported by a variety of studies of humans and other mammals (Hines, 2010). As the exposure to gender-atypical androgen levels is likely to affect the faces of people of different races to a similar degree, it is likely that their facial features are equally revealing of sexual orientation. (ebd.)

An dieser Stelle werden insbesondere zwei Aspekte deutlich: Einmal die Einschränkung der Studie in Bezug auf Race und zweitens die mangelnde Generalisierbarkeit, die trotz Vermutungen über eine eingeschränkte soziale Verteilung der untersuchten Plattform, unter Rückgriff auf biologische Studien, beibehalten wird. Die rassifizierte Einschränkung des Datenmaterials und damit der untersuchten Gruppe zeigt sich auch in den nicht belegten Vermutungen über unterschiedliche Levels an Toleranz von Gruppen, die sich in Bezug auf Race unterschiedlich positionieren, und ihre (Nicht-)Nutzung von Online-Datingplattformen. Aus sozialwissenschaftlicher Perspektive stellt sich die Frage, ob nicht vielmehr die ausgewählte Plattform eine bestimmte Zielgruppe hat.

Bekannt ist zudem, dass die eingesetzten Algorithmen mit Abbildungen von nicht-weißen Menschen zu weniger akkuraten Ergebnissen kommen. Wie viele kritische Studien bereits deutlich gemacht haben, sind Machine-Learning-Algorithmen häufig rassistisch vorbelastet (*racial bias*). Wie z.B. Boulamwini und Gebru (2018) zeigen, können die algorithmischen »classifiers« am besten mit Gesichtern mit heller Haut arbeiten und insbesondere als Menschen mit dunkler Haut klassifizierte werden benachteiligt. Dies steht im Zusammenhang mit der geringen Diversität von Bilddatensätzen, die Algorithmen wie dem von Wang und Kosinski eingesetzten f++ zugrunde liegen.⁸

In einem weiteren Schritt wurden die Abbildungen der Gesichter, die frontal der Kamera zugewandt sein mussten, um nicht aussortiert zu werden, mit Hilfe eines ersten DNN-classifier (namentlich f++) standardisiert, indem die »facial features«, d.h. aus der Biometrik bekannte Vermessungspunkte im Gesicht, extrahiert wurden, wodurch die Kopfhaltung und der Hintergrund vernachlässigt werden konnte. Mit dem DNN-classifier VGG-face wurden schließlich bestimmte »landmarks«

8 Der *racial bias* in Machine Learning Software, der z.B. im medizinischen Bereich zu einer schlechteren Gesundheitsversorgung von Nicht-Weißen führen kann (vgl. z.B. Guo et al. 2021) ist mittlerweile weitgehend bekannt und es wird von verschiedenen Seiten an weniger diskriminierenden Verfahren gearbeitet (vgl. Das et al. 2018; Serna et al. 2022).

extrahiert. Für die statistische Analyse wurde die sexuelle Orientierung (gay/heterosexual) als abhängige Variable und die ›landmarks‹ als unabhängige Variable verwendet (vgl. Wang/Kosinski 2018: 249).

Zur Unterstützung des algorithmischen Verfahrens wurden sogenannte ›Amazon Mechanical Turks‹ (AMTs)⁹ beauftragt (vgl. ebd.: 248). Diese menschlichen Mikro-Arbeiter*innen prüften die von den Nutzer*innen gemachten Angaben. Das (binär gedachte) Gender, die (ebenfalls binär gedachte) sexuelle Orientierung sowie das Alter (18–40) wurden anhand der von Nutzer*innen gemachten Angaben übernommen. Anschließend wurde das ›tatsächliche‹ Geschlecht durch die AMT überprüft und die Race der Abgebildeten klassifiziert (vgl. ebd.). Es wurden nur ›kaukasisch‹ aussehende Gesichter in die Studie einbezogen. Die Amazon-Arbeiter*innen haben damit einen aktiven Part an der Zusammenstellung des Untersuchungsmaterials. Deutlich wird hier eine soziale Verwobenheit mit materiellen Verhältnissen, was sowohl Fragen in Bezug auf die Neutralität als auch die zugrundeliegende Forschungsethik des Verfahrens aufwirft: Die unter prekären und schlecht bezahlten Bedingungen arbeitenden Personen werden hier genutzt wie Maschinen, was sich auch in ihrer Bezeichnung widerspiegelt. Auch die soziale Einordnung des Geschlechts anhand der Abbildungen sowie das ageistische Aussortieren von als ›zu alt‹ kategorisierten Gesichtern ist an dieser Stelle zu kritisieren. Damit der Algorithmus überhaupt arbeiten kann, werden in der Studie von Wang und Kosinski sowohl ethisch als auch epistemologisch fragwürdige Sortierungen vorgenommen.

2.3. Statistische Analyse

Nach der Zusammenstellung des Datenmaterials wurden mit selbst lernenden DNNs¹⁰ sowohl die auf den Abbildungen zu sehenden morphologischen Features (Abstand zwischen den Augen, Nasenlänge etc.) als auch das Styling und die

9 Der Begriff geht auf den sog. *Schachtürken* zurück, einem scheinbar schachspielenden Automaten aus dem 18. Jahrhundert, der jedoch – vor dem Publikum versteckt – tatsächlich von einem Menschen gesteuert wurde (vgl. Jank 2014: 112ff.). Die von der Firma Amazon angebotenen Dienstleistungen der ›Amazon Mechanical Turks‹ beschreibt der Konzern wie folgt: »Amazon Mechanical Turk (MTurk) is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually. [...] While technology continues to improve, there are still many things that human beings can do much more effectively than computers [...]« (Amazon 2022). Dazu siehe auch den Beitrag von Natalie Sontopski in diesem Band.

10 Diese ›Künstlichen neuronalen Netzwerke‹ werden meist als den menschlichen Nervenzellen nachempfundenen beschrieben, so auch von Wang und Kosinski: »DNNs mimic the neocortex by simulating large, multilevel networks of interconnected neurons. DNNs excel at recognizing patterns in large, unstructured data such as digital images, sound, or text, and analyzing such patterns to make predictions.« (Wang/Kosinski 2018: 247)

Kleidung etc. statistisch untersucht und auf Muster untersucht. So sollten Korrelationen zwischen der angegebenen sexuellen Orientierung und den Eigenschaften des Bildes herausgearbeitet werden, was der KI auch tatsächlich gelang. Um die von den DNN erkannten Muster mit der Vorhersagefähigkeit von Menschen zu vergleichen, wurde anschließend in einem experimentellen, d.h. nicht der sozialen Realität entsprechenden Setting, die Fähigkeit zur Erkennung von Homosexualität der KI mit der menschlichen verglichen. Hierzu sollten die bereits erwähnten Amazon-Arbeiter*innen und die KI jeweils zwischen zwei Bildern entscheiden, welches von ihnen eher jemanden mit einer homosexuellen Orientierung abbildet.¹¹ Der Vergleich zeigte, dass die Algorithmen die Homosexualität »besser erkennen« konnten als die*der jeweils eingesetzte Arbeiter*in. Mit der steigenden Anzahl von Bildern wurde der Algorithmus darüber hinaus immer besser (vgl. ebd.: 249f.).

Um ihre Ergebnisse zu testen – zur Vermeidung des »overfitting« – haben Wang und Kosinski die Algorithmen anschließend über ein weiteres, vom ersten unabhängiges Datenset laufen lassen. Als Kontrollgruppe dienten Gesichtsbilder von ausschließlich männlichen Facebook-Usern, die aus bestimmten Gründen, wie der Selbstbezeichnung, aber auch Zugehörigkeit zu Facebookgruppen, als homosexuell eingeordnet wurden. Die Vorhersagekraft der KI konnte dabei bestätigt werden (vgl. ebd.: 253f.). Erneut unberücksichtigt bleiben dabei die sozialen Komponenten der Nutzung dieser Plattformen und die unterschiedliche, kontextabhängige Expression der eigenen sexuellen Orientierung.

Die Komplexität der eingesetzten Algorithmen führt überdies dazu, dass ihre Ergebnisse nicht direkt nachvollzogen werden können. Dieses als »Black Box-Phänomen« bezeichnete Problem, das auch unter dem Begriff »opacity« (Carabantes 2020) oder der »explainability« (Amann et al. 2022) von KI verhandelt wird, entsteht durch die komplexe Analyse der großen Datenmengen und die verschiedenen, geschichteten Trainings- und Klassifizierungsprozesse, die diese durchlaufen. Im Verfahren der DNNs werden zunächst Trainingsdaten anhand zufälliger Werte klassifiziert, die in Prüfungsdurchläufen nach ihrer Vorhersagekraft weiter gewichtet werden, bis eine Menge optimaler Werte enthalten ist. Diese optimalen Werte werden schließlich zum Ausgangspunkt der weiteren Analyse (vgl. Carabantes 2020: 313f.), d.h. die gefundenen Korrelationen werden zum Ausgangspunkt für weitere Korrelationen. Um sich einer Erklärbarkeit der Ergebnisse anzunähern, mussten Wang und Kosinski einen weiteren Durchgang der Studie durchführen.

11 Ähnlich wie durch die Vorarbeit der AMTs wird durch diese Versuchsanordnung die Komplexität der Realität reduziert: Die Möglichkeiten 1) hier ist niemand homosexuell, und 2) hier sind beide homosexuell, werden ausgeschlossen. Eine 50 %-Trefferquote wird somit in das Verfahren eingebaut, obwohl es zumindest z.Zt. keine reale Situation gibt, in der diese so angenommen werden könnte.

Erst durch das Verdecken von unterschiedlichen Bereichen der Gesichtsabbildungen, also durch Komplexitätsreduktion auf der Ebene der Datenlage, konnten sie herausfinden, welche Eigenschaften der Bilder zu den Ergebnissen geführt hatten. Um herauszufinden welche Areale des Bildes als bedeutsam erachtet wurden, wurden ›Heatmaps‹ erstellt (vgl. Wang/Kosinski 2018: 250f.).

Schließlich erstellten sie aus den Durchschnittswerten, idealtypische oder ›archetypische‹ Bilder von ›homo- und heterosexuellen Gesichtern‹, die sie wiederum in die Beschreibung der Ergebnisse mit einfließen ließen. Es zeigen sich hier zwei physiognomische Momente der Untersuchung: einmal in der Beschreibung der Unterschiede der Gesichter (siehe hierzu das Eingangszitat) und auf der anderen Seite in dem Verfahren des zusammengesetzten Fotos selbst. Der Eugeniker und Rassenkundler Galton (1878)¹² entwickelte dieses Verfahren, um die typischen Eigenschaften von Gesichtern zu finden, anhand derer z.B. eine Anfälligkeit für Kriminalität festgemacht werden könnte, und bezeichnete sein Verfahren wegen der mechanischen statt subjektiven Erstellung als besonders präzise (ebd.: 97). Aus heutiger Sicht können Galtons Verfahren und seine Disziplin aber nur als rassistisch verstanden werden. Die von Wang und Kosinski erstellten, zusammengesetzten Bilder, erwecken den Anschein von Realität, da die Daten von Menschen als Gesichter erkannt werden, obwohl sie maschinell erstellt wurden. Die Bezeichnung als archetypisch täuscht in erhöhtem Maße darüber hinweg, dass es sich hierbei um eine aus einem bestimmten Sample (*weiße, junge Menschen, die Online-Dating praktizieren*) errechnete Bilder handelt. Hier von universellen Abbildungen auszugehen, reproduziert rassistische und kolonialistische, ageistische und andere soziale Ausgrenzungsstrukturen.

2.4. Biologischer Rückschluss

Neben den angesprochenen, aus sozialwissenschaftlicher Perspektive äußerst problematischen methodologischen Schwächen, die auch durch den zweiten Durchgang mit einer Kontrollgruppe nicht behoben wurden, sind auch die weiteren Schlussfolgerungen in Wang und Kosinskis Studie extrem fragwürdig: Statt von statistischen Häufungen und Korrelationen zu sprechen, wie es in einer validen sozialwissenschaftlichen Studie der Fall wäre, wird mit dem Rückgriff auf eine biologische Theorie eine Universalität der Ergebnisse unterstellt. So werden nicht Aussagen über das konkrete Datenset – nämlich die inszenierten Gesichts-Abbildungen einer US-amerikanischen Online-Datingplattform, die zuvor in Bezug auf Geschlecht und Race kategorisiert wurden – getroffen, sondern verallgemeinernde biologische Rückschlüsse gezogen. Wang und Kosinski beziehen sich auf die

12 Eine gute Übersicht zum Werk sowie diverse Originalquellen von Galton bietet die Website <https://galton.org>.

bereits als Ausgangsprámisse verwendete Pránatale Hormontheorie (PHT) (vgl. Wang/Kosinski 2018: 254).

In Übereinstimmung mit der PHT schließen Wang und Kosinski aus ihrer ersten Unterstudie, Homosexuelle hätten eine ›gender-atypische‹ Gesichtsmorphologie: Schwule Männer sollten demnach tendenziell mehr feminine Gesichtszüge als heterosexuelle Männer haben, während Lesben tendenziell mehr maskuline Merkmale als heterosexuelle Frauen ausbildeten. Maskulinität zeige sich demnach in der markanteren GröÙe von Kinn, Kiefer, Augenbrauen, Nase und Stirn. Bei schwulen Männern seien diese Bereiche weniger groß, bei lesbischen Frauen wiederum seien diese größer als für Frauen anzunehmen sei. Auch der von ihnen entdeckte geringere Bartwuchs und die hellere Haut bei homosexuellen Männern und weniger Augen-Make-Up, dunklere Haare, bedeckende Kleidung und das seltenere Lächeln bei lesbischen Frauen sowie das Tragen von Baseballkappen bei heterosexuellen Männern und lesbischen Frauen werden von Wang und Kosinski als ›gender-atypisch‹ bezeichnet (vgl. ebd.: 252). Verbildlicht und dadurch scheinbar auch für uns Menschen beobachtbar werden diese Annahmen durch die ›archetypischen‹ Gesichter, die Wang und Kosinski aus durchschnittlichen Werten der Daten erstellten (vgl. ebd.: 251).

Zum Schluss konstruieren Wang und Kosinski also ›normale‹ Gesichter und ›typische Abweichungen‹, die ›abweichende‹ Begehrensformen anzeigen sollen. Sie stehen damit genau in der Tradition der Physiognomik, der sie sich zu Beginn der Studie nicht zugehörig fühlen wollten. In die physiognomischen Beschreibungen der ›normalen‹ und der ›abweichenden‹ Gesichter gehen sowohl biologische bzw. biometrische wie auch sozial konstruierte bzw. subjektiv hergestellte Merkmale ein. Sie eignen sich daher nicht für biologische Rückschlüsse oder das Testen von biologischen Hypothesen. Im Gegenteil: Die gemeinsame Betrachtung von biometrischen und sozialen Gesichtspunkten legt die Untersuchung auf die soziale Herstellung von Geschlecht und den performativen Ausdruck von Begehren fest, biologische Rückschlüsse sind hier gar nicht möglich.

2.5. Konkrete soziale Praxis statt biologischer Essenz

Die Aussagekraft der Ergebnisse von Wang und Kosinskis Studie ist aus sozialwissenschaftlicher Perspektive also als stark eingeschränkt anzusehen. Das Studiendesign macht es unmöglich, zu biologischen oder auch zu universellen Aussagen zu gelangen. Selbst wenn keine methodischen Fehler vorliegen würden, begingen Wang und Kosinski immer noch einen logischen Fehler, der alltagssprachlich als unzulässige oder vorschnelle Verallgemeinerung und epistemologisch als Induktionsproblem bezeichnet wird (vgl. Ritsert 2003: 98f.): Von den untersuchten Eigenschaften einer bestimmten Gruppe (Nutzer*innen einer bestimmten Datingplattform), die Teil einer größeren, nicht untersuchten Gruppe sind (homosexuelle Menschen),

wird auf die Eigenschaften dieser größeren Gruppe geschlossen. Damit wird entweder Repräsentativität – die aber nicht gegeben ist – vorausgesetzt, oder die Verallgemeinerung wird durch die Wahrheit einer Theorie gerechtfertigt, die (wie die PHT) bereits Allgemeingültigkeit beansprucht. Letzteres sollte allerdings Gegenstand der Untersuchung sein und kann daher nicht vorausgesetzt werden, da es sich ansonsten um einen Zirkelschluss handelt.

Stattdessen sind die Ergebnisse auf ein konkretes Datenset zu beziehen und damit in diesem Fall auf die soziale Herstellung von Geschlecht und sexueller Orientierung *weiß*er Personen in den USA. Datingplattformen oder andere Social Media beinhalten – wie alle sozialen Räume – Normen und Ideale, an die sich Anwender*innen mehr oder weniger stark annähern oder von denen sie sich mehr oder weniger stark abgrenzen. Einfluss auf die äußere Erscheinung haben zudem vergeschlechtlichte Moden und Lebensstile, die gemeinsam mit »Männlichkeits- und Weiblichkeitscodes« (Degele 2004: 189) bewusst und/oder unbewusst zum Ausdruck kommen können.

Außerdem ist von einer sozial ungleichen Nutzung und einer entsprechenden Zielgruppendefinierung von Datingplattformen (oder auch Facebook) auszugehen, so dass die Daten auch nicht als repräsentativ für die ausgewählte Altersgruppe und Race anzusehen sind (vgl. zu Auswahlverfahren in der empirischen Sozialforschung z.B. Häder 2010: 139ff.). So ist auch die Auswahl einer Datingplattform zur Datenerhebung nicht als repräsentatives Verfahren anzusehen, da diese bestimmte Zielgruppen ansprechen wollen und ökonomische Interessen verfolgen.

Statt von verallgemeinerbaren oder gar biologischen Ergebnissen zu sprechen, sollte aus sozialwissenschaftlicher Perspektive also eher von Untersuchungsergebnissen gesprochen werden, die eine konkrete soziale Praxis betreffen, die eben nicht von der gesamten menschlichen Bevölkerung ausgeübt wird.

3. Wiederverzauberung und Essenzialisierung

Aus einer geschlechtertheoretischen Perspektive wirken Studien wie die von Wang und Kosinski, in Anlehnung an Campolo und Crawford (2020) wie ein »verzauberter Verstärker« von bestehenden Strukturen. Von verzaubert wird gesprochen, da hier der Anschein erweckt wird, die KI habe »über-menschliche« Fähigkeiten zur Erkennung von Wahrheit.

Bleiben soziale Strukturen bei wissenschaftlichen Studien unberücksichtigt, werden diese jedoch bloß reproduziert und nicht erklärt.¹³ Im Fall von Wang und

13 Auf diesen Zusammenhang wies Adorno (1979) anlässlich der potenziellen Theorielosigkeit von empirischer Sozialforschung hin. Ohne Gesellschaftstheorie »bleibt die wissenschaftliche Spiegelung in der Tat bloße Verdopplung, verdinglichte Apperzeption des Dinghaften,

Kosinskis Studie wird die heterosexuelle Geschlechtermatrix (vgl. Butler 1991) verfestigt: Es gibt zwei Geschlechter, die natürlicherweise jeweils das andere begehren. Damit wird eine Übereinstimmung von Gender und Begehren zugrunde gelegt. Abweichungen hiervon sollen ebenfalls der Matrix entsprechen. Was nicht passt, ist prekär und nur als Abweichung von der Norm erfassbar, nicht aber als etwas Eigenständiges oder wirklich anderes.¹⁴ Es handelt sich bei der Studie also um den Versuch, sexuelle Identitäten als Essenz der Person festzuschreiben und dabei wird sich bei der stereotypen Vorstellung der ›Verweiblichung‹ und ›Vermännlichung‹ der ›echten‹ Geschlechter bei Homosexualität bedient. Weiblichkeit wird dabei, wie bei der historischen Vorstellung von homosexuellen Männern, als Abweichung vom Männlichen konzipiert. So beschreibt Krafft-Ebing, wie bereits anfangs erwähnt, in seiner einflussreichen *Psychopathia Sexualis* (1894) die dreistufige »Conträre Sexualempfindung« bei Männern in der ersten Stufe als »Evirato« (Entmännlichung) (vgl. ebd.: 204ff.), in der zweiten als »Effeminatio« (Verweiblichung) (vgl. ebd.: 269ff.) und schließlich in der dritten als »Androgyny « (Androgynität) (vgl. ebd.: 275ff.). Auch wenn es bei Wang und Kosinski weder zu einer Pathologisierung noch zu einer direkten Abwertung von Weiblichkeit oder Homosexualität kommt, wird das heterosexuelle männliche Subjekt als Norm aufrechterhalten und Weiblichkeit sowie Homosexualität werden als etwas verstanden, das von dieser Norm abweicht, biologische Ursachen hat und sich daher auch im Körper manifestiert.

Abgesehen von wissenschaftstheoretischen Schwierigkeiten und deren Auswirkungen auf das Verständnis von Geschlecht und sexueller Orientierung, die eben nicht kohärent oder festgeschrieben sein müssen und sich vor allem nicht durch Gesichtsvermessung erkennen lassen, muss auch die potenzielle Gefährdung von Menschen kritisiert werden. Zwar ist es wichtig, vor physiognomischen Verfahren zu warnen, da diese bereits in Anwendung sein könnten. So ist das ›Targeted Advertising‹ aktuell eine wichtige Komponente des Online-Marketings. Hierfür werden große Datenmengen nach Mustern analysiert, um im Umkehrschluss Individuen anhand der statistischen Merkmale zu kategorisieren und ihnen auf Websites und in Social Media (vermeintlich) passgenaue Produktwerbung anzuzeigen (vgl. Richter/Kaminski 2016: 45; aus Perspektive einer wirtschaftswissenschaftlichen Marketing-Forschung vgl. z.B. Sivadas et al. 1998; Provost et al. 2009; Goldfarb/

und entstellt das Objekt gerade durch die Verdopplung, verzaubert das Vermittelte in ein Unmittelbares.« (ebd.: 204)

- 14 Butler beschreibt diesen Zusammenhang wie folgt: »Die Gespenster der Diskontinuität und Inkohärenz, die ihrerseits nur auf dem Hintergrund von existierenden Normen der Kohärenz und Kontinuität denkbar sind, werden ständig von jenen Gesetzen gebannt und zugleich produziert, die versuchen, ursächliche oder expressive Verbindungslinien zwischen dem biologischen Geschlecht, den kulturell konstituierten Geschlechtsidentitäten und dem ›Ausdruck‹ oder ›Effekt‹ beider in Darstellung des Begehrens in der Sexualpraxis zu errichten.« (Butler 1991: 38)

Tucker 2011). Auch das politische ›Micro-Targeting‹ bedient sich dieser Prinzipien (vgl. O’Neil 2018: 244). Wang und Kosinski weisen zwar selbst darauf hin, dass Verfahren, wie das von ihnen vorgestellte, von autoritären Staaten und Firmen genutzt werden könnten (vgl. Wang/Kosinski 2018: 255), dennoch ist ihre Studie selbst Teil einer potenziellen Gefährdung, da sie hierfür eine Anleitung liefern (vgl. Keyes et al. 2021: 168, FN). Homosexualität ist in konservativen Gesellschaften nach wie vor tabuisiert oder sogar verboten, und Menschen, denen man diese Orientierung nun (vermeintlich) maschinell ›ansehen‹ kann, könnten je nach Kontext Repressionen, Kriminalisierung und Verfolgung ausgesetzt werden. Aus ethischer Perspektive stellt sich nicht zuletzt deshalb die Frage, warum es gewollt sein sollte, die sexuelle Orientierung am Körper zu verorten, sie damit zu objektivieren und festzuschreiben bzw. zu essenzialisieren und dafür auch noch technische Verfahren zu verwenden, die Menschen potenziell gefährden.

Die Verbindungslinien und Kontinuitäten zwischen alter und neuer Physiognomik per Gesichtserkennungssoftware und Deep Learning Algorithmen konnten in diesem Artikel nur angerissen werden und bedürfen einer weitergehenden Untersuchung. Besondere Berücksichtigung sollte dabei auch der Zusammenhang von Rassismus, Biologismus und Esoterik (vgl. Wölflingseder 1995) finden. Auch die Zusammenhänge zwischen dem Aufkommen der statistischen Physiognomik und aktuellen gesellschaftlichen Entwicklungen müssen untersucht werden. Wie Celis Bueono analysiert, können die Verfahren zur maschinellen Gesichtserkennung und den damit zusammenhängenden Kategorisierungen als Versuch der Einhegung von Individuen in der Kontrollgesellschaft verstanden werden (vgl. 2020: 73). Auch die wachsende Quantifizierung und die damit zusammenhängende besondere Machtkonzentration durch statistische Verfahren im Kapitalismus (vgl. Joque 2022), die rassistische und vergeschlechtlichte Ungleichheiten (re-)produzieren, müssen in Zukunft noch stärker in den Fokus der sozialwissenschaftlichen Analysen gerückt werden, nicht zuletzt, um in einen kritischen Dialog mit Forscher*innen treten zu können, die quantitativ-technisch arbeiten.¹⁵

Literaturverzeichnis

- Adorno, Theodor W. 1979. Soziologie und empirische Forschung. In *Soziologische Schriften I. Gesammelte Schriften*, 196–216. Frankfurt a.M.: Suhrkamp.
- Amann, Julia, Dennis Vetter, Stig Nikolaj Blomberg, Helle Collatz Christensen, Megan Coffee, Sara Gerke, Thomas K. Gilbert, Thilo Hagendorff, Sune Holm,

15 Ich danke Malte Fuchs für die kritische Kommentierung des Manuskripts in seinen verschiedenen Fassungen und den Herausgeber*innen für ihre wertvollen Anmerkungen.

- Michelle Livne, Andy Spezzatti, Inga Strümke, Roberto V. Zicari und Vince Istan Madai. 2022. To explain or not to explain? Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health*, 1(2): 1–18.
- Amazon. 2022. *Amazon Mechanical Turk*. <https://www.mturk.com/>. Zugegriffen: 13.01.2022.
- Bendel, Oliver. 2018. The Uncanny Return of Physiognomy. In *The 2018 AAAI Spring Symposium Series*: 10–17.
- Buolamwini, Joy und Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities. Commercial Gender Classification. *Proceedings of Machine Learning Research* 81: 1–15.
- Butler, Judith. 1991. *Das Unbehagen der Geschlechter*. Frankfurt a.M.: Suhrkamp.
- Campolo, Alexander und Kate Crawford. 2020. Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science, Technology, and Society* 6: 1–19.
- Carabantes, Manuel. 2020. Black-box artificial intelligence. An epistemological and critical analysis. *AI & SOCIETY* 35(2): 309–317.
- Celis Bueno, Claudio. 2020. The Face Revisited. Using Deleuze and Guattari to Explore the Politics of Algorithmic Face Recognition. *Theory, Culture & Society* 37(1): 73–91.
- Das, Abhijit, Antitza Dantcheva und Francois Bremond. 2018. Mitigating Bias in Gender, Age and Ethnicity Classification. A Multi-Task Convolution Neural Network Approach. Paper präsentiert bei der *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*: 1–13.
- Degele, Nina. 2004. *Sich schön machen*. Wiesbaden: Springer VS.
- Galton, Francis. 1878. Composite Portraits. *Nature* : 97–100.
- Gebarek, Werner E. 2003. Physiognomik und Phrenologie. Formen der populären Anthropologie im 18. Jahrhundert. In *Medizin in Geschichte, Philologie und Ethnologie. Festschrift für Gundolf Keil*, Hg. Gundolf Keil und Monika Reiningger, 35–49. Würzburg : Königshausen & Neumann.
- Goldfarb, Avi und Catherine Tucker. 2011. Online Display Advertising. Targeting and Otrusiveness. *Marketing Science* 30(3): 389–404.
- Gould, Stephen Jay. 1988. *Der falsch vermessene Mensch*. Frankfurt a.M.: Suhrkamp.
- Guo, Lisa N., Michelle S. Lee, Bina Kassamali, Carol Mita und Vinod E. Nambudiri. 2021. Bias in, bias out. Underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection-A scoping review. *Journal of the American Academy of Dermatology* 87(1): 157–159.
- Häder, Michael. 2010. *Empirische Sozialforschung. Eine Einführung*. Wiesbaden: Springer VS.
- Harvey, Adam und Jules LaPlace. 2021. Researchers Gone Wild. Origins and Endpoints of Image Training Datasets Created »In the Wild«. In *Practicing*

- Sovereignty: Digital Involvement in Times of Crises*, Hg. Bianca Herlo, Daniel Irrgang, Gesche Joost und Andreas Unteidig, 289–309. Bielefeld: transcript.
- Jank, Marlen. 2014. *Der homme machine des 21. Jahrhunderts. Von lebendigen Maschinen im 18. Jahrhundert zur humanoiden Robotik der Gegenwart*. Paderborn: Wilhelm Fink.
- Joque, Justin. 2022. *Revolutionary Mathematics. Artificial Intelligence, Statistics and the Logic of Capitalism*. London: Verso Books.
- Keyes, Os, Zoë Hitzig und Mwenza Blell. 2021. Truth from the machine. Artificial intelligence and the materialization of identity. *Interdisciplinary Science Reviews* 46(1-2): 158–175.
- Kosinski, Michal. 2021. Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific Reports* 11(1): 100.
- Krafft-Ebing, Richard. 1894. *Psychopathia sexualis, mit besonderer Berücksichtigung der conträren Sexualempfindung. Eine klinischforensische Studie*. 9. Stuttgart: Ferdinand Enke.
- Lavater, Johann Caspar. 1775. *Physiognomische Fragmente, zur Beförderung der Menschenkenntniß und Menschenliebe*. Leipzig/Winterthur: Weidmann und Reich/Steiner.
- O’Neil, Cathy. 2018. *Angriff der Algorithmen. Wie sie Wahlen manipulieren, Berufschancen zerstören und unsere Gesundheit gefährden*. Bonn: Bundeszentrale für Politische Bildung.
- Provost, Foster, Brian Dalessandro, Rod Hook, Xiaohan Zhang und Alan Murray. 2009. Audience selection for on-line brand advertising: privacy-friendly social network targeting. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD’09*. New York: Association for Computing Machinery: 707–716.
- Rasmussen, Stig Hebbelstrup Rye, Steven Ludeke und Robert Klemmensen. 2022. Using deep learning to predict ideology from facial photographs: Expressions, beauty, and extra-facial information. *PsyArXiv Preprints*.
- Regener, Susanne. 2010. *Visuelle Gewalt. Menschenbilder aus der Psychiatrie des 20. Jahrhunderts*. Bielefeld: transcript.
- Richter, Philipp und Andreas Kaminski. 2016. »Before you even know ...«. Big Data und die Erkennbarkeit des Selbst. *International Review of Information Ethics* 24: 44–54.
- Ritsert, Jürgen. 2003. *Einführung in die Logik der Sozialwissenschaften*. Münster: Westfälisches Dampfboot.
- Rössler, Beate. 2001. *Der Wert des Privaten*. Frankfurt a.M.: Suhrkamp.
- Schmölders, Claudia. 2007. *Das Vorurteil im Leibe: Eine Einführung in die Physiognomik*. Berlin: De Gruyter.
- Serna, Ignacio, Aythami Morales, Julian Fierrez und Nick Obradovich. 2022. Sensitive Loss: Improving Accuracy and Fairness of Face Representations with Discrimination-Aware Deep Learning. *Artificial Intelligence* 00: 1–18.

- Seubert, Harald. 2011. Privatsphäre. In *Handbuch Angewandte Ethik*, Hg. Ralf Stoecker, Christian Neuhäuser, Marie-Luise Raters und Fabian Koberling, 219–222. Stuttgart: J.B. Metzler.
- Shrider, Emily A., Melissa Kollar, Frances Chen und Jessica Smegea. 2021. *U.S. Census Bureau. Current Population Reports. Income and Poverty in the United States: 2020*. Washington, DC: U.S. Government Publishing Office.
- Sivadas, Eugene, Rajdeep Grewal und James Kellaris. 1998. The Internet as a Micro Marketing Tool. Targeting Consumers through Preferences Revealed in Music Newsgroup Usage. *Journal of Business Research*, 41(3): 179–186.
- Wang, Yilun und Michal Kosinski. 2018. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology* 114(2): 246–257.
- Weber, Jutta und Corinna Bath. 2003. Technowissenschaftskultur und feministische Kritik. In *Turbulente Körper, soziale Maschinen. Feministische Studien zur Technowissenschaftskultur*, Hg. dies., 9–26. Opladen: Leske und Budrich.
- Wölflingseder, Maria. 1995. Biologismus – »Natur als Politik«. New-Age und Neue Rechte als Vorreiter einer (wieder) etablierten Ideologie. In *Biologismus, Rassismus, Nationalismus. Rechte Ideologien im Vormarsch*, Hg. Gero Fischer und Maria Wölflingseder, 22–36. Wien: Promedia.

