

Alexander Supertramp/Shutterstock

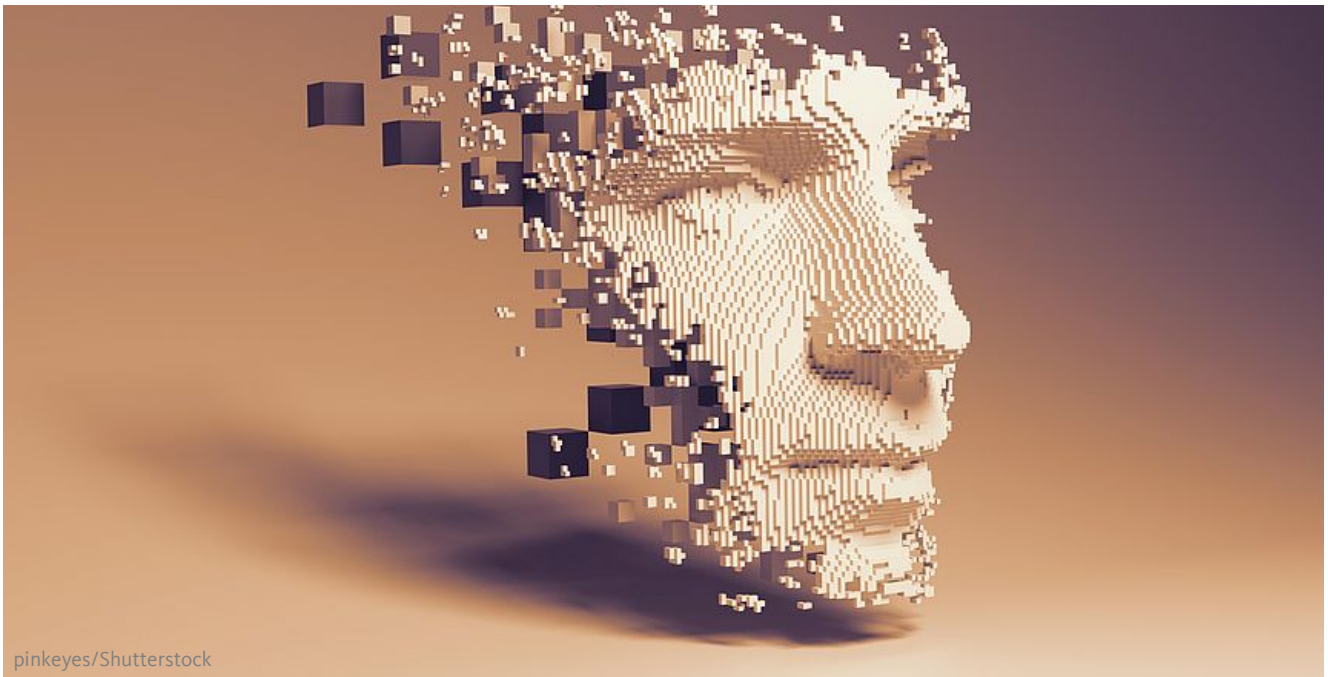
Öffentliche Anhörung

Künstliche Intelligenz und Mensch-Maschine-Schnittstellen

Online-Veranstaltung • 25. Februar 2021, Berlin

Diskutieren Sie mit:
[#menschundmaschine](#)

Die [Arbeitsgruppe Mensch und Maschine](#) des Deutschen Ethikrates beschäftigt sich derzeit mit den gesellschaftlichen Auswirkungen der Digitalisierung und den damit verbundenen ethischen Fragen des Verhältnisses von Mensch und Maschine. Im Mittelpunkt stehen Entwicklungen auf dem Gebiet der Künstlichen Intelligenz und die immer stärkere Durchdringung aller Lebensbereiche mit digitalen Technologien.



Im Rahmen der öffentlichen Anhörung wird der Deutsche Ethikrat vier Sachverständige, die sich mit der Entwicklung algorithmischer Systeme, maschinellem Lernen und möglichen Schnittstellen und Interaktionen zwischen Menschen und komplexen Maschinen beschäftigen, zum aktuellen Stand und zu absehbaren Entwicklungen ihrer Forschung befragen. Dabei wird es unter anderem um folgende Schwerpunkte gehen:

- Was sind wichtige Unterschiede zwischen Menschen und Maschinen mit ihren jeweiligen materiellen und funktionalen Merkmalen?
- Welche Gemeinsamkeiten, Annäherungen und Schnittstellen zwischen Mensch und Maschine sind zu beobachten oder zu erwarten?
- Welche Rolle spielen verschiedene Konzepte von (künstlicher) Intelligenz für die Entwicklung aber auch für die Bewertung (teil-)autonomer Systeme?
- Welche ethischen Herausforderungen gibt es mit Blick auf konkrete Forschungs- und Praxisfelder?

09:30 Begrüßung

Alena Buyx · Vorsitzende des Deutschen Ethikrates

09:40 Ethik und KI: Was kann die Technik (nicht) leisten?

Ulrike von Luxburg · Eberhard Karls Universität Tübingen

Rückfragen der Ratsmitglieder

10:10 Adaptive Intelligenz: Aktuelle Entwicklungen und Disruptionspotenzial des Maschinellen Lernens

Matthias Bethge · Eberhard Karls Universität Tübingen

Rückfragen der Ratsmitglieder

10:40 Kaffeepause

10:55 Aktuelle Entwicklungen bei Kognitiven Systemen und im Bereich der Mensch-Maschine-Interaktion

Tanja Schultz · Universität Bremen

Rückfragen der Ratsmitglieder

11:25 Aktuelle Entwicklungen in der KI im Bereich der Neurowissenschaften

Stefan Remy · Leibniz-Institut für Neurobiologie Magdeburg

Rückfragen der Ratsmitglieder

11:55 Kaffeepause

12:10 Diskussion mit den Ratsmitgliedern

13:40 Schlusswort

Alena Buyx · Vorsitzende des Deutschen Ethikrates

13:45 Ende

Moderation: Judith Simon · Deutscher Ethikrat



Vorsitzende des
Deutschen Ethikrates
Alena Buyx



Moderatorin der
Veranstaltung
Judith Simon

Ulrike von Luxburg

Eberhard Karls Universität Tübingen



Beruflicher Werdegang

- Seit 2015 Professorin für Theorie des Maschinellen Lernens an der Eberhard Karls Universität Tübingen
- 2012-2015 Heisenberg-Professorin für Maschinelles Lernen an der Universität Hamburg
- 2007-2012 Forschungsgruppenleiterin am Max-Planck-Institut für intelligente Systeme in Tübingen
- 2005-2006 Leiterin der Gruppe „Data Mining“ am Fraunhofer-Institut für Integrierte Informations- und Publikationssysteme (IPSI)
- 2004 Promotion in Informatik an der Technischen Universität Berlin
- 2001 Diplom in Mathematik (nach Studium in Konstanz, Grenoble, Tübingen)

Ausgewählte Mitgliedschaften

- Seit 2020 Mitglied der Nationalen Akademie der Wissenschaften Leopoldina
- Seit 2019 Sprecherin des Exzellenzclusters „Maschinelles Lernen: Neue Perspektiven für die Wissenschaft“
- Seit 2017 Max Planck Fellow am Max Planck Institute for Intelligent Systems

Ausgewählte Publikationen

- Lohaus, M.; Perrot, M.; von Luxburg, U. (2020): Too relaxed to be fair. ICML 2020, <https://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/LohausPerrotLuxburg2020.pdf> [18.02.2021].
- Garreau, D.; von Luxburg, U. (2020): Explaining the Explainer: A First Theoretical Analysis of LIME. AISTATS 2020, <https://arxiv.org/pdf/2001.03447.pdf> [18.02.2021].
- Rendsburg, L.; Heidrich, H.; von Luxburg, U. (2020): NetGAN without GAN: From Random Walks to Low-Rank Approximations. ICML 2020, https://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/RensburgLuxburg_Netgan_without_Gan2020.pdf [18.02.2021].
- von Luxburg, U. (2020): Wie funktioniert maschinelles Lernen? Eine Einführung für jedermann und -frau. http://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/luxburg_wie_funktioniert_ml.pdf [18.02.2021].
- von Luxburg, U. (2020): Vorlesungen zum Thema „Machine Learning and Society“, <https://www.youtube.com/watch?v=5CH2qcZQrpk&list=PLo5umP7R6ij2XCvrRzLokX6EoHWaGAzC&index=48> [18.02.2021].
- Tang, C.; Garreau, D.; von Luxburg, U. (2018): When do random forests fail? NeurIPS 2018, <https://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/TangGarreauLuxburg2018.pdf> [18.02.2021].

Ulrike von Luxburg

Eberhard Karls Universität Tübingen

Ethik und KI: Was kann die Technik (nicht) leisten?

Was ist das Besondere an maschinellem Lernen? Es macht Vorhersagen, ohne Modelle zu entwickeln. Es extrahiert „Wissen“ aus Daten, die niemals für diesen Zweck erhoben worden sind. Es verblüfft uns mit scheinbar kreativem (intelligentem?) Verhalten, und gleichzeitig verunsichert es uns, indem es einen weiten Raum an Möglichkeiten – guten sowie schlechten – eröffnet. Die Debatte über die ethischen Anforderungen und Beschränkungen des maschinellen Lernens ist in vollem Gange.

In meinem Vortrag werde ich mich darauf konzentrieren, die technische Seite dieser Debatte zu beleuchten: Was kann die Wissenschaft zum jetzigen Zeitpunkt darüber sagen, ob „ethische Anforderungen“ überhaupt technisch umsetzbar sind? Und welche ethischen Dilemmata lassen sich prinzipiell nicht durch technische Lösungen aus dem Weg räumen?

Matthias Bethge

Eberhard Karls Universität Tübingen



Beruflicher Werdegang

- Seit 2009 Professor in Computational Neuroscience and Machine Learning an der Eberhard Karls Universität Tübingen
- 2019-2021 Amazon Scholar im AWS Lablet “Research on Deep Learning” bei Amazon in Tübingen
- 2007-2009 Independent Junior Group Leader in Computational Neuroscience am Max-Planck-Institut für Biologische Kybernetik in Tübingen
- 2005-2006 Senior PostDoc in Computational Neuroscience am Max-Planck-Institut für Biologische Kybernetik in Tübingen
- 2003-2005 Direktor des Tübingen AI Centers PostDoc in Computational Neuroscience am Redwood Neuroscience Institute / Redwood Center for Theoretical Neuroscience der University of California in Berkeley (USA)

Ausgewählte Mitgliedschaften

- Seit 2019 Mitbegründer und Vorstandsmitglied der ELLIS Society
- Seit 2018 Direktor des Tübingen AI Centers
- Seit 2017 Sprecher des Sonderforschungsbereichs SFB 1233 Robustes Sehen
- Seit 2014 Stellvertretender Vorsitzender des Bernstein Netzwerks für Computational Neuroscience
- 2018 Mitgründer des Spin-offs Deepart UG haftungsbeschränkt
- 2016 Mitgründer des Spin-offs Layer7 AI GmbH

Ausgewählte Publikationen

- Geirhos, R. et al. (2019): ImageNet-trained CNNs are biased towards texture; increasing shape bias improve accuracy and robustness. ICLR 2019, <https://arxiv.org/pdf/1811.12231v2.pdf> [18.02.2021].
- Schott, L. et al. (2019): Towards the first adversarially robust neural network model on MNIST. ICLR 2019, <https://openreview.net/pdf?id=S1EHOsC9tX> [18.02.2021].
- Brendel, W.; Rauber, J.; Bethge, M. (2018): Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. ICLR 2018, <https://openreview.net/pdf?id=SyZloGWCZ> [18.02.2021].
- Mathis, A. et al. (2018): DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. In: Nature Neuroscience, 21 (9), 1281-1289.
- Baden, T. et al. (2016): The functional diversity of retinal ganglion cells in the mouse. In: Nature, 529, 345-350.
- Theis, L.; van den Oord, A.; Bethge, M. (2016): A note on the evaluation of generative models. ICLR 2016, <https://arxiv.org/pdf/1511.01844.pdf> [18.02.2021].
- Gatys, L. A.; Ecker, A. S.; Bethge, M. (2016): Image style transfer using convolutional neural networks. CVPR 2016, https://rn-unison.github.io/articulos/style_transfer.pdf [18.02.2021].
- Gatys, L. A.; Ecker, A. S.; Bethge, M. (2015): Texture synthesis using convolutional neural networks. NeurIPS 2015, <https://papers.nips.cc/paper/2015/file/a5e00132373a703100ofd987a3c9f87b-Paper.pdf> [18.02.2021].

Matthias Bethge

Eberhard Karls Universität Tübingen

Adaptive Intelligenz: Aktuelle Entwicklungen und Disruptionspotenzial des Maschinellen Lernens

Seit Beginn der industriellen Revolution konnte durch die zunehmende Automatisierung ein enormer Produktivitätszuwachs erreicht werden. Diese Automatisierung erforderte aber eine massive Umstellung der Arbeitsprozesse, zum Beispiel in Form von Fließbandarbeit oder dem Großmaschineneinsatz in der Landwirtschaft mit vielfältigen negativen Konsequenzen für Mensch, Natur und Umwelt. Der Grund dafür liegt in der sehr begrenzten Fähigkeit der klassischen regelbasierten KI- und Automatisierungstechnik, sich an eine variable Umwelt anzupassen.

Das letzte Jahrzehnt war vom Siegeszug des maschinellen Lernens gekennzeichnet. Auch wenn die lernende KI weit davon entfernt ist, ausgereift zu sein, nimmt der Grad an Flexibilität und Robustheit kontinuierlich zu. Ein entscheidendes Ziel nachhaltiger, menschenzentrierter KI-Forschung sollte daher sein, auf Basis des maschinellen Lernens eine Trendwende zu schaffen, damit sich in Zukunft Maschinen an die Bedürfnisse von Mensch, Natur und Umwelt anpassen und nicht umgekehrt.

Tanja Schultz

Universität Bremen



Beruflicher Werdegang

- Seit 2015 Professur für Kognitive Systeme im Fachbereich Mathematik und Informatik an der Universität Bremen und Direktorin des Cognitive Systems Laboratory
- 2007-2015 Professur für Kognitive Systeme an der Fakultät für Informatik und Karlsruher Instituts für Technologie und Direktorin des Cognitive Systems Laboratory
- Seit 2002 Research Scientist / Research Professor (seit 2014: Adjunct Professor) am Language Technologies Institute der Carnegie Mellon University in Pittsburgh, Pennsylvania (USA)
- 2000 Post-Doc an den Interactive Systems Laboratories der Carnegie Mellon University in Pittsburgh, Pennsylvania (USA)
- Bis 2000 Promotion zum Doktor der Ingenieurwissenschaften an der Universität Karlsruhe (2000); Diplom in Informatik an der Universität Karlsruhe (1995); 1. Staatsexamen (Lehramt an Gymnasien) in Mathematik, Sport und Erziehungswissenschaften an der Ruprecht-Karl Universität Heidelberg (1989)

Ausgewählte Mitgliedschaften

- Seit 2020 Fellow am Institute of Electrical and Electronics Engineers
- Seit 2017 Fellow der European Academy of Science and Arts
- Seit 2016 Fellow der International Speech Communication Association
- 2013-2015 Gewählte Präsidentin der International Speech Communication Association
- 2004-2012 Vorstandsmitglied der International Speech Communication Association
- Weitere Sprecherin für den Wissenschaftsschwerpunkt "Minds, Media, Machines" der Universität Bremen und des Beirats der UKRI Centre for Doctoral Training in Speech and Language Technologies; Mitglied des Direktoriums des Leibniz-WissenschaftsCampus Digital Public Health, des ISCA International Advisory Board, des Kuratoriums der Fraunhofer IDMT, der Brain Computer Interface Society und der Gesellschaft für Informatik

Ausgewählte Publikationen

- Mason, C. et al. (2020): From human to robot everyday activity. IROS 2020, https://www.csl.uni-bremen.de/cms/images/documents/publications/mason_iros2020.pdf [18.02.2021].
- Schultz, T. (2018): Technik und Demenz. In: Gebhard, D.; Mir, E. (Hg.): Gesundheitsförderung und Prävention für Menschen mit Demenz. Berlin, 239-255.
- Schultz, T. et al. (2017): Biosignal-based spoken communication: a survey. In: IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25 (12), 2257-2271.
- Herff, C.; Schultz, T. (2016): Automatic speech recognition from neural signals: a focused review. In: Frontiers in Neuroscience, 10, 429.
- Denby, B. et al. (2010): Silent speech interfaces. In: Speech Communication, 52 (4), 269-380.
- Schultz, T.; Kirchhoff, K. (2006): Multilingual Speech Processing. Orlando, Florida (USA).
- Schultz, T.; Waibel, A. (2001): Language independent and language adaptive acoustic modeling. In: Speech Communication, 35 (1-2), 31-51.

Tanja Schultz

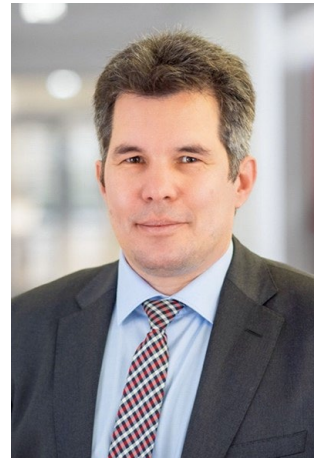
Universität Bremen

Aktuelle Entwicklungen bei Kognitiven Systemen und im Bereich der Mensch-Maschine-Interaktion

Die Entwicklung autonomer, intelligenter technischer Systeme, kurz Kognitive Systeme genannt, verläuft rasant. Durch das Zusammenspiel von Maschinellen Lernverfahren, Big Data und Super Computing erreichen sie in Spezialgebieten bereits Human Parity. Von zukünftigen Kognitiven Systemen wird erwartet, dass sie das individuelle Wohlbefinden ihrer Nutzer*innen steigern und Aufgaben des täglichen Lebens übernehmen. Dazu benötigen sie Wissen und Kenntnisse über ihre Nutzer*innen und die Umwelt. Zu diesem Zweck werden Mensch-Maschine-Schnittstellen kreiert, die globale IT-Player jedoch primär zur Akquise persönlicher Daten ausnutzen. Dabei werden Menschen rund um die Uhr mit vernetzten Sensoren vermessen: Zu Hause stehen Kamera und Mikrofon, für unterwegs wird das Handgelenk zum Datenhub, ausgestattet mit Inertialsensoren und Elektroden. Der Trend geht unter die Haut, denn neuartige Sensoren und Chips könnten injiziert oder implantiert werden. Eine differenzierte Auseinandersetzung mit individuellen, gesellschaftlichen und politischen Konsequenzen, die durch die Nutzung derartiger Systeme entstehen, ist von gewinnorientierten IT-Firmen und Überwachungsstaaten nicht zu erwarten. Vielmehr sollten demokratische Regierungen durch öffentlich geförderte Programme eine kritische Erforschung und Entwicklung von Kognitiven Systemen unterstützen, die Grundrechte respektieren, Nutzer*innen befähigen, ihre Daten zu kontrollieren, Transparenz und Erklärbarkeit der Algorithmen als Open Source umsetzen sowie bestmöglich vor potenziellem Missbrauch und Risiken schützen.

Stefan Remy

Leibniz-Institut für Neurobiologie Magdeburg



Beruflicher Werdegang

- Seit 2020 Geschäftsführender Direktor des Leibniz Instituts für Neurobiologie in Magdeburg und Professor für molekulare und zelluläre Neurobiologie an der Otto von-Guericke-University Magdeburg
- 2013-2019 Teilbereichskoordinator „Brain Function“ im Helmholtz Programm Neurodegenerative Erkrankungen im Bereich Gesundheit
- 2013-2019 Professor und Forschungsgruppenleiter “Neuronal Netzwerke” am Deutschen Zentrum für Neurodegenerative Erkrankungen in der Helmholtz Gemeinschaft in Bonn
- 2011-2013 Juniorprofessor und Nachwuchsgruppenleiter “Zelluläre Substrate kognitiver Dysfunktion” am Deutschen Zentrum für Neurodegenerative Erkrankungen in Bonn
- 2008-2009 BONFOR Gruppenleiter in der Klinik für Epileptologie der University of Bonn
- 2005-2007 Feodor Lynen Stipendiat der Alexander von Humboldt Stiftung am Department of Neurobiology and Physiology der Northwestern University in Chicago, Illinois (USA)
- 2003-2004 Arzt im Praktikum/Assistenzarzt in der Klinik für Epileptologie der University of Bonn

Ausgewählte Mitgliedschaften

- Seit 2020 Direktor des Center for Behavioral Brain Sciences - Leibniz WissenschaftsCampus Magdeburg
- Weitere Mitglied der Society for Neuroscience und der Neurowissenschaftlichen Gesellschaft

Ausgewählte Publikationen

- Luxem, K. et al. (2020). Identifying Behavioral Structure from Deep Variational Embeddings of Animal Motion. In: *BioRxiv*, DOI:10.1101/2020.05.14.095430.
- Giovannetti, A. et al. (2018): Restoring memory by optogenetic synchronization of hippocampal oscillations in an Alzheimer’s disease mouse model. In: *BioRxiv*, DOI:10.1101/363820.
- Justus, D. et al. (2017): Glutamatergic synaptic integration of locomotion speed in the medial entorhinal cortex. In: *Nature Neuroscience*, 20 (1), 16-19.
- Fuhrmann, F. et al. (2015): Locomotion, theta oscillations, and the speed-correlated firing of hippocampal neurons are controlled by a medial septal glutamatergic circuit. In: *Neuron*, 86 (5), 1253-1264.
- Müller, C. et al. (2012) Inhibitory control of linear and supralinear dendritic excitation in CA1 pyramidal neurons. In: *Neuron*, 75 (5): 851-864.

Stefan Remy

Leibniz-Institut für Neurobiologie Magdeburg

Aktuelle Entwicklungen in der KI im Bereich der Neurowissenschaften

Invasive Technologien zur Auslesung und Stimulation von Nervenzellen im intakten Gehirn ermöglichen in Tierversuchen in der Grundlagenforschung bereits eine dichte Aufzeichnung und Stimulation der Aktivität von Tausenden Nervenzellen pro implantierter Sonde. Für den Einsatz am Patienten zugelassene invasive Systeme sind technologisch auf deutlich niedrigerem Entwicklungsstand und werden bisher nur in Fällen eingesetzt, in denen ein klarer diagnostischer Vorteil für den Patienten erkennbar und der Eingriff somit ethisch vertretbar ist. Die auf den Menschen gerichtete Technologieentwicklung wird allerdings im kommerziellen Bereich massiv vorangetrieben. Erklärte und durch höheren Ressourceneinsatz erreichbare Ziele sind höhere Bandbreite, kabellose Signalübertragung und bessere Langzeitverträglichkeit sowie die Entwicklung von Operationsverfahren mit minimaler Gewebeschädigung. Anzuzweifeln ist, dass in absehbarem Zeitraum teilweise aus Marketinggründen erklärte Anwendungsziele erreicht werden. Diese reichen von Erleichterungen bei Blindheit, Taubheit, Lähmungen, Angststörungen, Depressionen oder neurodegenerativen Erkrankungen bis zur Anwendung zur Expansion kognitiver Ressourcen, zur Verhaltensmodulation und als digitale Kommunikationsschnittstelle. Hierfür fehlen noch neurobiologische Erkenntnisse über die Kommunikation von Nervenzellen im menschlichen Gehirn und neuronale Plastizität, was durch den Einsatz von KI-basierten Methoden kompensiert werden soll. Vorhersagen über die Wirkung der Stimulation weniger Nervenzellen auf ein spezifisches menschliches Verhalten brauchen aber verlässliche Modelle, die auf neurobiologischem Erkenntnisgewinn basieren. Die dafür notwendige Forschung ist nicht nur technisch schwierig, sondern auch ethisch umstritten, denn sie kann insbesondere beim Eingriff in autobiografische Gedächtnisareale die Persönlichkeit eines Menschen dauerhaft verändern. In mittelfristig absehbarer Zeit, mithilfe neuerer Technologien und dem Einsatz von KI erreichbar erscheint jedoch die Verbesserung der Auslesung neuronaler Aktivität bis hin zur Interpretation von Gedächtnisinhalten. Hier müssen ethische Aspekte frühzeitig einbezogen werden.

