



# Künstliche Intelligenz im ärztlichen Alltag Ethische Überlegungen

**Univ.-Prof. Dr. med. Dr. phil. Eva Winkler,**

Vorsitzende der Zentralen Ethikkommission der Bundesärztekammer (ZEKO)

Heisenberg-Professorin und Leiterin der Sektion für Translationale Medizinethik an der Universität Heidelberg,

Geschäftsführende Direktorin am Nationalen Centrum für Tumorerkrankungen (NCT), Heidelberg

19.10.2023

# Zentrale Ethikkommission bei der Bundesärztekammer (ZEKO)

---

- Seit 1994 – **unabhängig & interdisziplinär Zentralen Ethikkommission (ZEKO)**
- Aufgaben – Stellungnahmen/ Handlungsorientierung
- Gestern 100. Sitzung – bislang 28 Stellungnahmen
- Abrufbar unter [www.zentrale-ethikkommission.de](http://www.zentrale-ethikkommission.de)



# Zentrale Ethikkommission bei der Bundesärztekammer (ZEKO)

- Seit 1994 – **unabhängig & interdisziplinär Zentralen Ethikkommission (ZEKO)**
- Aufgaben – Stellungnahmen/ Handlungsorientierung
- Gestern 100. Sitzung – bislang 28 Stellungnahmen
- Abrufbar unter [www.zentrale-ethikkommission.de](http://www.zentrale-ethikkommission.de)



BUNDESÄRZTEKAMMER

Bekanntmachungen

## Stellungnahme

der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin  
und ihren Grenzgebieten (Zentrale Ethikkommission) bei der Bundesärztekammer

**„Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz“**

8/21

# 1. Potential von KI-basierten Systeme zur Entscheidungsunterstützung Clinical Decision Support Systems (CDSS)

## 1. Diagnostik: Bilderkennung – ca 70 %

- Screening (Retina, Knochendichte, Derma: Melanome)

## 2. Therapie: Entscheidungshilfen – ca 30%

- Behandlungsoptionen mit Wahrscheinlichkeiten (Ranking)
- Vorbefundung
- Priorisierung – Dringlichkeit

## 3. Prognose – allgemeine + spezielle Risikofaktoren

## 4. Prädiktion – Erkrankungsrisiken



Autonomous AI  
that instantly  
detects disease

Increase patient access to early disease  
detection.

[Learn More](#)

Idx-DR - Website

# 1. Innovationspotenzial KI-basierter Entscheidungsunterstützung

## *Vielfältige Chancen zur Verbesserung der Patient:innenversorgung*

- ✓ Hypothesengenerierend: Große Datensätze weisen auf Assoziationen hin
- ✓ Verbesserte Diagnostik
- ✓ Info zu personalisierte Therapiewahl
- ✓ Entlastung von ÄrztInnen /  
des Gesundheitssystems



**DGK.**

Deutsche Gesellschaft für Kardiologie  
– Herz- und Kreislaufforschung e.V.

Grafenberger Allee 100  
40237 Düsseldorf  
Telefon: +49 (0) 211 / 800 892-43  
Fax: +49 (0) 211 / 600 892-10  
E-Mail: [presse@dgk.org](mailto:presse@dgk.org)  
Web: [www.dgk.org](http://www.dgk.org)

### **Künstliche Intelligenz erkennt Myokardinfarkte im EKG zuverlässiger als Kardiologen**

**Prof. Dr. Hisaki Makimoto, Universitätsklinikum Düsseldorf**

Für diese Studie haben wir eine künstliche Intelligenz (KI) erschaffen und sie EKGs lernen lassen. Diese KI konnte in der EKG-Interpretation einen innerhalb der vergangenen 3 Monate abgelaufenen Myokardinfarkt zuverlässiger erkennen als ein Facharzt für Kardiologie.



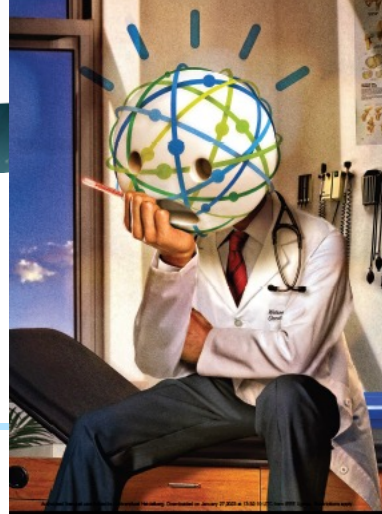
## 2. Sorge vor Schaden durch fehlerbehaftete KI

EXCLUSIVE

IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show



By [Casey Ross](#) and [Ike Swetlitz](#) July 25, 2018



STAT+

TechTarget

HEALTH  
IT ANALYTICS  
xelligent HEALTHCARE MEDIA

Population Health

Precision Medicine

Quality & Governance

Home

News

Features

Interview

Tools & Strategies

Focus on AI

### AI May Be More Prone to Errors in Image-Based Diagnoses Than Clinicians

New research indicates that AI may be more prone to making mistakes than humans in image-based medical diagnoses because of the features they use for analysis.

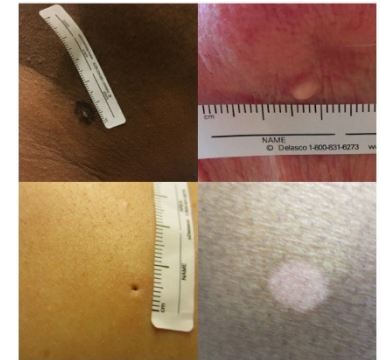


Bundesärztekammer

# Sorge vor Diskriminierung (in den Daten & Anwendung)

## Dermatologische State-of-the-art-KI und Bias

	Erfolgsrate (ROC AUC)	
	Hellhäutigen	Dunkelhäutigen
<b>ModelDerm</b>	0,64	0,55
<b>DeepDerm</b>	0,61	0,50
<b>HAM10000</b>	0,72	0,57



Roxana Daneshjou et al., Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci. Adv.* 8, eabq6147 (2022). DOI:10.1126/sciadv.abq6147

➔ KI machen (teilweise diskriminierende) Fehler

### **3. Was bedeutet KI aus ethischer Perspektive...**

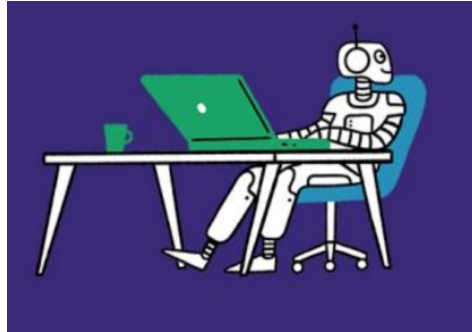
**... für die Interaktion von Ärztinnen und Ärzten mit der KI?**

**... für das Ärzt:in-Patient:innen-Verhältnis?**

**... für Trainings- und Validierungsdaten einer KI?**



# Interaktion von Ärzt:innen und KI



<https://ki-campus.org/>

## 3.1. Potential für bessere Qualität

### Zusammenarbeit: Entscheidungsunterstützung (Augmentation)

Vergleich 3 KI-Algorithmen für die Mammographie – Brustkrebsdetektion

- Eine der 3 Algorithmen besser als einige Radiolog:in
- Beste Performance: Radiolog:in plus bester KI-Algorithmus
- Fit für den prospektiven Einsatz

Salim et al 2020 JAMA Oncology

➔ **Gesicherte Qualität der KI**

➔ **Verbesserte Ergebnisse bei Zusammenarbeit von  
Ärzt:innen & KI**

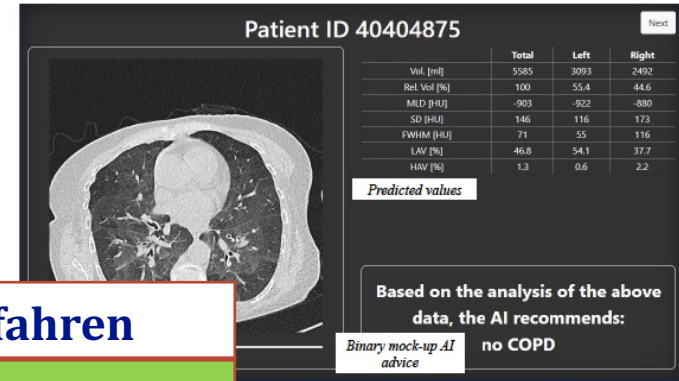
# 3.1. Risiko für schlechtere Qualität?

## Zusammenarbeit mit fehlerhafter KI

CT Bilder: Frage COPD – Info KI 90% zuverlässig – tatsächlich 50% Fehlerquote

47 Medizinstudierende mit keiner/wenig Erfahrung

12 Erfahrene Radiologinnen



		Unerfahren	Erfahren
		81%	83 %
KI System	Richtig	87%	83%
	Falsch	46%	67 %

# 3.1. Risiko für schlechtere Qualität!

## Zusammenarbeit mit fehlerhafter KI

---

### Interessante Befunde:

- Zielgruppe, die am meisten profitiert, ist am vulnerabelsten für Fehler
- Auch erfahrene Ärzt:Innen setzen sich nicht über falsche KI-Hinweise hinweg
- Erfahrene Ärzt:innen nehmen richtige Systemhinweise eher nicht an
- Erfahrene Ärzt:innen fallen unter ihr Niveau bei gelegentlich falschem Systemhinweis

**⇒ Robuste Validierung & Qualitätssicherung**

**⇒ Forderung nach Erklärbarkeit/ Explainability**

## 3.2. Interaktion– KI: Erklärbarkeit/Nachvollziehbarkeit



### Inferenz auf die bestmögliche Erklärung

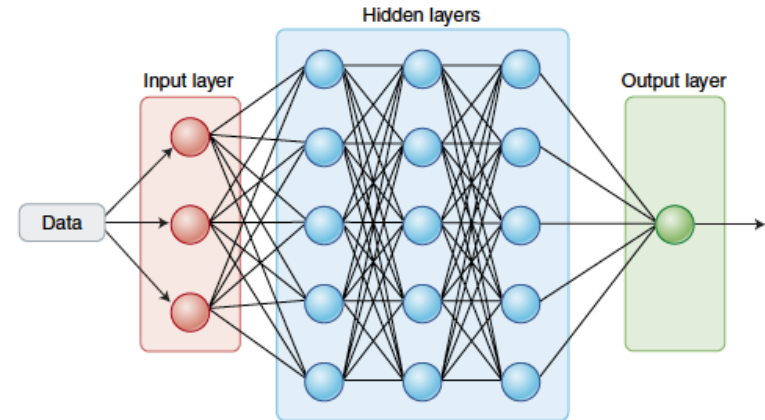
Ausgangspunkt: Befunde, Symptome und Indikatoren gewichten

Prüfung der Angemessenheit konkurrierender Erklärungshypothesen

→ Schlussfolgerung, die die Symptome am besten erklärt.

Dragulinescu, S. (2016). Interferenz to the best explanation  
Theoretical Medicine and Bioethics, 37(3), 211–232.

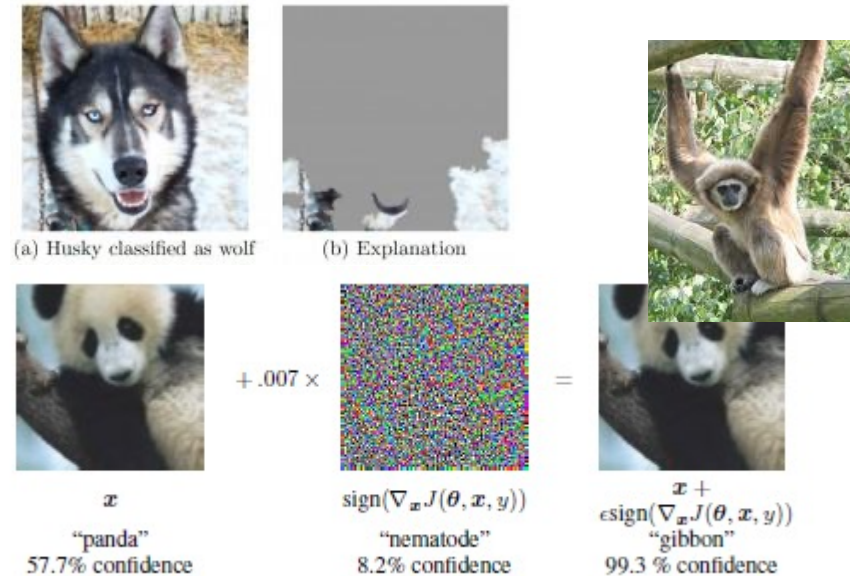
Aufgrund der Black Box ist nicht nachvollziehbar, wie eine Empfehlung zu Stande kommt



## 3.2. Interaktion- KI: Erklärbarkeit

- KI kann aus den falschen Gründen zum richtigen Ergebnis kommen (bspw. Husky / Derma-App)
- KI kann Bilder falsch klassifizieren

➔ Der Algorithmus soll seine Entscheidung begründen (**Interpretierbarkeit**)



Goodfellow ICLR 2015

## 3.2. Erklärbarkeit

### *Weiterführende Gedanken: Keine echte Interpretierbarkeit*

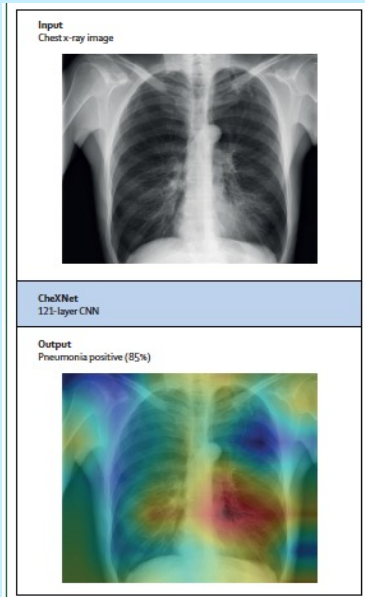


Figure 1: Heat map produced by a post-hoc explanation method for a deep learning model designed to detect pneumonia in chest x-rays. Brighter colours (red) indicate regions with higher levels of importance according to the deep neural network, and darker colours (blue) indicate regions with lower levels of importance. Reproduced with permission from Rajpurkar et al.'s CNN-convolutional neural network.

Häufig verwendet – Heat Maps – post-hoc-explainability (wie stark eine bestimmte Region zu einer Diagnose beigetragen hat)

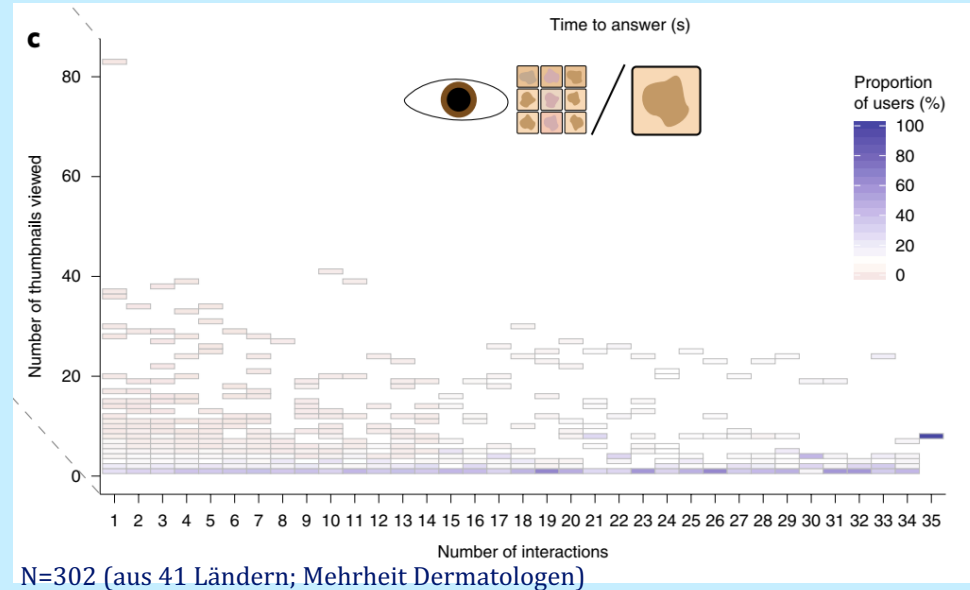
- Nur Annäherung – Heuristik und Performanz wiederum nicht klar
- Falsche Sicherheit / Konfirmations-Bias
- Ablenkung von der “Vogelperspektive“



Zeigt nur ein **was** kein **warum** (keine Begründung)  
(Ghassemi M, Oakden-Rayner L, Beam AL. Lancet Digit Health. 2021)

## 3.2. Erklärbarkeit

Weiterführende Gedanken: „Interpretability Fatigue“?

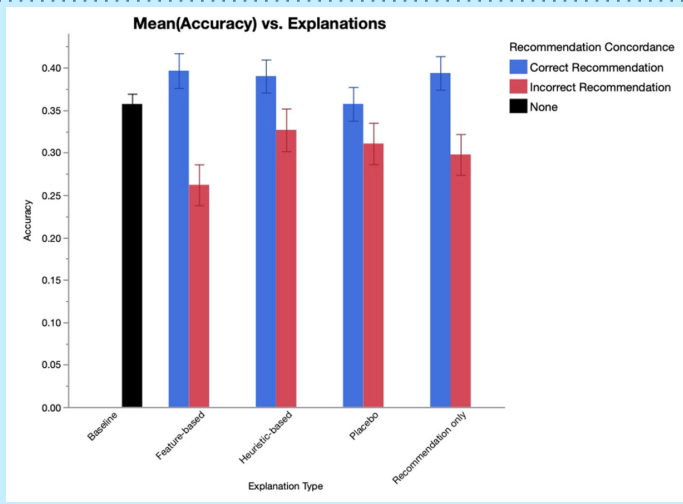


➡ Explainability-feature wurde zunehmend weniger genutzt



# 3.2. Erklärbarkeit

Weiterführende Gedanken: Interpretierbarkeit erhöht Automation Bias?



N=220 (US Kliniker mit Erfahrung)

➔ Interierbare KI führte zu schlechterer Performance bei fehlerhafter Behandlungsempfehlung als Black Box KI

## 3.2. Erklärbarkeit

*Weiterführende Gedanken: Ärztliche Letztverantwortung?*

---

Erklärbarkeit für Verantwortungsübernahme notwendig, aber:

- Echte Explainability (Angabe von Gründen) – derzeit - nicht möglich
- Explainability features nicht konsequent verwendet
- Explainability führt zu übermäßigem Verlassen auf KI
- Automatisierungsbias (mit und ohne Explainability)

Letztverantwortung



Letztentscheidung

**➔ Ärztliche Letztentscheidung ist gefordert**

## 3.3 Ärztliche Autonomie, Kontrolle und Letztentscheidung

---

### Autonomiebezogene Risiken:

- Automation bias
- Alert fatigue
- „self fulfilling prophecy“ / Sich selbst erfüllende Voraussage
- Überdiagnostik oder -behandlung



- ➔ **Präventive Maßnahmen und Schulungen im Umgang mit KI:**
- ➔ **Ärzte/Ärztinnen wichtige Rolle: Kontrolle & Letztentscheidung**

# 3.3. Ärztliche Autonomie: Verantwortung & Letztentscheidung

## Lokalisation von Verantwortung

- Arzt/Ärztin (Mikroebene) - Sorgfaltspflichten / ärztliche Kontrolle / Letztentscheidung
- Institution (Mesoebene). - Auswahl, Evaluation, Mitarbeiterschulung
- Gesetzgeber (Makroebene) – Zulassung / Fachgesellschaften

➔ Alle 3 Ebenen relevant bei der Sicherung von **Systemvertrauen**

# Das Ärzt:in-Patient:innen-Verhältnis



## 3.4. Patientenautonomie und gemeinsame Entscheidungsfindung

---

- Anspruch auf optimale Behandlung (ggf KI-unterstützt)
- Aufklärung und Einwilligung über KI-Einsatz  
--> insbesondere bei Neulandmethode
- Gefahr: „Datafizierung“ der Versorgung  
„Computer-Paternalismus“



- ➔ **Stärkung der gemeinsamen Entscheidungsfindung**
- ➔ **Große Bedeutung von Kommunikation und Transparenz**
- ➔ **Werte-sensitives Design von KI – Einbeziehung von Patientenpräferenzen in die Entwicklung von KI**

## 3.5 Das Ärzt:in-Patient:innen-Verhältnis

### *Medizin und wertebasierte Entscheidungen*

Medizinische  
Entscheidung



Schlussfolgerung aus  
biologischem Prozess

- Medizinische Entscheidungen sind **wertebasiert**
- Werte lassen sich oft erst im **Gespräch** explorieren
- konkrete **Lebenssituation** von PatientInnen im **biographischen Kontext**



Es kann in diesem Sinne **keine automatisierte  
Behandlungsentscheidung** durch KI geben



## 3.4 Ärztliche Erfahrungswissen als Plausibilitätskontrolle

---

- Mögliche Entwertung ärztlichen **Erfahrungswissens**
- „**tacit knowledge**“ und „**embodied knowledge**“ sind nicht abgebildet in KI-Trainingsdaten
- Verlernen/ **Verlust dieses Wissens** für zukünftige ÄrztInnen (EKG lesen, Bilder interpretieren, Abhören... )

➔ **KI bietet keinen Ersatz, sondern unterstützt die Entscheidung von ÄrztInnen**

➔ **Unbedingte Plausibilitätskontrolle ärztlicherseits**



# KI- Trainings und Validierungsdaten



## 3.5 KI-Trainings- & Validierungsdaten

### 3.5.1 Datenqualität

---

#### Daten essentiell als Trainingsmaterial

1. Algorithmen replizieren unsere Biases, Stereotypen und Vorurteile  
(Bsp: AI für die Triage von Bewerbern –Triage Algorithmen in der Medizin?)

#### → **Bereinigung der Daten von implizite Biases**

2. Selbst Datensätze mit einer hohen Qualität können einer Ungleichbehandlung Vorschub leisten  
(Bsp: Derma App geschult an hellhäutigen Patienten → Datenqualität und Aussagekraft für dunkelhäutige weniger valide)

#### → **Datengerechtigkeit, Repräsentativität und Qualität entscheidend**

## 3.6 KI-Trainings- & Validierungsdaten

### 3.5.2 Datenquellen

---

**Daten werden überwiegend von Patienten im Gesundheitssystem generiert**

1. Insbesondere bei personenbezogenen – oder beziehbaren Daten

→ **Informierte Zustimmung oder andere rechtliche Basis der Datennutzung**

2. Gesundheitsdaten aus dem solidarfinanzierten Gesundheitswesen als öffentliches Gut – Bereitstellung für private Unternehmen?

Bsp: Google Deep Mind – trainiert mit den Daten aus dem NHS – Powles et al Health Technol. 2017)

→ **Klar definierte Nutzungsregeln für „Trainingsdatensätze“  
in Public-Private-Zusammenarbeit**

# Aspekte einer vertrauenswürdige KI

---

1. Wahrung der Autonomie
2. Ärztliche Letztverantwortung
3. Erklärbarkeit (Explainability)
4. Ärzt:in-Patient:innen-Verhältnis
5. Hohe Datenqualität



## 3.4 Das Ärzt:in-Patient:innen-Verhältnis

### 3.4.2 Weiterführende Gedanken: Trustworthy AI oder Systemvertrauen?

- „Vertrauen“ Begriff aus **menschlicher** Praxis
  - Vertrauen in **rein probabilistische Prozesse** nicht möglich
  - KI ist anfällig für völlig unverständliche Fehler, die sich **nicht vorhersagen** lassen
- ➔ Vertrauen ist nur als **Systemvertrauen**, nicht als **Vertrauen in die KI** rational



Rathkopf u. Heinrichs. 2023.  
Cambridge Quarterly of Healthcare Ethics

## 4 Empfehlungen für Standards: ZEKO Empfehlung

### „Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz“

---

1. Die **Grenze zwischen Entscheidungsassistenz und automatisierter Entscheidung** darf nicht überschritten werden.
2. Zusammenwirkung von **CDSS und ärztlicher Expertise** vielversprechend.
3. **Fehleranfälligkeit: Zulassung & Plausibilitätsprüfung**
4. Wichtig ob Neulandmethode oder medizinischen Standard → **Neulandmethode erfordert Aufklärung**
5. Mitverantwortung für die **Qualitätssicherung** in Bezug auf die Patient:innenversorgung
6. **Stigmatisierung und Diskriminierung** durch Validierung der Modelle, Trainingsdaten und Ergebnisse entgegenwirken.
7. Erheblicher Forschungsbedarf besteht bei der Entwicklung von Maßnahmen zur **Erklärbarkeit und Zusammenarbeit von KI-basierten CDSS und ÄrztInnen**

# Vielen Dank !

---

