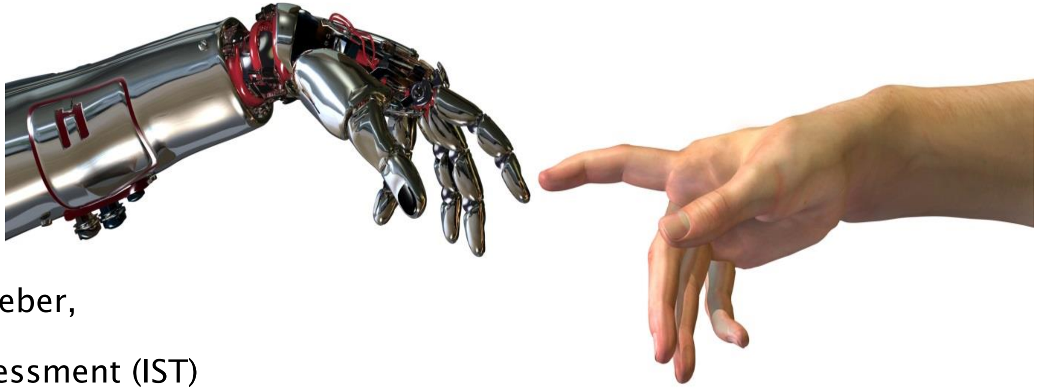
OSTBAYERISCHE
TECHNISCHE HOCHSCHULE
REGENSBURG

Evaluation of different stakeholder-values and their weight in AI-based technologies development, construction and usage-embedment

4th European Technology Assessment Conference

Bratislava, 4th-6th November 2019



Arne Sonar (M.A.), Prof. Dr. phil. habil. Karsten Weber,

Institute for Social Research and Technology Assessment (IST)
Regensburg Center of Health Sciences and Technology (RCHST)
Ostbayerische Technische Hochschule (OTH) Regensburg
Arne.Sonar@oth-regensburg.de, Karsten.Weber@oth-regensburg.de

Project: AI & Ethics

funding period: 10.2018 – 09.2020
funded by: Bayerisches Staatsministerium für Wissenschaft und Kunst

Starting point:

- AI-approach (neural network, deep learning) for analysis of medical image data

  ➔ detection & diagnostic support for esophageal carcinomas (cancer)

Critical aspects (e.g.):

- Back-box, responsibility, changes in doctor-patient relationship & care structures, accessibility ...

Aim:

- stakeholder perspectives on AI-supported medical decisions

➔ development of ethical guidelines for the medical use of AI systems

Additional problem :

- Divergence of ethical normativity (ought) / empiricism of clinical reality (is)

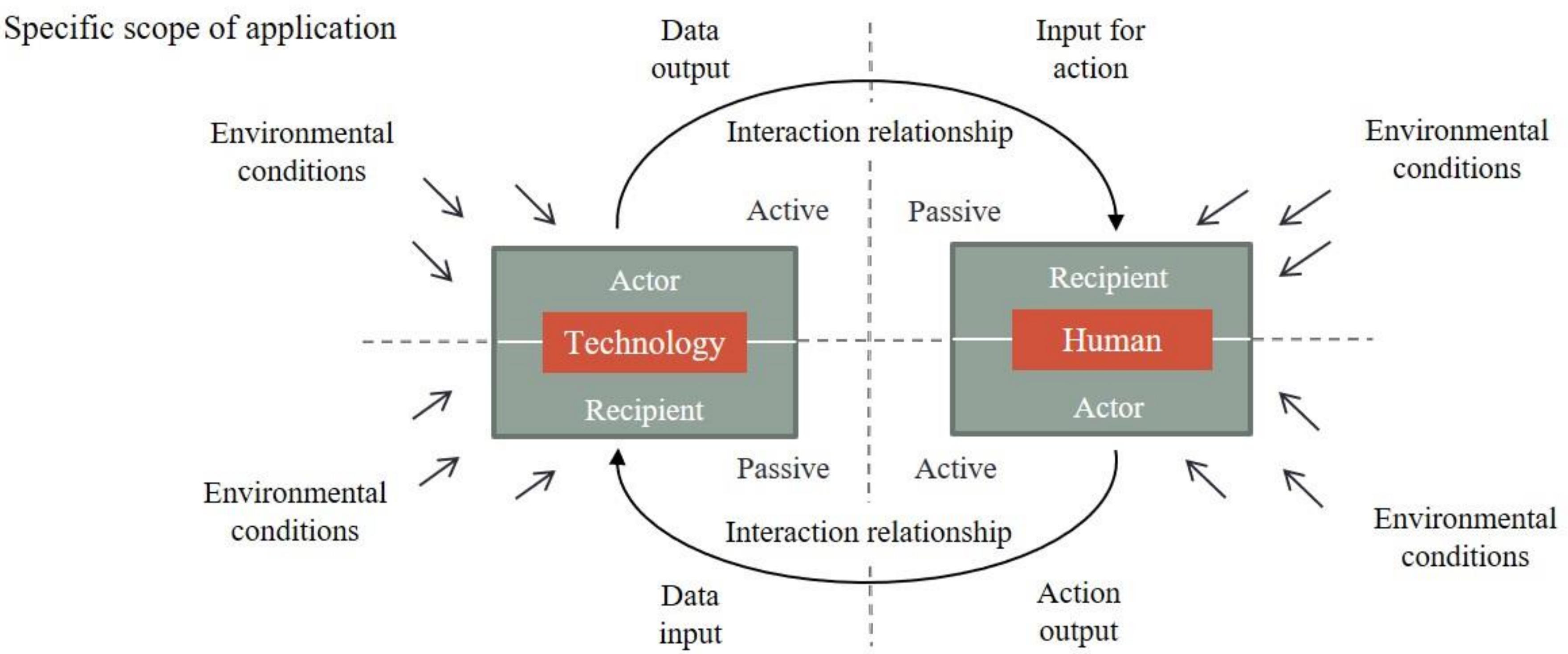
(Some) potential benefits of AI-applications in medicine

- Consolidation of health care costs & ensuring quality of medical care (Elsner et al. 2018, Kearney et al 2018)
- Help/support with detection/diagnosis, therapeutic decisions, forecasts or result predictions, monitoring (Combi 2017)
- Increasing temporal capacities on the physicians' side (for difficult cases) & extended access to decision-relevant information (de Bruijne 2016, Lebedev et al. 2018, Rampasek/Goldenberg 2018)
- Improvement of health care also in rural regions (Mayo/Leung 2018)
- A more comprehensive assessment with cost reduction (Madabhushi/Lee 2016)
- Reduction of time, labour and personal costs (Thompson et al. 2018)

(Some) critical aspects of AI-applications in medicine

- Controllability of results & defining gold standards for critical diagnostic cases (Schneider/Weiller 2018)
- comparability of human medical experts and technical capabilities (e.g. judgments, performance) (Jörk 2018, Müschenich 2018, Schneider/Weiller 2018)
- Legitimating AI-regulation & human corrections/interventions if AI outperforms human performances? (Thompson et al. 2018)
- norm of Individual medical empathy & personal presence = increasing physicians self-understanding? → correlative relation between worries - usage competences & trust (Jörk 2018)
- Patients interests (e.g. data souverains, data protection) vs. economic interests (Jörk 2018)

Semiotic & socio-technical/techno-social characteristics



(Sonar et al. 2017, 237)

Why taking on such considerations?

“As several authors have criticized AI and speculated how it may affect our lives, **it becomes doubly important to address the history of AI technology** from a non-deterministic stance (…). Rather than asking how AI has affected society and changed the nature of work, it is vital to ask instead what choices have been made in the type of AI technology that has been developed (…). This view accords more control to the individual, **thus making the technology appear less sinister and less mystifying and ultimately more open to criticism**. Such an approach can form a natural ally to the “human centered” debate in information systems and AI.”

(Adam 1990, 237)

Analysis of AI`s history

1st period – *AI-Pioneer´s Assumptions*

- equal meaning of term „intelligence“ for human and machines
- singularity of information processing

limited terms can`t grasp full complexity

lacking of semantical abilities

GPS – critique

IPS – critique

demands

- objectification of human beings

or

- differentiation in terms of intelligence
- limiting application areas

2nd period – *Expert Systems*

- plurality in terms of „intelligence“
- machines = talented or intellectual
- weak AI-approaches = functional tools

delegation of decision & responsibility

trustworthiness of output: → traceability & comprehensibility

value of support abilities: → reliability & robustness

domain-specific – critique

knowledge-based – critique

demands

- partnership-based linkage reflecting medium-term & long-term effects
- openness towards support in awareness for boundaries
- avoiding blind faith by critical reflection

3rd period – *Connectivism & Neural Networks*

- preparation of larger amounts of data
- multi-media ability of applications
- distributed AI

penetrating action & decision-making spaces

reductionism in images of humans and morality

tendencies of unconscious thinking

marginalization of differences in intelligence

multi-agent-systems – critique

data-based PDP – critique

demands

- transparency in derivation for morally relevant recommendations
- increasing user orientation & recognition of design and implementation aspect
- criteria for responsible design & use

4th period – *Big Data, IoT, Machine Learning*

- embodiment & situadedness
- highly divergent subcomponents
- locally unbound connectivity

autonomous & independent optimization → responsibility

weigh of different degrees of socially acceptable options for action

obfuscation of omnipresence of algorithmic action

data-driven power distribution → data-protection & sovereignty

cyber-physical-systems – critique

data-fication – critique

demands

- reflecting applications as socio-technical & techno-social entities
- aligning on consensual principles orientated on stakeholders values

Identified areas of problematization

(1) Understanding / perception of AI

- E.g. intelligence terms/criteria, positive/negative narrative, underestimating complexity (approving attitude) / overestimating possibilities (skeptical)

(2) Basic technical aspects / extended relationships to data

- E.g. more technology (data-dependent) = higher susceptibility to errors?

(3) Design & implementation aspects (focusing accountability & responsibility)

- E.g. special or general design & implementation criteria, dimensions of accountability/responsibility

(4) Relationship between humans & technology in decision-making

- E.g. dimensional limits of technical decisions/automation, Loss of decisions without technical support?

(5) Extended requirements for future systems & research benefits

- E.g. control challenges (legal framework conditions, data protection), health economic aspects (standard benefit, accessibility)

Conclusion

- combining of a theoretical analyzing of literary mentioned threats as initial insights into the normative landscape with stakeholder-interviews, provides empirical access to normative debates
- Implementation of AI as a social challenge leads to the necessity of an consciously construction on the basis of concrete, value-oriented principles, which should be obtained with the involvement of all stakeholders and through consensual procedures
- Reconstruction of AI`s history and the accompanying normative discourses as well as the involvement of stakeholders using social science methods are only two ways to support the formulation of ethical principles of medium range

References

Adam, A. E. (1990): What can the history of AI learn from the history of science? In: AI and Society. Vol. 4, Springer, London, p. 237.

Brödner, P.

Combi, C. (2017): Artificial Intelligence in medicine and the forthcoming challenges; in: Artificial Intelligence in Medicine, 76, 37-39.

De Bruijne, M. (2016): Machine learning approaches in medical image analysis: From detection to diagnosis; in: Medical Image Analysis, 33, 94-97.

Elsner, P., M. Fischer, S. Schliemann, J. Tittelbach (2018): Teledermatologie und künstliche Intelligenz: Potenziale für die Optimierung von Diagnostik, Therapie und Prävention bei Versicherten mit Berufskrankheiten der Haut; in: Trauma Berufskrankheit, 20, 103-108.

Fink / Janneck

Jörk, J. (2018): Digitalisierung in der Medizin: Wie Gesundheits-Apps, Telemedizin, künstliche Intelligenz und Robotik das Gesundheitswesen revolutionieren. Berlin.

Kearney, V., J. W. Chan, G. Valdes, T. D. Solberg, S. S. Yom (2018): The application of artificial intelligence in the IMRT planning process for head and neck cancer; in: Oral Oncology, 87, 111-116.

Lebedev, G., H. Klimenkoa, S. Kachkovskiyb, V. Konushinc, I. Ryabkov, A. Gromov (2018): Application of artificial intelligence methods to recognize pathologies on medical images; in: Procedia Computer Science, 126, 1171-1177.

Madabhushi, A., Lee, G. (2016): Image analysis and machine learning in digital pathology: Challeng-es and opportunities; in: Medical Image Analysis, 33, 170-175.

Mayo, R. C., J. Leung (2018): Artificial intelligence and deep learning: Radiology's next frontier?; in: Clinical Imaging, 49, 87-88.

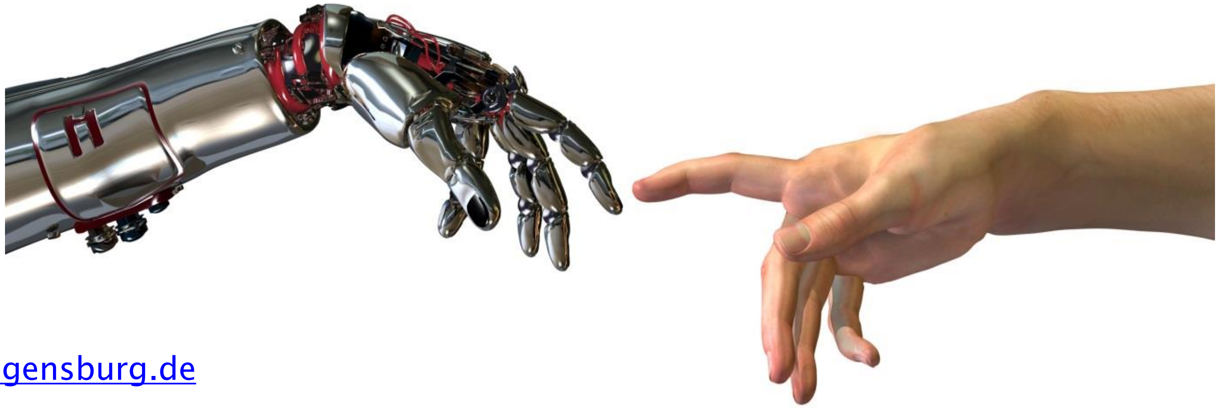
Rampasek, L., A. Goldenberg (2018): Learning from everyday images enables expert-like diagnosis of retinal diseases; in: Cell, 172, 893-895.

Schneider, F., C. Weiller (2018): Big Data und künstliche Intelligenz; in: Nervenarzt, 89, 859-860.

Sonar, A., Bleyer, B., Heckmann, D. (2017): Zur Synergie von reflexiver Technikbewertung und E(L)SA-Begleitforschung: Eine Bewertungstheorie sozio-technischer Systemgefüge im Rahmen der Digitalisierung. In: Bavarian Journal of Applied Sciences, Vol. 3, 234-247.

Thompson, R. F., G. Valdes, C. D. Fuller, C. M. Carpenter, O. Morin, S. Aneja, W. D. Lindsay, H. J.W.L. Aerts, B. Agrimson, C. Deville Jr., S. A. Rosenthal, J. B. Yu, C. R. Thomas Jr (2018): Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation?; in: Radiotherapy and Oncology, 129, 421-426.

Thank you for your attention!



Arne Sonar (M.A.): Arne.Sonar@oth-regensburg.de

Prof. Dr. phil. habil. Karsten Weber: Karsten.Weber@oth-regensburg.de

Institute for Social Research and Technology Assessment (IST)
Regensburg Center of Health Sciences and Technology (RCHST)
Ostbayerische Technische Hochschule (OTH) Regensburg

Conclusiones

- reducing the scope of AI-approaches for a more small-scale application orientation leads to an expansion (and specification) of aspects to be considered

- consideration of growing number of connected small technical factors: sensors, data, algorithms (and their training)

➔ higher potentials for error susceptibility implie necessity of human supervising?

➔ reflecting applications in more than just one dimension of their use (tool, machine, system)

- modern attributes of AI and especially their growing techno-social dimensions, increase the necessity of critical reflections in development and implementation of such specific operational AI systems

- functional tools with increasing effects on workflows
- development of technical applications as further parts of developing application areas (and future society)

➔ development and implementation as inter- and transdisciplinary approaches concerning all relevant stakeholder