

## Neue Neurotechnologien – Ethik, Recht und Gesellschaft

Donnerstag, 19. März 2026, 13:00–17:30 Uhr

Online

Begrüßung .....	2
Helmut Frister · Vorsitzender des Deutschen Ethikrates .....	2
Einführung.....	2
Aldo Faisal · Deutscher Ethikrat .....	2
Block 1 .....	3
Consumer neurotech: Integrating the brain into the digital ecosystem? .....	3
Laura Bernáez Timón · Centre for Future Generations .....	3
Neurotechnologie und Human Enhancement: Akzeptanzfaktoren und psychosoziale Risiken .....	7
Bertolt Meyer · Technische Universität Chemnitz.....	7
AI as neurotechnology: Tools and self-diagnosis in young people .....	11
Ilina Singh · University of Oxford .....	11
Neurotechnologien und ihre ethischen Herausforderungen für zwischenmenschliche und Mensch- Technik-Beziehungen .....	15
Saskia Nagel · RWTH Aachen University.....	15
Diskussion.....	20
Block 2.....	34
Mentale Integrität und Vulnerabilität als ethische Herausforderungen bei Neurotechnologien.....	34
Philipp Kellmeyer · Universität Mannheim .....	34
Nichtmedizinische Neurotechnologien: Ethische Grenzen, mentale Selbstbestimmung und Governance .....	38
Marcello Ienca · Technische Universität München .....	38
Gedankenfreiheit als Fixpunkt der Regulierung von Neurotechnologien.....	42
Jan Christoph Bublitz · Universität Hamburg.....	42
Diskussion.....	46
Zusammenfassung.....	62
Eva Winkler · Deutscher Ethikrat .....	62

Hinweis: Bei dem folgenden Text handelt es sich nicht um eine wörtliche Transkription. Der Text wurde lektoriert, um eine gute Lesbarkeit herzustellen.

## Begrüßung

### Helmut Frister · Vorsitzender des Deutschen Ethikrates

Guten Tag, meine Damen und Herren, im Namen des Deutschen Ethikrates darf ich Sie herzlich zu unserer öffentlichen Sitzung begrüßen. Gegenstand dieser Sitzung ist eine Sachverständigenanhörung zur Neurotechnologie. Ich danke den Sachverständigen dafür, dass Sie uns Ihre Expertise zur Verfügung stellen, und freue mich auch über das Interesse am Zuschauen im Livestream.

Der Deutsche Ethikrat befasst sich seit einem guten Jahr mit dem Thema Herausforderungen durch die rasante Entwicklung der Neurotechnologien, die insbesondere durch die Entwicklung der künstlichen Intelligenz angetrieben wird. Wir haben dazu schon eine öffentliche Anhörung durchgeführt<sup>1</sup>. Diese Anhörung befasste sich vor allem damit, dass wir uns einen Überblick über den gegenwärtigen Stand der Entwicklung der Neurotechnologie und über zu erwartende künftige Entwicklungen verschaffen wollten.

Heute soll es um die ethischen und möglichen rechtlichen Konsequenzen gehen: Wie ist dies ethisch zu bewerten und wie könnte man es eventuell regulieren? Das wird Gegenstand unserer heutigen Anhörung sein.

Die Anhörung wird überwiegend in deutscher Sprache durchgeführt. Zwei Sachverständige werden in englischer Sprache referieren. Dazu wird es sowohl im Saal als auch im Livestream eine Simultanübersetzung geben.

Jetzt übergebe ich an Aldo Faisal. Er ist Sprecher der Arbeitsgruppe Neurotechnologie und wird inhaltlich in das Thema einführen.

## Einführung

### Aldo Faisal · Deutscher Ethikrat

Guten Tag, liebe Ratsmitglieder, liebe Sachverständige, liebe Bürger an den Geräten online, ich begrüße Sie sehr herzlich zu unserer öffentlichen Anhörung des Deutschen Ethikrates zum Thema Neue Neurotechnologien, heute mit dem Fokus auf Ethik, Recht und Gesellschaft.

Ich freue mich sehr, dass wir heute hier sind und dass uns so viele Menschen online folgen. Das zeigt: Wir arbeiten hier an einem Thema, das längst kein fernes Zukunftsthema mehr ist, sondern Neurotechnologie hält bereits in unserem Alltag Einzug und wird in den kommenden Jahren noch viel stärker an Bedeutung zunehmen.

Wenn viele Menschen bei Neurotechnologien zuerst an medizinische Anwendungen denken – zum Beispiel die Anwendung bei Parkinson, um Patienten wieder die Bewegungsfähigkeit zu verleihen, oder bei Neuroprothesen, die uns nach dem Verlust einer Hand die Handfertigkeiten wiedergeben –, dann müssen wir jetzt darüber reden, dass Neurotechnologien den klinischen Raum bereits verlassen haben. Neurotechnologien halten Einzug in den Konsumbereich, in die Arbeitswelt, in die Bildung, in die Freizeit und in den Wellnessbereich, der unreguliert ist. Solche alltagsnahen neurotechnologischen Systeme sind nur die Vorboten einer Entwicklung, in der Technik immer näher an unsere Sinne, an unsere Bewegungsfähigkeiten, an unsere Entscheidungen, an unsere Gefühle und an unser Denken heranrückt. Deshalb müssen wir uns das als Ethikrat genauer anschauen.

Manche glauben, dass wir Human Augmentation, also die Fähigkeit zur menschlichen Steigerung

---

<sup>1</sup> Neue Neurotechnologien – Entwicklungen und Trends. Öffentliche Anhörung vom 18.09.2025.

brauchen, um gegenüber technologischen Entwicklungen konkurrenzfähig zu bleiben. Andere glauben das Gegenteil: dass wir uns die Menschlichkeit bewahren müssen, um von der Technologie unabhängig zu sein.

Diese Entwicklung wird heutzutage hauptsächlich durch das Potenzial der künstlichen Intelligenz dargestellt unter Verknüpfung von unterschiedlichsten Datenquellen. Neurotechnologie beschränkt sich nicht nur auf neuronale Daten, also Daten von Hirnzellen, sondern nutzt die Zusammenführung von Datenquellen aus unserem Alltag, aus Geräten, die wir tragen können, die unser Verhalten beobachten, um bessere Lösungen darzustellen. Heutzutage geht es nicht mehr darum, was biologisch oder technisch möglich ist, sondern was ethisch verantwortbar ist, was gesellschaftlich wünschenswert ist und was rechtlich geregelt werden sollte.

Wenn es um Technologien geht, die unser Denken, unsere Willensfreiheit beeinflussen, dann geht es um Selbstbestimmung, Gedankenfreiheit, Verantwortung und Fairness, und zwar nicht nur für den Einzelnen, sondern auch für das freie gesellschaftliche Miteinander.

Dazu haben wir eine Reihe von Experten geladen, die uns etwas zu diesen Aspekten erzählen werden – von den ökonomischen Entwicklungen in diesem Bereich bis hin zu rechtlichen, philosophischen, soziologischen und psychologischen Komponenten dieser Entwicklung – und die wir danach befragen werden.

Ich freue mich sehr auf diese Diskussion und möchte jetzt an meine Kollegin Eva Winkler übergeben, die die erste Sitzung leiten wird.

## **Block 1**

### **Eva Winkler**

Herzlichen Dank auch von meiner Seite und herzliche Grüße an alle, die uns im Livestream folgen. Wir beginnen in diesem ersten Block mit vier Vorträgen, die wir nacheinander hören werden. Danach haben wir ausreichend Zeit für die Diskussion.

Ich darf als Erstes Frau Laura Bernáez Timón begrüßen. Wir beginnen mit der Perspektive der Consumer-Neurotechnologie, also Anwendungen, die zunehmend in Alltagsgeräte integriert werden. Frau Timón ist Neurotechnologie-Forscherin am Center for Future Generations und promoviert im Bereich Computational Neuroscience am Max-Planck-Institut in Erlangen und Frankfurt. Sie arbeitet an der Schnittstelle von Wissenschaft, Marktanalyse und Politik und wird uns einen Überblick über die aktuellen Entwicklungen und Markttrends geben.

### **Consumer neurotech: Integrating the brain into the digital ecosystem?**

#### **Laura Bernáez Timón · Centre for Future Generations**

Thank you very much to the German Ethics Council for the invitation. It is a pleasure for me to be here and to be speaking about some of the research that we have been conducting at the Centre for Future Generations. Today I am going to focus on consumer neurotechnology, which is neurotechnology that is commercialised outside of medical markets. I am going to be covering three things. First, I am going to go through some recent market trends that we have studied. Second, I am going to be talking about some of the opportunities and risks that emerge in this sector. And third, I am going to be speaking about a few policy measures that we think could help protect consumers and their rights.

(Slide 2)

Neurotechnology is often pictured as complex and bulky medical devices used in hospitals and laboratories. But increasingly, neurotechnology is also appearing in consumer devices. These are products that measure brain activity or that stimulate it and that are directly sold to consumers. So, these products prioritise comfort and long-term wearability, and are designed mainly for entertainment and for wellness. Here are some examples. For example, fatigue monitoring caps with EEG used at the workplace, also headphones or earbuds with EEG to track focus or productivity. But also we are seeing prototypes of brain-computer interfaces that are connected to AI systems such as LLM models and that adapt the content shown based on mental states. And this is a field that is more broadly known as neuroadaptive computing or passive BCIs. And there are also glasses with EEG that continuously track mental states. So as you see, some of these devices look entirely like everyday gadgets. But as neurotechnologies, they raise the characteristic questions about data privacy, consent, etc. So if we want to think about societal impact, maybe one of the first things we want to know is how big this market is, actually. And that is what we did at the Centre for Future Generations. We conducted a study of nearly three hundred neurotechnology companies worldwide. And I'm going to show you today some of these trends.

(Slide 4)

Our analysis shows that the consumer neurotechnology sector has grown very rapidly. Consumer neurotech companies have actually more than tripled in the last decade, going from just forty to around one hundred and fifty today. And in fact now, consumer firms represent sixty per cent of all neurotechnology companies globally. Since 2018, they have outnumbered medical firms. And

this means that nowadays in the market, we have more consumer neurotech companies than medical ones. There are several technological advances that have driven this growth, including the miniaturisation of sensors, dry electrodes, artificial intelligence, but also the cloud ecosystem and smartphones.

(Slide 5)

When we look at technologies, we see that most of these companies use technologies that are also used in medical markets, for example, electroencephalography, or EEG. In fact, EEG is used by roughly sixty-four per cent of these consumer companies and it is attractive because of a number of reasons. It is relatively low cost, it is safe, it is portable, and it is established scientifically since its invention one hundred years ago, actually here in Germany. We are also seeing the emergence of multimodal systems, so combining EEG with other signals such as functional near-infrared spectroscopy. And, of course, all of this is usually powered by AI.

(Slide 6)

One important trend is that the consumer market is strongly driven by the wellness sector. We have applications for stress management, for sleep optimisation, for focus enhancement. Geographically, we observe important differences as well. In North America, consumer neurotech companies tend to focus more on wellness and lifetime, lifestyle applications, while here in Europe, there is relatively more activity in research and development tools. We also see that wellness companies tend to attract more private capital, which creates two innovation ecosystems. So, one that is centred in North America, focused on wellness and driven by venture capital, and another one in Europe that is focused on research and more often backed by public funding.

(Slide 7)

Consumer neurotechnology companies also tend to commercialise their products faster and that is, of course, a consequence of going through the regulatory approval in medical markets. So, a larger proportion of consumer firms already have products on the market as compared to medical companies that take longer to do that. And interestingly, our analysis also shows that consumer companies are able to generate comparably the same revenue with just a quarter of the funding, at least in the short term. We believe that this may tempt investors away from medical innovation and towards faster returns in the consumer market.

(Slide 8)

Another important development is the growing involvement of large technology companies. So major firms we know, such as Apple, Meta, Samsung and Amazon, are investing in neurotechnology. They are partnering with different consumer and medical companies or they are actually developing their own technologies. And one example of that is the Meta neural wristband that was released a couple of months ago in the US. This means that if neurotechnology is integrated in mainstream digital platforms through the products that these companies develop, its reach could scale very quickly.

(Slide 9)

So, we do actually see some opportunities. If developed responsibly, we believe that consumer neurotechnology could offer valuable and, importantly, new capabilities.

(Slide 10)

One of them is continuous and remote monitoring outside of the clinic. We know that the brain is in constant dialogue with the environment and also

that many neurological conditions unfold over time. However, you go to the hospital, you take a scan, and then miss the information about what is happening in between sessions. Unlike clinical systems, consumer devices often prioritise comfort, portability and usability, which makes them suitable for long-term use in real-world settings. This could enable the collection of longitudinal brain data, potentially contributing to research, to prevention, to biomarker identification, or for clinical trial management. We think that consumer and medical neurotechnology could potentially co-evolve, with consumer devices complementing clinical tools that prioritise safety and efficacy with comfort and with wearability. And, in fact, we have some examples of that. So, there are some medical devices like brain stimulation for depression that are now CE-cleared and that are used at home by consumers without a medical prescription.

(Slide 11)

But we also see some challenges.

(Slide 12)

As consumer neurotechnologies become more powerful, the boundary between medical and consumer tech is dissolving. Some consumer devices are now able to support functions with clinical relevance, like preventing insomnia or burnout. However, in consumer markets, these companies do not need to show evidence of efficacy. Under current regulatory frameworks, whether a device is classified as medical often depends largely on how it is marketed and the language they use, and not so much on what the device can actually do. So small differences in wording, such as saying that a device supports mental health rather than treats it, can determine whether medical device regulations apply or not. And then companies may refer to symptoms associated with medical

conditions such as anxiety, depression or burnout, while avoiding explicit medical claims that would trigger stricter regulation. And, as a result, some products now, many of them occupy a grey, also called borderline area between wellness and medicine, and this naturally can create potential confusion for consumers.

(Slide 13)

Another challenge concerns evidence and transparency. In an upcoming study, we analysed one hundred consumer wellness neurotechnology companies and found that roughly half of them do not clearly disclose whether their products, actually their products are not medically certified, even in their terms and conditions. And nearly forty per cent provide no evidence at all for the claims they make. At the same time, for the remaining companies, the level of evidence is extremely heterogeneous, both in quality and in quantity. And interestingly, half of consumer neurotech companies were mistakenly classified as medical by an AI model based on the claims that they make on their websites. So this illustrates how blurred the boundary between wellness and medical devices is, and how blurred it can actually become.

(Slide 14)

And these dynamics raise several concrete potential risks, one of them being medical substitution. Consumers might unknowingly use devices that lack clinical validation, especially where mental health services are difficult to access or remain unaffordable for many. The second risk is hype, and consequently, the erosion of trust. So if companies borrow the language and imagery of medicine and then fail to deliver on their promises, public trust in the entire field can erode and that can lead to fear, to conspiracy, to backlash, and potentially to overregulation as well. And this

could hamper responsible medical innovation, which is showing progress along different lines. The third risk is in how these devices interpret and shape mental states. Some devices translate complex mental states such as stress or attention into sometimes simplified metrics that are not clinically validated. And even if inaccurate, these metrics could influence how people interpret their own health, raising concerns about autonomy and also about meaningful consent. And finally, and this is something that at CFG we are particularly concerned about, is the potential for large-scale cognitive manipulation. So if neurotechnology becomes embedded in widely used consumer devices through one of the products that we already use, even basic signals such as attention or fatigue could potentially be used to optimise advertising, persuasion or disinformation strategies. And for that, you do not really need to read thoughts or to stimulate anything. Just knowing when someone is vulnerable, for instance, when someone's tired or when someone is paying attention, may be enough to influence them. And if this information about mental states goes through our smartphones, it also means that it is potentially hackable. This is not science fiction or very complex technology, all the necessary elements of the tech ecosystem are already there.

(Slide 15)

Which takes us to the question on regulation. How should policymakers respond? And as someone with a background in biomedical engineering and neuroscience, I particularly acknowledge my team in this section.

(Slide 16)

We have been exploring policy approaches that could reduce evidence and transparency gaps, while also promoting the different opportunities that these devices present, and particularly for

healthcare. We believe that the main problem is in the regulatory grey area. So, outside of medical regulations, consumer devices are governed by a patchwork of non-specific laws that increasingly do not fully protect consumers. But they do not help companies either, as they bring regulatory uncertainty and complicate compliance sometimes. We propose recognising a separate class of health-adjacent devices. These devices would not be primarily intended to diagnose or treat disease, but could still contribute to healthcare. For example, by enabling continuous monitoring outside the clinic, by generating longitudinal data, we spoke about it before, or to support research on clinical trials. These devices may also be used by consumers to manage aspects of their health, particularly about prevention, but with clear scientific grounding and transparency. To support this ecosystem, we propose several measures. The first one is an EU-level oversight to avoid fragmented national approaches and to improve monitoring of emerging devices. The second, clearer transparency on labelling, so users understand what level of evidence and validation supports a product. Third, stricter rules around health-related marketing claims, particularly when products target vulnerable populations. Fourth, practical compliance and guidance to help companies navigate existing frameworks such as the MDR, the AI Act and the GDPR. And finally, enabling safe interoperability with healthcare systems so that clinically relevant data could potentially integrate into health records. I will leave it here. Thank you very much for your attention.

(Slide 17)

### **Eva Winkler**

Herzlichen Dank, Frau Timón. Wir werden sicher an die Diskussion, was möglich ist, und auch Ihre Policy-Vorschläge im Anschluss anknüpfen. Das

interessiert uns sehr, gerade in diesem Borderline-Bereich.

Als nächsten Redner darf ich Herrn Bertolt Meyer ankündigen. Er ist Professor für Arbeits- und Organisationspsychologie an der TU Chemnitz. Wir wenden uns jetzt Fragen zu, wie diese Technologien gesellschaftlich wahrgenommen und genutzt werden, insbesondere im Kontext von Medizin und Enhancement. Er hat maßgeblich in dem Projekt „Hybride Gesellschaften und Mensch-Technik-Interaktionen“ geforscht. Wir sind gespannt auf die Ergebnisse und Ihre Expertise, Herr Meyer.

### **Neurotechnologie und Human Enhancement: Akzeptanzfaktoren und psychosoziale Risiken**

**Bertolt Meyer · Technische Universität Chemnitz**

(Folie 1)

Herzlichen Dank für die Einladung. Es ist eine große Ehre, heute hier sein zu dürfen.

Ich beginne mit einem persönlichen Wort. Ich bin selber mit einer sichtbaren Körperbehinderung zur Welt gekommen. Mir fehlt von Geburt an der linke Unterarm. Ich komme also selber in den Genuss, zu erleben, wie der technologische Fortschritt im Bereich der körpernahen Technologien voranschreitet. Das ist nicht unbedingt Neurotech, aber ich trage seit September ein neues Modell, in dem ein Mustererkennungsalgorithmus läuft, der versucht, die Signale an der Hautoberfläche in Echtzeit verschiedenen Bewegungsmustern zuzuordnen. Das ermöglicht mir, die Hand zu steuern. Das setzt nicht am Gehirn an, sondern am EMG-Signal, aber es ist wirklich beeindruckend. Man kann damit Sachen machen, die über die biologischen Grenzen des Körpers hinausgehen. Das ist mein persönlicher Bezug zum Thema.

(Folie 3)

Ich möchte einsteigen mit einer Begriffsklärung, die wir in unserer Forschung zu diesem Themenbereich verwenden: Das ist der Unterschied zwischen Human Enhancement, Human Augmentation und assistiver Technologie. Wir fokussieren hier auf Human Augmentation. Darunter verstehen wir in Anlehnung an De Boeck und Vaes gerätebasierte Technologien, die menschliche Fähigkeiten erweitern, als Integration oder Erweiterung zum menschlichen Körper. Assistive Technologien, die sich an Menschen mit Beeinträchtigungen richten, sind eine Untergruppe von Human Augmentation.

Human Augmentation ist wiederum ein Unterbegriff von Human Enhancement, der gerätebasierte und nicht gerätebasierte Methoden wie zum Beispiel Geneditierung, Medikamente, Drogen oder operative Eingriffe einschließen kann. Aber solange es sich um gerätebasierte, körpernahe oder im Körper getragene Geräte handelt, die auf menschliche Fähigkeiten und Fertigkeiten einzahlen, sprechen wir von Human Augmentation Technologies, kurz HATs oder Hats.

(Folie wechselt,  
nicht Teil des Präsentations-PDF)

Die Human Augmentation Technologies sind in aller Munde. Ich möchte Ihnen kurz ein YouTube-Video anspielen von Elon Musks Neuralink-Firma, die in den letzten Jahren einige Innovationen im Bereich von implantierbaren EEG-Elektroden erzielt hat.

Die Idee ist: Eine Fadenelektrode wird durch einen operativen Eingriff auf den Kortex einer Person aufgebracht und kann dann live, ohne dass man von außen etwas anbringt, die Hirnsignale aufzeichnen.

Das ist eigentlich nichts bahnbrechend Neues. EEGs kennen wir seit Jahrzehnten. Aber das Besondere ist hier, dass das mit einer KI verbunden wird. Die wird trainiert auf das spezifische EEG-Muster einer Person und ist sozusagen der Übersetzer. Aus den Signalen werden durch eine KI die entsprechenden Features extrahiert. Die Innovation liegt in dem Zusammenkommen von hochauflösender Messung direkt am Gehirn plus KI, die das in Echtzeit entschlüsseln und kategorisieren kann.

(Folie 4)

Vor zwei Jahren sah das Produkt *so* aus.

(Folie 5, Video;  
im Präsentations-PDF nicht sichtbar;  
siehe Videomitschnitt)

Hier sieht man es im Tierversuch. Sie sehen einen Makakenaffen, dem dieses Produkt implantiert wurde.

Hier lernt das Implantat oder die KI, während der Affe dieses Spiel spielt, Der Affe spielt das Spiel und bewegt den Joystick so, dass der Ball auf dem orangen Feld landet. Dabei werden die Hirnströme aufgezeichnet, und die KI lernt, die Hirnströme mit den Muskelbewegungen und den Intentionen zu verknüpfen.

Hier ist der Joystick abgebaut. Der Affe bewegt den Joystick nicht mehr, sondern das Spiel wird ausschließlich durch das Neuroimplantat gesteuert, also durch die Hirnwellen und die Entschlüsselung mit der KI. Das geht so schnell und so gut, dass es reicht, um live ein Computerspiel zu spielen.

Wenn Sie sich fragen, wofür dieses Röhren da ist: Da kommt immer ein Schluck Bananenshake raus als Belohnung, wenn das Äffchen es richtig gemacht hat. Die Geschichte ist allerdings ein bisschen traurig, denn wir wissen, dass dieses Äffchen inzwischen verstorben ist.

Aber ich denke, das gibt einen guten Eindruck in die Art von Technologie, mit der wir es hier zu tun haben: nämlich eine hochauflösende Signalerfassung und Messung am Körper plus eine Verknüpfung mit einer KI zur Signalentschlüsselung und -detektion.

(Folie 6)

Die verschiedenen Möglichkeiten der Human-Enhancement-Technologien lassen sich nach dem Dafürhalten von mir und vielen meiner Kolleginnen und Kollegen sehr gut durch dieses Human-Augmentation-Framework von De Boeck und Vaes klassifizieren. Sie schlagen vor, dass man in diesem breiten Technologiecluster unterscheiden muss: Welche Fähigkeit des Menschen soll augmentiert werden und mit welchem Zweck?

Der Zweck ist: Soll eine Fähigkeit, die Menschen normalerweise haben, zum Beispiel mit der linken Hand greifen, wiederhergestellt werden, wenn diese Fähigkeit fehlt oder abhandengekommen ist? Hier sind wir im Bereich der assistiven Technologie, die Unterstützung für Menschen mit Behinderungen und Beeinträchtigungen zur Verfügung stellt. Das steht in der linken Spalte: „Replicate“.

In der Mitte steht „Supplement“: unterstützen, erhalten. Das ist das Exoskelett beim Industriearbeiter, das den Körperablauf unterstützt und dafür sorgt, dass die Arbeit weniger belastend, weniger schädlich ist. Das sind durchaus präventive Einsätze: Geräte, die mir eine Fähigkeit, die ich schon besitze, leichter machen oder sie länger über die Lebensspanne erhalten.

In der rechten Spalte kommen wir in den Bereich „Exceed“, also übertreffen. Das sind Technologien, die darauf abzielen, die natürlichen Fähigkeiten des Menschen zu erweitern, Fähigkeiten zur Verfügung zu stellen, die über das natürliche Maß an menschlichen Fähigkeiten hinausgehen.

Für diese drei verschiedenen Dimensionen kann man nach der Fähigkeit unterscheiden: körperlich, sensorisch, kognitiv, sozial.

Wenn man eine ethische Bewertung dieser Technologien vornehmen möchte, liegt es nahe, dass die sich unterscheiden werden, vor allen Dingen nach der Dimension.

(Folie 7)

Dazu forschen Menschen in meiner Gruppe. Ich möchte kurz zeigen, dass die meisten Fähigkeiten des Menschen normal verteilt sind im Sinne von: Es gibt eine durchschnittliche Fähigkeit. Wenn wir durch eine Human-Augmentation-Technologie eine unterdurchschnittliche Fähigkeit in den durchschnittlichen Bereich kriegen, sprechen wir von *restore*, wenn wir eine durchschnittliche Fähigkeit erhalten, von *maintain*, und wenn wir eine durchschnittliche Fähigkeit verbessern in den Bereich, der noch menschlich ist, von *improve*. Wenn es in den übermenschlichen Bereich geht, sprechen wir von *expand*.

Meine Doktorandin Theresa Wehrle hat eine Studie gemacht, wo 1200 Personen eine Beschreibung vorgesetzt bekamen: Es wurde immer ein Brain-Computer-Interface [BCI] beschrieben, ähnlich wie die Technik, die wir gerade in dem Video gesehen haben. Der Beschreibungstext, der diese Technologie vorgestellt hat, variierte danach, ob dadurch eine verlorene Fähigkeit wiederhergestellt wird, ob eine Fähigkeit erhalten wird oder hinzukommt. Der Text variierte auch danach, was für eine Art von Fähigkeit dieses Implantat unterstützen soll: motorische oder kognitive Fähigkeiten.

(Folie 8)

Dann hat sie unterschiedliche Akzeptanz- und Bewertungsfaktoren gemessen, und es stellte sich heraus: Die Menschen erwarten die stärksten Zugewinne und den stärksten Nutzen einer solchen

Technologie im Bereich der körperlichen Unterstützung, der motorischen Unterstützung und vor allem im Bereich der restaurativen, also der assistiven Technologien. Aber in dem Bereich, wo es ins Übermenschliche geht – das, was *expand* ist, das sind die hellgrauen Balken –, da ist heute die Erwartung, dass es mir was bringt, am geringsten. Sehr gering ausgeprägt ist diese auch bei Social Influence; das ist die rechte Grafik. Hier geht es darum: Wie stark erwarten die Menschen, dass es eine Art normativen Druck geben wird, so etwas zu nutzen?

Was ich interessant finde: In der rechten Abbildung deutet der dunkle Balken bei „Motor“ darauf hin, dass bei assistiven Technologien, zum Beispiel einer Prothese, eine Art normativer Druck wahrgenommen wird: Wenn es so eine Technik gibt, dann solltest du sie auch verwenden, wenn sie dir nützen würde.

(Folie 9)

Sie hat auch eine zweite Studie gemacht zu einer möglichen Akzeptanz eines Brain-Computer-Interface, das im Auto verbaut ist, zum Beispiel in der Kopfstütze. Wenn man sich ins Auto setzt, könnte ja das Auto die Hirnwellen messen. Hier hat sie verschiedene Dimensionen der Technologieakzeptanz gemessen.

Es zeigt sich, dass die Menschen es eher akzeptieren, wenn sie wissen, dass es für Menschen mit Behinderung konzipiert ist. Wenn sie hingegen denken, dass es für den Alltagsgebrauch konzipiert ist, dann ist die Akzeptanz – und zwar unabhängig von der Art der Technologieakzeptanz – signifikant geringer.

Sie merken: Es zeichnet sich ab, dass die Akzeptanz gegenüber assistiven Technologien höher ist als die Akzeptanz gegenüber Technologien, die eher in den Bereich „Exceed“ hineingehen.

(Folie 10)

Das überspringe ich aus Zeitgründen.

(Folie 11)

Wir haben noch eine Studie gemacht, die sich damit beschäftigt, welche Auswirkungen diese Technologien auf Stereotypisierung und Stigmatisierung haben werden.

Stereotype kommunizieren vor allem Informationen darüber, wie gut oder böse Menschen sind – das ist die Wärmedimension: gut, schlecht, warm, kalt – und unabhängig davon, wie stark, schwach, wie kompetent oder inkompetent Menschen wahrgenommen werden; das ist hier die Horizontale. Das ist das sogenannte Stereotype Content Model der Sozialpsychologie.

(Folie 12)

Die Zuschreibungen von Wärme und Kompetenz sind universell über alle Kulturen und Epochen hinweg und determinieren, was wir gegenüber anderen Menschen fühlen und wie wir sie behandeln: Zustimmung und Unterstützung, Gleichgültigkeit und Ignoranz, Mitleid oder Ablehnung und Neid.

Wir haben eine Studie gemacht, in der wir geschaut haben: Was sind die Zuschreibungen von Wärme und Kompetenz gegenüber verschiedenen Arten von Behinderungen und gegenüber verschiedenen, wenn Sie so wollen, Upgrades?

(Folie 13)

Das überspringe ich.

(Folie 14)

Die Folie wollte ich Ihnen unbedingt zeigen. Sie sehen in der Horizontale die zugeschriebene Kompetenz gegenüber verschiedenen gesellschaftlichen Gruppen und in der Vertikalen die zugeschriebene Wärme. Wärme: gut, böse; Kompetenz: stark, schwach.

Schauen Sie mal: Körperbehinderte. Das findet sich etwas oben links, bei Querschnittslähmung. Das ist das typische Stereotyp gegenüber Menschen mit Körperbehinderung: nett, aber inkompetent. Das ist das Stereotyp des netten Behinderten. Das ist das Stereotyp, das ich erlebe, wenn ich ins Restaurant gehe und ein Stück Fleisch bestelle, und wenn es kommt, ist es schon kleingeschnitten, wie bei einem Kind, weil der Kellner mir nicht zutraut, dass ich mit Messer und Gabel essen kann.

Jetzt schauen Sie mal, wo sich das hinbewegt, wenn man das Label verändert von „körperbehindert“ hin zu „körperbehindert mit bionischen Prothesen“. Dann verschiebt sich das ganz weit nach rechts, ganz nah an „nicht behindert“ heran. Das heißt, eine Technologie wie diese ist in der Lage, wenn sie als assistiv wahrgenommen wird, das Stigma zu kompensieren. Die Technologie bietet also nicht nur einen funktionalen, sondern auch einen sozialpsychologischen Nutzen.

Aber es kommt extrem auf das Label und auf die Nutzungsintention an. Denn schauen Sie mal, wo Nichtbehinderte mit Upgrades und Cyborgs landen: Die landen unten rechts. Das sind die Bösen. Das fühlt sich kalt und bedrohlich an. Und weil diese Zuschreibungen beeinflussen, was wir gegenüber anderen Menschen fühlen und wie wir uns ihnen gegenüber verhalten, zeigt sich hier, dass es soziale Kosten haben wird, wenn wir anfangen, uns im großen Stile upzugraden. Wir gewinnen vielleicht Funktionen hinzu, aber wir verlieren Ansehen in den Augen unserer Mitmenschen, die uns negativ stereotypisieren werden.

Für alles Weitere habe ich jetzt keine Zeit mehr. Gerne noch in der Diskussion. Ich bedanke mich für die Aufmerksamkeit.

## **Eva Winkler**

Herzlichen Dank, Herr Meyer. Viel Stoff für die Diskussion nachher und auch für die gesellschaftliche Wahrnehmung.

Wir switchen jetzt in den Onlineraum und richten den Blick auf die Perspektive junger Menschen und auf Fragen der mentalen Gesundheit. Ich darf Frau Ilina Singh begrüßen, Professorin für Neuroscience and Society, also Neurowissenschaften und Gesellschaft, in Oxford und eine international führende Forscherin zu den ethischen gesellschaftlichen Dimensionen von Neurowissenschaften und KI, insbesondere auch in der partizipativen Forschung mit jungen Menschen. Sie berichtet über Ergebnisse ihrer Studien zur KI-basierten Anwendung in diesem Kontext. Vielen Dank, dass Sie heute bei uns sind.

## **AI as neurotechnology: Tools and self-diagnosis in young people**

### **Ilina Singh · University of Oxford**

Thank you und danke. Es tut mir leid, dass ich nicht dabei sein kann. And I'm going to switch to English, if I may. Thank you to the team for sharing my slides.

(Slide 1)

I'm going to talk mainly about one use case today just for the sake of time, which will be companion apps. But I also want to introduce a way of thinking about these problems and how we can approach them ethically with young people. And that, I think, is a more generalisable approach.

(Slide 2)

I want to start with an argument that we should be thinking about certain AI tools as a neurotechnology. Broadly speaking, neurotechnologies are technologies that act on the brain. When I was doing early work with medication, I used to say,

well, nobody is thinking about psychopharmacology as a neurotechnology, but of course you can take this to an extreme and say, well, spectacles are a neurotechnology. But I want to take a case here where we can think about some of these tools as endangering, having certain kinds of ethical risks, and endangering young people in particular in certain kinds of ways. So, classic neurotechnology, as we have been hearing and we will hear more this afternoon, can be either invasive or non-invasive. It often has a kind of materiality to it. It can be used for all kinds of purposes, including treatment and enhancement and learning and adaptation.

And there are standard ethical considerations that have been really beautifully dissected over many years now in the ethics community. Artificial intelligence also acts on the brain, but it has a range of forms, and those forms are still emerging. And the use case I am going to be talking about today, companion AIs, are really rapidly changing in their ability to actually be embodied. So when I asked AI to come up with a summary table of the cognitive impacts of how AI acts on the brain, it came up rapidly with this summary table, which I thought I would present to you both as an example of what AI can do very, very quickly, but also as a reasonably decent summary of the cognitive impacts, both the positives and the negatives, that AI can have on the brain. I think with these kinds of tools, it's always good to consider both the risks and the opportunities. And certainly when we talk to young people, we hear very much about the benefits as well as the downsides. Next slide, please.

(Slide 3)

So, my use case is generative AI, which, at the broadest level, is AI tools that can generate text, image and sound. And most of us are now very familiar with large language models, which is a

subset of generative AI, across a range of industries, sorry, across a range of, of these AI corporations. In Oxford we have adopted ChatGPT. Personally, I tend to use Claude AI for research. There is Google Gemini, which generated the table that you saw in the previous slide. There have been several generations of these LLMs, but these are generally used for research and conversation. But I want to talk specifically about companion AIs, which I think, among adults who I know, are less often used. And these are tools that really are built to be companions to people. And the data on young people's use in the UK was actually quite surprising to me. If we go to the next slide, please.

(Slide 4)

I think that is on the slide following this. But let me here introduce you to one kind of an AI companion that you can build. The idea here is that these AI companions are almost completely customisable. I've just taken a screenshot here, where you can design what the AI looks like. You can also design their cognitive and their affective features, you can design their embodied typologies. And they do not have to be human, they can be other than human. They can be more avatar-like, more manga-like. And the interesting thing to me about when I looked at how this technology is moving is that there is an increasing move towards embodying these companions so that you can actually also take them with you. I'm reminded sometimes, since we are in a German context, of Goethe's homunculus. That was an independently created spirit, a minded spirit. But now these homunculi, if I can put it that way, can be transported with a young person. And so they will sit on a windowsill or on a desk and be in a kind of holographic embodiment at the moment. They could take on other forms in the future. This is important because it is not just a, if you will, a rela-

tionship with one's computer, it is going to become increasingly an embodied relationship with an entity that is created in one's own image, only in a certain sense, in that it is a sense of what kind of companionship, what kind of dialogue, what kind of cognitive scaffolding, what kind of affective scaffolding, one might want in the future. Next slide, please.

(Slide 5)

It is important to say that humans seek AI companions for many reasons and some of those are perfectly legitimate reasons and many of them are about supporting our mental health and wellbeing. And we have heard from the first speaker in particular that people are looking for neurotechnological tools to enhance their wellness experiences, to understand their mental health better, to treat mental health difficulties. And so companion AI is a further such tool. I think we need to be reasonable in our understanding of that some of this is beneficial to people and to young people. But there are, of course, additional risks of these AI chatbots that I think, particularly in these embodied forms, become more of a substantive risk when we are talking about young people. We know that they are designed to be addictive and we know that that has implications for a loss of agency and for compromised autonomy. And this is, of course, again, particularly important for people who are already vulnerable to certain kinds of addictions. We also know that they are designed to be sycophantic, and this is part of the addiction. The AI companion is designed to mirror the human, to support the human, to tell them they are wonderful, to increasingly progress their own goals. And we have seen cases where this can, again, exacerbate vulnerabilities, particularly for young people who have a desire to self-harm or to harm others. There is also a risk of being drawn into AI hallucinations. A kind of, if you

will, folie à deux can happen where the AI hallucinates reality, the user engages in that hallucinated reality, and we see this particularly with young people who are already at risk for psychosis, and this can be rapidly very, very worrying and harmful. And then, again, for me this is one of the more important aspects, this loss of an embodied boundary. So, as these AIs become more embodied, there is a risk that people will lose that bounded sense of self and begin to merge in a certain way with the AI. And if you speak to young people today, you will probably be hearing them say already, "I don't see so much of a difference between the physical world and the virtual world. I move seamlessly among the two." We can see that this is a real risk. And then the fact that we do not know what kind of relationship we ought to have with a companion AI. Human care norms are applicable for us in relation to this AI, but not from the AI back, because the AI is not a human. So, what are the obligations in this relationship with an AI companion? Next slide, please.

(Slide 6)

As I said, more young people are using these AI companions than certainly I would have expected. And this is just UK data showing that up to a third of teenagers in this study are using the AI on a monthly or a weekly basis. So it's more than curiosity. Next slide, please.

(Slide 7)

The question is: What should we do? And I just wanted to give some quick definitions here about the kinds of terms I am going to be talking about. Epistemic paternalism is constraining a person's access to information or inquiry 'for their own good'. Epistemic justice recognises, on the other hand, the user's capacity and agency to

know and interpret their intentions and experiences. But I want to be clear that if we say we want to respect epistemic agency, that is not equivalent to ignoring vulnerabilities. Next slide, please.

(Slide 8)

A way forward we propose is a consent-based partnership with young people. We have seen with the, certainly in the UK but also more widely, the proposal for social media bans or AI bans as a way of managing the risks around AI. You know, we think that this displays a kind of excessive paternalism, and we think that this does not do justice to young people's agency and their ability to actually engage and, frankly, to teach us about the risks and harms of these kinds of technologies. And we think that, of course, we cannot do nothing because we do have vulnerabilities and we do already have significant use cases of harms. So what is a consent-based partnership? It balances the tension between autonomy and harm prevention, and it also grounds this human–AI interaction in an understanding of relationship that is modelled on care and collaboration. Next slide, please.

(Slide 9)

So, we think that young people can go through a process of self-assessment with trusted and supported others. And we have started doing this with young people, taking account, for example, of their personal factors, their mental health, vulnerabilities, what kinds of cognitive and emotional factors they have and where they want support and where they think the risks are, and also what kinds of educational and training support they want. Next slide, please.

(Slide 10)

I have about three slides left. Some of this is already in the engineered design factors in the tier

one. These are default safeguards that are rapidly emerging as well, as industry best practices are emerging. And I have just given some exemplars of the core protections here that, as I say, already exist across a lot of these platforms. Next slide, please.

(Slide 11)

The tier two safeguards are where the interest comes in. This is a proposal that the users, and this includes young users, should be able to activate safeguards at their discretion. These are opt-in and out safeguards that emerge over time because, of course, the users themselves are going to be changing over time. And the idea here is that this is decided in partnership with trusted Bublitz supporters. And again, we have given some idea about potential design tools that could be built into the design of these AI models. Next slide, please.

(Slide 12)

So, we do this in our lab, which is the Design Bioethics Lab, which I will not go on to talk about at any length, but you are welcome to look up online. And the next slide.

(Slide 13)

And this is an example of work that we are doing exactly in this area which is led by Dr Madeline Reinecke in my group, where we are working with the government to actually create a threat model that is generated by UK adolescents so that they can tell us about the risks and harms that they perceive and what kinds of factors they want to have built into these tools to help keep them safe. Next slide, please.

(Slide 14)

This is just the concluding thought, which is that we really should be thinking about a positive future for young human–AI relationships. We should be asking what kinds of relationships are sustainable, are positive, are beneficial and keep people from harm, and how the designs themselves can be adaptable and personalisable to accommodate both the risks and vulnerabilities, but also the desires and the goals of young people. Thank you very much.

### **Eva Winkler**

Herzlichen Dank, Frau Singh, für den Überblick und die Perspektive auf die partizipativen Ansätze mit jungen Menschen. Darauf werden wir in der Diskussion sicher noch zurückkommen.

Zum Abschluss dieses ersten Blocks wenden wir uns jetzt grundlegenden ethischen Fragen der Mensch-Technik-Beziehung zu. Ich darf ganz herzlich Frau Saskia Nagel begrüßen. Sie ist Professorin für Angewandte Ethik an der RWTH in Aachen und arbeitet an der Schnittstelle von Philosophie, Kognitionswissenschaft und Technik mit einem Schwerpunkt auf Themen wie Autonomie, Verantwortung und Vertrauen. Wir sind gespannt auf Ihre Expertise.

Wir hatten schon in der Pause darüber gesprochen, dass das eine Diskussion ist, die schon vor zehn Jahren intensiv geführt wurde, aber jetzt mit einer ganz anderen Dringlichkeit, weil die Technik so dynamisch voranschreitet.

### **Neurotechnologien und ihre ethischen Herausforderungen für zwischenmenschliche und Mensch-Technik-Beziehungen**

**Saskia Nagel · RWTH Aachen University**

(Folie 1)

Vielen Dank für die Einladung.

(Folie 2)

Ich beginne mit drei Vorbemerkungen.

[1] Neuere Neurotechnologien erlauben insbesondere neuere Formen der Interaktion mit Technologie: Wir interagieren nicht mehr über Maus und Tastatur, sondern haben eine direkte Interaktion mit der Technologie. Das ist, glaube ich, einer der ethisch relevanten Unterschiede.

Es ist aber nicht alles neu. Wir können aus vielen Debatten, die wir in der Bioethik, Medizinethik, Technikethik schon lange haben, lernen. Das ist mir immer ganz wichtig, damit wir nicht den Eindruck haben, auch in der Ethik das Rad neu erfinden zu müssen. Es gibt viele Debatten, von denen wir lernen können.

[2] Wir haben eine sehr heterogene Gruppe von Technologien: Wir haben nichtinvasive, die EEG-Technologien, die die Kollegin am Anfang auch als die große Gruppe der Neurotechnologien vorgestellt hat, bis zu hochinvasiven Technologien. Auch Neuralink würde ich vielleicht dazu fassen.

Das bedeutet auch, dass wir unterschiedliche Personen benötigen, um eine Neurotechnologie zu nutzen. Manche Technologien können Sie zu Hause selbst bauen; dann haben Sie ein Do it yourself. Andere Technologien brauchen immer noch medizinische Unterstützung. Das ist ein Riesenschied, weil da auch die Professionsethik teilweise mit reinspielt.

[3] Letztlich haben wir sehr heterogene Kontexte. Wir sprechen natürlich über sehr vielversprechende und segensreiche Anwendungen in der Medizin. Ich glaube, da gibt es viel weniger Fragezeichen und Sorgenfalten. Es gibt aber auch Anwendungen – das ist das, was heute im Fokus steht – in der Bildung, im Marketing, beim Gaming usw. Es ist natürlich zu unterscheiden, in welchen Bereichen wir uns da bewegen.

(Folie 3)

Ich habe nach den Gesprächen mit der Geschäftsstelle einen Gemischtwarenladen mitgebracht. Wir können beim Thema Neurotechnologie unglaublich viele ethische Fragen stellen. Ich bin in der Diskussion gern bereit, zu allen möglichen Fragen Rede und Antwort zu stehen. Ich habe drei Schlaglichter mitgebracht [Folie: „1) Enhancement, 2) Arbeitswelt, 3) Verantwortung“], die sich gut ergänzen mit den anderen Beiträgen, die wir hier haben.

(Folie 4)

Ich beginne mit dem Thema Enhancement und habe hier eine kleine Definition mitgebracht, sage aber gleich: Diese Definition ist nicht so gut, wie wir sie gerne hätten. Wir haben seit 20 Jahren Communitys zum Thema Enhancement. Bis jetzt hat niemand eine konsensfähige Definition vorgeschlagen. Das ist eine, die so lala funktioniert: „Erweiterungen oder Steigerungen von Fähigkeiten Gesunder durch Medikamente oder Methoden“ (oder Geräte, wie der Kollege vorgeschlagen hat) „die zur Heilung Kranker entwickelt wurden.“

Jetzt haben wir die Schwierigkeit, dass wir da in den Graubereich zwischen Gesundheit und Krankheit hineinfallen und uns überlegen müssen: Wo genau ziehen wir die Grenze? Was ist noch krank, was ist nicht krank? Ich komme gleich dazu, wo da eine ethische Problematik ist.

Wir können auch versuchen, Enhancement zu definieren, indem wir sagen: Es geht über das Natürliche hinaus. Dann kommen wir näher an das transhumanistische Versprechen, die Überschreitung biologischer Fähigkeiten. Oder wir sagen: Es geht über das Normale hinaus. Das Normale, das Natürliche und Gesundheit/Krankheit bringen alle einen Strauß an ethischen Fragestellungen und Herausforderungen mit sich. Das heißt, wir

sind definitorisch mit dem Begriff Enhancement noch nicht weiter.

Ich glaube, wo man sich einig sein kann, ist, dass generell versucht wird, über Selbstgestaltung jeglicher Art, Selbstoptimierung – es muss nicht immer das Optimum sein, aber Richtung Optimum, Streben nach Vollkommenheit – in irgendeiner Weise in Richtung gutes Leben zu kommen.

(Folie 5)

Die erste normative Frage im Kontext Enhancement ist: Was sollen wir unter normal und Normalität verstehen?

(Folie 6)

Das ist im Ethikrat kein neues Thema. Es gibt ein sehr schönes Papier des Ethikrates von vor ein paar Jahren, das Normalität, Normalisierung und Normativität auseinanderdröselte. Deswegen mache ich das hier nicht und beziehe das jetzt nur auf das Thema Enhancement.

Die Norm zu übersteigen ist Ziel vieler Enhancement-Interventionen, sowohl im Marketing als auch in der Selbstbeschreibung. Man versucht, über das Normale hinauszugehen.

Dabei haben wir oft Normalisierungseffekte. Vorher ungewöhnliche Fähigkeiten werden zum Normal. Das heißt, wir verschieben einfach die Gauß-Kurve.

Natürlich haben Vorstellungen und Beschreibungen von Normalität ethische und soziale Implikationen. Oft klingt es so, als wäre das eine Beschreibung, die man einfach statistisch erklären kann: Das ist Normalität. Das ist aber überhaupt nicht so. Es ist eine enorm wertgeladene Beschreibung.

Ich finde nach wie vor das Buch von Jonathan Glover aus dem Jahr 1984 mit dem Titel „What Sort of People Should There Be?“ ganz zentral für diese Fragestellung: Wer sagt eigentlich und wer

hat die Macht, zu entscheiden, welche Art Mensch wir sein sollen?

(Folie 7)

Zweiter Punkt: Eng verbunden mit dem Thema Normalisierung ist die Frage der Medikalisierung. Vormals „normale“, gesunde mentale Zustände werden als medizinisch zu behandelnde Probleme beschrieben.

Dann haben wir so etwas wie Condition Branding. Es gibt plötzlich Begriffe für medizinische Kategorien, die sich den technologischen Entwicklungen anpassen.

Wir haben unter anderem im Kontext von ADHS breit diskutiert, wie das funktioniert und wie auch die Behandlungsgrenze im Kontext dieser Medikalisierung gesenkt wird. Vormals natürliche Prozesse wie zum Beispiel kindliche Entwicklungsprozesse, Aufmerksamkeit, Stimmungen werden dann als behandlungsbedürftig beschrieben.

Ein Risiko ist eine Stigmatisierung. Aber so einfach ist es nicht. Denn auf der anderen Seite ist gerade das Potenzial einer Medikalisierung eine Destigmatisierung. Man kann auch sagen, dass das Labeling das Stigma reduziert. Es ist dann Aufgabe der Psychologen, zu überlegen, an welcher Stelle es für eine Stigmatisierung und wo für eine Destigmatisierung sorgt.

Relativ deutlich ist, dass wir hier einen Techno-Solutionismus riskieren, also mit technologischen Mitteln versuchen, Situationen oder Phänomene zu behandeln oder darauf zuzugehen, die eigentlich sozial anzugehen wären.

(Folie 8)

Ein Risiko beim Thema Enhancement von Neurotechnologien ist das Thema Manipulation. Dazu habe ich insbesondere Fragen mitgebracht.

Eine der wichtigsten Fragen, die mich in unserer Forschung sehr beschäftigt: Wie lässt sich eigentlich eine subtile Einflussnahme, die wir nicht gutheißen, abgrenzen von einer legitimen Verhaltens- oder Entscheidungsunterstützung, die wir sehr wohl gutheißen? Wo ist da die Grenze? Da muss man sich mit dem Thema Autonomie beschäftigen: Ab wann fühlen wir uns in unserer Autonomie eingeschränkt? Können wir das objektiv beschreiben?

Daran anschließend: Unter welchen Bedingungen kann Einwilligung überhaupt als informiert und reflektiert gelten, wenn neuronale Daten genutzt werden, um Verhalten zu modulieren, insbesondere wenn mir das nicht bewusst ist?

Ein wichtiges Thema, und da kommen wir schon Richtung Anwendung auch im Bereich KI: Inwieweit sind die Mechanismen neurotechnologischer Systeme für Betroffene überhaupt nachvollziehbar? Das brauchen wir für unsere normalen Prozesse der Einwilligungsfähigkeit. Wir müssen ja irgendwie die Nachvollziehbarkeit sicherstellen. Das wird immer schwieriger, je komplexer die Systeme werden. Wenn wir zum Thema KI kommen, ist das eines der Hauptprobleme.

Dann haben wir Gerechtigkeitsfragen. Gerechtigkeit habe ich mitgebracht beim Thema Manipulation, weil mir sehr wichtig ist, dass man sich auch die Machtstrukturen anschaut. Wir haben die Big-Tech-Firmen schon an verschiedener Stelle heute genannt, und die Frage ist natürlich: Wer hat warum Zugang zu potenziell manipulativen Technologien?

(Folie 9)

Ich wusste, dass Ilina Singh heute da ist und vor mir spricht. Deswegen in aller Kürze: Wenn wir über Enhancement reden, sind Kinder und Jugendliche *die* vulnerable Gruppe, die wir uns besonders anschauen müssen und wo wir besonders

überlegen müssen: Welche Arten von Regulierung brauchen wir und welche werden Kindern und Jugendlichen gerecht?

Das sind sensible Phasen von Autonomieentwicklung und Identitätsbildung. Wenn wir da neuronal intervenieren, kann das eine ganz andere Macht und Auswirkungen auf die Entwicklung der Kinder haben, als wenn wir das bei Erwachsenen machen.

Natürlich sind auch die Anforderungen an Einwilligung bei Kindern andere und viel sensibler anzugehen. Der Schutz vor Überwachung und der Schutz vor Manipulation, bei Neurotechnologien wie auch bei anderen Technologien – das ist wieder mein Punkt vom Anfang. Manche Diskussionen, zum Beispiel zum Thema Social Media, müssen wir hier ganz ähnlich führen, zum Thema Schutz vor Manipulation.

(Folie 10)

Ein zweites Schlaglicht, das mir wichtig ist und das im Diskurs bleiben sollte: Neurotechnologien am Arbeitsplatz.

Es gibt viel Potenzial, den Arbeitsplatz zu verbessern und für Arbeitnehmer eine Sicherheit zu schaffen. Wenn Sie sich vorstellen, es gibt EEG-Headsets, die vielleicht Baggerführern oder Minensuchenden helfen, zu navigieren oder die Aufmerksamkeit zu halten. Es wird auch im Bereich Pflege diskutiert, wie Neurotechnologien da hilfreich eingesetzt werden können. Da gibt es Vorteile. Gleichzeitig gibt es Vorteile, dass es Werkzeuge zur Selbstregulation werden können, also dass Sie selbst feststellen können: „Jetzt meldet mir mein Gerät, dass ich schon zu lange dabei bin oder langsam gestresst bin.“ Dann hätten Sie die Möglichkeit, Ihren Arbeitsalltag oder Ihre Arbeitsrhythmen anzupassen.

Das ist die schöne Welt. Die nicht so schöne Welt ist, dass das auch zu Optimierungsdruck, entweder Selbstoptimierungsdruck von sich aus oder Optimierungsdruck von außen führen kann.

Was ich nicht unterschätzen möchte, ist das Thema Abhängigkeit von den Geräten und Technologien. Wenn wir Neurotechnologien zum Beispiel in Verwaltungen standardmäßig einsetzen würden, sind wir abhängig von diesen Geräten, den Updates, den Unternehmen.

Das ist die andere Seite: Wir beeinflussen auch die Arbeitserwartungen, die wir stellen. Auf eine Folie der Kollegin stand: „Work smarter, not longer.“ Ja, das klingt erst mal gut. Das heißt aber auch: Es ist zu erwarten, dass wir smarter arbeiten.

Das bringt mich zum letzten Punkt von Neurotechnologien am Arbeitsplatz: Auch Fragen zu Überwachung und Macht müssen hier diskutiert werden. Chilling-Effekte, Abschreckungseffekte sind immer ein Thema, wenn es um Privatsphäre geht, ebenso Neuro-Ableismus, dass eben eine bestimmte Art von Arbeitsfähigkeit erwartet wird.

(Folie 11)

Die Grundfrage, die wir uns stellen, ist: Welche Werkzeuge wollen wir für welchen Zweck in welchen Arbeitskontexten einsetzen? Das ist eine wichtige und interessante Frage, und grundlegend natürlich: Wie soll überhaupt Arbeit aussehen?

(Folie 12)

Letztes Thema: KI und Neurotechnologien. Ich glaube, weil wir KI haben, wird das Thema Neurotechnologien jetzt so spannend. Die Fragen nach Privatsphäre stellen sich anders; die Frage nach Beeinflussung von Entscheidungen stellt sich anders.

Ich glaube auch, dass wir – noch nicht jetzt, aber in naher Zukunft – viel weniger über Werkzeuge nachdenken, sondern eher über Schnittstellen und über Mensch-Technik-Hybridsysteme, also nicht mehr *wir* als Menschen handhaben diese Technologien wie ein Werkzeug. Das war bis vor Kurzem meine präferierte Sicht auf Technologiegebrauch: Ich als autonomer Mensch handhabe ein Werkzeug. Das wird mit KI immer weniger so sein. Das wird immer mehr eine hybride Mensch-Maschine-Systeme-Sicht sein, und das sorgt dafür, dass wir eine Reihe von neuen Fragen haben, die in der KI-Ethik – das ist eine wachsende Community – diskutiert werden.

(Folie 13)

Ich bringe nur ein Thema mit, weil es eines meiner Lieblingsthemen ist, aber auch, weil ich glaube, dass es in diesem Kontext ein besonders wichtiges Thema ist: Das ist das Thema Verantwortung.

Wenn wir Neurotechnologien einsetzen, die KI-gesteuert sind oder mit KI kombiniert funktionieren, dann haben wir eine Diffusion von Verantwortung, weil wir so viele Player und ein komplexes Geflecht an möglichen Verantwortungsträgern haben. Wenn dann zum Beispiel etwas schiefgeht oder nicht so funktioniert wie geplant, können wir nicht mehr sagen: „Diese oder jene Stelle war verantwortlich.“ Wir haben eine Diffusion.

Seit ein paar Jahren wird ein spezieller Fall der Diffusion diskutiert, die sogenannten Verantwortungslücken. Sie entstehen, wenn es keinen Menschen gibt, den wir verantwortlich machen würden. Wir würden eigentlich die Technik verantwortlich machen wollen, aber weil Technik kein

moralischer Agent ist, können wir das nicht. So funktioniert unsere Zuschreibung von Verantwortlichkeit nicht. Deswegen entsteht eine Lücke, die Responsibility Gap, Verantwortungslücke genannt wird.

Das ist ein ernstes Problem, nicht nur in Bezug auf Akzeptanz, sondern auch in unserem Verständnis dieser Technologien. Denn es kann passieren, dass niemand dafür, was in einem bestimmten Fall passiert, verantwortlich gemacht werden kann.

Ich betone: Das sage ich im Jahr 2026. Vielleicht sage ich 2030 etwas anderes. Aber im Moment würde ich sagen: Ganz pragmatisch sollten wir extrem darauf achten, dass wir klare Verantwortungsschemata haben, wenn wir diese Neurotechnologien zulassen. Alles, was Regulierung beachten muss, muss immer Verantwortungsschemata mitbedenken und überlegen: Welche Entscheidungsarchitektur habe ich? Wer hat an welcher Stelle als menschlicher Akteur, als moralischer Akteur eine Entscheidung getroffen, sodass ich auch sagen kann: Hier liegt eine Verantwortung für einen bestimmten Schritt oder für eine bestimmte Entscheidung.

Ich glaube, dass das nur funktionieren wird – das haben wir gerade kürzlich publiziert<sup>2</sup> –, wenn wir so etwas denken wie einen Fokus auf Verantwortungsübernahme, also aktiv Verantwortung zu übernehmen. Das mag nicht jeder, und insbesondere wenn es um Institutionen geht, wird das nicht immer gern gesehen. Ich sage das auch bewusst mit einem Blick auf die Rechtswissenschaftler im Raum: Ich rede jetzt über moralische Verantwortung, also moralische Verantwortung zu übernehmen.

---

<sup>2</sup> Folie: Conradie, N. H.; Nagel, S. K. (2025): Taking responsibility for the outcomes of autonomous technologies. In: *Ethics and Information Technology*, 27 (2), 21.

(Folie 14)

Ich schließe mit zwei Schlusskommentaren. Das sind aber nur Denkanstöße, die ich gern in die Debatte über Neurotechnologien geben würde. Ich glaube, wir müssen uns Gedanken machen über die strukturellen Rahmenbedingungen von Neurotechnologien in Europa. Obwohl es der Deutsche Ethikrat ist, würde ich es trotzdem europäisch denken wollen.

Wir haben ein sehr kompetitives Umfeld bei neurotechnologischer Entwicklung. Wir sind alles andere als im luftleeren Raum, und die Standards, die wir in Europa haben, sind andere als anderswo in der Welt. Wenn wir so etwas wollen, was wir wollen sollten, wie technologische Souveränität, dann müssen wir überlegen, wie wir das gut hinkommen, ohne den Anschluss zu verlieren.

(Folie 15)

Der letzte Punkt steht vielleicht nicht auf der Agenda, aber ich glaube, man sollte es berücksichtigen: Das ist das Thema der Doppelverwendbarkeit, Dual Use, Mehrfachverwendbarkeit. Das können wir bei Neurotechnologien nicht ausblenden.

Man kann natürlich bei jeder Regulierung und bei jeder Stellungnahme sagen: „Das blenden wir hier aus und behandeln es ein andermal.“ Aber wir können es nicht ignorieren. Und genau so habe ich es auch mitgebracht, nur als Hinweis: Wir müssen darüber nachdenken. Ob das in diesem Kontext oder in dieser Stellungnahme ist, das entscheiden andere. Aber es ist ein wichtiges Thema, weil es auch praktisch relevant wird.

(Folie 16)

Damit schließe mit einem Fazit: Wir müssen die zentralen Werte klären. Wir müssen Transparenz, Einwilligung, Schutz vulnerabler Gruppen mit Partizipation ermöglichen. Wir müssen ethische Fragen – das ist sowieso mein Credo – möglichst

früh in die Diskussion mit einbringen, und wir müssen verantwortungsvolle Forschung in Europa fördern. Ich danke Ihnen.

## **Diskussion**

### **Eva Winkler**

Herzlichen Dank, Frau Nagel, auch Ihnen und allen Sachverständigen im ersten Block. Wir haben jetzt knapp eine Stunde Zeit für die Diskussion. Ich bitte die Ratsmitglieder hier im Saal, per Handzeichen oder im Onlineraum mit Handheben zu zeigen, wenn Sie Fragen haben. Vielleicht sagen Sie auch, an wen die Frage primär gerichtet ist.

### **Nils Goldschmidt**

Ich habe eine Frage an Herrn Meyer. Ich fand das einen sehr spannenden Vortrag und vor allem spannend zu sehen, wie sehr wir auch in den Blick nehmen müssen, welche Rolle die gesellschaftliche Wahrnehmung dabei spielt.

Dabei haben wir gesagt, dass es unterschiedliche Muster gibt, was Akzeptanz usw. angeht. Aber müssen wir nicht auch in den Blick nehmen oder wie versuchen Sie es mit einzuholen, dass sich in diesem Prozess auch die gesellschaftliche Wahrnehmung verändern wird? Also dass vielleicht in drei Wochen irgendwann das Normalnull wird und man dann ein anderes Verhältnis dazu entwickelt? Man kann ja die Bewertung nicht allein auf den Status quo der gesellschaftlichen Wahrnehmung beziehen.

### **Bertolt Meyer**

Sie haben vollkommen recht. Das ist schwer vorherzusagen. Die Forschung zu Akzeptanz von Neurotechnologien – wie Forschung allgemein zur Akzeptanz von Technologien, die noch keine weite Verbreitung haben – krankt immer daran, dass Menschen sich Dinge vorstellen müssen. Das heißt, wir erzählen ihnen Geschichten über

zukünftige Technologien. Meine Doktorandin Frau Wehler hat zum Beispiel aufwendige Videos produziert, die das dann so im Video zeigen.

Aber natürlich haben die meisten Menschen noch keine umfangreichen Erfahrungen mit diesen Technologien. Manche haben noch nicht einmal von ihnen gehört. Und natürlich kann es sein, dass wir ein ähnliches Phänomen erleben wie bei der Einführung des Mobiltelefons. Die Älteren von uns können sich erinnern: Als die Handys neu waren, war das auch extrem negativ stigmatisiert. Da war man der letzte Idiot, wenn man es gewagt hat, in einem öffentlichen Verkehrsmittel an ein klingelndes Handy zu gehen. Das ist heute überhaupt nicht mehr so. Vielleicht ist das in Zukunft auch so bei Implantaten.

Nur: Wenn man sich anschaut, welche moralischen Dimensionen hier bewertet werden, und zwar elementare Vorstellungen von Fairness, dann glaube ich schon, dass sich das nicht komplett nivellieren wird, wenn die Technologien weit verbreitet sind. Denn die Verhaltenszuschreibung ist eine der wesentlichen Funktionen der zwischenmenschlichen Wahrnehmung in unserer sozialen Umwelt: Eine andere Person verhält sich, und ich versuche mir zu erklären, woher dieses Verhalten kommt. Wir sind alle die ganze Zeit Verhaltenswissenschaftlerinnen und Verhaltenswissenschaftler. Theory of Mind.

Und wenn ich mir ein anderes Verhalten erklären soll, das etwas mit Leistung zu tun hat, und die naheliegende Erklärung ist, dass das keine Leistung ist, die in diesem Menschen ist, sondern die durch etwas hinzugekommen ist, dass der Mensch sich zum Beispiel hat kaufen können, was ich mir nicht kaufen kann, dann halte ich es für naheliegend, dass das Bewertungen nach sich zieht, die sich auch in Zukunft nicht ändern werden. Denn was sich hier abzeichnet, ist ein neues oder ein weiteres Spielfeld, auf dem sich diejenigen mit

Zugang zu Ressourcen auf eine ganz andere Art und Weise funktional ausdifferenzieren können als diejenigen mit wenig Zugang zu Ressourcen. Das ist heute auch schon ein bisschen so, aber auf eine andere Weise. Ich glaube, besser kann ich es nicht beantworten.

### **Kerstin Schlögl-Flierl**

Vielen Dank an alle vier Referent:innen für die tollen Einblicke, die wir bekommen haben. Ich habe eine Frage an Frau Timón. Sie haben formuliert, es müsse verantwortungsvoll entwickelt worden sein. Da würde mich mal interessieren, was die Kriterien sind, dass Neurotech verantwortungsvoll entwickelt wird.

An Frau Nagel hätte ich die Frage: Man spricht immer von Vertrauenswürdigkeit bei KI. Mich würde interessieren, wie die Vertrauenswürdigkeit bei Neurotechnologie – wo da Schnittstellen oder Unterschiede sind. Es ist in Ihrer Publikation sehr klar aufgeführt, deswegen würde es mich interessieren. Vielleicht könnten wir da Impulse für unser Papier sammeln. Danke.

### **Laura Bernáez Timón**

Thank you very much for the question. Coming from an engineering perspective, I think that there is a lot of advice and methodologies on how to develop technologies from a technical perspective; there are different TRL levels and different steps that supposedly need to happen at all of these different stages. But I think more broadly and specifically for neurotechnologies, we do not have that type of framework concerning ethics. So, from the developmental perspective, it is not clear which type of ethical questions should be posed and answered at each time of this development process. Something that could also be done probably in that regard is to have some clarity for developers, some guidelines. Speaking with them, we found that there are many different aspects in

which they feel they lack this guidance, both on the regulatory perspective as well, but also on ethics, how to embed ethical principles early on in the development. So, this could be a useful contribution – a set of guidelines and ethical questions and, actually, when is the right moment to ask for these questions.

### **Eva Winkler**

Vielen Dank. Dann die Frage der Vertrauenswürdigkeit an Frau Nagel.

### **Saskia Nagel**

Vertrauenswürdigkeit ist aus meiner Sicht ein Begriff, der auf den Menschen reduziert bleiben sollte. Den sollten wir nicht für KI nutzen. Das wird gemacht. Auch die High-Level-Expert-Group-Empfehlung der Europäischen Kommission nutzt den Begriff Trustworthy AI, und ich würde da aus verschiedenen konzeptionellen und normativen Gründen gegenhalten wollen.

Vertrauenswürdigkeit ist sowohl sozial-epistemologisch als auch ethisch ein geladenes Konzept, und nach allem, wie es bis jetzt benutzt wird – sowohl alltagssprachlich als auch im philosophischen Diskurs –, erfordert es moralische Agenden, die die vertrauenswürdig sind und die Vertrauen geben können. Ich möchte den Begriff des Vertrauens und die Vertrauenswürdigkeit zwischen Menschen halten, und würde für als einfache Variante sagen: Für die Technologien haben wir den Begriff der Verlässlichkeit. Das reicht. Wenn wir über Technologien reden, können wir sagen: Die sind verlässlich. Und wenn wir über Menschen reden, reden wir über Vertrauen.

Jetzt kommt das, was aus meiner Sicht gut funktioniert: Wenn wir jetzt über Neurotechnologien in Kombination mit KI nachdenken – da sind wir als Menschen die, die im Umgang mit diesen Technologien vertrauenswürdig sein müssen. Also Menschen oder auch Firmen, Institutionen,

aber wir müssen an der Stelle die Verantwortung übernehmen, auch wenn es möglicherweise diese von mir skizzierten Verantwortungslücken gibt. Das heißt, wir übernehmen Verantwortung und werden dadurch vertrauenswürdig auch für die Neurotechnologien, die mit KI arbeiten.

So ist im Moment mein Konstrukt, und ich finde, dass das noch ganz gut funktioniert. Aber mir kann auch jemand das Gegenteil zeigen.

### **Eva Winkler**

Wir haben ja noch ein bisschen Zeit für die Diskussion, vielen Dank. Frau Betsch.

### **Cornelia Betsch**

Herzlichen Dank für die spannenden Einblicke. Aldo Faisal hat am Anfang gesagt, es geht um ein freies, gesellschaftliches Miteinander, und ich würde behaupten, dass dazu so etwas wie Prosozialität gehört.

Meine Frage geht an alle, weil diese Frage aus allen Bereichen gefüttert wird. Mir drängt sich ein bisschen der Eindruck auf, dass einer der Marktmechanismen darin liegt, dass es die Beziehung zwischen Mensch und Maschine besonders gut gestaltet. Wir hatten die Abhängigkeitsmechanismen. Ich gehe eine Beziehung ein. Gerade gab es ein neues Paper darüber, wie künstliche Intelligenz mir schmeichelt. Diese Art der Beziehung macht etwas mit den Menschen. In der einen Studie kam auch heraus, dass Menschen gerne mit der KI diese Probleme besprechen.

Wir hatten heute den AI Companion und dass dabei ein *selfish behavior* herauskommt, ein *loss in reciprocity*, also dass die Leute sich nicht mehr so stark auf das Gegenüber einlassen und mehr *selfish* werden.

Damit haben wir eine Verlagerung; der Techno-Solutionismus wurde angesprochen: Probleme,

die eigentlich sozial geklärt werden müssen, werden technisch geklärt. Das macht auch etwas damit, wie wir menschlich miteinander umgehen.

Deswegen würde ich gern Ihre Einschätzung hören. Wenn wir Weiterentwicklungen im Bereich der Neurotechnologie haben, wo ja die KI stark mit einbezogen ist, dort werden sicherlich Muster mit eingebaut, die die Nutzung verstärken. Was macht das mit der Akzeptanz, aber auch mit der Gestaltung der Beziehung zwischen den Menschen, vor allem im Bereich der Prosozialität?

### **Eva Winkler**

Vielen Dank. Die Frage würde ich zunächst an Herrn Meyer geben und dann an Frau Singh mit der Beziehungsgestaltung und dem KI-Schwerpunkt.

### **Bertolt Meyer**

Ich finde, in der Frage steckt eine sehr wichtige Differenzierung, nämlich zwischen Technologie und Geschäftsmodell. Es wäre mir ein wichtiges Anliegen, wenn wir diese Differenzierung hier aufgreifen könnten.

Das eine ist die Frage: Welche Affordanzen, also welche Fähigkeiten der Technologie beeinflussen die Interaktion und vielleicht auch die User Experience mit der Technologie? Welche Eigenschaften der Technologie führen also dazu, dass man sich gegebenenfalls von einer Technologie abhängig macht oder sich in eine Abhängigkeit begeben, sie exzessiv nutzt?

Und zum anderen: Welche Eigenschaften eines Geschäftsmodells führen dazu, dass eine Technologie besonders abhängig macht oder besonders häufig benutzt wird?

Da gibt es einige der Anbieter, die in dem Vortrag von Frau Singh genannt wurden, zum Beispiel Replica AI. Das ist auch nur ein Sprachmodell wie ChatGPT. Die Technologie ist die gleiche,

aber das Geschäftsmodell dahinter ist ein ganz anderes: Die verkaufen einem eine virtuelle Freundin, einen virtuellen Freund. Ich habe mich mal etwas intensiver mit Replica AI beschäftigt. Da gibt es auch ein Monatsabo, teilweise sehr teuer. Die virtuellen Partnerinnen und Partner, die da geschaffen werden, sind darauf angelegt, so etwas wie ein Abhängigkeitsverhältnis zur Nutzerin oder Nutzer herzustellen, dadurch, dass sie sich selber melden und so etwas sagen wie: „Hast du mich vergessen? Denkst du nicht mehr an mich?“ usw.

Das sind Dinge, die im Geschäftsmodell dieser Plattform verortet sind, was durch diese Technologie ermöglicht wird. Dieselbe Technologie ermöglicht also unterschiedliche Geschäftsmodelle. Für die ethische Bewertung einer Technologie können wir den wirtschaftlichen Verwertungszusammenhang dahinter – wie soll damit Geld verdient werden? – nicht aus der ethischen Bewertung herauslassen, glaube ich. Das ist ein Beispiel dafür. Das war vielleicht ein bisschen vage, aber viel besser kann ich es gerade nicht beantworten.

### **Eva Winkler**

Ein guter Anfang. Ich gebe weiter an Frau Singh.

### **Ilina Singh**

I mean I agree this is a very complex problem and I think particularly when we think about young people, of course, we want to be thinking about the entire developmental spectrum. But I suppose when we speak with young people about these concerns, about dependency and trust readiness, and what is engineered into the AI, you know, one of the things they say is that humans are terribly untrustworthy and they do all sorts of things to make me depend on them or to manipulate me. And so, to some extent, they see AIs, particularly where they can design the features of the AI, the cognitive and affective features, as potentially

more trustworthy and a better friend, somebody who can also be a bit more objective. Of course, I do not want to minimise the real harms here, but I just want to say that I think we are moving rapidly into a place where young people are going to be increasingly looking to these technologies for the kinds of relationships that something like Replika offers, but also for the kind of information that you need if you are going to think about your mental health and whether or not you need a diagnostic criterion, etc. And I think this question of trustworthiness is actually very central, and I agree with Saskia that we need to be looking at it very, very carefully. I also think that in terms of the AI design features, I mean, one of the, you know, this is a short presentation, so I could not give you the whole package, but the idea that young people, or anyone really, could be able to toggle the design of these interactions in a way that reflects their particular cognitive and affective needs and vulnerabilities and that they would be supported to understand what those are in the first place is, I think something that I have not seen so much. And I wonder if colleagues who understand about design more than I do would be able to have some reflection on whether that is even a feature that is possible and something that AI companies would want to include because, of course, it goes against some of the more addictive and sycophantic features that might already be included to promote use.

### **Eva Winkler**

Vielen Dank, Frau Singh. Vielleicht darf ich da eine Frage an Sie anschließen, nämlich zur partizipativen Forschung mit jungen Menschen. Sie haben gesagt, dass es da Kipppunkte gibt. Wo schlägt es um von Empowerment zum Risiko aus Sicht der jungen Menschen? Gibt es da andere Dinge, an die wir nicht denken? Oder was sind die

wichtigsten drei Einsichten, wenn man die Benutzer selbst fragt?

### **Ilina Singh**

So, we are still in the process of developing this data. In the project that I featured by Gracie Reincke, where we have asked the young people themselves to come up with what we are calling a threat model, which is where they are looking at different risks and opportunities in relation to AI and themselves prioritising that, but also trying to understand what the pathways to risk are. So, they do not have a static understanding of risk. This is a risk, this is not a risk. They see it very much that the risk is embedded in the kind of user and the kind of user pathway of the technology. So, it is a very dynamic kind of risk assessment that they want to produce so that we can understand the risks, but we can also understand the pathways to how those risks become severe risks. So, if we take the example of an AI companion, that might be risk-free at the outset, but for a particular kind of person in a particular kind of situation, the pathway to risk, if it has certain features, may then turn that into a very high risk. And they are very interested in protections along the way of that pathway so that we not only understand the risks, but we understand those important pathways. And some of the protections along the pathway can be designed into the feature, such as the kind of protections I was talking about. But many of them will also need to be provided by humans. And a graph that I did not show you, which is from that same data set, where a third of the kids said that they were using companion AIs infrequently, a majority of those children also said that they wanted parental support, which was also surprising to us. I mean, these are adolescents who we generally think of as not wanting parental support, but they did want parental support, and it was parents they wanted, not teachers and not peers. And

so, of course, there's a whole parental literacy programme here that needs to happen, because if you are, like I am, I mean, I'm far behind my children in AI literacy.

### **Eva Winkler**

Vielen Dank. Das erinnert mich auch an unsere Social-Media-Diskussion, wunderbar.

### **Judith Simon**

Das passt gut, darauf wollte ich nämlich hinaus. Ich habe auch eine Frage an Frau Singh. Als Information an die anderen: Wir arbeiten nicht nur an der Neurotechnologie-Stellungnahme, sondern auch an einer Ad-hoc-Stellungnahme zu Social Media und Kinder- und Jugendschutz in der digitalen Welt. Eine Sache, die wir auch im Blick hatten, ist, dass nicht nur die sozialen Medien, sondern auch generative KI und Chatbots ähnliche Probleme mit sich bringen. Da ist die Bitte an Frau Singh, dass ich wahrscheinlich noch einmal auf Sie zukomme und dass alles, was Sie an Literatur und Information mit uns teilen könnten, total super wäre, weil wir das auch an anderer Stelle brauchen. Das wollte ich nur einmal definieren. Vielen Dank.

Ich fand auch alle anderen Beiträge fantastisch. Ich habe nur einen Punkt, der an die Diskussion um Vertrauen und Vertrauenswürdigkeit anschließt. Ich glaube, das Grundproblem ist, dass die wahrgenommene und die tatsächliche Vertrauenswürdigkeit so weit auseinandergehen und dass weniger die ontologische Frage, ob KI vertrauenswürdig ist, relevant ist, sondern die Tatsache, dass offensichtlich Nutzerinnen und Nutzer Technologien vertrauen, weil sie sie als vertrauenswürdig in Bezug auf Dinge einschätzen, die es aber nicht sind. Und in diesem normativen Spalt zwischen einer angemessenen Etablierung von Vertrauen liegt der Hund begraben. Das kann man nicht gut ins Englische übersetzen, oder?

Ich glaube, in diese Frage müssen wir noch ein bisschen reinbohren, dass viele von den ethischen Problemen da auftauchen, wo Vertrauenswürdigkeit angenommen wird, die tatsächlich aber aus ökonomischen oder technischen – ich glaube, beides ist der Fall: Auf der einen Seite sind es die ökonomischen Fragen mit Abomodellen und mit dem Datensammeln, auf der anderen Seite aber auch die Simulation von Empathie, die sozusagen in der Technologie selber liegt. Da werden wir sicherlich noch reinbohren.

Das war ein Kommentar und eine Bitte um Hilfe und um Literatur.

### **Eva Winkler**

Vielen Dank. Sie sehen, wir gehen mit unseren Sachverständigen sehr effizient um und nutzen sie gleich für mehrere Stellungnahmen. [Lachen]

### **Armin Nassehi**

I hope the translation of „the dog is buried“ was successful. [Lachen]

Vielen Dank für die wunderbaren Vorträge. Ich würde auch auf die Frage der Vertrauenswürdigkeit kommen und ähnlich wie Judith die Differenz starkmachen: Wir diskutieren Vertrauenswürdigkeit aus einer ethischen Perspektive völlig anders als aus einer praktischen oder pragmatischen Perspektive. In den Vorträgen bei Frau Singh war das für mich am deutlichsten. Sie hat im Prinzip die Antwort auf die Frage gegeben, die ich noch gar nicht gestellt habe. Bei der Vertrauenswürdigkeit ist es so, dass sich während der Praxis die Formen, was riskant ist und was nicht riskant ist, verändern.

Das ist nichts Neues, das haben wir nicht nur bei neuen Technologien. Ich will jetzt nicht wieder anfangen damit, dass das alles vor Hunderten von Jahren mit dem Buchdruck auch schon diskutiert wurde: Das Lesen ist wahnsinnig riskant für

Leute, weil sie mit Dingen zusammenkommen, die sie völlig verwirren.

Aber das haben wir eigentlich bei fast allen technologischen Veränderungen und auch bei fast allen sozialen Veränderungen. Es ist unglaublich riskant, durch eine Großstadt zu gehen. Aber wir empfinden es nicht mehr so. Wir finden es riskanter, uns dort zu bewegen, wo es nicht so riskant ist, zum Beispiel dort, wo man sich kennt.

Das Spannende ist (und jetzt komme ich zum Thema zurück): Wenn wir aus einer ethischen Perspektive auf diese Sachen schauen (Frau Nagel, Sie haben auf die Normalisierung hingewiesen), dann werden wir eigentlich immer bereits mit einer Lösung dieser Frage durch die Praxis konfrontiert.

Meine Frage ist: Wie kommen wir da eigentlich ethisch hinterher? Frau Timón, Sie haben es wunderbar beschrieben, wenn es um Konsumenten geht. Denn Konsumenten haben die Entscheidung immer schon getroffen und müssen dafür andere Gründe angeben als wir, wenn wir es aus einer ethischen Perspektive betrachten. Ich würde das, was Judith gerade als einen Kommentar formuliert hat, als Frage stellen: Wie gehen wir ethisch-technologisch damit um, aus dieser Differenz etwas zu machen? Denn ich habe das Gefühl, dass wir da total hinterherhinken.

Noch eine Bemerkung zu Herrn Meyer: Mir hat vor allem eine Bemerkung wahnsinnig gut gefallen: dass die Dinge, wenn sie assistiv sind, eine höhere Anerkennung bekommen. Das könnte ja eine Lösung sein, und wenn man das auf andere Fragen überträgt, stellt sich die Frage: Ist es eigentlich möglich, einen Diskurs darüber zu führen, wie Dinge, die jetzt nicht aus dem Bereich stammen wie dieses Beispiel, aber ob wir bei der empirischen Beobachtung solcher Geschichten nicht so einen Katalog bekommen: Was wird Anerkennung bekommen und was nicht?

Dann stellt sich die Frage, ob Anerkennung eigentlich die Währung ist, um die es geht, oder ob wir Kriterien brauchen, die außerhalb dieser Empirie liegen. Dann haben wir ethische Fragen, an die man nicht richtig rankommt, weil wir die Dinge in ihrer Pragmatik nicht verändern können.

### **Bertolt Meyer**

Aus den Befunden, die sich in meiner Gruppe abzeichnen und die ich auch in der Literatur kenne, zeichnet sich ab, dass dieses Framework von De Boeck und Vaes, das ich in meinem Vortrag gezeigt habe, zu den verschiedenen Nutzungsdimensionen hinsichtlich: Ist es wiederherstellend, ist es unterstützend oder ist es erweiternd?, eine sehr nützliche Differenzierung ist, weil sie offensichtlich für jegliche Form der subjektiven Bewertung dieser Technologien eine große Rolle zu spielen scheint. Vielleicht kann sie deswegen auch nützlich sein für das, was hier erarbeitet wird.

### **Eva Winkler**

Vielen Dank. Dann Frau Timón zu den Marktmechanismen, an die Herr Nassehi andockt hat.

### **Laura Bernáez Timón**

Thank you for the question. Concerning neurotechnologies, I think that what is probably different from other technologies is that the users or consumers really need to trust. Because people are very concerned about the brain, maybe more than they are concerned about other parts of the body. Companies know that. And, for many companies with whom we spoke, for them it is also an issue of company liability. They think of that along those terms. So, if there were to be any problem with brain data in the future, for them, that would be extremely dangerous as a company. So, in that regard, it is also probably very complicated for the companies that are sort of expanding the field because they are dealing with all of these

issues. They are expanding or creating new opportunities, but at the same time, they are also managing all this risk. I think that is slightly different in neurotechnology compared to other medical technologies. Also, I was thinking that for, and this applies maybe more to implantables, so, for other, let's say, for example, for motor abilities, we have in research good animal models that allow us to sort of predict the consequences of the implant and the unintended side effects. But when we talk about cognitive processes, that becomes more difficult because we do not have animal models for, let's say, intelligence or decision-making that are comparable to how humans address these problems, nor do we have a language to talk about that. So it is very difficult as well to quantify metrics that relate to cognitive processes.

### **Eva Winkler**

Vielen Dank, dann Jutta Allmendinger.

### **Jutta Allmendinger**

Auch von meiner Seite vielen Dank für die hervorragenden Vorträge, und auch danke an jene, die das kuratiert und genau Sie eingeladen haben. Das will ich auch mal so sagen.

Ich gehe auch sehr effektiv mit den Sachverständigen um und frage zunächst mal Frau Singh, wie sie ihre Forschung über jüngere auf ältere Personen übertragen würde. Denn wir sitzen gerade an einem Papier über Langzeitpflege für Ältere, und das, was Sie beschrieben haben, scheint mir teilweise genauso auf Ältere zuzutreffen. Wo sind da die Abgrenzungen, die Sie machen?

Die zweite Frage geht an Herrn Meyer. Da möchte ich anschließen an das, was Sie mit Herrn Goldschmidt diskutiert haben, mit den Mobiltelefonen. Ja, die sieht man ja. Oder das, was Herr Nassehi gerade sagte: Wir können nicht sehen, inwieweit ein bestimmtes Tool für Enhancement ist – diese drei Kategorien, die Sie öffnen, sind für

den Betrachter unsichtbar. Das ist eine Unterscheidung, die mir theoretisch absolut einleuchtet. Aber praktisch gar nicht. [Lachen]

### **Eva Winkler**

Dann machen wir kurz den Ausflug zu den Älteren und kommen dann noch mal zu den Neurotechnologien. Frau Singh.

### **Ilina Singh**

Ja, danke. This also allows me to address a previous point about the ethical norms of trust and trustworthiness versus design norms. And I think that one of the things we are missing as ethicists in this area is to start from a position of relationship rather than the individual being protected, autonomy considerations, privacy considerations, etc. Because I think that we are talking about a relationship with these technologies and, particularly in the use case that I was talking about, a very intimate relationship sometimes. And so our thinking about trust needs to come through relationship. And I don't think we've done a very good job yet of articulating that, and we are trying to do some of that in our team. And I would say the same pertains to how we work with older people in these technologies. We have looked a bit at care bots, for example, where again, some of the ethical writing around that has been very much about these kind of individualist, principalist ethics. But not about that relationship to a person who is also changing cognitively, quite significantly sometimes, into a period of more vulnerability. And so we think that some of what we are talking about with young people, who are immensely capacitated in technology, is applicable here. In relation to thinking about how we can think of them still as epistemic agents, if they are, to build a care plan around them as in cases of diminished capacity and competence, and to enable them to, in a sense, program into the, you know, if it is a companion or a care bot, for example,

what their cognitive and affective and relational goals are. And concretely, one of the things we have started doing is that our young people's advisory group meets with our older people's advisory group. We call it a YOPAG, a young and old people's advisory group. And we enable conversations between them about technologies and what their kind of understanding of the risks and benefits are and what kinds of goals they have around technology use. And, that has been, again, more successful than I would have thought. There is a lot of really mutual learning and knowledge sharing and listening. So that is a concrete example.

### **Eva Winkler**

Vielen Dank. Jetzt zur Sichtbarkeit und Unsichtbarkeit.

### **Bertolt Meyer**

Human-Orientierungstechnologien können nur dann für zwischenmenschliche, soziale Prozesse wirksam sein, wenn sie wahrgenommen werden und im Zusammenhang mit dieser Wahrnehmung eine Zuschreibung erfolgt. Das heißt konkret: Wenn eine Person mit Demenz einen Neurostimulator implantiert bekommt, der die Funktion des Hippocampus unterstützt, damit das Gedächtnis wieder besser funktioniert, dann ist das für eine außenstehende Person im Idealfall nicht sichtbar.

### **Jutta Allmendinger**

Man kann es sich ja schon mal präventiv einpflanzen lassen.

### **Bertolt Meyer**

Ja, das könnte man im Moment. Aber das ist in der Situation nicht sichtbar, und ich würde sagen: Für diesen Anwendungsfall muss es das auch nicht sein. Denn in diesem Fall sind wir im Bereich dessen, was nicht so sehr Gegenstand ist,

nämlich die assistive Technologie. Auch ein implantierter Herzschrittmacher muss nicht sichtbar sein. Aber der gleiche Chip kann den Arbeitsspeicher in unserem Kopf massiv erweitern. Wir haben das schon im Tiermodell gesehen: Derselbe Chip, einer gesunden Ratte implantiert, steigert die Gedächtniskapazität der Ratte um 70 Prozent.

Wenn ich jetzt mit einer Person interagiere, die so etwas eingebaut hat, und ich kann das nicht sehen, dann ist es, normativ gesprochen, Schummel. Von außen sieht es so aus, als wäre es eine in Anführungszeichen natürliche Fähigkeit. Faktisch ist sie das nicht.

An der Stelle muss man natürlich fragen: Ist das ethisch vertretbar? Das kann ich nicht beantworten. Aber es gibt viele dieser Technologien, die eben doch sichtbar sind. Und da zeigen die Erfahrungen: Es kommt am Ende darauf an, was die Leute subjektiv glauben, was es kann und wofür es da ist, für eine Zuschreibung, die dann wirksam wird.

Eine Kollegin von mir hat eine Interviewstudie gemacht mit Gärtnern, die im Außenbereich tätig sind für die Stadt, die mit Exoskeletten experimentiert haben. Da war die Annahme, das würde sich für die vielleicht unangenehm anfühlen, weil das so cyborgmäßig als Übermensch rüberkommt, ähnlich wie unsere Forschung das vorher sagen würde. Wenn wir dann mit den Leuten gesprochen haben, hat sich gezeigt: Denen war das auch unangenehm, aber aus einem völlig anderen Grund. Sie hatten nämlich die Befürchtung, dass das nach außen signalisiert, dass sie schwach sind und Unterstützung brauchen, und das war ihnen vor den Kollegen unangenehm.

An dem Beispiel zeigt sich, dass es im Endeffekt die subjektive Zuschreibung ist, wofür es da ist, die dann dazu führt, welche Auswirkungen es auf die Interaktionen und die Zuschreibung hat.

**Kerstin Schlögl-Flierl**

Wir wollen ja in der Stellungnahme auch anthropologisch und anthropologisch-ethisch argumentieren und haben festgestellt, dass dafür das Thema der personalen Identität sehr spannend ist. Inwieweit sehen Sie, dass die neueren Neurotechnologien auch die Frage der personalen Identität verändern? – Eine offene Frage für alle vier. Es kann auch nur einer antworten.

**Saskia Nagel**

Ich habe einen Schnipsel einer Antwort. Es gibt die Diskussion um personale Identität. Das ist etwas anderes als Persönlichkeit. Da geht es wirklich um Identitätsfragen.

Es gibt einen Befund, der zurückführt zum Thema sich verantwortlich fühlen, dass sich der Sense of Agency – und zwar der Sense, nicht die Agency an sich – beim Einsatz von Neurotechnologien massiv verändern kann. Denn ein reduzierter Sense of Agency sorgt dafür, dass ich mich auch weniger verantwortlich fühle.

Das ist ein Hinweis darauf, dass es Auswirkungen auf die personale Identität geben kann. Denn der Sense of Agency ist zentral für die personale Identität. Aber das ist nur ein Mosaikstein für das, was Sie wahrscheinlich eigentlich brauchen.

**Eva Winkler**

Ja, das Gefühl der eigenen Handlungsfähigkeit, wie auch immer das zu übersetzen wäre. Ein zweiter Schnipsel.

**Bertolt Meyer**

Ich würde gerne aus Sicht der sozialen Identität und mit Bezug zu Human Augmentation Technologies (das heißt Technologien, die körpernah

am Körper eingesetzt werden und auf die körperlichen Funktionen wirken) die Frage beantworten: Beeinflussen solche Technologien die soziale Identität? Mit anderen Worten die Frage: Wer bin ich und zu welchen gesellschaftlichen Gruppen der Anwenderinnen oder Anwender gehöre ich? Ja, das tun sie.

Mein Postdoc Maximilian Bretschneider hat eine qualitative Studie durchgeführt,<sup>3</sup> in deren Rahmen er in Anführungszeichen Power User von bionischen Armprothesen, wie ich sie trage, interviewt und mit ihnen darüber gesprochen hat, was die Versorgung mit dieser Technologie mit ihnen in ihrem Alltag und vor allen Dingen mit ihrer Identität und ihrem Selbstverständnis macht. Da sieht man *profunde* Auswirkungen. Denn insbesondere Technologien, die nah am Körper sind und stark auf die körperlichen Fähigkeiten einzuwirken, haben eine massive Auswirkung auf die soziale Identität. Denn die beiden Dimensionen der interpersonalen Wahrnehmung, die ich in meinem Vortrag angerissen habe, nämlich Wärme und Kompetenz, spielen auch eine zentrale Rolle für die Selbstwahrnehmung. Das bedeutet: Alle Technologien, die an der Kompetenz ansetzen, können einen massiven Einfluss auf die Selbstzuschreibung der Kompetenz haben.

Ich erlebe das selbst im Vergleich zu früher, wo ich mit Prothesen durch die Gegend gelaufen bin, die die Anmutung eines Sanitätshauses der siebziger Jahre ausgestrahlt haben. Da war soziale Interaktion am besten mit dem Wort *awkward* beschrieben: Die Leute merken das und merken, da ist irgendwas komisch, es ist irgendwie unangenehm, und wenn man Pech hat, geht es in Richtung Ekel.

---

<sup>3</sup> Bretschneider, M.; Meyer, B.; Asbrock, F. (2023): The impact of bionic prostheses on users' self-perceptions: a qualitative study. In: Acta Psychologica, 241, 104085.

Und wenn Sie so etwas tragen [zeigt auf seine Handprothese] und die Leute merken das, dann bekommen Sie eine völlig andere Rückmeldung in der sozialen Interaktion. Junge Leute, Kinder in der Straßenbahn kommen zu mir und sagen: „Das ist ja cool. Was hast du denn da?“ Und cool ist das Gegenteil von behindert.

Allein dieses Beispiel zeigt: Die Auswirkungen können profund sein.

### **Eva Winkler**

Frau Singh hatte sich auch zu der Frage gemeldet.

### **Irina Singh**

Ja, I think this is a really interesting question about identities and how they are shifting. In bioethics, we used to talk about whether the self was discovered or created, and that was sort of the binary option. And I think now, with young people in particular, I think that there is an increasing not just acceptance, but desire to continually self-create and an understanding that multiple identities are possible, particularly if you're moving in and out of virtual realms, but also physical contexts. Certainly also in terms of what we see with young people adopting neurodiversity as a strength, we see this understanding of identities as something that is creative and dynamic and shifting. And I think there is an open question as to whether this poses benefits at the end, whether we end up with young people who are sort of disintegrating because there is no set core sense of self or whether it is a process of strengthening identity at the end. And I think that is something that at the moment we cannot answer because these are rapidly changing times as we know, not just technologically.

### **Eva Winkler**

Vielen Dank. Eine zentrale Frage. Herr Ienca, Sie arbeiten dazu schon lange.

### **Marcello Ienca**

Ja, danke. Weil wir gerade das Thema Phänomenologie breit definiert besprechen, also Identität, Agency, Selbstwahrnehmung usw., möchte ich kurz über unsere Studie berichten. Mit Dr. Bublitz und zwei weiteren internationalen Kollegen haben wir eine Studie durchgeführt, die „Hybrid Minds“ heißt. Da haben wir versucht, diese phänomenologischen Veränderungen zu verstehen.

Warum? Ich bin völlig mit Herrn Meyer einverstanden: Die Auswirkungen scheinen profund zu sein, aber bis jetzt haben wir nur einzelne Patientenberichte. Was fehlt, ist wirklich systematische Evidenz. Wir haben eine Studie mit über hundert BCI- und Neurostimulation-Nutzer:innen in drei Ländern durchgeführt. Unsere präliminäre Datenanalyse scheint das zu bestätigen.

Interessanterweise sind viele Patient:innen der Ansicht, dass ihre Identität und Agency von der Krankheit zerstört wurde und dass die Technologie dabei helfen kann, diese Identität zu rekonstruieren. Das haben wir auch ex negativo durch eine andere Studie mit australischen Kollegen bestätigt. Da haben wir eine Interviewstudie mit einer Patientin durchgeführt, der ein Gehirnimplantat explantiert werden musste, weil die Firma, die dieses Implantat produziert hat, insolvent gegangen ist. Und wenn das passiert, dann besteht für die Patientin keine Alternative zur Explantation. Warum? Weil das Risiko für Blutungen und Infektionen durch ein *unmaintained device* zu hoch ist.

Nach der Explantierung hat sie berichtet, dass sie Veränderungen der phänomenologischen Sphäre gespürt hat, und sie konnte sich nicht mehr mit der Person vor der Explantation erkennen. Das würde die These vertreten, dass Identität kokreiert ist in

dem Sinne, dass die Interaktion mit diesen Geräten auch zum Co-Design von phänomenologischen Variablen beitragen kann.

### **Eva Winkler**

Sehr interessante Ergebnisse und Beobachtungen. Wir haben jetzt noch drei Redner und 14 Minuten Zeit. Zunächst Frau Gräß-Schmidt.

### **Elisabeth Gräß-Schmidt**

Zunächst auch von mir herzlichen Dank für die großartigen Vorträge. Meine Frage bezieht sich weniger auf das Selbstverständnis und die Identität als konkret auf das Gehirn.

Frau Nagel hat die Hybridisierung von Mensch und Maschine thematisiert, also dass das mehr und mehr durchlässig wird. Unter dem Aspekt der Normalisierung oder der neuen Normalität würde mich interessieren, ob sich bei dieser zunehmenden Durchlässigkeit – die ja auch mit dem Unterschied von Mensch und Maschine, Natur und Technik zusammenhängt – erkennen lässt, dass diese Hybridisierung direkte Auswirkungen auf das Gehirn und auf die Gehirntätigkeit hat und ob diese Auswirkungen auch die Selbstwahrnehmung beeinflussen und das, was als normal und was als nicht normal gilt. Inwiefern führt also die Hybridisierung zu einer neuen Normalität? Und bildet sich das unter Umständen auch im Gehirn ab und ergeben sich dadurch Verstärkungseffekte?

### **Eva Winkler**

Das ist eine Frage an die Neurowissenschaftler. Frau Timón, können Sie dazu etwas sagen?

### **Laura Bernáez Timón**

Thank you. That's a very good question. So, I think, how the brain will react is also difficult to assess, because the brain is slightly different for every person and it's also different across life, across lifetimes. So, there is plasticity, there are

experiments showing that actually the way or the places in which certain functions are mapped can change over time. And what we know or what, like the experiments that we have in which we can rely on to predict what is going to happen, which likely come as well from preclinical studies and research, also show us that it really changes how different things are encoded in the brain if you are in very constrained laboratory settings versus whether you are in the real world. So usually, experiments in which neuroscientists study things like decision-making happen under very, very controlled scenarios. So, you have, for example, a virtual reality dome and there is a monkey there learning to do some task. But that's not how we operate in the real world, right? So that makes it really, really difficult to predict what is going to happen when something is implanted into the brain. There is also the question of how we know where to implant for each person. There are some areas that are very well mapped, such as the motor cortex, that we understand very well. But then when we go deeper as well, things are a bit more complicated and different across individuals. And now, what I mentioned before as well, so there are no perfect, let's say, animal models to study these things. So one potential consequence of all of this is that we will get to know along the way what these effects are on things such as perception, which is, of course, concerning.

### **Eva Winkler**

Vielen Dank, Frau Timón. Wir haben jetzt noch zwei Fragen und fünf Minuten.

### **Hans-Georg Dederer**

Herzlichen Dank auch von meiner Seite für diese phänomenalen Impulse. Ich habe an Frau Nagel eine Frage zur Verantwortlichkeit. Sie haben gesagt, Sie wollen sich bewusst von der juristischen Betrachtung der rechtlichen Verantwortlichkeit

abgrenzen. Sie haben gesagt, dass Sie bei der moralischen Verantwortung Lücken sehen. Können Sie das näher beschreiben? Denn für die Juristen ist es schon interessant, ob sich diese Lücken zum Beispiel in einer Haftungslücke im Recht spiegeln.

Man hat es ja beim autonomen Fahren auch diskutiert: Wer haftet eigentlich? Am Ende scheint man sich, meine ich jedenfalls, im Recht doch einig zu sein: Im Moment haben wir keine Haftungslücken. Wir haben eine Herstellerhaftung, das ist eine objektive Unrechtshaftung. Man haftet für Fehler. Dann gibt es die Fahrzeugführerhaftung, das ist im Grundsatz eine Verschuldenshaftung. Dann gibt es die Fahrzeughalterhaftung. Das ist eine Gefährdungshaftung ohne Verschulden, ohne Rechtswidrigkeit, einfach weil das Objekt inhärent gefährlich ist, eine typische Gefährlichkeit hat, die sich jederzeit realisieren kann. Das ist ein spezieller Bereich, der uns bei der Neurotechnologie nicht direkt interessiert, aber die Frage: Wo sehen Sie aus Ihrer philosophischen Perspektive Verantwortlichkeitslücken?

Wenn ich eine Frage an Herrn Meyer anschließen darf: Ihr schönes Schema, auf der X-Achse kompetent/inkompetent und auf der Y-Achse warm/kalt, da hatten Sie gesagt: Die Wahrnehmung geht Richtung Cyborg, je aufgerüsteter die Menschen sind, wenn ich es so vereinfacht sagen kann. Da könnte man die Frage anschließen: Wie disruptiv ist das in der Wahrnehmung innerhalb der Gesellschaft? In unserer Gesellschaft, gerade in unserer demokratischen Ordnung, nehmen wir uns als gleich wahr. Ausgangspunkt ist: Wir sind alle Menschen, frei und gleich geboren. Das ist auch der Grund für die gleiche Teilhabe an der

Ausübung von Staatsgewalt, das Recht auf gleiche Einflussnahme. Diese Annahme ist sozusagen eine Konstitutivbedingung.

Inwiefern wird das infrage gestellt? Oder ist das aus Ihrer empirischen Forschung heraus eigentlich völlig spekulativ und vielleicht gar nicht relevant, weil diese Aufrüstung des Einzelnen gar nicht wahrnehmbar ist? Die ist ja implantiert oder im Mikrobereich, dass man gar nicht mehr sehen kann, ob jemand ganz ungleich ist und dieser kalte, aber hochbefähigte Cyborg ist.

### **Eva Winkler**

Jetzt brauchen wir listenhafte Antworten. Eine Liste von Lücken an Frau Nagel, und eine kurze Antwort auf eine konzeptionelle Frage. Das ist nicht so einfach.

### **Saskia Nagel**

Die Sorge um die Lücke wird seit ungefähr zehn Jahren diskutiert. Das ist nicht so mein Beritt. Ich bin auch nicht sicher, ob es diese Lücke gibt. Ich interessiere mich mehr für die Verantwortungsdiffusion, weniger für die Lücke. Ich würde sagen: Wir können Verantwortung übernehmen, dann schließen wir die Lücke.

Aber es geht ja nicht nur um das, was ich vertrete, sondern auch um das, was im Raum ist. Der Diskurs ist so, dass die Lücke entsteht, sobald wir KI ins System bringen, also autonome Systeme, über die wir keine vollständige Kontrolle und kein vollständiges Wissen haben. Denn dann haben wir die beiden Bedingungen Kontrolle und Wissen bei moralischer Verantwortung an einer Stelle nicht mehr.

Der Ursprung der Debatte ist ein Paper von Andreas Matthias, das schon relativ alt ist.<sup>4</sup> Seitdem

<sup>4</sup> Matthias, A. (2004): The responsibility gap: Ascribing responsibility for the actions of learning automata. In: *Ethics and Information Technology*, 6 (3), 175–183.

wird hoch und runter diskutiert, wie genau und wann diese Lücken entstehen und ob sie sich schließen lassen. Das ist einfach der Diskurs. Man kann darüber streiten, ob es diese Lücke gibt. Auf jeden Fall ist es die Sorge, dass es solche Situationen geben kann.

Ein Beispiel, das passt nicht hier in den Raum, aber wird immer diskutiert: selbstfahrende Autos. Ich sitze da als Fahrerin drin. Das Auto entscheidet ohne mich. Ich habe nicht die Kontrolle und auch überhaupt nicht das Wissen darüber, wie die Entscheidung vom Auto getroffen wird. Jetzt gehen wir ein bisschen in die Zukunft. Wenn das so wäre, dann kann ich nicht verantwortlich dafür gemacht werden, wenn ein Unfall passiert. Dann wird in der moralischen Verantwortungsperspektive gesagt: Da entsteht eine Lücke.

Aber wie gesagt, das ist strittig. Es gibt Kollegen, die sagen, da ist keine Lücke. Ich würde sagen, man weiß es nicht so genau, aber wir können immer noch die pragmatische Lösung finden und Verantwortung übernehmen.

### **Bertolt Meyer**

Eine schwierige Frage. – Wenn man den Begriff des Cyborgs so versteht, wie er in der Enhancement-Literatur geführt wird, dann ist der Cyborg der Prototyp des so weit erweiterten Menschen, dass er außermenschliche Fähigkeiten besitzt. Und genau das ist das Ziel der Extremisten in der Enhancement-Bewegung, die nämlich in diesen Technologien, die das Überwinden der Grenzen des menschlichen Körpers ermöglichen, ein quasi-religiöses Heilsversprechen sehen. Dort ist auch häufig davon die Rede, dass das ultimative Ziel all dieser Technologien das Bezwingen des Todes ist, und all das, was wir hier machen, sind nur Zwischenschritte. Das sind extreme Ansichten, die aber durchaus von Leuten wie – ich denke, ich trete Nick Bostrom nicht zu nahe,

wenn ich seinen Namen als einen der Vertreter dieser Strömung hier nenne.

Mit anderen Worten: Das Ziel ist die Überwindung der egalitären Vorstellung von „Wir sind alle gleich, denn am Ende müssen wir alle sterben.“ Das Ziel von einigen ist tatsächlich, sich da rauszunehmen und zu sagen: „Wir aber nicht.“ Wir können unseren Körper in extremo so upgraden, dass die Grenzen der Biologie einschließlich der extremsten, nämlich des Todes, überschritten werden.

Das ist natürlich ein Problem, und ich halte es, mit Verlaub, auch für eine sehr narzisstische Fantasie, weil sie häufig von Menschen begründet wird mit: „Mein Wissen, meine Erkenntnisse, im Grunde genommen der Inhalt meines Kopfes ist viel zu schade, um durch meinen Tod vernichtet zu werden, und müssen der Nachwelt erhalten bleiben.“ Das ist das Motiv, das ich da unterstelle.

Aber noch einmal: Wir bewegen uns hier in dem sehr extremen Diskurs der Transhumanisten, die sich aber im Umfeld der Enhancement-Debatten bewegen, und deswegen müssen das hier auch spiegeln.

### **Eva Winkler**

Vielen Dank auch für die Projektion in Richtung derjenigen, die sich in der Marktlogik und besonders im Rahmen der Techkonzerne tummeln.

Wir sind jetzt mit dem ersten Block am Ende. Im zweiten Block werden wir Herrn Hans-Georg Dederer, auch Mitglied des Ethikrats, als Moderator der Diskussion mit einem Schwerpunkt im ethisch-rechtlichen Bereich haben. Alle Fragen, die noch nicht gestellt oder beantwortet wurden, haben dann dort ihren Platz. Vielen Dank.

## Block 2

### Hans-Georg Dederer

Meine Damen und Herren, verehrte Kolleginnen und Kollegen, wir wollen die Anhörung zu den ethisch-rechtlichen Fragen der Neurotechnologie fortsetzen. Im zweiten Block haben wir drei Impulsvorträge von Philipp Kellmeyer, Marcello Ienca und Christoph Bublitz. Wir werden alle drei Vorträge hintereinander hören und danach in die Diskussion einsteigen.

Ich stelle zunächst Herrn Kellmeyer, der online zugeschaltet ist, kurz vor. Er ist Professor für Responsible AI und Digital Health an der Universität Mannheim und befasst sich mit den ethisch-rechtlichen und regulatorischen Fragen beim Einsatz von Neuro- und KI-Technologien. Hierzu hat er einen neurologischen und neurowissenschaftlichen Hintergrund. Herr Kellmeyer, Sie haben das Wort.

### Mentale Integrität und Vulnerabilität als ethische Herausforderungen bei Neurotechnologien

#### Philipp Kellmeyer · Universität Mannheim

(Folie 1)

Vielen Dank für die Einladung. Ich freue mich auf den Austausch. In meinem Vortrag möchte ich mentale Integrität und Vulnerabilität als zentrale ethische Herausforderungen bei Neurotechnologien herausstellen.

(Folie 2)

Sie haben in den anderen Vorträgen schon viel gehört über die verschiedenen Arten von Neurotechnologien. Das sind zum einen die Gehirn-Computer-Schnittstellen, die Brain-Computer-Interfaces, die mittlerweile eine Reihe interessanter Anwendungen ermöglichen, sei es im Bereich der Kommunikation oder der motorischen Unterstützung,

im Bereich des Sehens, aber auch für therapeutische Anwendungen, für eine Vielzahl neurologischer psychiatrischer Erkrankungen.

(Folie 3)

Ein wesentlicher Bestandteil von Neurotechnologien sind Neurostimulationsverfahren. Das sind Methoden, mit denen wir entweder von außen, zum Beispiel durch Gleichstrom, durch elektromagnetische Stimulation Hirnaktivität beeinflussen können oder direkt im Gehirn durch die Tiefenhirnstimulation und andere Stimulationsverfahren. In diesem Bereich tut sich immer wieder etwas. Wir werden hier in den nächsten Jahren neue Verfahren sehen, das heißt, neben der Möglichkeit, Gehirnaktivität abzuleiten und mithilfe von KI-Algorithmen auch entsprechend zu interpretieren, haben wir wahrscheinlich zunehmend die Möglichkeit, diese Aktivität gezielter mit Stimulationsverfahren zu modulieren.

(Folie 4)

Ein wichtiger Aspekt ist dabei, dass die Fortschritte im Bereich der KI und in der Ableitung der Signale der Elektroden uns immer mehr ermöglichen, in Echtzeit, abhängig von der jeweiligen Hirnaktivität, bestimmte Stimulationsmuster ins Gehirn einzubringen über sogenannte Closed-Loop-Verfahren, Verfahren im geschlossenen Regelkreis. Diese können, abhängig von der Aktivität des Gehirns, bestimmte Stimulationsparameter in Echtzeit anpassen und werden heutzutage oft unterstützt durch KI-Verfahren.

(Folie 5)

Wir haben heute schon gehört, was Herr Musk mit der Firma Neuralink plant, wo neben dem Neurochip auch Implantationsgeräte hergestellt werden sollen und wir neben dem schon vorgestellten Bereich der verbraucherorientierten Neurotechnologie mit Geräten, die von außen die Hirnaktivität messen, möglicherweise in ein Zeitalter eintreten,

wo es implantierbare Chips gibt, bei denen nicht immer ganz klar ist, ob dies nur für streng medizinische Anwendungen genutzt werden soll oder auch für verbraucherorientierte oder medizinnahe Anwendungen im Bereich Wellness, Wellbeing usw.

(Folie 6)

Und um das Ganze ethisch zu bewerten, müssen wir verschiedene Ebenen anschauen: [1] einmal die Mikroebene der individuellen Auswirkungen der Mensch-Neurotechnologie-Interaktion, sei es im Bereich individueller Vulnerabilität oder im Hinblick auf Autonomie oder, was vorhin zur Sprache kam, der personalen Identität, [2] aber auch gruppen- und populationsspezifische Effekte. Das, was im Moment im Kontext der Diskussion um Social Media und Auswirkungen auf Mental Health bei Kindern und Jugendlichen stattfindet, sind unter anderem Effekte der hochskalierten weitläufigen Verbreitung der Technologien.

[3] Aus der Perspektive von Public Health und Global Brain Health ist natürlich auch die Frage: Wer hat eigentlich in Zukunft Zugriff auf solche Innovationen, auf solche Neurotechnologien? Verschärfen Neurotechnologien die bereits bestehende Krise in der Verteilung von Gesundheitsgütern und dem Zugang zu neuen innovativen Verfahren auf einer globalen Skala? Oder trägt Neurotechnologie zur Verbesserung der Situation bei? Hier können auch globale Gerechtigkeitsfragen mit ins Spiel kommen.

(Folie 7)

Heute möchte ich aber über mentale Integrität und Vulnerabilität sprechen und über die Arbeitsdefinitionen aus der vielfältigen Forschung zu dem Thema im Bereich der medizinischen Ethik. Mentale Integrität im Bereich der Neuroethik wird verstanden als „Schutz der Unversehrtheit von

Gehirnfunktionen und den damit verbundenen mentalen Erfahrungen gegen ungerechtfertigte, nicht einwilligungsbasierte oder manipulative Eingriffe“. Wir haben schon gehört, dass das ein sehr breites Spektrum ist. Genannt werden können Psychopharmaka bis hin zu adaptiven Neurostimulationsverfahren.

Vulnerabilität – als ethisch-anthropologische Querschnittsdimension – ist die erhöhte Anfälligkeit für Schädigung, Ausbeutung oder unfaire Einflussnahme, die oft situativ, manchmal auch chronisch ist, die dynamisch sein kann, sich über die Zeit verändert und die auch – was wichtig ist – intersektional ist. Das heißt, es kann sein, dass man in verschiedenen Dimensionen vulnerabel ist, was beispielsweise die Gesundheit oder sozioökonomische Faktoren betrifft.

Das ist besonders relevant für Neurotechnologien, denn hier haben wir es oft mit Konstellationen von eingeschränkter oder unklarer Einwilligungsfähigkeit im Bereich neurodegenerativer Erkrankungen oder schwerer psychiatrischer Erkrankungen zu tun. Wir haben es mit Themen wie Abhängigkeiten, Leidensdruck, therapeutischen Fehlvorstellungen oder manchmal auch sozialem Druck im Consumer- oder Wellness-Bereich zu tun.

Wir haben schon gehört, dass im Bereich der Hirnentwicklung, der kognitiven, digitalen Verletzbarkeit, zum Beispiel bei Kindern und Jugendlichen, bei Menschen mit Behinderung oder bei Menschen mit Demenz eine besondere Vorsicht geboten ist.

(Folie 8)

Wir haben dazu unter anderem mit den Kolleginnen und Kollegen, die heute hier sind, in den letzten Jahren gemeinsam viel gearbeitet zu den konzeptuellen Grundlagen von Begriffen wie Mental Integrity oder Mental Privacy, die im Moment

eine große Rolle in der internationalen Debatte spielen.

Mental Integrity, hatten wir gesagt, kann man analog zur körperlichen Integrität verstehen als Schutz vor bestimmten Formen des Eingriffs.

(Folie 9)

Vulnerabilität ist etwas, was im Bereich Mensch-Technik-Interaktion eine zunehmende Relevanz hat und wird insofern in unserer Arbeit verstanden als übergreifende anthropologisch-ethische Dimension in der Interaktion zwischen Mensch und Technologie. Dazu gehört auch, dass neue Technologien je nach Verbreitungsgrad und je nach Zugriffsmöglichkeiten, wie ich gesagt habe, bestehende strukturelle Ungerechtigkeiten noch verschärfen können.

(Folie 10)

Neurointerventionen bei Personen mit erhöhter Vulnerabilität sind stark kontextabhängig und von der Art der Intervention abhängig. Kinder und Jugendliche wurden schon als besonders schutzbedürftige Gruppe genannt, ebenso Menschen mit Behinderung, Menschen mit einer Demenzerkrankung oder kognitiven Einschränkungen im höheren Alter, aber auch der große Bereich der psychischen Erkrankungen mit den damit einhergehenden Einschränkungen, beispielsweise der Einwilligungsfähigkeit oder der emotionalen Steuerung.

Auch andere Aspekte können eine Rolle spielen. Deswegen ist es mir wichtig zu betonen, dass wir diese Vulnerabilität immer sehr genau und kontextspezifisch bestimmen und schauen sollten, welche Risikokonstellation vorliegt, und dass wir, wie es in der ethischen Debatte ja sein soll, auch eine Nutzen-Risiko-Abwägung für die verschiedenen Anwendungskontexte machen sollten.

(Folie 11)

Vorhin fiel der Begriff der personalen Identität. Das hängt aus meiner Sicht eng mit mentaler Integrität zusammen, weil Neurotechnologien nicht nur gezielt bestimmte Funktionen wie Gedächtnis oder Sprache verändern können, sondern möglicherweise auch längerfristige Dispositionen wie Präferenzen, affektive Lagen oder die Selbstwahrnehmung beeinflussen können. Das heißt, es kann zu einer Beeinträchtigung oder Verlusterfahrung der Selbstkohärenz kommen. Wir haben von Agency als mögliche Dimension der Beeinträchtigung gehört.

Diese Gegenüberstellung von: wo ist es Therapie und wo geht es in den Bereich der Identitätsveränderung? ist kritisch zu sehen. Man kann sie in vielerlei Hinsicht auch analog zur Enhancement-Debatte und den schwierigen Bestimmungen und Übergängen betrachten.

Die Implikation wäre, wenn wir jetzt mentale Integrität als besonders schützenswert betrachten und sagen: Personale Identität ist eigentlich ein Bereich der mentalen Integrität, sodass ein effektiver Schutz mentaler Integrität im Idealfall auch personale Identität schützen sollte. Aber man sollte es als bestimmten Bereich natürlich besonders anschauen.

(Folie 12)

Wir haben das in unserer Arbeit auch stark mit der Perspektive der Phänomenologie und phänomenologischen Interviewmethoden verknüpft, die uns hierfür nützlich erscheinen.

(Folie 13)

Die regulatorische Umgebung für Neurotechnologie ist auf der einen Seite geprägt von dem stark regulierten Bereich der CE-zertifizierten Medizingeräte für die engere klinische Anwendung. Auf der anderen Seite haben wir den großen, wachsenden Bereich der Consumer-Geräte für

Lifestyle-Anwendungen und dazwischen diesen Graubereich der medizinnahen Anwendungen für Wohlbefinden, Stressabbau bis hin zu kognitiver Leistungssteigerung oder affektiven Anwendungen.

Hier muss mitunter eine unterschiedliche Regulatorik zum Einsatz kommen. Dieses komplexe Geflecht aus regulatorischen Anforderungen stellt natürlich auch erhebliche Anforderungen beispielsweise für sinnvolle Innovationen für kleinere und mittlere Unternehmen, für Start-ups, aber auch im klinischen Bereich.

(Folie 14)

Mit der UNESCO-Empfehlung [zur Ethik der Neurotechnologie, 2025] gibt es ein weltweit wirksames Soft-Law-Instrument, wo unter anderem der Schutz der mentalen Integrität genannt und starkgemacht wird.

(Folie 15)

Aus einem Projekt heraus, wo wir intensiv mit Menschen mit Behinderungen als Co-Forschenden zum Thema Neurotechnologie und mentale Privatheit arbeiten, haben wir das Ganze noch mal geprüft und gesagt: Okay, bestimmte Aspekte, die bereits zum Beispiel in der UN-Behindertenrechtskonvention aufgeführt sind, könnte man in diese Empfehlungen noch integrieren.

(Folie 16)

Denn wie auch schon in Ilina Singhs Vortrag gesagt wurde, ist das Thema Partizipation und Teilhabe ein wesentlicher Faktor in der Innovation von Neurotechnologie, auch in der Akzeptanz und in der ethischen Bewertung. Denn neben den verschiedenen Dimensionen von Teilhabe ist insbesondere der Bereich der sozialen Teilhabe, aber

auch die anderen Bereiche in besonderer Weise wichtig, wenn sie zum Beispiel von assistiven Systemen effektiv gefördert oder verbessert werden können.

(Folie 17)

Häufig ist es aber in Forschungs- und Entwicklungsprojekten sehr schwierig, einen hohen Grad an Partizipation zu erreichen. Das ist mit enormem Aufwand verbunden und mitunter auch – das darf man nicht vergessen – mit einer emotionalen, moralischen Belastung für die Beteiligten.

(Folie 18)

Wir versuchen, das in diesem Projekt<sup>5</sup> so zu gestalten, dass wir die Menschen mit Behinderung von vornherein als Co-Forschende in das Projekt mit einbeziehen.

(Folie 19)

Als Ausblick: Wenn es um die Regulierung geht, haben wir mit der risikobasierten Regulierung im AI Act schon eine Vorlage, wo wir sagen können:

(Folie 20)

Gibt es in anderen Regulierungsbereichen, wie der Medizinprodukteverordnung, schon risikobasierte Ansätze, die vielleicht helfen können, zum Beispiel implantierbare Neurotechnologien oder eher niedrig risikobehaftete Neurotechnologien zu unterscheiden?

(Folie 21)

Letztlich braucht man für die Governance von Neurotechnologie aber allgemeine Prinzipien, nämlich einen menschenrechtsbasierten Ansatz, der technologisch und wissenschaftlich fundiert und möglicherweise auch regulierungsneutral ist,

<sup>5</sup> Folie: „Projekt PIVETDIS: Partizipative Forschung mit Menschen mit Behinderung zu Neurotechnologie“

damit es auf globaler Ebene keine Diskontinuitäten in der Regulierung gibt.

(Folie 22)

Staaten können viel tun, indem sie Human Rights Impact Assessments und Mental Impact Assessments integrieren, indem sie klare rote Linien festlegen für Praktiken, die auch aus ethischer Perspektive nicht akzeptabel sind, und indem sie klare Rechtsbehelfe schaffen; darüber kann uns Christoph Bublitz sicherlich noch mehr sagen.

(Folie 23)

Man muss auch überlegen, ob implantierbare Neurotechnologien für nichtmedizinische Anwendungen überhaupt erlaubt werden sollten oder ob wir hier nicht viel mehr Forschung brauchen, um zu verstehen, was die langfristigen Auswirkungen auf kognitive, affektive Funktionen, das Wohlbefinden und die personale Identität sind.

(Folie 24)

Auf der Empfehlungsebene würde ich sagen: Wir brauchen eine vulnerabilitätssensible Nutzen-Risiko-Abwägung, nicht eine pauschale Freigabe oder ein pauschales Verbot; eine gestufte Einwilligung, einen Fokus auf Zugangs- und Teilhabeerechtigkeit, einen besonderen Fokus auf die Frage des Einsatzes zu Hause. Gerade bei älteren Personen und Personen mit erhöhter Vulnerabilität. Und die Förderung partizipativer Entwicklung.

(Folie 25)

Im nichtmedizinischen Bereich brauchen wir möglicherweise höhere regulatorische Schwellen, als wir sie im Moment haben: strikte Regeln bis hin zu roten Linien.

(Folie 26)

Was man dafür braucht, ist ein Instrument für die Bewertung, zum Beispiel ein Mental Impact As-

essment, wo man anhand des Kontexts, bestimmter Red Flags und bestimmter Kriterien, die noch zu definieren wären, eine risikobasierte Einordnung von Neurotechnologien machen kann. Das ist nur ein Entwurf, ein Work in Progress, an dem wir im Moment arbeiten, aber das wäre eine operationalisierbare und sichtbare Möglichkeit, das in Zukunft zu machen.

(Folie 27)

Damit bedanke ich mich ganz herzlich und freue mich auf den Austausch und die Diskussion.

### **Hans-Georg Dederer**

Herzlichen Dank, Herr Kellmeyer, für den spannenden und detailreichen Vortrag.

Der nächste Impuls kommt von Herrn Marcello Ienca. Herr Ienca ist Professor für Ethik und künstliche Intelligenz und Neurowissenschaften an der Technischen Universität München und außerordentlich ausgewiesen in den Bereichen der ethisch-rechtlichen Fragen der Neurotechnologie. Er war auch ein maßgeblicher Mitautor des Konzepts von Neurorechten, Neuro Rights, und in der Ad Hoc Expert Group der UNESCO, die die UNESCO-Empfehlung zur Ethik der Neurotechnologie ausgearbeitet hat, das Dokument, das heute bereits mehrfach erwähnt wurde und im November 2025 publiziert wurde.

Bitte schön, Herr Ienca.

### **Nichtmedizinische Neurotechnologien: Ethische Grenzen, mentale Selbstbestimmung und Governance**

**Marcello Ienca · Technische Universität  
München**

Herzlichen Dank. Es ist mir eine Ehre, hier sprechen zu dürfen. Mein Vortrag befasst sich mit dem Thema nichtmedizinische Anwendungen der

Neurotechnologie mit dem Ziel, das Feld zu kartieren, ethische Prioritäten zu identifizieren und mögliche Governance-Lösungen aufzuzeigen.

Ich möchte definitorisch anfangen. Nichtmedizinische Neurotechnologien bezeichnen Anwendungen von Neurotechnologie außerhalb des Gesundheitsbereichs.

(Folie 2)

Dieser Bereich ist klar zu unterscheiden von medizinischer Neurotechnologie, also Anwendungen, die für gesundheitsbezogene Zwecke eingesetzt werden, etwa zur Therapie, Diagnose, Prävention, Rehabilitation oder als assistive Technologie, wie wir vorher gehört haben. Meine zentrale These ist heute, dass diese Unterscheidung von großer ethischer und gesellschaftlicher Bedeutung ist.

Medizinische Anwendungen der Neurotechnologien stellen einen wichtigen Innovationsmotor für Deutschland und Europa dar. Sie öffnen auch neue Lösungswege für Millionen Menschen mit neurologischen oder psychiatrischen Krankheiten weltweit, um ihr Leben zu verbessern. Und sie verbessern schon das Leben von vielen Menschen weltweit.

Neurotechnologie ist auch eine der wenigen Zukunftstechnologien, wo Europa immer noch wettbewerbsfähig mit den USA und China ist. Wir haben eine starke Positionierung von Europa in dem breiteren neurotechnologischen Bereich. Demgegenüber zeigt die nichtmedizinische Anwendung der Neurotechnologie bislang ein begrenztes Innovationspotenzial. Sie stützt sich häufig auf wissenschaftlich fragwürdige Grundlagen und bringt zugleich neue Risiken mit sich, die ich im Verlauf meines Vortrags erläutern werde.

(Folie 3)

Nichtmedizinische Neurotechnologien umfassen natürlich keine Verfahren wie MRT-Scans, deren großes Gewicht und Einsatzmöglichkeiten für die außermedizinische Anwendung einfach nicht geeignet sind. Vielmehr beziehen sich nichtmedizinische Neurotechnologien auf neuronale Schnittstellen, das heißt personalisierte Geräte, die Gehirnaktivität aufzeichnen und/oder stimulieren können.

(Folie 4)

Laut UNESCO-Daten hat sich die Gesamtzahl neuer neuronaler Schnittstellen in den letzten 20 Jahren um den Faktor sechs erhöht, während sich die Marktgröße in den vergangenen fünf Jahren verdoppelt hat. Wie wir vorher gehört haben, machen nichtmedizinische Anwendungen heute den größten Teil des Wachstums des Neurotechnologiemarkts aus.

(Folie 5)

Nichtmedizinische Neurotechnologien finden heute in mehreren Bereichen Anwendung. Die Hauptbereiche sind militärische Anwendungen, Arbeitsplatz, Bildung, Unterhaltung, Gerätesteuerung, Selbstvermessung und Leistungssteigerung und bald auch nichtverbale Kommunikation.

(Folie 6)

Hier sind einige Beispiele für militärische Anwendungen. Ich möchte betonen: Bei zivilen Neurotechnologien geht es nicht um eine Zweckentfremdung. Ganz im Gegenteil. Einige der ersten Gehirn-Maschine-Schnittstellen aus den Siebziger- und Achtzigerjahren gingen aus der militärischen Forschung hervor. Die Dynamik zwischen militärischer und ziviler Neurotechnologie ist daher nicht einseitig, sondern zirkulär, wie wir auch in unserem Artikel dargelegt haben.

(Folie 7)

Auch Anwendungen am Arbeitsplatz sind bereits im Einsatz. In Berufen mit höherer Verantwortung berichten einige staatlich unterstützte Unternehmen in China, insbesondere im Bereich Sicherheit für Atomkraftwerke, dass Beschäftigte verpflichtet sind, EEG-basierte Neurotechnologien, also Neuromonitoring-Headsets zu tragen.

Vergleichbare Berichte liegen aus Europa oder Nordamerika bislang nicht vor. Allerdings führen KI-Unternehmen heute bereits verpflichtende Anwendungen von KI ein, was einen Präzedenzfall für zukünftiges Neuromonitoring schaffen könnte.

(Folie 8)

Hier ist ein Beispiel aus dem Bildungsbereich. Da haben einige Unternehmen, etwa die amerikanische Firma BrainCo, Pilotversuche mit tragbaren neuronalen Schnittstellen zur Überwachung der Konzentration in Grundschulen in einigen chinesischen Provinzen durchgeführt. Das Gerät, das Sie sehen, funktioniert mit einer Art Ampelsystem. Grünes Licht heißt: Das Kind ist konzentriert. Rot heißt: Es ist abgelenkt.

(Folie 9)

Hier sehen Sie andere Anwendungen im Unterhaltungsbereich, sogenannte Neurogaming-Anwendungen, also Videospiele, die mit dem Gehirn gesteuert werden, und auch Kommunikation, also Gerätesteuerung, hauptsächlich Drohnen.

(Folie 10)

Das sind Geräte, die schon auf dem Markt sind. Wie Sie sehen, sind die meist EEG-basiert. Es gibt auch einige Technologien, die für nichtinva-

sive Neuromodulation geeignet sind. Interessanterweise hat Apple eine miniaturisierte Elektroenzephalografie für die nächste Generation AirPods patentiert.

Das habe ich schon in der Pause gesagt: Aus meiner Sicht wird der ChatGPT-Moment von Neurotechnologie nicht aus rein neurotechnologischen Anwendungen stammen, sondern aus der Integration von neurotechnologischen Kapazitäten in Alltagstechnologien, die wir heute schon verwenden.

(Folie 11)

Wir haben vor einigen Jahren eine Studie durchgeführt, einen systematischen Überblick über den nichtmedizinischen Markt von Neurotechnologien.<sup>6</sup>

(Folie 12)

Ein interessanter Fall ist die Firma Neuralink. Sie verkauft im Moment keine nichtmedizinischen Neurotechnologien, aber ist meines Wissens die einzige Firma weltweit, die schon öffentliche Behauptungen über die Nutzung von invasiven, also implantierten Neurotechnologien für die gesamte Bevölkerung macht, also nicht nur für Patienten, sondern auch für gesunde Menschen.

Gestern habe ich im Zug die Nachricht aus China gelesen, dass die erste invasive BCI für kommerzielle Nutzung in China genehmigt wurde.

(Folie 13)

Nichtmedizinische Neurotechnologien sollten nicht als medizinische Technologien mit möglichen Anwendungen im Consumersektor verstanden werden, sondern vielmehr als eine Fortsetzung der datengetriebenen Ökonomie der großen Big-Tech, also der Technologieunternehmen, mit

---

<sup>6</sup> Ienca, M. et al. (2018): Brain leaks and consumer neurotechnology. In: Nature Biotechnology, 36 (9), 805–810.

anderen Mitteln. Die gesamte Datenökonomie kann als eine Art Wettrüsten um den menschlichen Geist verstanden werden, als der Versuch, Absichten, Präferenzen, Wünsche und Verhaltensweisen algorithmisch vorherzusagen. Bisher war dies nur auf der Grundlage von indirekten sogenannten Proxydaten möglich. Nichtmedizinische neuronale Schnittstellen können heutige algorithmische Vorhersagemodelle mit direkten neuronalen Korrelaten von Absichten, Präferenzen und ähnlichen mentalen Zuständen verbinden.

Vor diesem Hintergrund überrascht es nicht, dass Big-Tech-Firmen wie Apple und Google begonnen haben, in diesem Bereich zu arbeiten, dass OpenAI, ein BCI-Schwesterunternehmen, genannt Merge Labs, gefördert hat und dass Meta ein BCI-Unternehmen namens Control Labs übernommen hat.

(Folie 14)

Im letzten Teil meiner Präsentation möchte ich fünf Prioritäten hervorheben, die aus meiner Sicht im Moment am wichtigsten sind.

(Folie 16)

Zuerst sind kontextabhängige Genehmigungs- und Einwilligungsregeln zu entwickeln. Es ist sehr wichtig, dass für diese nichtmedizinischen Kontexte strengere Schutzstandards gelten sollten. Ich habe in den letzten Jahren die These vertreten, dass diese Anwendungen prima facie eingeschränkt werden und nur unter kumulativer Erfüllung folgender Voraussetzungen zulässig sein sollten: eine nachweisbare und legitime Notwendigkeit, das Fehlen weniger eingriffsintensiver Alternativen und die Durchführung eines unabhängigen Human Rights und Data Protection Impact Assessment.

Mit Data Protection Expert aus den Niederlanden haben wir 2022 das Konzept von Mental Data

Protection Impact Assessment eingeführt, um ein spezifisches Data Protection Impact Assessment für Neurotechnologie zu haben.

(Folie 17)

Zweite Priorität ist, den Schutz der Privatsphäre und die Sicherheit der neurotechnologischen Infrastruktur zu stärken. Ich habe vorher unsere Nature-Biotechnology-Studie zu nichtmedizinischen Anwendungen von Neurotechnologie erwähnt. Darin haben wir herausgefunden, dass diese Technologien meist ungeschützte Datenübertragungskanäle verwenden. Daher können kontextreiche Informationen mit Big-Data-Analysen genutzt werden, um datenschutzsensible Schlussfolgerungen auch aus nicht sensiblen Daten zu ziehen. Das passiert ohne institutionelle Aufsicht durch Ethikkommissionen und mit begrenzter Nachvollziehbarkeit und Überprüfbarkeit der Schlussfolgerungen durch Black-Box-KI. Dafür werden also nicht transparente KI-Modelle verwendet.

(Folie 19)

Zum Thema Privatheit gehört auch das Problem des Gedankenlesens. Hier können wir lange darüber sprechen, ob die Metapher „Lesen“ epistemisch relevant ist oder nicht. Wichtig ist, dass wir heutzutage KI-Modelle zur Dekodierung von semantischen und visuellen Inhalten aus mentalen Zuständen verwenden können. Das ist ein Ansatz, der Stable Diffusion heißt, wodurch Bilder rein anhand von Gehirndaten rekonstruiert werden.

Was Sie hier oben sehen, ist der ursprüngliche visuelle Stimulus, also was die Leute in einem Video gesehen haben, und unten ist das, was der Algorithmus rekonstruiert hat.

(Folie 20)

Ich möchte dafür plädieren, dass die Metapher des Gedankenlesens nicht als unmittelbarer Zugriff

auf Gedanken oder als eine Form der Telepathie zu verstehen ist, sondern vielmehr analog zu dem, was wir unter Lesen beim Buchlesen verstehen, das heißt ein semiotischer Prozess, wodurch Signale in Bedeutungen umgewandelt werden. Ich glaube, das ist die beste epistemische Interpretation.

(Folie 22)

Dritte Priorität ist, die KI für Neurotechnologie erklärbar zu machen. Aus der Forschung mit meiner Doktorandin Laura Schopp haben wir festgestellt, dass weniger als 10 Prozent aller KI-gestützten Ansätze in der Neurotechnologie erklärbare KI verwenden. Wir haben eine Interviewstudie gemacht, um die Gründe dafür zu verstehen. Es gibt natürlich technologiebezogene Einschränkungen, aber es scheint, dass die Einstellung der Stakeholder, hauptsächlich der Kliniker, ganz wichtig ist.

(Folie 23)

Vierter Punkt: Nutzerzentrierung und Standardisierung ausbauen. Wir haben im Moment wenig Forschung zu menschenzentrierten Designvoraussetzungen für Neurotechnologien.

Ich habe vorhin gesagt, es fehlen noch systematische Studien. Wir haben versucht, mit Dr. Bublitz und zwei weiteren Kollegen die erste systematische Studie mit über hundert Patient:innen durchzuführen. Wir brauchen aber viel mehr Studien dieser Art.

(Folie 24)

Und last but not least ist es sehr wichtig, Menschen vor unrealistischen Erwartungen und irreführenden Praktiken zu schützen. Hier ist ein Tweet von Elon Musk: „First Neuralink product will enable someone with paralysis to use a smartphone with their mind faster than someone using thumbs.“

Das ist natürlich noch nicht der Fall. Diese Erwartungen, die wir im Bereich der Neurotechnologie generieren, können direkten Einfluss auf die Gesundheit von Menschen haben. Wir haben mit einer Doktorandin einen Case Report analysiert von einer 66-jährigen Patientin mit parkinsonischem Typ, Multisystematrophie, die ein DBS[Deep Brain Stimulation]-Implantat bekommen hat und sehr hohe Erwartungen an die Nutzung dieser Technologie hatte, besonders durch Online-Informationen, die sie gefunden hat. Und dann war sie so desillusioniert und enttäuscht, dass dies zu einem Suizidversuch geführt hat.

Ich möchte jetzt hier stoppen. Das Thema UNESCO-Empfehlung können wir gern in der Diskussion besprechen, wenn Sie Fragen dazu haben. Vielen Dank.

### **Hans-Georg Dederer**

Auch Ihnen, Herr Ienca, herzlichen Dank für die spannenden Einsichten und Ihre anschaulichen Schlussfolgerungen.

Der dritte Redner ist Christoph Bublitz. Er ist Rechtswissenschaftler und Rechtsphilosoph an der Universität Hamburg. Ich glaube, man kann sagen, Herr Bublitz, dass es unter den Juristen, Juristinnen in Deutschland, aber auch international niemanden gibt, der profiliert wäre auf dem Gebiet des Rechts der Neurotechnologien. Sie haben dazu viel publiziert, wie wir schon gehört haben, mit Herrn Kellmeyer und mit Herrn Ienca.

Bei Ihnen geht es oft um Freedom of Thought, Gedankenfreiheit, und das wird auch das Thema Ihres Impulses sein. Bitte schön.

### **Gedankenfreiheit als Fixpunkt der Regulierung von Neurotechnologien**

**Jan Christoph Bublitz · Universität Hamburg**

Vielen Dank für die netten Worte und für die Einladung. Ich habe keine PowerPoint. Das heißt, Sie

können jetzt Ihre Augen schließen und ich würde dann versuchen, Sie 15 Minuten lang vom Schlafen abzuhalten.“

Ich bin Rechtswissenschaftler, und mein Forschungsgebiet sind die Bezüge von Recht und Psyche. Das hat natürlich mehrere Ebenen, aber eine Frage ist: Welche Regeln haben wir für den Umgang mit der Psyche, gerade im Vergleich zu den Regeln über den Umgang mit Körpern, die ja sehr unterschiedlich sind? Neurotechnologien sind da ein interessanter Fall, weil sie mit relativ geringen körperlichen Einwirkungen große psychische Effekte erzielen können. Das heißt, es stellt sich die Frage nach einem eigenständigen Schutz der Psyche.

Ich glaube, dass das ein generelles Problem in der Regulierung von digitalen Technologien ist, dass wir keinen eigenständigen Schutz der Psyche haben und deswegen viele psychische Folgen digitaler Technologien unter dem regulatorischen Netz durchfließen und nicht auftauchen. Dazu möchte ich vielleicht später noch etwas sagen. Hauptsächlich möchte ich meine paar Minuten vor Ihnen aber für den Hinweis auf ein rechtliches und ethisches Prinzip nutzen, das bei Neurotechnologien von großer Relevanz sein sollte, derzeit aber noch nicht ist, weil es noch nicht hinreichend anerkannt und ausgearbeitet ist: Das ist die Idee der Gedankenfreiheit. Das ist ein großer und vielschichtiger Begriff, der ideengeschichtlich ganz schön aufgeladen ist. Es ist aber auch ein bereits existierendes Menschenrecht, etwa in der Europäischen Menschenrechtskonvention. Sie alle haben also das Recht auf Gedankenfreiheit. Das entstand 1948 in der Allgemeinen Erklärung der Menschenrechte. Was es schützen soll, wurde grob umschrieben als Raum der Gedanken, innerer Denkraum, die innere Freiheit, die Freiheit des Geistes. Das sind vage Definitionen. Genauer wurde es damals nicht besprochen.

Aber es war klar: Es geht irgendwie um den Schutz der menschlichen Innenwelt als zentraler Bestandteil des Personenseins. Und es war auch klar, dass dieser Schutz als absoluter Schutz ausgestaltet ist. Das heißt, Eingriffe in die Gedankenfreiheit kann man – anders als bei vielen anderen Menschenrechten – nicht gegen andere Ziele abwägen. Das war der Gedanke.

Damals ging es natürlich nicht um Neurowissenschaften, aber nicht um ganz andere Themen: Da gab es die Besorgnis vor der aufkommenden Massengesellschaft, Massenmedien, all die Einflüsse und Einwirkungen der modernen Welt, gegen die der Einzelne nicht viel auszurichten hat. Das war damals das Denken.

Als Gegensatz zu diesen ganzen Einflüssen der Moderne wurde das Recht auf Gedankenfreiheit konzipiert, und ganz grob verkörpert es die Idee, dass gewisse Bereiche des menschlichen Geistes, *wherever they precisely are*, prinzipiell Eingriffen des Staates, aber auch von Privaten entzogen sein sollen. Das ist die Idee dahinter.

Aber diese Idee – und das Recht – ist dann lange in Vergessenheit geraten, und es hat jedenfalls im Recht keine praktische Relevanz entfaltet. Im Grundgesetz findet es sich auch nicht ausdrücklich. Manchmal nehmen aber Gerichte darauf Bezug, etwa der Bundesgerichtshof in einer Entscheidung aus dem Jahr 2011. Da hat die Polizei das Selbstgespräch eines Beschuldigten abgehört, das er in seinem Auto mit sich selbst geführt hat. Der BGH hat gesagt: Das ist nicht zulässig. Das darf auch nicht verwertet werden, weil die Gedanken absolut geschützt sind, und das laute Denken, das Mit-sich-selber-Sprechen jedenfalls im nicht-öffentlichen Bereich, muss auch darunterfallen. Indirekt wird dieses Recht also auch im deutschen Recht anerkannt.

Aber dieses Anerkenntnis ist fragil. Es fehlt an einem tieferen Verständnis dieses Rechtes und an

einer robusten Rechtsprechung. Es bedarf also der theoretischen Konkretisierung in Zusammenarbeit von Recht und Ethik.

Deswegen müsste man heute fragen: Ist die Idee eines absolut geschützten Innenbereichs überhaupt noch attraktiv? Und ist sie auch überzeugend? Und wie ließe sie sich genauer fassen?

Und wenn man so fragt, erheben sich sofort die Einwände: Natürlich ist der Mensch einer Vielzahl von Einflüssen und Einwirkungen ausgesetzt. Unser Denken ist nicht hermetisch abgeriegelt, und im Grunde genommen wollen wir auch keine kognitiv isolierten Individuen. Das ist nicht das Modell der Gedankenfreiheit, was besonders attraktiv ist. Das heißt aber nicht, dass die Idee und dass andere Einwirkungen alle gleich sind. Also kommt es vielleicht darauf an, den richtigen Arten von Einflüssen auf die richtige Art und Weise ausgesetzt zu sein.

Man könnte vielleicht in Referenz an Jürgen Habermas sagen, dass alle kommunikativen Einwirkungen keine Eingriffe in die Gedankenfreiheit darstellen, solange sie ohne körperliche Zwangsmittel oder andere Methoden, die den anderen als eigenständigen Denker nicht respektieren, auskommen. Solange sie ohne sie dahin kommen, könnte man sagen, dass diese Eingriffe nicht darunterfallen sollen.

Man müsste erklären, worin genau der Eingriff in die Gedankenfreiheit liegt. Das Abhören – haben wir gerade gesehen – wird vom Bundesgerichtshof anerkannt. Und wenn das Abhören schon unzulässig ist, dann wäre es das Auslesen mit dem Gedankenlesen gegen den Willen des Betroffenen sicherlich auch.

Ich würde sagen, es kommt nicht darauf an, dass man den Inhalt auslesen kann. Meiner Meinung nach würde es schon reichen, wenn man die Art

des Denkens auslesen könnte, etwa ob eine Person gerade tagträumt – wie Sie vielleicht – oder ob man sich erinnert oder rechnet. Der Bereich der Gedankenfreiheit beginnt dort, wo man den Typus der mentalen Handlung auslesen könnte, würde ich sagen. Das müsste man im Einzelnen alles klären.

Ich möchte auch nicht sagen, dass es heute schon möglich ist, aber das ist der Fluchtpunkt der Entwicklung. Laborstudien, haben wir gesehen, deuten an, dass das zumindest möglich sein könnte.

Die andere Art von Eingriffen sind Eingriffe, die das Denken verändern, etwa durch Hirnstimulationsverfahren. Das muss die Gedankenfreiheit natürlich auch umfassen.

Dann kommt immer das Gegenbeispiel: Ja gut, aber wir machen auch psychiatrische Zwangsbehandlungen, wo zwangsweise Mittel verabreicht werden, die zum Beispiel psychotisches Denken verändern. Und weil wir das weiterhin machen wollen, ist das Recht auf Gedankenfreiheit so nicht zu begründen.

Ich würde sagen, das ist ein schnelles Gegenbeispiel, das letztlich nicht überzeugt. Der Schlüssel liegt darin, verschiedene Dimensionen von Freiheiten zu unterscheiden, wie Sie das auch anderswo kennen: Es gibt die Freiheit *vor* Einwirkungen von außen, aber auch die Freiheit *zum*, die positive Freiheit zum freien Denken. Das muss man gehaltvoller verstehen. Das ist unabhängiges, eigenständiges, rationales Denken, traditionell in der Begriffsgeschichte, aber heutzutage sicherlich auch kreatives oder künstlerisches Denken.

All diese Denkformen haben Voraussetzungen, und man kann nicht einfach die negative und die positive Freiheit zugunsten einer Seite gewinnen lassen. Man muss sie miteinander ins Verhältnis setzen. Wenn es staatliche Maßnahmen gibt, die

solche notwendigen kognitiven Fähigkeiten fördern oder ermöglichen, und wenn das bei Menschen eingesetzt wird, die vielleicht nicht ganz autonom sind, wie es häufig in der Psychiatrie der Fall ist, dann kann es sein, dass das nicht unter ein absolutes Verbot der Gedankenfreiheit fallen kann. Das lässt sich gut begründen, glaube ich. Man bräuchte also solche Erwägungen, um dieses Recht genauer zu umfassen. Meine Idee ist, dass das den Fixpunkt bilden soll – so auch der Titel meines Vortrags – und dass wir das als oberstes Prinzip nehmen, das einen absolut geschützten Bereich abschichtet. Das ist die rote Linie.

Aber viele Eingriffe und Einblicke der Neurotechnologien – das meiste, was heute möglich ist – dürfen nicht darunterfallen, die sind schwächer. Dafür bräuchten wir andere Grundrechte, und die gibt es schon: in Deutschland das allgemeine Persönlichkeitsrecht, und auf europäischer Ebene gibt es interessante Rechte, weil sie explizit die Psyche schützen: das Recht auf Mental Integrity – auf Deutsch etwas enger übersetzt: auf geistige Unversehrtheit – in der Grundrechtecharta. Auch in der EMRK [Europäische Menschenrechtskonvention] wird so etwas anerkannt.

Diese Rechte sind da, aber noch nicht weit ausgebaut. Die Frage wäre, ob sie neurotechnische Eingriffe erfassen können, die nicht zu Gesundheitsschäden führen. Ich glaube, dass diese Rechte dafür offen sind. Sie sind entwicklungs offen, und dafür sind sie eigentlich da. Das könnte gut passen. Noch ist das aber nicht die etablierte, feststehende Rechtslage.

Dann müsste man die Gedankenfreiheit im Verhältnis zu Privaten schützen. Dafür brauchen wir einfachgesetzliche Normen. Datenschutzrecht ist ein Ansatzpunkt, Strafrecht wäre ein anderer.

Dann kommen wir zu dem, was Herr Kellmeyer gerade gesagt hat, zur Regulierung von Neurotechnologien. Er sprach von einem Mental Impact

Assessment. Das ist eine gute Idee. In einem Mental Impact Assessment könnte man auch ein Freedom of Thought Impact Assessment integrieren. Das wird übrigens von einigen Institutionen auch so gefordert. Worum geht es da? Darin soll systematisch untersucht werden, wie diese Technologien – und das lässt sich auf alle digitalen Technologien ausweiten – eigentlich auf die Psyche wirken.

An der Stelle kurz zur Regulierung. In der EU ist es so: Wenn Sie ein Produkt auf den Markt bringen wollen, muss dieses Produkt sicher sein. Das ist EU-geltendes Recht. Dazu brauchen wir nicht in die Medizinprodukteverordnung zu gehen.

Das Problem ist nur: Was heißt „sicher“? Da sagt die EU-Kommission: Natürlich betrifft das auch psychische Fragen. Mental Health wird ausdrücklich darunter gefasst. Das Problem ist, dass diese Effekte subtil und nicht so einfach zu konzeptualisieren sind. Dafür brauchen wir ein Mental Impact Assessment, das diese Dinge auf der Operationalisierungsebene angeht. Wir brauchen also keine neuen Gesetze. Wir müssen nur klarer fassen: Was sind denn psychische Risiken dieser Geräte? Dann können wir viele Probleme auf dieser Ebene lösen. Das ist der Gedanke hinter dem Mental Impact Assessment.

Ungeachtet dieses Assessments, das ja Unternehmen durchführen müssen oder vielleicht eine Regulierungsbehörde und je nachdem, wie scharf man das stellen möchte, müssen Risiken minimiert werden. Trotzdem bleibt die Frage, ob man nicht bestimmte Anwendungen gerade im Verbraucherbereich kategorisch verbietet, etwa das Auslesen von Gedanken. Frankreich hat solche Verbote. Im AI Act gibt es Emotionserkennungsverbote in gewissen Bereichen. Das könnte man auf andere Bereiche des Geistes ausdehnen. Andere Länder denken auch darüber nach. Das ist

vielleicht eine Frage, die zuerst ethisch zu beantworten ist, bevor die Juristen die Verbote fordern.

Was heißt das jetzt alles für den Ethikrat?

Gedankenfreiheit ist eine große Idee. Sie ist theoretisch anerkannt. Es gibt sie als Menschenrecht seit 1948. Aber jetzt wäre es Zeit, dass die Idee aus den Büchern auf die Straße kommt, die Bücher verlässt und in der wirklichen Welt wirksam wird. Dafür braucht es zweierlei: zuerst ein ausdrückliches Anerkenntnis der Gedankenfreiheit. Das gibt es zwar im Vagen, aber so richtig nicht. Zweitens ihren tieferen inhaltlichen Ausbau. Zu beidem kann der Ethikrat ja beitragen.

Darauf könnte vielleicht auch der Bundestag aufbauen und vielleicht ein Gedankenfreiheitsgesetz erlassen. Heute ist der 19. März. In zwei Jahren, noch in dieser Legislatur, könnte man 180 Jahre Märzrevolution standesgemäß mit einem Gedankenfreiheitsgesetz angehen. Vielen Dank.

## Diskussion

### Hans-Georg Dederer

Auch Ihnen herzlichen Dank, Herr Bublitz, für den anschaulichen Vortrag. Alle drei Vorträge haben dem Ethikrat schon Empfehlungen gegeben. Das wird für die Arbeit sehr hilfreich sein.

Damit würde ich gern in die Diskussion einsteigen. Wir wollen die Diskussion etwas absichten: Zunächst wollen wir Fragen an die drei Referenten Herrn Kellmeyer, Herrn Ienca und Herrn Bublitz richten, damit diese drei Redner die Möglichkeit haben, ihre Positionen zu klären, zu schärfen oder nachzulegen. Danach soll die Diskussion geöffnet werden für das gesamte Podium aller sieben Expertinnen und Experten. Damit würde ich die Diskussion gern eröffnen. Die erste Wortmeldung von Aldo Faisal, bitte.

### Aldo Faisal

Herzlichen Dank für die spannenden Vorträge auch im zweiten Teil. Ich möchte mit einer Frage an alle anfangen: Wenn wir über neuronale Daten reden, und zwar Gehirndaten, Verhaltensdaten, die unmittelbar mit Gedanken und Gefühlen verbunden sind, müssten wir darüber anders oder schärfer nachdenken als über generelle Gesundheitsdaten? Dazu würde ich gern ein Statement von den drei Sprechern hören.

### Marcello Ienca

Erst mal muss man neuronale Daten definieren. Wir haben in der UNESCO-Empfehlung eine breite Definition von neuronalen Daten als Daten über die Aktivitätsstruktur und Funktion des menschlichen Nervensystems. Aber ob diese Daten anders als Gesundheitsdaten behandelt werden sollten, hängt von vielen Faktoren ab.

Erst mal der Anwendungsbereich, wenn sie außerhalb des Gesundheitssektors und für nichtmedizinische Zwecke verwendet werden. Die Interpretation von *allen* neuronalen Daten als Gesundheitsdaten ist natürlich schwierig, aber dazu kann Herr Bublitz mehr sagen.

Aus einer ethischen Perspektive – die These, die ich vertreten habe, ist: Das, was aus einer normativen Perspektive wichtig ist, ist nicht Datenarten oder Daten klassifizieren, sondern die KI-gestützten Schlussfolgerungen aus diesen Daten. Man kann aggressive Schlussfolgerungen ziehen über menschliche Gedanken oder Affekte und Emotionen anhand von nichtneuronalen Daten, zum Beispiel Gesichtserkennungs-KI. Aber die Integration von neuronalen Daten und Standardmodellen wird diese Schlussfolgerungen noch stärker und akkurater machen.

Meine kurze Antwort wäre: Der Fokus liegt nicht auf den Daten, sondern auf den Schlussfolgerungen, die wir basierend auf diesen Daten machen können.

### **Jan Christoph Bublitz**

Neurale Daten sind derzeit noch keine besonders geschützte Kategorie. Die erste Forderung, die man gerade im Digital-Omnibus-Verfahren gut einbringen kann – ich mache das auch –, ist zu sagen: Lass uns die erst mal in Artikel 9, also die besonders geschützte Datenkategorie, einbringen.

Dann ist die Frage: Soll es noch darüber hinausgehen? Ich habe gewisse Schwierigkeiten mit der Rhetorik der Einzigartigkeit neuraler Daten, die man in dem Diskurs oft trifft. Unter einer risikobasierten Abschätzung sehe ich nicht, dass es höhere Risiken gibt als meine sexuelle Orientierung und die anderen Entitäten in Artikel 9. Ich halte das für eine Kategorie: Einige Daten mögen genauso schlimm sein, viele andere aber – wenn ich zum Beispiel an das periphere Nervensystem denke – sind es nicht. Nur ich glaube, dass der Datenschutz in der Debatte ein bisschen als Allzweckinstrument benutzt wird. Man wird die Privacy-Probleme nicht auf Ebene des Datenschutzrechtes und der Kategorisierung von Datentypen lösen können. Ich glaube, da treffen wir uns auch. Man muss bestimmte Erhebungsarten, etwa im Bereich des Gaming, explizit verbieten, wie es zum Beispiel die KI-Verordnung macht.

### **Philipp Kellmeyer**

Ich schließe mich da gerne an. Ich wäre auch vorsichtig, Neurodaten hier als strikt eigenständige Kategorie zu behandeln, weil die Entwicklung im Gesundheitsdaten-, KI-Bereich zeigt, dass sich selbst die klassische Kategorie von Gesundheitsdaten zunehmend auflöst, weil wir Verhaltensdaten, Smartphone-Daten, Plattforminteraktionen mit Machine-Learning-Modellen so analysieren

können, dass damit Rückschlüsse auf mentale Zustände, kognitive Profile und prädiktive Modelle möglich sind.

Das heißt, Daten werden nicht durch ihre Natur sensibel, sondern durch die Auswertbarkeit und den Verwendungszweck. Deswegen würde ich es auch für zielführender halten, nicht spezifisch bestimmte Datenkategorien wie Neurodaten oder Gesundheitsdaten zu regulieren, sondern die Verwendungszwecke und die daraus entstehenden Risiken.

Das würde zum Beispiel im Kontext von KI bedeuten, dass es relevant ist, ob die Daten zur Diagnose einer bestimmten Disposition oder Krankheit benutzt werden, zur Verhaltensbeeinflussung oder zur Profilbildung, personalisiertes Marketing und solche Geschichten; nicht, ob sie ursprünglich mal von der Provenienz her Neurodaten oder Gesundheitsdaten sind, weil sich all diese Kategorien aus meiner Sicht zunehmend auflösen. Ich wäre auch dabei zu sagen, lieber ein stärker risikobasierter, kontextsensitiver Ansatz, der auch anschlussfähig ist an die Logik des AI Acts, statt sich auf zunehmend poröse Datenkategorien zu konzentrieren. Denn es können ja immer wieder neue Devices kommen, neue Device-Daten. Dann sind es irgendwann Quantendaten, andere Kategorien von Daten. Ich glaube, da läuft man dann immer hinterher.

### **Hans-Georg Dederer**

Herzlichen Dank. Wir haben noch eine Nachfrage von Aldo Faisal.

### **Aldo Faisal**

Der Hintergrund ist jetzt weniger, dass andere Einblicke in meine Daten bekommen, sondern dass de facto zurzeit der Datenschutz erst mit der Geburt anfängt und mit dem Tode erlischt. Und was passiert dann, wenn sozusagen Simulationen

von meinem Gehirn laufen können, die von Plattformbetreibern einfach geerbt werden, also wenn das Persönlichkeitselement erlischt, aber das Gehirn noch weiter benutzt werden kann?

### **Jan Christoph Bublitz**

Ich sehe jetzt nicht den Unterschied zu anderen persönlichen Daten, etwa mein Tagebuch, was Sie ja auch nach meinem Tod noch sehen können. Wenn neurale Daten während meiner Lebenszeit nicht besonders sind, dann sind sie es vielleicht nach meinem Tod auch nicht.

### **Cornelia Betsch**

Ich habe zwei Fragen an Sie, Herr Ienca. Wenn die anderen auch dazu etwas sagen wollen, gerne.

Sie haben angedeutet, das Militär war zuerst da, und wir werden uns darauf nicht fokussieren. Trotzdem haben Sie auch gezeigt, wie alles mit allem zusammenhängt. Vielleicht können Sie uns mal ein Gefühl dafür geben, wie groß der Eisberg unter dem Wasser ist und was aus Ihrer Sicht da die relevantesten Entwicklungen sind.

Das andere betrifft die Frage – wir gucken ja viel: Was macht KI mit den Neurodaten? Vielleicht können Sie was dazu sagen – das ist zwar ein bisschen eine Sci-Fi-Frage, aber das hat eine Kollegin aus der ersten Anhörung deutlich gemacht: Wir sollen auch Sci-Fi-Fragen stellen, und das tue ich deswegen hiermit – welches Potenzial sich für die KI aus dem Arbeiten mit so vielen Neurodaten ergibt. Kann die sich dadurch auch irgendwo anders hinentwickeln? Spielt das eine Rolle oder nicht?

### **Marcello Ienca**

Vielen Dank. Zwei sehr wichtige Fragen. Kurze Antwort zur ersten Frage: Wie groß ist der militärische Bereich? Das ist schwer zu sagen, weil viele Informationen klassifiziert sind. Aber wir

wissen, dass die DARPA [Defense Advanced Research Projects Agency] in den USA der größte Förderer für Neurotechnologie weltweit ist. Ich glaube, das ist schon ein Hinweis, dass das militärische Interesse in diesem Bereich sehr groß ist.

Zur zweiten Frage: Was kann die KI hier? Ich verwende nicht so gern das Wort Revolution in der Wissenschaft allgemein, aber ich glaube, in diesem Bereich ist es wirklich eine Revolution. Die Kapazitäten von Neurotechnologie sind heutzutage viel größer als vor 20 Jahren. Auf der Hardware-Ebene hat sich nicht viel geändert. Natürlich kleine Anpassungen, aber die große Revolution ist auf der Softwareebene. Wir haben jetzt viele KI-Modelle für Datenverarbeitung, die extrem effizient sind. Darüber hinaus ist die KI sehr wichtig für Gerätoptimierung, die Gehirn-Maschinen-Schnittstelle. Das sind die zwei großen Anwendungsbereiche.

Aber es gibt noch viel mehr. Zum Beispiel ist KI sehr wichtig, um die richtige Intensität von Neurostimulation vorherzusagen oder welche Gehirnareale stimuliert werden sollten, um am besten zu intervenieren. Es gibt also viele Anwendungsbereiche. Aber Geräte-Leistungsoptimierung und Datenverarbeitung sind die großen Anwendungsbereiche.

### **Hans-Georg Dederer**

Vielen Dank. Herr Kellmeyer würde gern ergänzen.

### **Philipp Kellmeyer**

Ja, ich wollte ergänzen, den Fokus darauf legen, dass ein wichtiges Potenzial die Möglichkeit ist, dadurch, dass wir zunehmend Neurodaten in großen Repositories zur Verfügung haben und immer mehr akkumulieren, möglicherweise auch durch hochskalierte Consumer-Anwendungen mehr Neurodaten anfallen, dass man solche großen

multimodalen Datensätze mithilfe generativer Architekturen auch in sogenannte neuronale Basismodelle oder Neural Foundation Models integrieren könnte, das heißt generative Basismodelle, die analog zu Sprach- oder Bildmodellen auf sehr großen, heterogenen Datensätzen vortrainiert werden und grundlegende statistische und funktionale Strukturen neuronaler Aktivität lernen. Das wäre ein Unterschied dazu, wie wir es im Moment machen, sehr klassisch auf spezifische Aufgaben trainierte Modelle.

Das Besondere liegt dann in der Multimodalität, wenn man Daten aus unterschiedlichen Methoden zusammenbringt. Das könnte eine ganz neue Ebene schaffen in der Integration von Grundlagenforschung, klinischer Anwendung und letztlich auch Anwendungen, die vielleicht ethisch problematisch sind. Die Möglichkeit von solchen Neural Foundation Models sollte man mitdenken, denn das, was man mit Bild- und Textdaten machen kann, könnte man auch mit Neurodaten machen, wenn man es richtig macht.

### **Hans-Georg Dederer**

Vielen Dank, Herr Kellmeyer, für die Ergänzung. Dann würde ich das Wort Kerstin Schlögl-Flierl geben.

### **Kerstin Schlögl-Flierl**

Vielen Dank für die drei spannenden Vorträge. Ich habe zwei Fragen.

Herr Bublitz, Herr Kellmeyer, Sie haben beide Artikel geschrieben zu einem Moratorium für implantierbare nichtmedizinische Anwendungen, und ich habe fast erwartet, dass Sie sagen, der Deutsche Ethikrat soll ein Moratorium fordern. Mich würde interessieren: Was wäre die Hauptbegründung? Was wäre die größte Verletzung? Also was ist so gefährlich an Verletzung, dass es ein Moratorium geben soll? Wie lange sollte es

dauern? Zeitperspektive. Und was müsste passieren, dass man sagen könnte, nun ist alles genügend reguliert und das Moratorium könnte wieder aufgehoben werden? Also eine Moratoriumsfrage.

Die zweite Frage geht an Herrn Kellmeyer. Der Hintergrund ist, dass wir im Deutschen Ethikrat schon viel zum Thema Vulnerabilität gearbeitet haben und Sie das Thema jetzt so prominent in die Neurotechnologie eingeführt haben. Aber was mir aufgefallen ist: Es ist eine starke Defizitperspektive, Vulnerabilität, und wir versuchen immer, in unseren Policy Papern auch eine Ressourcenperspektive von Vulnerabilität aufzuzeigen. Könnten Sie sich vorstellen, das auch so zu denken?

Von Ihnen her ist stark gedacht: Die besonders Vulnerablen müssen besonders geschützt werden. Ich würde sagen: Vulnerabilität hat noch viele andere Dimensionen, auch aus der Vulnerabilitätsforschung heraus. In dem Punkt würde ich Sie gern ein bisschen herausfordern. Danke.

### **Hans-Georg Dederer**

Eine Frage an zwei der Experten. Ich würde zunächst Herrn Bublitz die Gelegenheit geben, auf das Moratorium zu antworten. Dann wird Herr Kellmeyer ergänzen, denke ich.

### **Jan Christoph Bublitz**

Das Moratorium – wir wissen alle: Moratorien haben auch ihre Nachteile. Aber der Kerngedanke schließt an das an, was ich eben gesagt habe. Wenn wir sehen, wie schwer es ist, in digitalen Technologien die Risiken, vor allem die psychischen Risiken zu erfassen und zu konzeptualisieren, dann könnte man daraus ja lernen und sagen: Je näher wir an das Gehirn kommen – und da bei dem Moratorium geht es um implantierte Sachen, also um die Neuralink-Sachen im nichtmedizinischen Bereich, die vielleicht noch so zehn Jahre

von jetzt weg sind – dann ist die Frage: Was müssten die Bedingungen sein, damit wir so etwas überhaupt auf den Markt bringen wollen würden?

Dann schien uns zu sein: Der erste Punkt ist, dass wir es hinkriegen müssen, die psychischen Folgen zu erheben. Das sagt man so einfach, aber das ist für jeden, wir sind ja alle Wissenschaftler – psychische Folgen zu erheben, ist empirisch richtig schwierig. Was sind überhaupt die Methoden? Das sind dann andere Technologien. Da müssten wir, Philipp Kellmeyer und ich, mit den Phänomenologen sprechen, und da müssen die Psychologen mit ins Boot. Wenn man das ernst meint, ist das eine riesige wissenschaftliche Aufgabe, und die müsste man angehen, bevor wir solche Dinge auf den Markt lassen.

Das wäre auch die Antwort auf Ihre Frage: Was müssten die Umstände sein? Wir müssen zumindest ein wesentlich besseres Verständnis von diesen subtilen Effekten haben, und das Moratorium sollte sein, um einen Anreiz zu geben, das überhaupt zu entwickeln. Weil das sehr aufwendig ist und weil das viel Geld kostet und weil die Unternehmen mitspielen müssen, wäre es natürlich geschickter, das an so ein Moratorium zu binden. Das war eigentlich die Überlegung.

### **Philipp Kellmeyer**

Da würde ich mich komplett anschließen. Das ist der Aspekt der epistemischen Unsicherheit, die Herr Bublitz jetzt in den Fokus gestellt hat. Man könnte noch die Eingriffstiefe in die mentale Identität nennen. Wenn jetzt emotionale, affektive Zustände, Entscheidungsprozesse oder Präferenzstrukturen nachhaltig beeinflusst werden, wäre das auch ein Grund, genauer hinzuschauen. Wir haben halt eine riesige Evidenzlücke, was auch die langfristigen Folgen der Interaktion mit solchen Systemen betrifft. Deswegen wäre das eine Möglichkeit.

Wichtig ist, wie man das dann versteht. Es ist eine temporäre, klar begrenzte Aussetzung bestimmter Anwendungen, bis sich die Evidenzlage so verbessert hat, dass eine wissenschaftsbasierte rationale Regulierung, die Zulassung möglich ist. So würde ich das verstehen, oder so haben wir es auch versucht zu erklären.

Was jetzt die Ressourcenperspektive angeht: Da haben Sie natürlich recht, dass ich in meinem Vortrag nicht die Fähigkeiten und Handlungsspielräume in den Mittelpunkt gestellt hat, sondern eher den Schutz vor bestimmten Risiken.

Man könnte natürlich so argumentieren: Ein vertieftes Verständnis bestimmter Vulnerabilitäten im kognitiven, emotionalen, Bereich, beispielsweise bei neurodegenerativen oder psychischen Erkrankungen, also Vulnerability Assessment, hilft dabei, Hilfsbedarfe zu identifizieren, nämlich zu sagen: Genau in diesen Konstellationen, unter diesen Umständen sind Menschen besonders unterstützungsbedürftig, was die Ausübung von Autonomie angeht, was die Ausübung von Agency angeht.

Das heißt, eine Vulnerabilitätsperspektive kann insofern auch ressourcenaktivierend und enabling sein, wenn sie handlungsleitend ist. In der Forschung und Entwicklung sagen wir, wir priorisieren Neurotechnologien, die assistiv genau diese Vulnerabilitäten adressieren und vielleicht verbessern, dass Menschen mit eingeschränkter Möglichkeit zur Ausübung ihrer Autonomie, beispielsweise bei Locked-in-Syndrom, schweren Kommunikationsstörungen, Schwierigkeiten, verbal Einwilligung zu geben – dass ein Assistenzsystem, was das verbessert, im Sinne einer Autonomieprothese, wenn man so will, auch ressourcenstärkend sein kann.

Das habe ich nicht so in den Fokus gestellt, aber diese Perspektive würde ich sehr unterstützen und

würde mich freuen, wenn es im Bericht auch vorkommt, weil das ja eine zentrale Motivation ist in der klinischen Entwicklung, eben in diesen Bereichen besonders zu helfen.

### **Hans-Georg Dederer**

Vielen Dank, Herr Kellmeyer. Dann hätten wir eine Frage von Helmut Frister.

### **Helmut Frister**

Ich wollte mich zunächst dem Dank für die hervorragenden Referate anschließen und habe dann eine Frage an Herrn Bublitz.

Ich habe große Sympathien für die Idee eines absoluten Schutzes der Gedankenfreiheit. Aber absolut geschützt wäre dann nur der Zugriff oder der Eingriff ohne Einwilligung. Das heißt, es wäre grundsätzlich möglich, da eine informierte Einwilligung zu erteilen.

Das führt dann zu der Frage, welche Anforderungen an diese informierte Einwilligung zu stellen sind. Wir haben vom Deutschen Ethikrat ja eine Stellungnahme zu Suizidprävention und Suizidassistenten gemacht und da den Grundsatz aufgestellt: Je gewichtiger das Gut ist, um das es geht, desto höher müssen die Anforderungen sein. Da wäre meine Frage: Was bedeutet das hier, für diesen Fall? Und das vielleicht noch vor dem Hintergrund: Ich sehe eine besondere Schwierigkeit darin, dass man so eine informierte Einwilligung möglicherweise auch durch manipulative Techniken erzeugen kann, die ihrerseits vielleicht schon einen Eingriff in die Gedankenfreiheit darstellen.

Wenn ich noch eine zweite Frage anschließen darf: Sie haben berichtet, dass in einigen Staaten ein kategorisches Verbot erwogen wird. Das habe ich so verstanden, kategorisch auch mit Einwilligung. Wie könnte man das rechtfertigen? Ist das so eine patriarchalische Rechtfertigung oder vielleicht eine Rechtfertigung, die daraus resultiert, dass ein Zugriff auf Gedanken, auch wenn er mit

Einwilligung folgt, auch allgemeingesellschaftliche Konsequenzen hat, über die der Einzelne gar nicht verfügen kann?

### **Jan Christoph Bublitz**

Das wäre sicherlich eine Argumentationslinie. Der Paternalismuskurs, der da im Hintergrund schwebt, und die realen rechtlichen, auch in einem liberalen Verfassungsstaat gezogenen Grenzen für Einwilligung – das sind zwei Ebenen, die man auseinanderhalten kann. Wir haben zum Beispiel in der KI-Verordnung das Verbot subliminaler Einwirkungen. Das gilt generell, da können Sie auch nicht drin einwilligen. Die Idee ist: Wir wollen solche Technologien gar nicht auf dem Markt haben, und im Grunde genommen wollen wir auch nicht, dass sich die Menschen subliminal manipulieren lassen.

Das hat natürlich einen gewissen Freiheitsverlust für den, der sich subliminal – das kann man bestimmt irgendwie positiv für die Persönlichkeitsentwicklung einsetzen, mag sein. Aber den Freiheitsverlust nehmen wir hin, um uns solche Systeme vom Leib zu halten. Denn das könnten wir ja ähnlich tun: Emotionserkennung, wie gesagt, im Universitätsbereich oder im Arbeitsbereich. Gut, da kann man sagen: Im Arbeitsfeld kann man nie wirklich einwilligen, aber das ist ja auch nur eine theoretische Konstruktion.

Wir können also einfach sagen, wir machen ein Gedankenerkennungsverbot unter weiteren Umständen, und dann müssen wir halt hinnehmen, dass damit Freiheitsverluste einhergehen. Die sind dann durch den Wert der Gedankenfreiheit zu rechtfertigen, würde ich sagen. Ich weiß, dass diese Antwort der philosophischen Debatte um Paternalismus nicht gerecht wird.

### **Hans-Georg Dederer**

Vielen Dank. Wir haben jetzt drei weitere Fragende.

**Jutta Allmendinger**

Ich habe keine Frage mehr, weil wir jetzt schon so weit sind, dass man Gedanken lesen könnte. Also meine Fragen sind gestellt.

Ich hätte eine Nachfrage zu Ihrem [deutet zu Cornelia Betsch] Eisberg. Weiß man überhaupt, was das Militär alles weiß und macht?

**Marcello Ienca**

Nicht komplett, nicht ausführlich. Wie ich vorher gesagt habe: Viele Informationen sind klassifiziert. Wir wissen, was öffentlich gemacht wird.

DARPA war in den letzten Jahren relativ transparent. Wir wissen, woher militärische Investitionen in den neurotechnologischen Bereich kommen und wie hoch das ist. Man kann dasselbe nicht für andere Länder sagen. Wir wissen, dass die Rote Armee in China 2023 eine Einheit gegründet hat, die Cognitive Warfare heißt. Das fokussiert hauptsächlich auf Informationskampf, durch digitale Plattformen. Aber es gibt Berichte, dass auch neurotechnologische Anwendungen Teil dieses Programms sind. Das Recht kann man nur spekulativ analysieren, aber was ich mitteilen kann, ist, dass das UN Institute for Disarmament Research seit fünf Jahren zu diesem Thema arbeitet, und sie haben viele Anhörungen zu diesem Thema gemacht.

Eine Möglichkeit, die auch diskutiert wird, ist, dass neuroelektronische Anwendungen Teil der Genfer Konvention werden sollten. Es gibt Leute in diesem Bereich, die diese These vertreten haben.

**Philipp Kellmeyer**

Eine kurze Ergänzung zum Diskurs Neurotechnologie im Kontext militärischer Forschung, militärischer Auseinandersetzung: Hier gibt es nicht nur die klassischen Verfahren zum Auslesen und Stimulieren – also das, was an den eigenen Soldaten vielleicht erforscht und ausprobiert werden soll –

, sondern auch das ganze Spektrum offensiver, nicht tödlicher Waffen, die gezielt auf das Nervensystem ausgerichtet sind, die auch stark auf Neurotechnologie basieren: Sonic Weapons, andere Verfahren, um gezielt das Nervensystem anzugreifen, Gleichgewichtsstörungen zu erzeugen, andere Dinge. Auch das erfolgt durch neurotechnologische Geräte. Das spielt im zivilen Kontext keine Rolle, kommt aber im militärischen Kontext mit dazu: dass es nicht nur um Restoration von Soldatinnen und Soldaten geht oder um die Verbesserung kognitiver Fähigkeiten, sondern auch um die aktive Schädigung von anderen.

Ich bin kein Experte im internationalen Recht, aber da ist nicht alles von den entsprechenden Konventionen klar abgedeckt. Dieser Offensivbereich der Neurotechnologie im militärischen Kontext verdient auf jeden Fall eine Würdigung oder eine kritische Frage.

**Hans-Georg Dederer**

Vielen Dank, Herr Kellmeyer, für die Ergänzung. Dann würde ich Armin Nassehi aufrufen.

**Armin Nassehi**

Herr Bublitz, Sie haben auf Gedankenfreiheit abgestellt und, das fand ich ganz toll, am Ende auf 1848: „Die Gedanken sind frei“ – „wer kann sie erraten?“, geht das übrigens weiter – hingewiesen und dann mit einem kleinen Hinweis noch die Phänomenologie aufgerufen.

Wenn ich so daran denke – ich bin jetzt ein alter Mann und habe lange, lange gelehrt. Ich habe meinen Studis als Soziologe immer beigebracht: Kommunikation basiert darauf, dass wir wechselseitig intransparent füreinander sind. Das erzählen uns die Anthropologen, das erzählen uns die Phänomenologen, das erzählt uns der Nietzsche-Pragmatismus, das erzählt uns die systemtheoretische Soziologie. Das hat uns Ende der Neunzigerjahre Jürgen Habermas beigebracht, als es um

die Frage ging: Was ändert sich eigentlich im Personenstatus, wenn wir uns gegenseitig klonen können? Sie erinnern sich vielleicht an die Diskussion. Argumentiert wurde eigentlich immer: Sobald diese Intransparenz infrage gestellt wird, geht es nicht mehr nur um Gedankenfreiheit, sondern um die anthropologische Bedingung überhaupt. Dass wir hier miteinander reden können, liegt ja daran, dass wir nicht voneinander wissen, was wir übereinander denken. Sonst würden wir mit den meisten wahrscheinlich gar nicht reden. [Lachen]

Ich meine das nicht nur witzig, sondern das ist eine anthropologische Grundbedingung. Das geht noch darüber hinaus. Da kann man jetzt viel weiterdenken, auch in philosophische und bewusstseinsphilosophische Fragen rein, das will ich gar nicht tun.

Die spannende Frage ist: Hätte dieser Gedanke nicht auch Konsequenzen für die juristische Frage, aber auch für die Frage, wie wir mit diesen Dingen eigentlich umgehen? Denn bis jetzt haben wir immer nur punktuelle Fragen des Auslesens. Ja, dafür muss man eine Menge tun, aber wenn die Dinge weiterlaufen, dann ändert das tatsächlich die *Conditio humana*. Wie würde man sich dazu stellen? Was würden Sie dazu sagen?

### **Jan Christoph Bublitz**

Das ist ein sehr interessanter Aspekt. Ich würde sagen, erst mal stärkt das den Case vor Gedankenfreiheit. Das wäre auf rechtsphilosophischer Ebene ein sehr gutes Argument dafür, und das kann natürlich auch indirekt in die Auslegung fließen. Wir Juristen formen im Grunde genommen – und das ist schon sehr viel – unbestimmte Begriffe nach dem, was wir für richtig halten. Da muss man ein paar Grenzen einhalten. Und wenn es nun mal Gedankenfreiheit heißt – auch wenn es genauso wichtig wäre, Emotionsfreiheit oder

so zu schützen –, gibt es ja gewisse Grenzen darin, wie wir das Recht formen können.

Solche Argumente könnte man nutzen, um zu einem besonders breiten Verständnis zu kommen, was einige Institutionen auch tun. Aber ob das dann juristisch noch plausibel ist, würde ich in Frage stellen.

### **Armin Nassehi**

Ich habe mir immer vorgestellt, wenn ich mal Zeit dafür haben sollte – was ich nie haben werde – und etwas Literarisches schreiben würde, dann würde ich mir soziale Situationen vornehmen und versuchen, zu beschreiben, was passieren würde, wenn diese Intransparenz nicht da wäre. Ich habe mir immer so gedacht: Wir sind ja kommunikativ alle total neurotisch, weil wir das meiste, was wir sagen wollen, nicht sagen. Wir dürften dann das meiste, was wir denken wollen, nicht denken. Das muss man sich mal vorstellen, was das bedeuten würde, und wenn man das nicht nur als Bonmot sieht, sondern als eine Situation, in der wir Abwehrrechte, aber auch Abwehrpraktiken bräuchten gegen Technologien oder was auch immer, was da eine Rolle spielt, um uns vor denen zu schützen. Kommunikativ können wir das machen: Wir können lügen, wir können so tun, als wären wir freundliche Menschen und so. Aber das wäre dann nicht mehr möglich. Das wäre der Extremfall. Sorry, ich fand das gerade sehr spannend.

### **Hans-Georg Dederer**

Ja, vielen Dank. Der Dialog hat mir auch gerade Gelegenheit gegeben, kurz übers Kriegsvölkerrecht nachzudenken [Lachen] im Anschluss an Herrn Kellmeyers Bemerkung, weil es natürlich schwierig ist, die Genfer Konventionen aufzuschneiden, einen völkerrechtlichen Vertrag oder die beiden Zusatzprotokolle zu den Genfer Konventionen neu zu verhandeln.

Aber es gibt auch ganz allgemeine Auffangregeln, etwa die: Es gibt kein unbeschränktes Recht zur Schädigung des Feindes und das Verbot unnötigen Leidens, wenn es um Angriffe gegen Kombattanten geht. Und natürlich auch das Unterscheidungsgebot und Verhältnismäßigkeit im Kollateralschaden. Wir haben also zumindest Auffangregeln.

Ich habe mich selbst auf die Liste gesetzt. Danach kommt Aldo Faisal. Das greift auch diese Frage der Neurorechte und der Gedankenfreiheit auf, betrifft zunächst Herrn Ienca und Herrn Bublitz, vielleicht auch Herrn Ienca, was die Neurorechte angeht.

Wenn ich das richtig verstanden habe, war eigentlich im Ursprung daran gedacht, wirklich neue Rechte einzuführen. Aber jetzt habe ich den Eindruck, wenn ich die Seite der Neurorights Foundation lese, dass es doch eher so gedacht ist, dass man das, was als Neurorechte firmiert, eigentlich aus dem geltenden Rechtsbestand ableiten möchte. Eine wichtige Rolle dabei spielt natürlich die Gedankenfreiheit, die schon da sei.

Aber noch mal: Wenn man jetzt ins positive Recht schaut, Herr Bublitz, ist die Gedankenfreiheit immer gekoppelt mit der Glaubens- und Gewissensfreiheit und nie ein eigenständiges Recht. Und auch im Deutschen, gibt es, haben Sie richtig gesagt, die Gedankenfreiheit nicht explizit, aber es gibt das *forum internum*. Auch das denkt man primär bei der Glaubens- und Gewissensfreiheit, sonst bei der Meinungsfreiheit. Also die Frage: Ist es wirklich ein eigenständiges Recht, das wir haben? Oder müsste man nicht doch sagen – und Sie kennen ja das Bundesverfassungsgericht, das hat ja in technologischen Fortentwicklungen eine eigene Rechtsposition, eigene Grundrechte entwickelt.

Dann zur Frage der Gedankenfreiheit. Wenn man sagt, dass sie ein eigenes Recht mit absolutem

Schutz ist, dann heißt das für mich, dass man den Schutzbereich auch eng definieren würde. Dann könnte nicht jeder x-beliebige Gedanke – eine mathematische Operation oder was ich gerade so vor mir habe – darunterfallen können.

Und eine Frage an Herrn Kellmeyer und Herrn Bublitz: Dieses Impact Assessment – Mental Impact Assessment oder auch Human Rights Impact Assessment – taucht immer wieder auch in Ihren Publikationen auf. Dahinter muss natürlich ein Konzept stehen. Was wird hier abgeschätzt? Gibt es da schon eine Methodologie für diese Abschätzung, wie man es aus der Umweltverträglichkeitsprüfung kennt? Die ist ja letztlich auch ein Impact Assessment.

Dann wäre die Frage: Wo ist eigentlich der regulatorische Ort für dieses Mental Impact Assessment? Das muss ja irgendwo im Recht, in der Regulatorik aufgehängt sein, also im Produktsicherheitsrecht oder im Medizinprodukterecht. Wo soll das rein?

Juristisch gedacht wäre das für mich die Frage: Wo kommt das Impact Assessment eigentlich rein? Beim Human Rights Impact Assessment könnte ich mir vorstellen, dass es da bereits Vorlagen gibt, nämlich aus dem Kontext der Lieferkettenproblematik, wo es auch darum geht, dass unsere transnationalen Unternehmen den menschenrechtlichen Impact ihres transnationalen Handelns mitbedenken, wie es mit Menschenrechten in den ausländischen Teilen ihrer Lieferkette bestellt ist.

### **Marcello Ienca**

Ich kann mit der ersten Frage zum Thema Neurorechte anfangen. Als ich und Roberto Andorno das Konzept Neurorechte eingeführt haben, haben wir in dem ursprünglichen Paper von vornherein klargemacht, dass wir normativ agnostisch blei-

ben. Das heißt, für uns sind Neurorechte die Teilmenge von Menschenrechten, die mit dem Schutz von Geist und Gehirn des Menschen zu tun haben. Das ist ein Nebenprodukt der technologischen Evolution, ähnlich wie wir genetische Rechte in technologischen Entwicklungen im Bereich Geneditierung und Sequenzierung haben.

Wichtig für uns ist nicht so sehr die rechtliche Einordnung, sondern dass wir einen angemessenen rechtlichen Schutz haben. Ich glaube, das ist auch das, wo ich und Dr. Bublitz einverstanden sind. Wir haben, glaube ich, ein ganzes Jahrzehnt darüber debattiert, ob diese Rechte als neue Rechte zu verstehen sind oder als Auslegungsdimensionen existierender Rechte. Ich habe sogar Positionen gelesen, die sagen, Neurorechte sind Rechte des Gehirns, was ich gar nicht verstehen kann. Die Einheit des Rechts ist der Mensch. Aber ich glaube, all diese unterschiedlichen Positionen stimmen darin überein, dass wir einen angemessenen rechtlichen Schutz brauchen.

Ich glaube, diese Vorschläge gehen in die richtige Richtung. Ich möchte kurz betonen, dass wir in den letzten Jahren viele Soft-Law-Instrumente entwickelt haben, die versuchen, diesen Bereich zu regulieren. Die UNESCO-Ethikempfehlung ist die letzte in der Reihe. Das wurde erst im November 2025, also vor vier Monaten genehmigt. Aber schon 2019 hatte die OECD eine Empfehlung zu verantwortungsvollen Innovationen im neurotechnologischen Bereich, und diese Instrumente sind aus meiner Sicht eine gute Guidance in diesem Bereich.

### **Jan Christoph Bublitz**

Herr Dederer, Sie stellen genau die richtigen Fragen in der internationalen Debatte, und das gibt mir Gelegenheit, dieses Missverständnis aufzuräumen, dass die Gedanken-, Glaubens- und Gewissensfreiheit ein Recht ist. Das liest man so immer wieder, gerade in Deutschland, das ja nur die

Gewissens- und die Glaubensfreiheit hat. Das war international aber ausdrücklich nicht so gedacht. Die Gedankenfreiheit, 1948 eingeführt, greift auf die Tradition der Gewissens- und Glaubensfreiheit zurück. Die gibt es schon viel länger. Sie vollzieht sozusagen die Wende der Aufklärung in der menschenrechtlichen Sache, indem sie sagt: Vor dem Glauben und Gewissen steht noch das Denken. Und das ist sogar als das allgemeine Recht gedacht worden. Deswegen ist es an den Anfang gesetzt worden, und die Gewissens- und die Glaubensfreiheit sind als *derivatives* verstanden worden. Das allgemeine Recht ist eigentlich die Gedankenfreiheit in der Konzeption der Menschenrechtserklärung. Das wurde auch weithin so geteilt, obwohl das Recht nur eine Nebenrolle gespielt hat, muss man dazu sagen. Aber in den Dokumenten ist es klar, dass es darüber hinausgeht. Das sagen jedenfalls die zentralen Figuren in den Debatten, und es wird auch immer als absolut dargestellt. Das kann man rechtspositivistisch, wenn man das in dem Fall so sagen will, einfach mal so hinnehmen.

Sie haben natürlich recht, das hat Auswirkungen auf den Schutzzumfang. Da müsste man dann sprechen, wie man den angemessen begrenzt. Das Problem sehe ich ein. Der muss angemessen begrenzt werden.

Im Hintergrund ist die geistesgeschichtliche Idee, dass das Innere nicht in soziale Konflikte mit dem anderen gerät und das Recht nur dann legitim sein kann, wenn es einen externen Konflikt regelt. Das wäre so eine kantische Idee dahinter. Aber das führt jetzt zu weit.

Wo soll das Mental Impact Assessment hin? Ich würde das erst mal regulatorisch offenlassen. Wir haben das nicht entwickelt. Wir haben nur versucht zu markieren, was eigentlich das Problem ist und dass das ein Instrument sein könnte, um

viele Probleme zu lösen. Das muss man interdisziplinär entwickeln, und das ist total schwierig. Das können wir natürlich auch nicht.

Wir haben gerade einen Antrag gestellt dafür, der es uns ermöglichen würde, ein bisschen genauer reinzuschauen. Aber das ist eine Riesenaufgabe, die vor allem von der Industrie angegangen werden muss. Erst mal sind *die* dafür zuständig, dafür zu sorgen, dass die Produkte, die sie auf den Markt bringen, sicher sind.

Wenn man denen aber eine Testbatterie vorlegt und sagt: „Ihr müsst zumindest *das* aber prüfen“, würde man die ein bisschen in Zugzwang bringen. Wir wissen, dass die Firmen das machen, und dann verschwindet es in der Schublade, wenn die Ergebnisse nicht richtig sind. Das müsste man also ein bisschen strenger angehen, und dann könnte man das in alle Regelungsinstrumente einbauen, in die Medizinprodukteverordnung so wieso. Aber warum nicht auch in die allgemeine Risikoanalyse, die jeder durchzuführen hat, der etwas in den Verkehr bringt? Es reicht ja, wenn man das über Soft Law reinschreibt, dass das sozusagen der maßgebliche Stand der Technik heißt: Ihr überprüft das auf diese zehn Faktoren oder so etwas. So könnte man das einbauen.

Wenn wir das jetzt von vornherein zu regulatorisch auf irgendwas Bestimmtes festschreiben, dann wird das nie entwickelt, weil das dann nur für ein bestimmtes Verfahren gilt und sowieso nur in Europa usw. Wir würden versuchen, das ein bisschen allgemeiner denken zu wollen.

### **Philipp Kellmeyer**

Nur kurz als Ergänzung. Das würde ich genauso sagen, dass das ein Desiderat ist, was sinnvoll sein kann und woran man gemeinsam arbeiten sollte. Aber es ist auch diese Verschiebung weg von einer rein datenzentrierten hin zu einer mehr wirkungszentrierten Regulierung, denn wir haben ja

unter anderem das Problem, dass wir durch die rasante Innovationsdynamik und die technologische Konvergenz aus Neurotechnologie, KI und das komplexe Regulierungsgeflecht, das wir haben, nicht mehr hinterherkommen. Das heißt, wir brauchen effektive Regime für Post-Market Surveillance, vielleicht auch adaptive, smarte, Regulierungsregime, die wissenschaftsbasiert, datengetrieben sind. Das fände ich sehr wichtig.

Und wenn man so ein Assessment hat und das operationalisieren kann, so wie Herr Bublitz das gesagt hat, dann hat man eine bessere Systematik in der Bewertung, was die Eingriffe in bestimmte mentale Prozesse angeht. Insofern würde ich sagen, das hat regulatorisch keinen bestimmten Ort, sondern ist vielleicht eine Reform der Governance-Logik an der Schnittstelle bestehender Regime, sei es besonderer Datenschutz, Medizinproduktrecht und risikobasierte Ansätze, wie wir sie aus dem AI Act kennen.

### **Hans-Georg Dederer**

Herzlichen Dank. Ich habe drei Wortmeldungen notiert und möchte darauf hinweisen, dass jetzt Fragen an alle unsere Expertinnen und Experten gerichtet werden können. Sie müssen sich nicht auf die Experten des zweiten Blocks beschränken.

### **Aldo Faisal**

Ich möchte noch mal auf den „Gedanken sind frei“-Aspekt zurückkommen. Wenn man sich die Literatur in der Privacy-Forschung anschaut, wissen wir: Effektiv gibt es de facto keine Anonymisierung von Gesundheitsdaten, die nicht dekonstruiert werden kann, wenn man das möchte.

Genauso reden wir von Gedankenfreiheit. Da ist das Element der mentalen Privatheit Voraussetzung oder gegeben.

Wenn man in die Literatur geht, und auch Teile unserer eigenen Forschung zeigen, wie vorhersehbar Menschen sind, wenn man deren Daten

hat, um Verhalten auf kleiner Ebene und auf makroskopischer Ebene 95 Prozent oder mehr der Zeit vorherzusagen.

Das heißt, angesichts dieser Praktikabilitäten und der Menge an Daten, die wir über Datenlecks ständig in unsere Umgebung hinein- und hinausgeben, müsste man davon ausgehen, dass man vielleicht noch über Willensfreiheit reden kann, dass aber Gedankenfreiheit mit diesem Element der Privatheit de facto technisch unter Umständen nicht mehr existiert.

### **Jan Christoph Bublitz**

Es ist interessant, dass Sie die Schlussfolgerung ziehen, weil ja der Begriff der Willensfreiheit eigentlich viel anspruchsvoller ist. Da klingt ja noch eine metaphysische Freiheit mit. Bei der Gedankenfreiheit ist das nicht der Fall. Wir können vielleicht gedankenfrei sein, auch wenn wir vollständig determiniert sind usw. Insoweit würde ich sagen, wir sollten dann, wenn die Welt so ist, wie Sie sie beschreiben, vielleicht den Begriff der Willensfreiheit aufgeben, aber die Gedankenfreiheit behalten, weil sie dieses Element nicht trägt. Aber Sie haben natürlich recht: Die Daten und immer mehr wird vorhersehbar.

Aber was soll daraus folgen? Erst mal würde ich schon sagen, dass das, was Herr Nassehi eben diese wechselseitige Intransparenz genannt hat, noch da ist. Auch diese Verhaltensvorhersagen bleiben vor dem Gedanken irgendwie stehen. Ganz durchleuchtbar ist das alles noch nicht. Und Verhaltensvorhersage ist nicht unbedingt eine Mental-State-Vorhersage, oder wenn, dann nur eine sehr vage. Aber der Gedankeninhalt als solcher entzieht sich noch diesen Vorhersagen. Und solange das so ist, können wir versuchen, das zu verteidigen. Vielleicht müssen wir eines Tages aufgeben, und dann ist das so. Aber bis wir da sind, können wir ja noch versuchen, Grenzen zu ziehen.

### **Philipp Kellmeyer**

Eine kurze Ergänzung. Ich denke, dass viel der Fantasie, die jetzt im Hinblick auf Dekodierung mentaler Zustände in den letzten Jahren gekommen ist, aus der Richtung Dekodierung sprachbezogener Hirnsignale kommt. Denn das ist ein Bereich mit den implementierbaren Systemen, die über Regionen des Gehirns angebracht sind, wo sprachbezogene Aktivität stattfindet, dass man dort bis auf die Ebene einzelner Phoneme auch sprachbezogene vorgestellte innere Sprache, motorische Aspekte von Sprache mittlerweile sehr gut dekodieren kann.

Wir haben auf der konzeptionellen Ebene oft noch ein logozentrisches Verständnis von Gedanken, dass wir sagen: Gedanken sind Dinge, die wir artikulieren können, die ausgesprochen werden können, die anderen mitgeteilt werden können. Aber die Welt der Gedanken ist ja viel weiter. Sie umfasst auch affektive Zustände, amorphe Gedanken, denken Sie an Traumzustände, denken Sie an Bilder, an Fantasien, an Vorstellungen. Das rein Propositionale, Sprachbezogene ist unter bestimmten Umständen besser dekodierbar als je zuvor. Trotzdem ist dieser weitere Bereich der Gedanken, wie Christoph Bublitz eben auch geschildert hat, eine andere Hausnummer.

Deswegen würde ich sagen: Es gibt immer noch einen privilegierten Raum des nicht-propositionalen Denkens, was auch für uns selbst oft schwer fassbar ist in der Introspektion, was vielleicht nur durch phänomenologische Methoden erfassbar und beschreibbar ist. Wir sind weit davon entfernt, die genau auslesen zu können. Das ist etwas anderes als die Vorhersage von Verhalten.

Ich glaube, die beste Technologie zur Vorhersage des Verhaltens von jemand anderem ist immer noch die Ehe, weil da die Neurotechnologie noch nicht so weit ist. [Lachen]

**Helmut Frister**

Kurz zur Frage mit dem absoluten Schutz der Gedankenfreiheit und damit korrespondierend, dass man dann die Gedankenfreiheit eng definieren müsste. Herr Bublitz, Sie haben dem zugestimmt. Ich würde dem auch zustimmen, aber Herr Dederer hat dann so eine inhaltliche Definition gebracht, also bestimmte Gedanken, und andere Gedanken nicht. Da wäre ich anderer Meinung. Wenn ich jetzt während der Debatte in meinem Kopf ein mathematisches Rätsel löse, anstatt der Debatte zu folgen, möchte ich das auch nicht ausgelesen haben. [Lachen] Also da wäre die Frage: Kann man das wirklich so inhaltlich machen?

Dann hätte ich noch einen Kommentar. Ich fand das, was, Armin Nassehi gesagt hat, sehr wichtig, und ich glaube, daraus könnte man ein juristisches und auch ethisches Argument für die Unverfügbarkeit des Ganzen machen.

**Hans-Georg Dederer**

Armin Nassehi, wolltest du ergänzen?

**Armin Nassehi**

Genau, und ich würde es relativieren. Es ist natürlich toll, dass man das daran anschließt, aber ich würde es relativieren, denn gerade war vom Verhalten die Rede. Wir sind vom Gedanken zum Verhalten gekommen. Das ist ein interessanter Sprung, und wir würden ja vom Verhalten eigentlich immer auf Gedanken schließen. Das ist natürlich eine Zurechnungsfrage und nichts, was wir auslesen.

Aber was doch demütigend ist, ist, dass wir mit elektronischen Mitteln – das ist nicht KI, sondern das sind klassische statistische Methoden mit großen Datenmengen, mit Computerunterstützung – unglaublich genau vorhersagen können, nicht welche Individuen, aber welche Aggregate oder Gruppen sich wie verhalten. Ich finde es immer noch irre, dass man nach der Befragung von 2000

Leuten die Bundestagswahl vergleichsweise genau vorhersagen kann. Das ist schon fast Voodoo. Aber es ist eben kein Voodoo, weil das Aggregat relativ formiert ist.

Das müsste im engeren Sinne auch für die Gedanken gelten. Da ist der Punkt, dass da ethisch schon eine ganze Menge dran ist und dass man nicht naiv hingehen und sagen sollte: „Eigentlich haben wir die Idee einer völligen Freiheit in unserem Innenleben.“ Das müssen wir wahrscheinlich normativ hochhalten, aber empirisch gibt es das nicht. Wir würden wahrscheinlich von mentalen Stilen oder von Formen reden, die sehr erwartbar sind.

Ich weiß nicht, ob es in der Psychologie – da kenne ich mich nicht aus – Forschung darüber gibt, wer welche Assoziationen in welchen Kontexten hat, aber mich würde wundern, wenn das nicht sehr musterhaft wäre. Vorurteile, Stereotype, all diese Dinge gehören wahrscheinlich dazu.

**Hans-Georg Dederer**

Herr Bublitz, wollen Sie noch kurz?

**Jan Christoph Bublitz**

Kurz zum absoluten Schutz: Ich gebe Ihnen recht. Ich würde das auch nicht nach der Qualität oder Wichtigkeit der Gedanken unterscheiden. Das sagen einige in der Literatur, aber ich würde sagen, es muss so etwas wie der Denkraum sein. Das Denken vollzieht sich sprunghaft und fragmentarisch und wechselt die ganze Zeit von tiefschürfend bis zu völlig banal. Das nun noch zu unterscheiden, funktioniert nicht. Der Menschen darf nie das Gefühl haben, dass das irgendwie mitgelesen oder mitgeschnitten wird, weil sonst auch das eigene Denken zensiert wird. Deswegen würde ich sagen, es muss der Denkraum sein.

Aber man muss es natürlich trotzdem *narrow*, eng halten, und das muss man über das Element der

Eingriffe machen. Das können nur technologische Eingriffe sein, und da muss man aufpassen, dass da nicht zu viel drunter fällt.

### **Hans-Georg Dederer**

Ja, das ist ein juristischer Trick: den Schutzbereich offenlassen und das dann über den Eingriff definieren. Jetzt haben wir noch die Wortmeldung von Kerstin Schlögl-Flierl, bitte.

### **Kerstin Schlögl-Flierl**

Vielen Dank. Neurotechnologien sind ja nicht neu, und Sie sind eine Community, die schon sehr lange zusammenarbeitet. Was ist Ihr Learning daraus, dass sich durch KI etwas verändert? Können Sie aus der schon längeren Beschäftigung mit Neurotechnologie lernen durch die neue Situation, die KI-getriggert ist?

Ich würde gern Ihre Selbstreflexion auf die schon längere Beschäftigung lenken, damit wir etwas mitnehmen können. Dass Sie sagen: „Das Neuro-Enhancement war gut, aber das hätten wir verhindern können.“ Da geht es ja viel um die unrealistischen Erwartungen, die du, Marcello [Ienca], am Schluss angesprochen hast. Ich würde also gern etwas mitnehmen. An alle.

### **Hans-Georg Dederer**

Dann haben wir eine Frage an alle. Jetzt Frau Nagel, und dann würde ich allen Referentinnen und Referenten die Gelegenheit geben, sofern sie wollen, noch mal dazu Stellung zu nehmen.

### **Saskia Nagel**

Etwas ist mir auf dem Weg hierher und in den Gesprächen in der Kaffeepause aufgefallen: Wir arbeiten teilweise schon seit 15, 20 Jahren an diesen Themen. Vor 20 Jahren war das eher Science Fiction: Wir reden darüber, aber noch geht das nicht.

Heute sehen die Dinge anders aus. Die Realität zeigt: Es kommt viel mehr. Wie viel dann wirklich da ist und was immer noch Hype ist, muss

man filtern. Ich bin ein großer Fan davon, immer abzuschichten: Was ist Hype und was kommt? Das finde ich generell wichtig, auch als Technikethikerin.

Aber dieser Fall zeigt, dass mehr gekommen ist, als man vor 15, 20 Jahren gedacht hat, weil wir alle da KI noch nicht so richtig mitgedacht hatten. Und jetzt denkt jeder mit und alle sehen: Das ist jetzt das Thema, das diesen Technologiebereich komplett verändert. Das ist ein Learning.

### **Bertolt Meyer**

Ein Learning aus meiner Perspektive ist, dass Menschen bereit sind, eine Technologie, deren Funktionsweise sie nicht verstehen und der gegenüber sie vielleicht sogar große ethische Vorbehalte haben, einfach zu nutzen, sobald sie das Gefühl haben, dass ihnen dadurch ein unmittelbarer starker Nutzen entsteht. Und in dem Moment werden jegliche Bedenken über Bord geworfen, bis hin dazu, dass – ich sehe es in meinem Umfeld – stets und ständig eigentlich Datenschutz verletzt und gebrochen wird, weil Leute hochsensible berufsbezogene Dokumente an amerikanische Techkonzerne mit Servern außerhalb der EU übergeben, weil sie wissen: ChatGPT schreibt mir das Gutachten in einer halben Stunde, für das ich ansonsten einen ganzen Tag brauchen würde. Der Nutzen ist dann so groß, dass sämtliche Bedenken über Bord gehen, und das in einem Umfang, den ich nicht gedacht hätte. Ich glaube, das blüht uns auch bei anderen Technologien, die die Kompetenz der Menschen massiv erhöhen, solange sie den Eindruck haben, dass das so ist.

### **Hans-Georg Dederer**

Vielen Dank. Adressiert waren auch die Referenten des zweiten Blocks. Ich möchte aber auch Frau Singh und Frau Bernés Timón die Gelegenheit geben, noch auf den Aspekt von Frau Schlögl-Flierl einzugehen.

**Laura Bernáez Timón**

Thank you. I would like to add something else to the field of AI. I definitely agree that AI is playing a big role in how neurotechnologies are developing. But as we were speaking, I was also writing down some other fields that I think have significantly contributed to the advances of neurotechnology, both in the consumer and in the medical market. And I was writing about or thinking about Bluetooth technology, for instance, low latency to communicate between devices, biocompatible materials so that implants can stay there for years. The cloud infrastructure as well. The fact that all of us have smartphones nowadays, dry electrodes as well, robotic surgery to help in the implantations, and also batteries. So I think AI is there and it is really helping, but it could not be possible for a consumer to use a neurotech device if you have to use gel every time that you want to use it. Or, for example, maybe for a patient to have an implant there for years without the necessary materials or duration of the battery or the Bluetooth systems. I was thinking that beyond AI, neurotech is highly interdisciplinary and it would be great to include all these different perspectives into the conversation, because many different fields are actually impacting the development of neurotech.

**Hina Singh**

I have, I suppose, maybe two areas that I think we really should start thinking about now and I again very much appreciate Saskia Nagel's intervention there. You know, six months ago, when we started sending our human–AI-relationship papers to journals, we had a short paragraph at the end saying, of course, we are assuming the AI is not conscious. And that was just taken as read. Okay, all of your analyses assume the AI is not conscious. And now, suddenly, we are starting to get much more editorial pushback saying, well, what if the AI is conscious? Could you please elaborate on

that? And that over the space of six months, but already a big shift in thinking. And I do think that a conscious chatbot, for example, will fundamentally change the way we think about relational obligations. In part because we do not know what it is that we have a relationship to. And so that, to me, at the moment is a really open question, and I know that there are a lot of philosophers who have been thinking about this very productively. But one of the areas that has really struck me today is the difference between our desire for some normative ethics and the empirical realities. Because I think people are already treating these entities as conscious, in whatever their understanding of consciousness is. I think we also need to think about our risks and protections in that way. And the second is really related. It is a half a point, and we have written a bit about hybridisation and collective minds and swarm minds and hive minds, all of which are conceptually distinct. But I think, again, you know, if we begin to think about these entities as having consciousness and mindedness that is apart from some human actor, then I think that kind of thinking about collective mindedness, particularly around autonomy and accountability, will be really important for us to do.

**Hans-Georg Dederer**

Herzlichen Dank. Jetzt möchte ich den drei Referenten des zweiten Blocks noch Gelegenheit geben, auf die Frage von Frau Schlögl-Flierl einzugehen. Wer möchte starten?

**Marcello Ienca**

Auf der technischen Ebene hat Frau Bernáez Timón diese Frage schon sehr gut beantwortet. Ich würde die Frage „Was hat sich in den letzten 20 Jahren geändert?“ auch auf einer soziologischen Ebene beantworten. Also nicht nur technisch, KI, Bluetooth usw., sondern die soziale und die sozioökonomische Infrastruktur von Neurotechnologie haben sich stark geändert. Vor 20

Jahren hatten wir nur wenige kleine Start-ups in diesem Bereich. Heute haben wir Big-Tech-Konzerne wie Google, IBM und Apple, die hier aktiv sind. Das ist wirklich ein Gamechanger.

Der reichste Mensch auf der Welt besitzt eine Neurotech-Firma und hatte mittlerweile auch eine Regierungsposition, wodurch er mit der Food and Drug Administration gekämpft hat, die ein Mandat zur Regulation dieser Technologien hatte. Das heißt, wir haben heute eine ganz andere Infrastruktur mit großen Problemen des Interessenkonflikts.

Ich glaube, wir müssen nicht nur die technischen Aspekte in Betracht ziehen, sondern auch die soziokulturellen Aspekte. Wir müssen uns auch die Frage stellen: Wie wird diese Welt in 20 Jahren aussehen? Retrospektiv wissen wir es, aber diese Veränderungen sind rasant. Wie sieht die Welt 2046 aus? Das ist eine sehr wichtige Frage für so ein Gremium.

### **Philipp Kellmeyer**

Die Frage betraf ja spezifisch den Impact von KI, und das ist fundamental und weitreichend. Das ist eine tiefgreifende epistemische Revolution, die die Neurowissenschaften und damit auch die Neurotechnologie erfasst hat. Wenn man nur mal das Beispiel von Deep Learning nimmt: Hier haben wir eine epistemische Zugriffsweise auf neurobiologische Daten, die sich fundamental von der hypothesengetriebenen klassischen Neurowissenschaft unterscheidet, wie ich sie noch gelernt habe, oder auch von linearen Klassifikationsverfahren, Machine-Learning-Verfahren, die wir früher hatten.

Heute haben wir Modelle, die in absolut hochdimensionalen multimodalen Datenräumen latente Strukturen identifizieren können, die sich eigentlich einer direkten Beobachtung und expliziten

Modellierung für uns entziehen, das heißt möglicherweise auch gar nicht auf unser eigenes konzeptionelles und Verständnissystem zu mappen sind. Diese Architekturen lernen nicht nur Korrelationen zwischen den Daten, sondern komplexe, nichtlineare Repräsentationen neuronaler Aktivität über verschiedene Skalen hinweg, von lokalen Signalen bis hin zu großskaligen Netzwerken. Das ist ein fundamentaler Paradigmenwechsel innerhalb der Neurowissenschaft.

Das erzeugt epistemische Spannungen. Man könnte sagen: So wie die technologische Konvergenz zwischen KI und Neurowissenschaft die ganze KI jetzt in die Neurowissenschaft importiert, importieren wir auch die KI-ethischen Probleme, weil nämlich die generierten Repräsentationen möglicherweise hochprädiktiv sind für bestimmte Krankheitszustände oder Verhaltensdispositionen oder andere Dinge, aber nicht unbedingt direkt interpretierbar, geschweige denn kausal erklärbar. Es ist nicht nur einfach eine Methode, die wir jetzt in der Neurowissenschaft anwenden, sondern es verändert das Selbstverständnis dieses ganzen Feldes tiefgreifend.

Insofern sind die Folgen auch in der Arbeit deutlich spürbar. Es gibt kaum jemand, der in dem Feld arbeitet, der nicht diese Methoden benutzt und dann aber unter bestimmten Umständen in genau diese Interpretierbarkeits- und Erklärbarkeitsprobleme reinläuft, mit den ganzen ethischen Folgen, die daran hängen. Man kann in dem Sinne Neuroethik und KI-Ethik in diesem Bereich gar nicht mehr so stark voneinander getrennt sehen. Insofern ist der Impact groß und relevant.

### **Jan Christoph Bublitz**

Was wir auch im KI-Bereich sehen, ist die Unregulierbarkeit geopolitischer Hintergründe. Vielleicht ist das der kurze Vorteil, den wir jetzt noch im Neurotech-Bereich haben. *Noch* reden wir ein bisschen über Science-Fiction-Sachen. Sobald die

Geräte einmal da sind, glaube ich, wird man das nur noch sehr schwer regulieren können. Derzeit wäre vielleicht bei den Big Playern – die USA vielleicht mal weg, aber vielleicht in China – noch die Bereitschaft da, sich auf vernünftige Regeln einzulassen. Die UNESCO-Sachen waren schon der richtige Schritt zur richtigen Zeit und haben ja geklappt mit China und Russland, vielleicht gerade deswegen, weil es noch nicht ganz da ist.

Vielleicht sollte man in den nächsten fünf Jahren weitere Schritte folgen lassen, bevor das endgültig in die Nichtregulierbarkeit entgleitet.

### **Hans-Georg Dederer**

Ganz herzlichen Dank für Ihr Schlusswort, Herr Bublitz. Damit möchte ich mich auch noch mal ganz herzlich bei allen bedanken, insbesondere bei den Expertinnen und Experten, für die sehr konzentrierte und fruchtbare Diskussion, die uns sehr weitergeholfen hat.

Damit möchte ich das Wort weitergeben an Eva Winkler.

## **Zusammenfassung**

### **Eva Winkler · Deutscher Ethikrat**

Meine Aufgabe ist es, alles zusammenzufassen. Wir hatten eine sehr differenzierte und vielschichtige Diskussion. Es gab einige Linien, Muster oder Themencluster, die man benennen kann. Es ist deutlich geworden, dass Neurotechnologie längst nicht mehr auf den medizinischen Kontext beschränkt ist, sondern dass das, worum wir uns ethisch und rechtlich in der Diskussion kümmern müssen, der kommerzielle oder wie Sie gesagt haben, Frau Timón, der Borderline-Bereich zwischen kommerziell und medizinisch ist.

Wir haben Überlegungen in vier Themenclustern. Das Erste ist die Entwicklung vom Werkzeug zur Schnittstelle (so hatten Sie es genannt, Frau Nagel), also die hybride Handlungsfähigkeit von der

Wiederherstellung von Funktionen bis zur Erweiterung unserer Fähigkeiten und der damit verbundenen Verantwortung. Je stärker menschliche Kognitionen und technische Systeme miteinander verschränkt sind, desto schwieriger wird die Zuschreibung von Kontrolle und Verantwortung.

Wir haben über mögliche Verantwortungslücken diskutiert, ob es sie gibt, wer sie schließen soll oder ob wir sie einfach proaktiv schließen, und über die damit verbundene Vertrauenswürdigkeit des Umgangs mit den Systemen, vielleicht nicht der Systeme an sich, und dass die gerade dann (das hatten wir zuletzt noch mal), wenn KI integriert ist, besonders schwer herzustellen ist.

Wir hatten gesehen, dass sich die Akzeptanz solcher hybrider Systeme entlang eines Spektrums ändert: Sie ist am höchsten, wenn sie restaurativ sind (also Fähigkeiten wiederherstellen, erhalten), und nimmt deutlich ab, wenn sie die Fähigkeiten erweitern in einen Bereich, der übermenschlich ist oder nicht typischen menschlichen Fähigkeiten entspricht.

Das ist auch deshalb interessant, weil der zweite große Themenbereich, den wir diskutiert haben, die gegenseitige Beeinflussung von individuellen Entscheidungen und Nutzenperspektiven mit den gesellschaftlichen Effekten war. Wir haben gesehen: Für den Einzelnen haben Neurotechnologien profunde Auswirkungen auf die persönliche Identität, aber auch auf die soziale Identität: Wo verorte ich mich als Mensch in einer Gruppe oder wo werde ich verortet? Und obwohl diese Anwendungen als freiwillige Entscheidung daher kommen und unterstützend erscheinen, wurde deutlich, dass es Formen von sozialem Druck zur Normalisierung gibt, denen man sich nicht gut entziehen kann, und dass vielleicht sogar intendiert ist, dass man sich denen nicht entziehen kann, weil es Marktlogiken sind, wo genau das im Grunde das Verkaufsmodell ist.

Insofern ist die Frage aufgekommen, ob die individuelle Einwilligung allein als normative Grundlage ausreicht oder welche Anforderungen sie erfüllen muss und ob man die im Verbraucherbereich überhaupt umsetzen kann. Also sind auch Fragen offengeblieben.

Wir haben zuletzt deutlich gesehen, dass Neurotechnologien grundlegende Fragen aufwerfen nach dem Selbst, nach der mentalen Integrität, Autonomie und Gedankenfreiheit. Hier haben wir in der Diskussion gelernt und herausgearbeitet, dass es nicht die Daten selbst sind, die wir schützen, sondern dass die Rückschlüsse aus den Daten für den Einzelnen riskant werden, weil es Ableitungen über das eigene Denken, über Präferenzen, über das, was man in Zukunft tun wird, ermöglicht. Das wäre der eine Grund, sie schützen zu wollen.

Auf der anderen Seite haben wir gesehen: Es gibt vielleicht einen tiefer liegenden Grund dafür, dass dieser Raum der Gedankenfreiheit – „Die Gedanken sind frei, wer kann sie erraten?“, war das Zitat – auch per se geschützt werden muss: damit wir auch privat bleiben können und nicht andere die Gedanken mitlesen können und weil nur so menschliches Miteinander funktioniert, weil man sich sonst das Denken verbieten müsste.

Ein bisschen mehr Klarheit, wie es reguliert werden soll, besteht, so haben wir gesehen, zwischen dem medizinischen und dem nichtmedizinischen Bereich, weil im medizinischen Bereich die Nutzen-, Wirksamkeits- und Evidenzanforderungen klarer geregelt sind. Vielleicht müssen wir für den kommerziellen und nichtmedizinischen Bereich mit anderen Instrumenten arbeiten – Mental Impact Assessment wurde genannt – und uns fragen, ob es da auch rote Linien gibt, Herr Kellmeyer, Sie hatten das auch erwähnt, und dass außerhalb des medizinischen Bereichs solche Technologie gar nicht zugelassen wird, bis wir nicht besser

verstanden haben, was es denn für individuelle, aber auch gesellschaftliche Implikationen und Folgen hat.

Diese vielen Aspekte und Diskussionen fließen alle in unsere Stellungnahme ein, die wir gerade im Deutschen Ethikrat vorbereiten. Wir haben auch viel gelernt von Ihren Gedanken zu den internationalen Diskussionen und Initiativen – UNESCO und OECD, haben wir gehört –, und wir sehen, dass wir in Deutschland mit diesen Herausforderungen nicht allein sind, wie man hier Governance, Menschenrechte und Schutz der mentalen Integrität stärker in den Fokus rückt und vielleicht regulieren kann, ohne die Chancen zu verbauen.

Ich möchte mich ganz herzlich bedanken bei allen Sachverständigen und bei den Mitgliedern des Ethikrates, die sich an der sehr differenzierten und anregenden Diskussion beteiligt haben. Damit beende ich die heutige Anhörung. Die Diskussion geht natürlich weiter. Herzlichen Dank.