How to achieve responsible AI in (health)care?

AI Blind Spots in (health)care

Knowledge Centre Data & Society

Artificial Intelligence Flanders

AI Blind Spots in (health)care
2025, Knowledge Centre Data & Society
www.data-en-maatschappij.ai

AI BLIND SPOTS IN (HEALTH)CARE

This card set has been designed to get you thinking about how you can use AI **responsibly and inclusively** in (health)care. The cards will help you uncover your **blind spots**: ethical pitfalls such as bias in the data. By identifying these early on, you increase the chance of AI having a truly **positive impact** on the standard of care provided.

This card set focuses on AI applications that support (health)care processes, such as **diagnostics**. The Knowledge Centre Data & Society also has a separate card set on Generative AI. Use this one if you'd also like to consider the risks and challenges of GenAI.

WHAT DOES THE CARD SET CONTAIN?

Each card describes a single blind spot. Per individual card, you'll get a key question, the category or categories to which the blind spot belongs a short description, a concrete example and some reflective questions. As a set, these cards therefore help you **identify and address potential risks** of AI applications in (health)care.

If you'd like to develop further **action points** based on this exercise, we recommend using the Guidance Ethics Approach. This method focuses on the 'how' question: how can we develop and implement an AI application in a responsible and inclusive way? This method was developed by ECP – Platform for the Information Society.

FOUR CATEGORIES

APPLICATION

Things to watch out for when it comes to the design, training and development of AI applications.

USE

Considerations when using AI applications.

OUTCOMES

Potential risks related to the analysis, forecasts and recommendations, etc. provided by an AI application.

ORGANISATION & SOCIETY

The broader impact of AI on (health)care recipients and organisations, as well as on society as a whole.

GETTING STARTED WITH THE CARD SET

You can use the card set for **different purposes**:

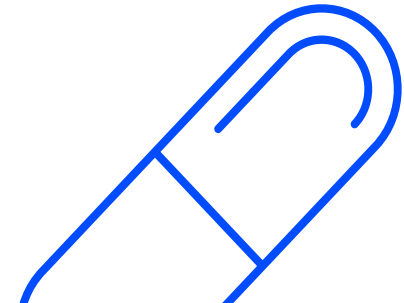
- to raise more awareness of the things to watch out for when using AI,
- as a guide for setting agreements for the use of AI (similar to a code of conduct),
- to identify risks during the development and implementation of an AI application,
- as a starting point for conversations about the potential impact of AI within your organisation.

The card set has been developed **for caregivers, innovation managers, AI professionals and researchers**. You can use it to reflect on blind spots and ethical dilemmas within your team. We also encourage you to involve (health)**care recipients, informal carers, policymakers or administrators** as this will ensure you consider a range of different, interdisciplinary perspectives. This is especially important if you want to develop a code of conduct or are thinking about using AI applications within your organisation.

On the Knowledge Centre Data & Society website you'll find **guidance** on how to use the card set in a workshop.

ACCESS TO DATA

“Do you have access to the data you need for the AI application?”



APPLICATION

DESCRIPTION

(Health)care data often contains sensitive information. It is therefore **highly protected and not freely accessible**. AI applications require a vast amount of data for training and use. So while an AI application may look good on paper, it might not work so well in practice.

EXAMPLE

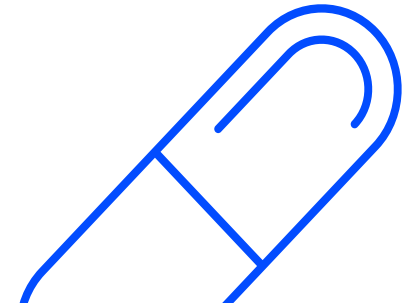
An AI application that can analyse **data about multiple sclerosis (MS) patient**s across Europe could lead to a major breakthrough in treatment. However, in many cases, the data is **not available** to train the AI application.

ASK YOURSELF THESE QUESTIONS

1. What **restrictions** apply when using the data?
2. What steps can you take to **use this data anyway** (e.g., through collaboration, an agreement, etc.)?
3. Can the various systems, devices and applications **exchange data**?

DATA MINIMISATION

“Which data is not needed for the AI application to work properly?”



APPLICATION

DESCRIPTION

Even if the data is available, not all of it needs to be processed by the AI application. This can have **unintended and unexpected consequences** for (health)care recipients, such as violation of their privacy, function creep (use for other purposes) and profiling, etc. Therefore, it is not only recommended but also a legal requirement to **collect and use as little data as possible**.

EXAMPLE

A hospital is carrying out an **experiment with diabetes patients** using a **lifestyle coaching app** created by a third party. The app collects a lot of personal data that is not needed for the experiment, including addresses, contact details and location information.

ASK YOURSELF THESE QUESTIONS

1. What data is being **collected**?
2. Is the collected data **necessary for the AI system to work properly**?
3. How long is the data **kept**, and is this necessary to ensure the AI application works properly?
4. Is the collection and use of the data (still) **justified**?

DATA PROTECTION & PRIVACY

“Do you handle personal data correctly and in a way that respects privacy?”

APPLICATION

DESCRIPTION

(Health)care data is often **personal** and legally protected under GDPR (General Data Protection Regulation) in Europe. So when using this data for an AI application, you must **proceed with caution**, paying attention to privacy, data control, and the grounds for processing.

EXAMPLE

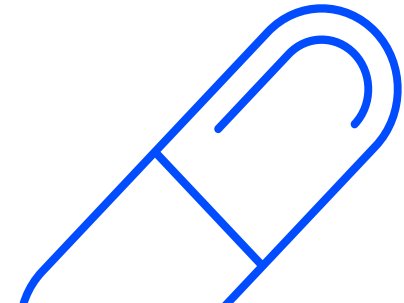
A hospital wants to use the **data in its electronic patient records** to train an AI application to predict the risk of heart failure. The data has been **pseudonymised** so that the hospital can still inform patients who are at risk. However, the patients are surprised to hear this as the hospital did not ask for their **consent**.

ASK YOURSELF THESE QUESTIONS

1. Are personal data processed in accordance with the **GDPR** and other relevant data regulations?
2. Is there a **data management policy**, including a data breach procedure?
3. Are individuals **properly informed** when their consent for data use is requested?
4. How can individuals **withdraw their consent** for their data to be used?

TRANSPARENCY & EXPLAINABILITY

“Do you know how the AI application was created and how it works?”



APPLICATION

DESCRIPTION

To properly **assess an AI application's output**, you need **clear information** about who developed it, what you can use it for, which data was used and which choices were made during development, how it reasons, etc.

EXAMPLE

When a **home care organisation** receives multiple **emergency calls**, an AI application indicates where a caregiver should go first. The application **prioritises automatically**, but it's unclear exactly how it makes this decision. This raises questions for both the caregivers and recipients.

ASK YOURSELF THESE QUESTIONS

1. Do you have information about the **design, processing, intended purpose, data used**, etc. of the AI application?
2. Can you find out how the AI application makes its **decisions**?
3. Has the **decision-making process** and intended purpose of the AI application been explained in a way that **everyone understands**?

GOAL

“Why do you want to use the AI application?”

USE

DESCRIPTION

The potential of AI can be so exciting that you might forget exactly **what your goal is**. Perhaps there are other, **more efficient ways** to achieve your objective?

EXAMPLE

A home care organisation installs **cameras** in an individual's home. These cameras use AI to detect **unusual events**, including falls. It seems promising, but the technology gives a lot of **false-positives**, and the individual feels like they're being constantly monitored. After a while, the organisation switches back to a **sensor that the individual can wear**. If they have a fall, they can activate it themselves by pressing a button.

ASK YOURSELF THESE QUESTIONS

1. Is the AI application the **best way to achieve your goal**?
2. Do the **benefits** of the AI application **outweigh the risks and changes** for the organisation and the individual?
3. What are the consequences of the AI application **not being (adequately) used**?

AUTONOMY

“Can people choose whether they want to use the AI application?”

USE

DESCRIPTION

An AI application can support a better standard of care, but not everyone is open to using AI for this purpose. **Give people the choice and provide alternatives**.

EXAMPLE

An AI application uses body sensors to suggest **personalised physiotherapy exercises** based on heart rate, physio needs, and other physical parameters such as muscle and fat mass. Those who prefer not to use the AI application still get **advice from the physiotherapist** overseeing their care.

ASK YOURSELF THESE QUESTIONS

1. Are there **alternatives** if someone doesn't want to use AI?
2. Are individuals given **multiple treatment options**?
3. To what extent do people truly have the **option to say no** to the use of AI without this significantly compromising the level of care they receive?

MONITORING AI APPLICATIONS

“Do you regularly monitor whether an AI application is still valid and being used effectively?”

USE

MONITORING AI APPLICATIONS

DESCRIPTION

The quality of an AI application can decline over time because the data used to train it becomes outdated, the supplier's policy has changed, or because employees are less likely to follow usage agreements. Ensure you have a **clear overview of all the AI applications** used by employees within your organisation and regularly **assess** them (and their use).

EXAMPLE

A home care AI application that uses **sensors to track movements and falls** is updated. As part of that **update**, **real-time location data** is now also being stored. This 'hidden' change in the way personal data is stored significantly increases the risk of **privacy breaches**. It is only discovered during an audit of the application several months after the update.

ASK YOURSELF THESE QUESTIONS

1. Do you have an **overview** of the AI applications being used within the organisation?
2. Do you regularly review the **quality** of the AI applications and whether they're **still required**?
3. How does the organisation **monitor the use** of the applications?

USE

HUMAN OVERSIGHT & ACCOUNTABILITY

“Is it clear who is responsible for any errors in the AI output?”

ORGANISATION & SOCIETY

USE

DESCRIPTION

AI applications can be very powerful, but also make mistakes. So it's important to regularly **check their output**. Before launching an AI application, consider how you will do this. Clearly define **who is responsible** for decisions made with the help of AI and who needs to take action in the event of errors.

EXAMPLE

An AI application advises a **general practitioner** on the **dosage for a medication**. But it suggests the wrong dosage for a patient who is taking multiple medications. The general practitioner didn't have time to check this.

ASK YOURSELF THESE QUESTIONS

1. Is it clear who will **monitor the output** of an AI application, and how?
2. Is the AI application **designed** so you have to confirm the output if it is critical or questionable?
3. Is it clear **who is liable** if the output of the AI application leads or contributes to a decline in the quality of care?

VALIDITY

“Is the output of the AI application relevant and representative for the intended goal?”

OUTCOMES

DESCRIPTION

When you're trying to draw conclusions about a specific topic using large amounts of data, ask yourself if you've **actually measured what you intended**. Does the AI application accurately represent the issue?

EXAMPLE

A mental health centre is **monitoring depression using wearables** and an **AI application that analyses sleep patterns**. Sleep can be negatively affected by depression, but also by many other factors. Therefore, the AI application doesn't provide **any relevant added value**.

ASK YOURSELF THESE QUESTIONS

1. Is there a risk that the AI application will produce something **close but unrelated to the topic**?
2. Have you analysed which data you need to **quantify the topic in a representative way**?
3. Is the result of the AI application based on **relevant characteristics** or only loosely related indications?

(INDIRECT) BIAS & DISCRIMINATION

“Can the AI application (indirectly) lead to bias and discrimination?”

OUTCOMES

DESCRIPTION

AI models can indirectly **reinforce biases** when the training data **lack diversity** or when certain **information**, for example about certain population groups or diseases, is **under- or overrepresented**. Even data such as postal codes can be indirectly discriminatory, requiring extra care during the training of the AI application.

EXAMPLE

A dietitian recommends an **AI application for weight-loss coaching**. However, it is **mainly trained on Western European diets** and cannot accurately assess the nutritional value of meals from other cultures. As a result, some users consistently receive incorrect advice.

ASK YOURSELF THESE QUESTIONS

1. Can you find out **which data** the application was trained on and which data was excluded?
2. Is the data the application is based on **accurate and representative of the context** (time, space, population, etc.) in which it will be used?
3. Is there a **strategy to detect and address inaccurate or missing data** that could lead to bias?

INCLUSIVE SERVICES

“With more AI-driven services, can we still rely on high-quality care for everyone?”

ORGANISATION & SOCIETY

DESCRIPTION

More technology- or AI-driven services can be a disadvantage for **those who are less digital**. They may have limited access to technology, fewer digital skills, or think less critically about the outcomes of digital or AI applications. This can mean they receive substandard care.

EXAMPLE

A hospital uses an AI application to **automatically schedule follow-up appointments** based on persons' medical records and doctors' availability. Patients receive an **email** with the proposed appointment and can confirm, reschedule or cancel it. But many people have trouble opening or understanding these emails, so they often miss their appointments or arrive late.

ASK YOURSELF THESE QUESTIONS

1. Who might be **negatively affected** by the AI application?
2. How will you identify these **people's wishes and needs** regarding the AI application (e.g., through a survey, workshops, etc.)?
3. What are the **alternative services** if they are unable or unwilling to use the AI application?

RAISING AWARENESS & SUPPORT

“How do you get caregivers on board with the changes that come with AI applications?”

ORGANISATION & SOCIETY

DESCRIPTION

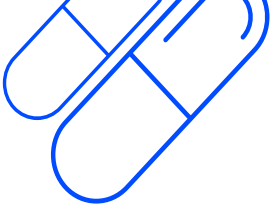
A new AI application can lead to new processes, extra responsibilities and additional training. Certain roles can also become obsolete. This can all **create unrest** in the workplace, so it's important to involve all those impacted by the AI application. **Gather their buy-in and input** during the implementation and monitoring of the application.

EXAMPLE

A hospital evaluates an application that optimises CT scans and finds that its benefits are valued mostly by the more innovation-minded radiologists. But the majority are more sceptical, mainly because it was forced on them and they had **no say in the choice of application**. Despite its proven quality, it is therefore rarely used.

ASK YOURSELF THESE QUESTIONS

1. How will you **raise awareness and communicate** the purpose and impact of the AI application to caregivers?
2. Is there a **plan to train them** on how to correctly use and interpret the application?
3. What will you do to **address reluctance and scepticism** about the application?



ECOLOGICAL SUSTAINABILITY

“Do you consider the ecological impact of the AI application during use?”

DESCRIPTION

Training and using AI applications can **consume significant amounts of energy, water and scarce resources**. The ecological impact can vary significantly between AI applications. Try to understand the impact and factor it in when implementing and using an AI application.

EXAMPLE

A home care organisation uses **wearables that continuously send heart rate and blood pressure data to the cloud for AI analysis**. The constant data flow keeps servers running 24/7, resulting in significant energy consumption and higher CO2 emissions.

ASK YOURSELF THESE QUESTIONS

1. Do you know the **ecological impact** of the AI application?
2. Could you choose a **more sustainable alternative** or change the AI application to reduce the ecological impact?
3. How can you **make users aware of this impact** and, if necessary, limit the application's use?

JOKER CARD

.....

.....

.....

.....?

APPLICATION

JOKER CARD

DESCRIPTION

.....
.....
.....
.....
.....

EXAMPLE

.....
.....
.....
.....
.....

ASK YOURSELF THESE QUESTIONS

1. .....
   .....
   .....
2. .....
   .....
   .....
3. .....
   .....
   .....

JOKER CARD

.....

.....

.....

.....?

USE

JOKER CARD

DESCRIPTION

.....
.....
.....
.....
.....

EXAMPLE

.....
.....
.....
.....
.....

ASK YOURSELF THESE QUESTIONS

1. .....
   .....
   .....
2. .....
   .....
   .....
3. .....
   .....
   .....

USE

JOKER CARD

.....

.....

.....

.....?

OUTCOMES

JOKER CARD

DESCRIPTION

.....
.....
.....
.....
.....

EXAMPLE

.....
.....
.....
.....
.....

ASK YOURSELF THESE QUESTIONS

1. .....
   .....
   .....
2. .....
   .....
   .....
3. .....
   .....
   .....

OUTCOMES

JOKER CARD

.....

.....

.....

.....?

ORGANISATION & SOCIETY

JOKER CARD

DESCRIPTION

.....
.....
.....
.....
.....

EXAMPLE

.....
.....
.....
.....
.....

ASK YOURSELF THESE QUESTIONS

1. .....
   .....
   .....
2. .....
   .....
   .....
3. .....
   .....
   .....

ORGANISATION & SOCIETY

COLOFON

The 'AI Blind Spots in (health)care' is a tool developed by the Knowledge Centre Data & Society (KCDS).

The content of this card set is based on our own insights as well as:

- our collaborations with imec-SMIT (Vrije Universiteit Brussel) and LiCalab (Thomas More University of Applied Sciences),
- takeaways from workshops facilitated by the KCDS on AI use in (health)care with experts and professionals.

This tool is based on the AI Blindspot card set by Ania Calderon, Dan Taber, Hong Qu and Jeff Wen, developed during the Berkman Klein Center and MIT Media Lab's 2019 Assembly programme.

The card set was created with the help of ChatGPT 4o to improve phrasing and generate examples of blind spots.

Publication year: 2025

Responsible publisher:
Pieter Duysburgh
Knowledge Centre Data & Society
Pleinlaan 2
1050 Brussels
info@data-en-maatschappij.ai

Knowledge Centre Data & Society

Artificial Intelligence Flanders

Knowledge Centre
Data & Society

Artificial
Intelligence
Flanders