Supported by:

Federal Ministry
of Health

on the basis of a decision
by the German Bundestag

ELNET

GIHF AI

German Israeli
Health Forum for
Artificial Intelligence

Policy Briefing "Communication & Trust"

Ethical guidelines for AI in medicine

The current policy briefing of the German Israeli Health Forum for Artificial Intelligence (GIHF-AI) sheds light on the use of AI in medicine with a focus on ethical principles and their influence on trustworthiness and acceptance among the population. In addition to the European Union's ethical guidelines for trustworthy AI, Israel's ethical framework on the use of AI in medicine will be presented as well as use cases.

Authors: Elisabeth Liebert (University Medicine Essen), Lea Ledwon (ELNET)

In Israel, a man who had been suffering from severe headaches for some time underwent a CT scan in a hospital. Immediately after the examination, the director of the MRI institute received an urgent alert from a new AI-based app on his smartphone. This indicated that the patient was suffering from a brain hemorrhage and contained the results of the CT scan as well as the patient's medical and personal data. The doctor immediately called the patient, who was fortunately still nearby and asked him to come to the hospital immediately.

Further imaging examinations confirmed that the man was indeed suffering from a brain hemorrhage. The following day, a life-saving operation was performed, during which a blood clot on his brain was also removed. The patient was discharged safely two days after the operation. This example of the benefits of AI in medicine is just one of many.[1]

However, the **use of artificial intelligence in medicine is not uncontroversial and harbors risks**. Regardless of whether resentment stems from existing problems, one's own imagination, or the fear of being replaced by the machine, **trust in the use of AI is required on the part of both practitioners and patients**.[2] The Policy Briefing of the German Israeli

Health Forum for Artificial Intelligence (GIHF-AI) on "Trustworthy use of artificial intelligence in healthcare"[3] already deals with this aspect. The definition of trust provided there, in particular the testing and regulation or certification of AI, also serves as a basic definition for this policy paper: "Trust denotes a specific relationship quality between a trust giver and a trust object".[4]

The **trustworthiness of an AI is essential for its acceptance by the population and its use in the European Union** (EU). Therefore, any **AI used in the EU** should be **subject to Europe-wide legal regulations such as the EU AI Act**. From an ethical perspective, the **reference to European fundamental values is important**, as the use of AI must not harm these fundamental values.[5] **Trust is created through transparency and guidelines** that both users and developers can refer to.

EU Ethics Guidelines for Trustworthy Artificial Intelligence

In line with this starting position, the EU High-Level Expert Group on AI (AI HLEG) proposes the **four ethical principles of respect for human autonomy, prevention of harm, fairness, and explicability** in

its guidelines on trustworthy AI.[6]

The AI HLEG derives seven requirements from these four principles:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability.[7]

With its guidelines, the AI HLEG provides a general approach to the ethical assessment of AI that can be applied to any field of application of AI and also serves as a basis for work on specific certification catalogs (e.g. the *Guideline for Designing Trustworthy Artificial Intelligence* or the Whitepaper *KI-Anwendungen systematisch prüfen und absichern of the Fraunhofer* IAIS).[8] The development of ethical guidelines and directives is about "ethical norms that are not necessarily legally binding yet crucial to ensure trustworthiness".[9]

Although **ethical norms can serve as a basis for the legal regulation of AI, their purpose goes beyond this**, as they encourage each person to critically examine the respective specific use case for its moral correctness. To this end, it is necessary to place situation-specific circumstances in an overall context, which cannot be achieved by general legislation per se.[10]

However, the debate around the use of AI in medicine is not a European one, but a global one. With the EU AI Act, **Europe is currently taking the lead in the regulation of AI**. Companies around the world will have to adapt to EU regulations if they want to use their AI-based products in Europe. In addition, other legislators could use the EU's regulatory approach as a model for their own AI regulations or align their regulations with EU standards to enable international trade and cooperation. For these reasons, **Israel** is also looking to European institutions[11], although it has **already adopted its own ethical guidelines in 2023**.

Israel's ethical guidelines on the use of AI in medicine

In December 2023, the Israeli Ministry of Innovation, Science, and Technology published a policy paper[12], in which AI principles, regulation, and ethics are described. Since then, this paper establishes a framework for adopting AI technology, which also serves the Israeli Ministry of Health. The most important points are listed below:

- **Holistic approach:** AI policy identifies seven main problems arising from the use of AI in the private sector: bias and discrimination, lack of transparency and human oversight, potential harms to privacy, the vulnerability of AI systems, safety concerns, concerns about accountability, and IP-related considerations.
- **Collaborative development:** The AI policy was formulated in collaboration with various interest groups, including the Ministry of Innovation, Science and Technology, the Ministry of Justice, the Israel Innovation Authority, the Data Protection Authority, the Israeli National Cyber Directorate, the Israeli National Digital Agency, leading AI companies in Israel, and academia.
- **Policy principles:** Following the OECD AI recommendations, Israel's AI policy outlines common policy principles, and practical recommendations to address challenges and promote responsible AI innovation.
- **Concept of responsible innovation:** By emphasizing the concept of "responsible innovation", AI policy aims to strike a **balance between innovation and ethics by viewing them as synergistic rather than conflicting goals**.[13]

In addition, the Ministry of Health adopted the Good Machine Learning Practice for Medical Device Development: Guiding Principles (GMLP) from the FDA, MHRA, and Health Canada. The aim was to adapt these principles for implementation in the Israeli healthcare system and to develop initial guiding principles:

- Incorporating multidisciplinary expertise throughout the product lifecycle

- Using best practices of software engineering and information security in building the model
- Proper representation of the target population within the framework of the clinical study and data sets
- Using independent databases to train and test the model
- Using accepted and best practices for developing a reference data set
- Building a model that is adapted to the latest available data and reflects the intended use of the product
- Focusing on the performance of the human-in-the-loop team
- Testing device performance under clinically relevant conditions
- Communicating essential information to users in a transparent and clear manner
- Monitor the performance of embedded models and manage the risks associated with real-world retraining.[14]

Israel's AI policy focuses on **regulations that are conducive to innovation**. It includes several recommendations worth mentioning, such as empowering sector-specific regulators, fostering international interoperability, and promoting multi-stakeholder cooperation. This can be explained, among other things, by the high-tech industry's high share of **18 percent of gross domestic product** and the **great interest in expanding the AI research sector**. At the same time, **human rights are given a high priority**, which is why Israel not only endorses the generally recognized ethical principles of AI but is also actively involved in setting international AI standards.[15]

Ethical challenges in digitalized medicine

As described in the example above, the use of AI in the medical context is primarily deployed as an element that supports humans, for example in image recognition. These **Clinical Decision Support Systems (CDSS) are already being applied in various areas such as diagnostics, therapy, and prognosis**.[16]

Due to the diversity of the healthcare system in terms of actors, possible uses of AI, and levels of responsibility, ethical guidelines are needed that take these circumstances into account.[17] Both the German Ethics Council and the Central Ethics Commission (ZEKO) of the German Medical Association therefore address the ethical challenges of using AI in medicine in their statements.[18]

The fundamental ethical question behind any use of AI is "For whom does it lead to extensions or reductions in authorship?" of his or her actions. This means that an **ethically justifiable use of AI in medicine** can **only be guaranteed if the free decision-making process and free development opportunities of people are not restricted or jeopardized**. Respect for autonomy applies to both practitioners and patients. The autonomy of the latter, which is necessary for informed consent, can only be guaranteed if the use of a CDSS is communicated appropriately by the doctors. In turn, the autonomy of doctors is inadmissibly restricted if the "boundary between assisted decision-making and the assumption of the system's decision"[20] blurs.

Especially the automation bias ("the tendency of humans to favor suggestions from automated decision-making systems and to ignore contradictory information, even if it is correct"[21]) poses a risk regarding this unquestioned assumption of AI-generated recommendations.

Trust and responsibility

Doctors' freedom of treatment and autonomy, which cannot be delegated to an AI system, is closely linked to their professional responsibility. For example, ZEKO believes that doctors have a micro-level responsibility to maintain people's trust in the medical profession as well as in medical care in general. To this end, **doctors must be able to rely on the fact that the AI systems they use have been developed and tested to the highest quality standards**.

As soon as an AI system is used, the respective doctors must still have the **relevant technical skills and understand the basic functioning of the AI**. This includes possible black box issues and potential biases in the AI training data sets. The plausibility check of

the results reported by the AI is essential. The responsibility for testing and certifying the AI systems lies at the meso level. In addition, the respective institution must ensure that its staff are provided with the necessary opportunities to acquire skills. At the macro level, the legislator is responsible for creating the necessary framework conditions, while medical societies should draw up appropriate guidelines on the benefits of AI systems.[22]

For the general use of AI in medicine, the German Ethics Council provides corresponding recommendations for action relating to certification, authorization and access options, data collection and data sovereignty, qualification of personnel, and the doctor's responsibility towards the patient. In principle, the **doctor's ultimate moral and legal responsibility for the diagnosis and treatment of the patient applies to the medical use of AI systems**.[23] Last but not least, patients' trust in the professional and interpersonal skills of doctors is important for successful treatment.

Delegation of treatment steps

The extent to which the specific use of AI in medicine raises moral concerns also depends on the degree to which the individual steps are delegated to the AI. If, for example, the AI is only used as another tool and takes over only a part of a single action step in the treatment chain, there is only a minor replacement of the human being, which is relatively unproblematic. However, it is important that the **doctor subjects the results to a plausibility check** and that automation bias is avoided.

More worrying is the far-reaching replacement of a human counterpart, as is already happening today in the form of apps or chatbots in the field of psychotherapy. The effects of this supposedly "direct, therapeutic relationship"[24] have not yet been fully clarified, but it is questionable whether the relationship of trust between therapists and patients necessary for successful treatment can be established in this way.[25] But even if, for example, chatbots only partially take over communicative steps in the treatment process, this could lead to

a neglect of the emotional interpersonal relationships between doctors and patients. However, **if,** for example, **AI is used to speed up diagnosis and make treatment processes more efficient doctors would ideally have more time for their patients**, which could strengthen the doctor-patient relationship.[26] The **use of such chatbots can also be very useful in cases where people cannot get a place in therapy or are still waiting, as well as in conflict regions**. German and Israeli institutions are currently working with such digital applications to help people in Ukraine and Israel.[27]

Another important aspect of successful treatment and doctor-patient relationships is the experience and knowledge of the medical profession. In contrast to the usually highly specialized AI systems, doctors can put the information they receive, such as the patient's biography or current life situation, into context. Comprehensive mechanization of diagnosis and treatment recommendations can lead to this knowledge being considered less valuable or not being acquired at all. It **remains to be seen whether the increased use of AI systems will not lead to a completely new type of experiential knowledge, as the entire workflow will potentially change**.[28]

Summary and outlook

Both the EU guidelines and the guidelines of the German Ethics Council and ZEKO refer to similar ethical requirements. They all focus on **people and their autonomy** and all call for **continuous quality control and the corresponding certification of the AI systems used**. Another key aspect is the **personal responsibility of humans: The AI systems used must not be relied on blindly**. People are always ultimately responsible and should critically compare the results of the system with their knowledge. This requires the appropriate technical skills to understand how an AI works and possible sources of error within the AI. Adequate handling of the data used is also necessary, whereby the framework conditions must be enacted by the legislator.

Furthermore, **what all guidelines have in common**

is that they are not legally binding. This raises the question of the extent to which context-specific ethical thinking can and should be translated into generally binding laws. **Israel's ethical guidelines also pave a similar path, although the issue of certification plays a subordinate role**, and they tend to be more flexible and more conducive to innovation.

The use of AI in such sensitive areas as medicine poses challenges. Given the rapid developments in the field of AI, the fundamental question arises as to how awareness of the ethical challenges can be raised without being overtaken by technical developments. The **embedded ethics** approach offers one way of doing this. **Here, ethicists are part of the development team from the very beginning and are involved in the entire AI develop-** ment process[29] In this way, ethical challenges can be addressed during development and the AI can be modified accordingly. The specific form of this cooperation is the responsibility of the respective developing companies. If necessary, consideration should be given to the extent to which embedded ethics can be legally supported.

To balance the benefits of AI in medicine and the challenges of using the new technology, an **exchange at the international level is necessary. Countries such as Israel, where the application is already more advanced, can use their experience to contribute to a constructive discourse**. Regulation must identify risks and challenges, but not ignore the great benefits for medical care. Ethical use of AI also means, above all, that AI is used when it improves treatment.

Bibliography

1. **The Jerusalem Post:** "Patient saved by AI app at Galilee Medical Center", 04.09.2023 in *https://www.jpost.com/health-and-wellness/article-757511*.

2. **ÄrzteZeitung:** "Die natürliche Angst vor dem Algorithmus", 04.11.2029 in *https://www.aerztezeitung.de/Wirtschaft/Die-natuerliche-Angst-vor-dem-Algorithmus-403227.html*.

3. **GIHF-AI:** "Vertrauensvolle Nutzung von Künstlicher Intelligenz im Gesundheitswesen", 14.10.2022 in *https://gihf-ai.eu/policy-briefing-2/vertrauensvolle-nutzung-von-kuenstlicher-intelligenz-im-gesundheitswesen/*.

4. Ibid.

5. **Europäische Kommission:** "Weissbuch. Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen", 19.02.2020 in *https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0065*.

6. **Hochrangige Expertengruppe für Künstliche Intelligenz:** "Ethik-leitlinien für eine vertrauenswürdige KI", 08.04.2019 in *https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai*.

7. Ibid.

8. **Fraunhofer IAIS:** "Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz", July 2021 in *https://www.iais.fraunhofer.de/de/forschung/kuenstliche-intelligenz/ki-pruefkatalog.html;* **Fraunhofer IAIS:** "KI-Anwendungen systematisch prüfen und absichern. Prüfwerk-zeuge und Prüfplattform zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz", May 2023 in *https://www.iais.fraunhofer.de/content/dam/iais/publikationen/studien-und-whitepaper/Fraunhofer_IAIS_Whitepaper_KI-Prueftools.pdf*.

9. **Hochrangige Expertengruppe für Künstliche Intelligenz:** "Ethik-leitlinien für eine vertrauenswürdige KI", 08.04.2019 in *https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai*.

10. Ibid.

11. **Scientific American:** "Europe's New AI Rules Could Go Global—Here's What That Will Mean", 07.02.2024 in *https://www.scientificamerican.com/article/europes-new-ai-rules-could-go-global-heres-what-that-will-mean/;* **Epstein Rosenblum Maoz:** "The AI Act is coming – Why Israeli businesses should care and start preparing for it", 21.02.2024 in *https://erm-law.com/news/ai-act-coming-israeli-businesses-care-start-preparing/*.

12. **Iapp:** "Proactive caution: Israel's approach to AI regulation", 10.01.2024 in *https://iapp.org/news/a/proactive-caution-israels-approach-to-ai-regulation/;* **Ministry of Innovation, Science & Technology:** "Israel's Policy on Artificial Intelligence", 2023 in *https://www.gov.il/BlobFolder/news/most-news20231218/en/Israels%20AI%20Policy%202023.pdf*.

13. Ibid.

14. **Ministry of Health:** „Guiding Principles for the Development of Machine Learning-Based Technologies", 03.04.2023 in *https://www.gov.il/en/Departments/General/digital-medical-technology-gmlp-1*.

15. **Iapp:** "Proactive caution: Israel's approach to AI regulation", 10.01.2024 in *https://iapp.org/news/a/proactive-caution-israels-approach-to-ai-regulation/*.

16. **ZEKO:** "Jahresbericht 2021 der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer (Zentrale Ethikkommission) in der 9. Amtsperiode 2019 – 2022", 12.07.2022 in *https://www.zentrale-ethikkommission.de/fileadmin/user_upload/zentrale-ethikkommission/Jahresberichte/ZEKO_Jahresbericht_2021.pdf*.

17. **Deutscher Ethikrat:** "Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz", 20.03.2023 in *https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf*.

18. **ZEKO:** "Jahresbericht 2021 der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer (Zentrale Ethikkommission) in der 9. Amtsperiode 2019 – 2022", 12.07.2022 in *https://www.zentrale-ethikkommission.de/fileadmin/user_upload/zentrale-ethikkommission/Jahresberichte/ZEKO_Jahresbericht_2021.pdf*.

19. **Deutscher Ethikrat:** "Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz", 20.03.2023 in *https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf*.

20. **ZEKO:** "Jahresbericht 2021 der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer (Zentrale Ethikkommission) in der 9. Amtsperiode 2019 – 2022", 12.07.2022 in *https://www.zentrale-ethikkommission.de/fileadmin/user_upload/zentrale-ethikkommission/Jahresberichte/ZEKO_Jahresbericht_2021.pdf*.

21. **Cummings, Mary:** "Automation Bias in Intelligent Time Critical Decision Support Systems", in 01.11.2014 in *https://web.archive.org/web/20141101113133/http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf*.

22. **ZEKO:** "Jahresbericht 2021 der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer (Zentrale Ethikkommission) in der 9. Amtsperiode 2019 – 2022", 12.07.2022 in *https://www.zentrale-ethikkommission.de/fileadmin/user_upload/zentrale-ethikkommission/Jahresberichte/ZEKO_Jahresbericht_2021.pdf*.

23. **Deutscher Ethikrat:** "Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz", 20.03.2023 in *https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf*.

24. Ibid.

25. Ibid.

26. **ZEKO:** "Jahresbericht 2021 der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer (Zentrale Ethikkommission) in der 9. Amtsperiode 2019 – 2022", 12.07.2022 in *https://www.zentrale-ethikkommission.de/fileadmin/user_upload/zentrale-ethikkommission/Jahresberichte/ZEKO_Jahresbericht_2021.pdf*.

27. **GIHF-AI:** "Report GIHF-AI Digital Mental Health: Psychische Gesundheit in Israel nach dem 7. Oktober", 05.01.2024 in *https://gihf-ai.eu/report/report-gihf-ai-digital-mental-health-psychische-gesundheit-in-israel-nach-dem-7-oktober/*.

28. Ibid.

29. **McLennan et al.:** "Embedded ethics: a proposal for integrating ethics into the development of medical AI." *2022 in BMC Medical Ethics, 23:6. doi: 10.1186/s12910-022-00746-3*.

ELNET

GIHF AI

German Israeli
Health Forum for
Artificial Intelligence

An initiative by ELNET:

GIHF-AI is an initiative by ELNET in Germany, a think tank and network organization in the context of European-Israeli relations. We work independently and across party lines on the basis of shared democratic interests and values. Better mutual understanding is promoted through networking and information exchange. Since its founding in 2007, ELNET has focused its work on the topics of foreign and security policy, antisemitism, and innovation.