

Balancing Ethics and Progress: An Engaging Approach to Reporting on AI in Healthcare

By Anne Ringgaard

Fellow at The Constructive Institute, 2022 - 2023

Journalist, Videnskab.dk

Introduction

Artificial Intelligence (AI) is gradually being applied to the healthcare sector: Globally billions of Dollars are being invested in machine learning models that are trained to assist physicians' decisions about diagnosis and treatment. Some say that the use of the data driven technologies has the potential to revolutionise the sector and solve the pressing problems of staff shortages and overload. In the best case, the application of clinical AI will benefit patients, increase productivity, reduce waiting lists, and release the workload of doctors and nurses.

However, the use of AI in healthcare also comes with significant risks. The use of the data driven technologies raises several ethical questions that should be discussed in public in order to ensure an acceptable and just application.

Constructive journalism focuses on solutions, nuances, and dialogue. It is an obvious approach to provide people with accurate knowledge and lay the foundation for informed conversations about AI ethics and regulations. This report is about the need for constructive journalism about AI ethics in healthcare, but it can also inspire a responsible coverage of the use of data driven technologies in other public and private sectors.

In the first part of the report, some of the risks and ethical dilemmas related to the use of AI in healthcare will be reviewed.

The second part of the report considers why and how journalists can report on AI and ethics in a constructive, critical, and engaging way.

Why should journalists report on AI in healthcare?

Every one of us is in contact with the healthcare system several times throughout our lives. Sometimes the encounter is a matter of life or death. Diseases, treatments, patient rights, pandemic responses, vaccines, and other health related issues matter to anyone everywhere in the world.

In Denmark, the population values healthcare as one of the most important topics for politicians to deal with, as shown by one poll after another before elections for the parliament. The health sector is also one of the largest expenditure items on the public budget.

Health has news value for national and local media. Journalists frequently cover the topic through investigative stories, uncovering flaws in the healthcare system and violations of patient rights. Additionally, breakthroughs in medical science are popular news, and stories about personal health generate significant interest among the audience of any media outlet.

The direct way to receive both attention and engagement from various stakeholders is by covering controversial bioethical topics such as euthanasia, prioritisation of medicine, abortion, gender change, or vaccines. Bioethics can truly ignite debates among politicians, citizens, and researchers. At the time of writing, ethical questions related to the use of AI in healthcare urgently needs to be dealt with.

Machine learning models are already essential tools in gaining new knowledge about human biology and health by analysing a continuously growing amount of data about genes, proteins, microbiomes, and public health. The AI wave is also expected to hit healthcare: After decades of research and training, algorithms are becoming ready to be utilised by doctors for predicting diagnoses, treatment plans, communicating with patients, and analysing scanning images and biological samples.

Significant investments in AI are being made globally to enhance the efficiency of the healthcare system and alleviate some of the workload from doctors and nurses. However, the use of data-driven technology raises several bioethical issues concerning patients' data security, their right to informed consent, accountability, and equitable access to treatment.

To ensure a morally just implementation of AI in hospitals, laboratories, and health clinics, it is imperative to provide people with accurate information and foster a nuanced understanding through open public dialogues that encompass diverse voices from all levels of society. In this regard, a free press plays a significant role. In fact, active engagement of the mass media is essential in addressing bioethical questions, as indicated in the book 'Bioethics and Journalism' published by the UNESCO:

“Without the active engagement of the media, not only would existing and future consumers of biomedical technologies have nowhere to obtain basic information, they would also be deprived of the opportunity to make truly informed decisions on critical, life-changing issues,” argue the authors, Pavel Tischenko and Boris Yudin.

“As such, it is imperative to maintain the bond between bioethics and the mass media if they are to achieve the common goal of helping the average citizen navigate the convoluted and often frightening world of newly emerging biomedical technologies,” they continue.

Artificial intelligence is likely to play a profound role in the future healthcare system. Thus, journalists should cover the development in a constructive manner, shedding light not only on the technology’s potentials but also on the ethical dilemmas, uncertainties, and limitations associated with its implementation. Before returning to recommendations for a constructive media coverage, we will look at the potentials, risks, and ethical dilemmas related to the use of AI in healthcare.

1. AI: A life jacket for a strained healthcare system?

Healthcare systems all over the world are challenged. There are too few physicians for too many patients. Waiting lists grow and nurses leave the sinking ship. Hospitals fail to provide patients with the care they are entitled to, according to one investigative media story after another.

Underneath the scandals in the healthcare system is an ageing population: young people have fewer children. The old ones live longer. New technologies lead to ever more efficient medical drugs meaning that previously untreatable rare genetic diseases can now be cured. People live many years with chronic diseases. The expenses grow and the demand for doctors and nurses increases.

Add to this an unequal distribution of healthcare providers across different regions and countries: in rural areas and low-income countries, vulnerable populations do not have access to timely and quality care.

Data driven AI technology could be used to increase the productivity in the healthcare sector, improve the quality of treatments, and relieve some of the work pressure from the employees.

“Unlike humans who get tired, have bad days, may get emotional, sleep deprived, or distracted, machines are steady, can work 24/7 without vacations, and don’t complain,” the American MD and Professor Erik Topol writes in his book *Deep Medicine*.

Globally, billions are invested in the development of new AI tools that are trained to assist doctors in predicting diagnoses, treatment plans, and remote monitoring. The market for AI in healthcare is

projected to grow from 14.6 billion US Dollars in 2023 to 102 billion by 2028, according to a [Global AI in Healthcare Market Forecast](#) from the research firm Markets and Markets.

In Denmark, the Government has invested almost 2 billion DKK in pilot projects with AI systems in the public sector, primarily in healthcare.

“A progression will happen from now on and far out in the future,” says MD and Ph.D. Benjamin Sørensen, who is the leader of the Centre for Clinical AI at The University of Southern Denmark

AI models based on machine learning can find patterns in endless amounts of data from medical records, gene banks, laboratories, scientific journals, scan images, and patient records. No human will ever be able to process all that data.

Large language models can be used for remote patient monitoring, medical management, mental health support, medical writing and documentation, according to the article [‘Revolutionizing Healthcare: The Top 14 Uses Of ChatGPT In Medicine and Wellness’](#) published in Forbes. In the US, doctors have also embraced the chatbots for another purpose: They use them as assistants to communicate with their patients in an emphatic way, according to [an article](#) in the New York Times.

In the Capital Region of Denmark among other places, AI [has been used](#) for some years to assist radiologists in analysing X-ray-images in mammography screening.

However, AI technologies are still far from living up to the expected potential as a life jacket for the healthcare sector. Technical and ethical challenges need to be dealt with before the data driven technologies can be broadly implemented, according to [a scientific review](#) published in the journal Nature Medicine in January 2022.

“As these challenges are addressed, AI’s potential may be realised, making healthcare more accurate, efficient and accessible for patients worldwide,” the review concludes.

News outlets often imply that specific AI systems that have performed impressively in controlled experiments can be seamlessly applied to a clinical practice and achieve the same results. But this is not the case. In order to avoid hype and exaggerations, journalists should be aware of the risks and technical limitations of clinical AI. These will be explained in the next chapter.

2. Potential hazards: The risks of using AI in healthcare

Even though AI has the potential to benefit patients and relieve the work pressure on physicians and nurses, the use of data driven technology is not without risks, and it cannot be seamlessly implemented in the healthcare sector. As mentioned, several technical and ethical issues need to be addressed.

Firstly, the best performing and most promising AI systems have a black box problem: Harmful dysfunctions in the systems might be hidden to the human eye. The black box systems are machine learning models built up by so-called deep neural networks that process information through artificial neurons arranged in several layers.

Through training and without human supervision, the models learn to make predictions about diagnoses or treatments based on millions of parameters in data from patient records, biological samples, scanning images, etc. The inner workings of the neural networks are so complex that not even the developers can decipher which parameters the systems base their predictions on. Over time, explaining how a deep learning model works becomes more and more complicated because it continuously refines its internal parameters as it encounters new data.

Potentially, the deep learning models can be of great benefit to patients and increase efficiency in the healthcare system: Unlike humans, they can make sense of massive amounts of health data stored in gene banks, patient records, public registers, and other sources. With this foundation, when all factors align, they could offer more accurate diagnoses and personalized treatment plans for each individual patient. However, deep learning models also pose significant risks due to their lack of transparency. In the following paragraphs, four risk factors related to the use of black box AI will be explained.

2.1. Biased data can lead to flawed predictions

One risk of black box AI is related to the datasets they are trained on. Health data are often biased, meaning that they do not represent the diversity of the patients in the real world. If AI models are trained on biased data, there is a significant risk of flawed predictions.

Bias can best be explained by an example: envision a deep learning model designed to predict diagnoses, trained solely on data from Scandinavians between 20 and 60 years old. If such a model is

used in a hospital, it will have problems predicting diagnoses or treatments for patients older than 60 who more often than younger people have multiple health problems. It will also face difficulties when used on people from other parts of the world with different genetics and lifestyles.

A deep learning model will always come up with a suggestion for a diagnosis if that is what it has been designed to do. But the accuracy of its prediction relies on the availability of comprehensive and diverse training data that encompasses various demographic, socioeconomic, and clinical factors. Without such data, the model may struggle to generalize well, and this could result in less reliable diagnoses when applied to new cases.

It is a well-known fact that certain population groups, such as ethnic minorities, women, older people, and individuals with multiple health conditions, are and have historically been underrepresented in medical trials. If AI systems are trained on data from these trials, the skewed data can lead to biased predictions and inaccurate outcomes for the underrepresented patients.

In addition to this, some countries have collected and stored much more health data about the population than others store. For instance, Denmark and other Scandinavian countries have recorded more information about their citizens' health, doctor visits, and so forth in public registers than any other country. Hence, researchers have access to extensive information about the Scandinavians. In low-income countries, on the other hand, the availability of such data is scarce. As a result, certain ethnic groups are often overrepresented in the datasets, clinical AI models are trained on.

The examples mentioned are just a few instances of skewed health data. There are several other factors contributing to biases. The main point of mentioning these examples is to illustrate that biases can impact the predictions made by clinical AI tools.

Efforts are being made to improve diversity in datasets. However, it is unlikely that biases in health data will be completely eradicated. Flawed results stemming from biases in data are particularly problematic in black box AI models since nobody can account for the parameters the predictions of the models are based on.

2.2. Predictions based on spurious correlations

There is also a risk that predictions of black box AI are based on spurious correlations. Spurious correlation refers to a situation where two variables in the data seem to be related or correlated, but

their connection is not meaningful or causative. It's a case of coincidence rather than a true cause-and-effect relationship.

A classic example of a spurious correlation is the case where the number of ice cream sales and the number of drownings at the beach increase at the same time during the summer. It might seem like more ice cream sales somehow lead to more drowning, but this is an example of a spurious correlation: the increase in both variables is likely due to a third factor, which is the hot weather during the summer. Hot weather can lead to more people buying ice cream and more people going to the beach and swimming, thus increasing the risk of drowning.

There is also a statistical correlation between baldness and cardiovascular disease: Bald men more often have cardiovascular diseases than men with hair. But this correlation is also spurious and not indicative of a direct causal relationship: baldness does not cause heart problems or vice versa.

Unlike human experts in statistics, AI models are not capable of making the thorough analysis needed to distinguish between spurious correlations and causal relationships in datasets. This means that the models risk basing their predictions on spurious correlations. This can lead to inaccurate or unreliable results and, in the worst cases scenario, insufficient or harmful treatments of patients.

In black box AI models, it is particularly challenging to identify the reasons behind the models' predictions, including whether the correlations they finds are spurious.

Researchers are working on improving data quality and developing methods to make black box models more transparent and reliable in order to address the problem of spurious correlations among others. However, the problem has not been solved and it is still an open question if it is possible to make the inner workings of the future AI systems more transparent and explainable.

2.3. AI systems are vulnerable to cyber-attacks

Machine learning systems are computer systems and, therefore, subject to the risk of cyber-attacks.

One potential way to hack an AI system is by manipulating the data entering the AI system so that it will deliver a wrong result. If such an attack happens, patients may receive a false diagnosis, potentially leading to unnecessary or insufficient treatment.

These so-called input attacks are challenging because, unlike traditional cyber-attacks on computer systems, they are not bugs in the computer code that are easy to detect and fix. Input attacks only require subtle changes in the input data and can be especially difficult to detect and address when targeting unexplainable black box systems.

2.4. Physicians may over-rely on predictions made by AI

In healthcare, AI is meant to be used as decision support. This means that the suggestions and predictions of an AI system should always be assessed by human medical specialists who have the final say about diagnoses or medical treatment. In fact, patients and other citizens in the EU have a juridical right to obtain an evaluation from a human specialist if they are affected by a decision made by an AI system, according to the General Data Protection Regulation (GDPR).

However, it is well known that humans – also medical experts - sometimes tend to over-accept computer outputs and that this can lead to errors of over-reliance. If an AI decision support system performs better than medical specialists in tests, physicians might start to doubt their own professional judgment when they use it. If so, there is a risk that they reverse their own correct decisions if the AI suggests something else. Such errors are referred to as automation bias.

Research shows that physicians are more likely to over rely on predictions about diagnoses made by an AI if they are insecure about their own diagnostics or if they are young and inexperienced.

Deskilling is another risk that comes with automation: if doctors and other healthcare professionals rely too much on decisions made by AI, they might lose their own ability to make professional judgements. Especially those who are new to the profession and lack experience might not properly train and develop their medical skills. As a result, in the worst-case scenario, automation could lead to less skilled human medical experts.

3. AI and the four pillars of medical ethics

Deep learning models and other AI systems hold the potential to improve healthcare, enabling more precise diagnoses and boosting productivity within the sector. Yet, addressing challenges such as inexplicability of the systems' inner workings, poor data quality, the risk of cyber-attacks, and concerns about potential deskilling of physicians remain ongoing discussions among academics and clinicians.

If these issues remain unresolved, the integration of AI in healthcare could potentially conflict with bioethical principles such as the four fundamental pillars of medical ethics. These four pillars serve as guidelines for doctors and healthcare professionals worldwide, helping them make morally justified decisions in patient care.

The four pillars of medical ethics:

- **Respect for Autonomy:** Patients have the right to make their own decisions about their health and treatment. Doctors should provide their patients with substantial and comprehensible information so that they are able to make informed choices and give informed consent.
- **Beneficence:** This principle focuses on doing good and acting in the best interest of the patient. Doctors should always seek to improve the patients' health and well-being and provide treatments that benefit them.
- **Nonmaleficence:** Doctors and healthcare professionals have the duty to do no harm. They should always prioritize the safety and well-being of the patients.
- **Justice:** This principle is about fairness and treating patients equally. It involves distributing healthcare resources fairly, avoiding discrimination, and ensuring everyone has access to appropriate medical care.

In the following paragraphs, we will consider how doctors' use of AI could potentially conflict with universal ethical principles, such as the four pillars of medical ethics.

4. Ethical questions related to clinical AI

As described above, the use of AI in healthcare can be beneficial but also problematic as it comes with risks of biases, deskilling, and cyber-attacks.

Overall, this raises important ethical questions such as:

- How many risks are we willing to accept to get the benefits of AI in healthcare?

- At what point are the risks significant enough to outweigh the benefits?

This chapter reviews several ethical dilemmas that arise with the use of AI in healthcare.

4.1. Ethical questions related to patients' autonomy

Physicians and healthcare professionals are obligated to give sufficient and comprehensible information about the benefits and the risks of medical treatments so that patients can make informed decisions about their own health. The use of black box AI leads to dilemmas about informed consent:

- Should physicians inform patients about the risks of flaws due to biases in training data and predictions based on spurious correlations?
- How much explanation of the inner workings of AI systems do patient need in order to be able to give informed consent? How can physicians provide such explanations if they themselves do not understand how AI systems come up with the results they do?
- Should physicians be obliged to disclose the risk of cyber-attacks to the patients?

4.2. Ethical questions related to justice

According to bioethical principles, healthcare should be equally and justly distributed, ensuring that everyone has access to the best available treatment. Remote monitoring and treatment through AI can provide easier access to sufficient and specialized healthcare, potentially reducing health inequalities.

However, the use of AI in healthcare could also perpetuate health disparities because patients from underrepresented population groups or marginalized communities may be disproportionately affected by biases in data used to train the machine learning models.

Furthermore, some population groups have more digital literacy than others. Older people, poor, or less educated individuals might have less access to the benefits of digital solutions such as remote monitoring and treatment simply because they do not have the skills to operate the systems, or they might not have internet connection or the right equipment.

There is also an imminent risk that AI will be of greater benefit to people from high-income countries than populations of low-income countries. This is because the populations of high-income countries are better represented in the datasets used to train the clinical AI systems. Furthermore, high-income countries have better digital infrastructure than low-income countries.

- Should we accept, that AI might perpetuate inequalities in health? If not, how do we ensure that all population groups benefit equally from the technologies?
- Are we willing to accept the risk of biases in data, knowing that they might be particularly damaging to minorities and marginalized groups that already lag when it comes to health? This question should be held up against the fact that decisions made by human physicians are also biased.
- Is it acceptable that some population groups might miss out on the benefits of remote healthcare due to a lack of digital literacy or poor digital infrastructure?

4.3. Ethical questions related to beneficence

As mentioned above, patients in the EU have the right to get decisions made by an AI system assessed by a human medical specialist. However, in the future, AI might perform better and more accurately in predicting diagnosis and treatments than specialised physicians. Furthermore, with the use of AI, physicians might over-rely on computers and be deskilled. This means that patients might be better off with a second opinion from another AI system than with an assessment from a human expert.

- Should patients have the right to choose to get a second opinion from another AI model instead of a human assessment if this could lead to a better outcome?
- Should all patients have the right to be offered an assessment from a relevant AI system if it exists?

4.4. Ethical questions related to nonmaleficence

As described previously, the inner workings of deep learning systems are hidden to the human eye. Hence, the AI predictions might unknowingly be based on spurious correlations or biases in data. In

the worst-case scenario, this can cause harm to the patient. However, predictions made by the so-called black box systems also have the potential to benefit the patients better than decisions made by doctors alone.

- In the context of black box AI, how can physicians effectively assess whether the benefits outweigh the risks when they lack insight into the potential flaws in the system's predictions?
- Should black box AI be used at all knowing that the harms of using them might be hidden?

4.5. Regulations might have blind spots

AI systems designed to be used in the healthcare industry are regulated. In the EU, they are subject to the General Data Protection Regulation (GDPR) and the Medical Devices Regulation. Furthermore, the EU has proposed an AI act with the purpose of regulating the use of data driven technologies. The act divides AI systems in three categories depending on their level of risks:

- Unacceptable risk
- High risk
- Limited risk

Clinical AI to be used in healthcare are placed in the high-risk category. This means that the systems will be assessed before being put on the market and throughout their lifecycle.

However, the ethical dilemmas listed above are not completely solved by the regulations. Furthermore, the regulations potentially have blind spots and unexpected risks might emerge as more AI systems are taken into use.

Several other ethical questions are related to AI in healthcare than the ones listed above. For instance, AI could affect the relationship between healthcare professionals and their patients in either beneficial or harmful ways: What consequences does the use of AI have for the principle of patient centred treatment?

Questions about accountability are also necessary to address: Who is responsible if a clinical AI system makes mistakes? Can AI be held responsible? The developers? The hospital or the doctor who uses the system?

Constructive journalism about AI as a solution to the healthcare crisis should take these risks and unsolved ethical questions into consideration. By doing so, the media can provide audiences with nuanced stories, enabling them to make informed decisions and actively participate in the important dialogue about regulation and use of AI.

In the second part of this report, we will go through some recommendations for nuanced news journalism about AI.

5. How to report constructively about AI

“Flocks of self-proclaimed tech experts incessantly chatter about Artificial Intelligence and robots in a cacophony noise that drowns out any informed discussion about the role of technology in a democratic society.”

This is how Professor Anne Gerdes expressed her irritation with the Danish media coverage of AI in an opinion piece published in February 2023 in the daily newspaper Politiken following the release of the large language model ChatGPT.

Anne Gerdes studies AI ethics in healthcare. She was irritated with statements in media outlets insinuating that the chatbot is conscious and has intelligence like a human. Those kinds of statements contribute to AI hype, she elaborated over the phone. The hype can have unfortunate consequences for the regulation and implementation of the data driven technology.

“When such nonsense is announced, it influences the public debate. Politicians who must make important decisions about how we should use AI in healthcare and in the public sector in general risk getting a wrong impression of what the systems can and cannot do. If the press sounds the alarm every time an AI lets out a fart, we will end up with a skewed public debate,” said Anne Gerdes, a professor at the Institute of Media, Design, Learning, and Cognition at The University of Southern Denmark.

In this chapter, we will look at some research-based recommendations for responsible AI-reporting.

5.1. Do not spread false fear and technochauvinism

In general, journalism about AI tends to be oriented towards the future, research shows.

Sometimes media outlets pass on speculations about a dystopian future in which AI will outperform human experts and take control over the world. Such speculations can create false fear and overshadow current challenges and ethical dilemmas that need to be dealt with, researchers warn.

In the report 'Portrayals and perceptions of AI and why they matter,' published by researchers from Cambridge University in the United Kingdom, it is argued that false fear can contribute to a misinformed debate and a decrease in confidence. This, in turn, can potentially impact AI research, funding, and regulation.

Exaggerated expectations for AI to solve any kind of problem, ranging from the shortage of healthcare professionals to climate crises, is another prevalent narrative according to the report. Such expectations can contribute to a hype bubble:

"If such a bubble bursts because the technology was unable to live up to the unrealistic expectations, public confidence in the technology and its advocates could be damaged," the authors write.

News outlets run the risk of shaping, mediating, and amplifying expectations that AI in the future will be a good solution to a myriad of problems, according to [a critical analysis](#) of selected news articles performed by researchers from Reuters Institute for the Study of Journalism at Oxford University in the United Kingdom.

In doing so, the outlets overestimate the capacity of the data driven technologies and promote technochauvinism, the researchers conclude. Technochauvinism refers to the belief that most, if not all, complex issues can be solved with computation and engineering.

“When AI is seen as a skeleton-key solution, it is easy to forget that there are myriad of *other* possibilities - some of which might not involve technologies,” the researchers conclude in their analysis published in the academic journal Journalism in 2022.

5.2. Avoid exaggerating results from controlled experiments with AI

“Google “robomedics” spot disease faster than doctors,” is the headline of one of the news articles analysed by the researchers from the Reuters Institute. The article from 2018 is about an AI system that was trained to detect more than 50 different forms of eye disease from retinal scans.

The system was developed by the Google owned company DeepMind and tested in an experiment made by researchers from University College London and Moorfield Eye Hospital.

In a press release about the test results, the collaboration announced that “AI was able to make the right referral recommendation more than 94 percent of the time, matching the performance of expert clinicians.”

Several British outlets chose to frame the experiment as a competition in which AI was pitted against human eye specialists according to the analysis from the Reuters Institute. The AI system performed better and faster than the physicians, the outlets reported. The competition frame is not surprising, the researchers note:

“Highlighting competition and conflict is a frequent narrative strategy across many types of journalism,” they write.

However, when reporting on AI, the competition framing can be problematic because it may suggest that data driven technologies can surpass and replace human specialists. The researchers from the Reuters Institute argue that, in the specific case mentioned, the media coverage supported expectations of the AI system’s potential as a solution to the lack of capacity in the eye hospital.

Even though the AI system from DeepMind had only been tested in a controlled experiment, none of the media outlets questioned whether the results could be replicated in a more complex real clinical setting. Additionally, the researchers argue that the media outlets did not inquire about potential losses in replacing human experts with AI systems, or the cost of doing so.

The challenges of replicating results from controlled experiments with medical AI in a real world clinical practice is highlighted in [a review](#) published in the scientific journal Nature Medicine. In the review, the authors conclude that while clinical AI systems demonstrate great potential in experiments, they often fail to meet expectations when used in real clinical settings, where they do not achieve the same level of accuracy as in the tests. This discrepancy can be attributed to factors such as insufficient or biased training data, which can lead to flawed results in a clinical setting.

Furthermore, unexpected challenges and expenses often arise when technological tools are implemented in the healthcare sector. The new technology might be incompatible with existing systems, creating additional complexity. Additionally, machine learning models require monitoring and updates by technical specialists, which can result in extra costs.

Journalists should consider these limitations before implying that an AI system has the potential to benefit patients and replace human experts

5.3. Get second opinions from independent experts

Another analysis from the Reuters Institute for the Study of Journalism shows that perspectives and sources from the technology industry were overrepresented in 760 news articles about AI brought in six British mainstream news outlets in the first eight months of 2018:

- Nearly 60 percent of the news articles across outlets were indexed to industry products, initiatives, or announcements.
- 33 percent of unique sources across all articles were affiliated with industry which was almost twice as many as those from academia, and six times as many as those from government.
- Nearly 12 percent of all articles referred to Elon Musk.

The intensive coverage of industry concerns, products, and initiatives can be explained by the fact that much of the research and development of AI occurs in the commercial sector. However, the Reuters

researchers find the media outlets' prioritisation of sources and initiative from the industry problematic for the following reasons:

1. Industry sources may exaggerate or promote the benefits of AI for their own interests, which can lead to an imbalance or skewed narrative surrounding AI technology.
2. By focusing on the industry perspectives, the outlets downplay the responsibility of politicians and other public representatives in addressing AI as an emerging public issue.
3. Outlets limit the range of voices included in the public conversation. "For example, every time that an article reports Elon Musk's extreme - if entertaining - opinions about AI, it misses an opportunity to bring in other, less familiar voices," the researchers write.
4. Industry sources regularly encourage outlets to position AI as a viable solution to a wide range of problems. But in doing so, the outlets rarely question the limits of AI's competency, ethical problems, or the potential risk of implementing the data driven technologies.

The researchers from the Reuters Institute recommend journalists to include a wider range of voices in stories about AI.

"Academics, activists, politicians, civilians, and civil servants, amongst others, can all contribute to a rich and sophisticated public debate around AI," they write in the report '[An Industry-Led Debate: How UK Media Cover Artificial Intelligence.](#)'

It is not achievable nor desirable to attempt to control the public narratives about AI. However, it is beneficial to widen "the body of available narratives, drawing in a wider range of authors and protagonists and creating spaces for public dialogues," researchers from Cambridge University also suggest in their report 'Portrayals and perceptions of AI and why they matter.'

There are numerous well-informed sources beyond industry representatives that journalists can talk to about AI. Artificial Intelligence is a scientific field that was coined in the 1950's by a group of American computer scientists. Since then, a wide range of academics from various disciplines, including computer scientists, engineers, philosophers, and economists from universities worldwide, have

studied not only the technical potentials and limitations, but also the ethical and other derivative effects of AI systems.

Human rights organisations, ethical committees, and representatives from international associations such as WHO and UNESCO are also well informed about AI ethics. Hence, there are plenty of specialists without associations to the industry that can assess the potential of new AI inventions, adding new perspectives to the news coverage.

5.4. Summary: How to report on AI constructively

Machine learning and other AI technologies could have potential as tools to solve real world problems such as the lack of personnel in the healthcare sector. In a liberal democracy, members of the general public are key stakeholders for the technology adoption and, to a certain extent, how AI is regulated by the politicians. The level of public engagement, confidence, and acceptance of the data driven technology as a solution to problems will likely be impacted by the media coverage. Hence, nuanced information, transparency, and a critical approach are needed to ensure that AI is used in a beneficial, appropriate, safe and fair manner.

Journalism that covers solutions in a nuanced way have the potential to counteract news avoidance and to empower members of the audience by giving them a feeling of agency, a review of research made by the Constructive Institute shows. Hence, constructive reporting can contribute to engagement and informed public conversations about the potentials, limitations, and risks of using AI, for example in the healthcare sector.

The following key principles can ensure a nuanced and righteous reporting on AI:

1. **Accuracy:** Avoid spreading false fear and unrealistic expectations.
2. **Balance:** Report on insecurities. Be honest about the costs of implementing AI, risks, and the limitations of initial test results.
3. **Diversity:** Add new perspectives and nuances to the public conversation about AI by including diverse sources. Get second opinions from independent researchers. Do not just talk to representatives from the industry.

The next chapter will explore methods to engage the audience and foster meaningful discussions about ethical questions concerning AI in healthcare.

6. Engage media users in dialogues about AI ethics

Debates about bioethics are often surrounded by controversies. This is unavoidable because new technologies spark anxiety and insecurity, according to UNESCO.

“In modern societies, technological innovation is valorised, as it promises new possibilities for the improvement of our lives. Nevertheless, these are often contested developments that are not equally enthusiastically received. The controversies that arise are rooted in an intermingling of anxieties: worries about the physical effects of the technologies and about their socioeconomic effects, and ambivalence regarding their moral character and the real uncertainties of the achievements of converging advances and results,” as noted in UNESCO’S guide for bioethical committees, [‘Bioethical Committees and Public Engagement.’](#)

In fact, modern societies are characterized by a fundamental tension between scientific knowledge and its applications. The resolution demands a public debate in which all voices are heard, the guide recommends:

“Civic process and social debate should support the democratization of decision-making procedures and ensure that the diverse perspectives of men, women, youth, the elderly, various ethnic groups, among others, are specifically taken into account and that no one is left behind when it comes to benefiting from the fruits of science and technology.”

Fair solutions to bioethical problems are dependent on a public dialogue where opinions from all levels of society, educational backgrounds, ethnic groups, genders, and generations are heard. National and local media are obvious arenas for such conversations that bring together multiple opinions, experiences and perspectives.

The Constructive Institute has developed a model called the Constructive House consisting of three pillars: Focus on solutions, cover nuances, and promote democratic conversation. In this chapter, we

will look at journalistic approaches to the third pillar: How can journalists engage the audience and promote democratic conversation about AI ethics?

6.1. Dialogue journalism: invite the audience to co-create the stories

The core principle of dialogue journalism is to listen to and invite the audience to actively participate in the production of media content. Associated professors from three European universities has developed the approach as an integrated method within the mindset of constructive journalism.

Three key principles in dialogue journalism:

- **Engaging** communities in conversations on solutions.
- **Creating** new ways of debating.
- **Looking** for nuances in dialogues with the audience.

In order to engage media audiences in debates about AI ethics and to make appealing stories about the topic, the approach is highly relevant.

In dialogue journalism, the audience takes an active part in the journalistic production process to various degrees. Members of the audience can be included in the whole journalistic process from developing ideas, editing the stories, and debating them on social media. Or the journalist can choose to have a dialogue with the audience in parts of the process for example by asking for ideas, answering questions, or by taking an active part in debating the story on social media.

Dialogue journalism is a core principle behind the Danish member owned media Zetland: The journalists of Zetland ask the members what they want to know and what curiosities they have. By doing so, they attempt to get better ideas and to make sure that their stories are relevant to their members. Sometimes the journalists even invite members to co-edit their stories before they get published. More information on how Zetland and other European media work with dialogue journalism can be found at [the homepage](#) of the project DIALOGUE.

Behind the DIALOGUE project are associate professors from The Danish School of Media and Journalism (DMJX), The School of Journalism at Windesheim University of Applied Sciences in The Netherlands, and Stuttgart Media University in Germany.

6.2. Frirummet (The Free Space): facilitate constructive debates

Media outlets could also engage their audiences in conversations about AI ethics by arranging events where lay people get the chance to discuss their worries, doubts, and opinions with politicians, developers of AI, scientists, and healthcare professionals. As discussions about bioethical questions are often polarised and, it is worth recommending the principles formulated by the Danish initiative “Frirummet” (The Free Space).

Frirummet has developed a new form of debate that attempts to lay the groundwork for solution initiatives by opening for mutual understanding, doubt, and reflection in discussions about conflict-filled and polarised political issues. In order to ensure constructive and solution-oriented conversations, Frirummet educates discussion moderators - so called debate pilots. The moderator’s job is to:

- Create a space where the opposite fronts listen to each other with mutual respect and understanding.
- Provide a foundation for debaters and the audience to collectively generate new solutions to problems.
- Ensure that all participants in the debate become more informed about the conflict and the viewpoint of the opponent.

Frirummet is a collective democracy initiative of the Danish Independent Schools, the Folk High Schools, and the Boarding Schools. More information about the initiative and the constructive debate form can be found on [Frirummet’s homepage](#).

6.3. Rethink science journalism: reflect and listen to the audience

Both AI and bioethics are scientific disciplines. Globally, academics engage in constant discussions and publish studies about technological breakthroughs, experiments, limitations, and ethical considerations related to the use of data-driven technologies. Since AI is predicted to transform the healthcare sector with consequences for both physicians and their patients, it is highly relevant and constructive to enrich and qualify the public debate with news stories about significant scientific results

and academic discussions within the field.

However, when reporting on science there are several pitfalls to be aware of. In order to avoid misleading the media users, the Danish popular science media Videnskab.dk has made an easily read guide with 11 pieces of advice on how to avoid blunders when reporting on science. [The guide can be downloaded here.](#)

Another challenge when disseminating science is complexity: Scientific processes and results are often abstract and difficult to relate to for a lay audience. Furthermore, in the public sphere certain scientific topics have become increasingly politicised and subject to polarised debates and conspiracy theories.

To ensure that lay people share and engage with accurate stories about AI instead of fake ones, journalists should have a sense of the audience's interests, level of understanding, worries and questions, according to the EU project Rethink Science Communication. To achieve this, science communicators and journalists should strive to engage citizens from all levels of society and provide them with the necessary knowledge to participate in the public conversation. Some pieces of advice from the Rethink project:

- **Be open-minded and try to understand the audience:** In order to counteract science scepticism and misinformation, many journalists fall into a 'knowledge-deficit trap' meaning that they attempt to change the beliefs of the audience by explaining the facts again. But this does not work because people make sense of science based on their own personal situation and social context. In order to convince the sceptics, it is necessary to reach out and be open to their values, emotions, experiences, background, and worldview.
- **Ask yourself:**
 - What assumptions do I have about my audience?
 - What am I not seeing?
 - With whom am I not interacting?
 - Should I engage with them differently?

How can I transform my practice, so that I can connect to the personal and contextual notion people have towards science?

- **Listen, be curious, and find common ground**

When confronted with people with different worldviews, emotions, and values regarding scientific information, it is easy for journalists to become dismissive of their opinions.

Nevertheless, it is useful to listen to the stories of the sceptics and be curious about them. Try to find a common ground and let that be the point of departure of the story.

Nanna Holst, editor for inclusion and dialogue at TV2 Kosmopol and a fellow at the Constructive Institute in 2022-2023, has also concluded that by listening to the audience and responding to their worries, news outlets become more relevant, relatable, and nuanced. Nanna Holst has published a tiny book with pieces of advice for journalists on how to listen better. She recommends:

- Perceive the media users as more than just receivers.
- Treat the users as competent partners.
- Reach out and ask the users for help and input.

6.4. Summary: How to engage media users in dialogues about AI ethics

A public dialogue that includes voices from people with different educational backgrounds, ethnicities, ages, and other demographic characteristics could be crucial to ensure an acceptable, just, beneficial, and successful use of AI in healthcare. Constructive journalism that is accurate, critical, and nuanced is an obvious approach to provide lay people with knowledge and engage them in such a dialogue.

However, in order to catch the media users' attention and engage them, the coverage of AI in healthcare should start from their interests, level of understanding, and questions about the topic.

Dialogue based journalism is a useful approach to engage the media users and make relatable stories. Media outlets could also take a step further than simple reporting by facilitating public discussions about AI in healthcare using the principles of Frirummet (The Free Space).

7. Inspiration from lectures at Aarhus University

The considerations about AI ethics discussed in this report are largely inspired and qualified by the university lectures attended during my fellowship at the Constructive Institute. Two lecture series were particularly relevant, one focusing on bioethics from a philosophical perspective by the lecturer postdoc Joonas Räsänen, and the other covering the background and implications of bioethics in medical science by the lecturer Kristian Hvidtfelt Nielsen.

Philosophical bioethics: The value of thought experiments

In the lecture series, “Bioethics and Politics”, taught by Joonas Räsänen, I was introduced to philosophical theories relevant to qualify bioethical discussion about the use of AI in healthcare. These include:

Utilitarianism: This theory emphasises maximising overall happiness and well-being. In bioethics, this can involve evaluating medical decisions based on the greatest benefit to the majority of people. Regarding AI in healthcare, utilitarianism is useful when considering whether the use of the technology is justified despite the risk of harm to some patients.

Deontology: Deontological ethics focuses on moral duties and obligations. In bioethics, it may involve following specific ethical rules or principles, regardless of the consequences. For instance, deontologists might argue that black box AI should never be used in healthcare even if a majority of patients benefit from them because they come with a risk of biases that can lead to unethical discrimination.

Principlism: Principlism is the framework that uses core ethical principles, such as autonomy, beneficence, nonmaleficence, and justice, to guide ethical decision-making in bioethical dilemmas. Hence, principlism is the foundation of the ethical principles that are used to guide physicians and healthcare professionals in their clinical practice.

In this report, I have approached ethical dilemmas related to AI in healthcare using the lens of principlism. However, the lectures by Joonas Räsänen taught me that it is also valuable to incorporate other philosophical theories to ensure a comprehensive consideration of all aspects of bioethical dilemmas.

Philosophical thought experiments can uncover undoing

The lectures also opened my eyes to the value of philosophical thought experiments. Philosophical thought experiments are imagined scenarios designed by philosophers to gain insights into complex concepts in order to challenge intuitions about what is morally justified. Such thought experiments can be used to spot undiscovered ethical undoing in current practice.

“Are we currently making serious mistakes, not just at the level of implementation, but also about the very values and considerations that guide our decision making in healthcare and biosciences?,” the Norwegian philosopher Ole Martin Moen asks in his paper [‘Why good work in philosophical bioethics often looks strange’](#) published in the journal *Theoretical Medicine and Bioethics*.

Thought experiments can also be used to consider unexpected, hypothetical ethical dilemmas that could arise in the future. A commonly used example of a philosophical thought experiment about AI in healthcare is the following:

Consider a situation where a patient is diagnosed with a severe illness and has the option to be treated by either an AI-powered medical system or a human doctor. An ethical dilemma arises if the AI system is known to have a higher success rate in treating the specific condition compared to the human doctor. However, the patient expresses a strong preference for being treated by a human doctor due to the need for empathy and emotional support.

This thought experiment raises questions about the balance between efficacy and human connection in healthcare. It exposes the role of AI in medical decision-making and the importance of considering patients’ emotional needs alongside clinical outcomes when integrating AI systems into healthcare practices.

In journalism, thought experiments could be employed to translate abstract ethical concerns into concrete scenarios, making them more relatable and engaging to the audience. With thought-provoking scenarios, journalists have the potential to stimulate the audience to ponder and reflect upon crucial ethical questions concerning the application of AI in clinical settings. This might encourage readers or viewers to engage with the complex issues in a more meaningful way and foster a deeper understanding of the implications.

7.2. The foundation of bioethics in medical science

In another lecture series about bioethics and the philosophy of science, taught by Kristian Hvidtfelt Nielsen, I gained valuable background knowledge about the foundation, development, and significance of bioethical principles. The course's emphasis on bioethics in science is particularly relevant for my work as a journalist at the popular science media Videnskab.dk. We also studied the history behind and the formulation of ethical:

The Hippocratic Oath: Ethical considerations in medicine and healthcare can be traced all the way back to ancient civilizations such as ancient Egypt, Greece, and Rome. An important ancient ethical code is the Hippocratic Oath, which is attributed to the Greek physician Hippocrates, who is considered the father of Western medicine. The oath outlines the moral and ethical duties of physicians towards their patients. It emphasises principles that are still used today such as patient confidentiality, nonmaleficence, and the importance of patient autonomy.

Four Principles Approach: In 1983, two American philosophers and ethicists, Tom L. Beauchamp and James F. Childress proposed the four principles of bioethics that are still widely accepted: autonomy, beneficence, nonmaleficence, and justice.

These and other bioethical principles have throughout history been formulated in response to technological development and moral wrongdoings to individuals. Constructive and critical journalism, which considers both the potentials and the risks, as well as the ethical dilemmas, could play a crucial role in ensuring a morally justified implementation of AI in healthcare and preventing harm from occurring.

References and sources

The content of this report is based on various interviews, lectures, oral presentations, academic papers, and books.

Lectures, Aarhus University

Artificial Intelligence, The Folk University, Aarhus, spring 2023.

Political Philosophy meets Bioethics, Aarhus University, spring 2023.

Bioethical and philosophy of Science, Aarhus University, autumn 2022.

Genetics and Personalized Medicine, Aarhus University, autumn 2022.

Presentations at The Constructive Institute and during excursions

Felix M. Simon, DPhil student, Oxford Internet Institute, Oxford University. Presentation about journalism and AI, 2023.

Hany Farid, professor, School of Information, Berkeley University. Presentation about AI, deep fakes and misinformation, 2023.

Jens Christian Bjerring, associate professor, School of Culture and Society, Aarhus University. Presentation on large language models and AI ethics, 2023.

Johanne Haagerup & Annette Holm, Danish School of Media and Journalism. Presentation on dialogue based journalism, 2022.

Julian Schüssler, Department of Political Science, Aarhus University. Presentation about AI and large language models, spring 2023.

Kjeld Møller Pedersen, professor in Health Economics, University of Southern Denmark. Presentation about the current state of the healthcare system, spring 2023.

KQED, San Francisco: Presentation about Cracking the Code - a project about users' engagement in science journalism.

Peter Busch Nicolassen, Ph.D. student at the Department of Political Science, Aarhus University. Presentation about science journalism, spring 2023.

SF Chronicle: Presentation on the engagement project, SF Next.

Staff at Constructive Institute: [Presentation](#) of a review of studies about the effect of constructive journalism, 2023.

Tina Klindt Bohni, Frirummet. Presentation on facilitating constructive debates, 2022.

Interviews

Anne Gerdes, professor, The Department of Media, Design, Education and Cognition, University of Southern Denmark, 2023.

Anne-Marie Søndergaard Christensen, professor of philosophy, University of Southern Denmark, 2023.

Benjamin Rasmussen, MD, Ph.D. and head of research, CAI-X Centre for Clinical AI, Odense University Hospital, 2023.

Sarah Coy and Nisha Shah, The Centre for Health, Law, and Emerging Technologies, Oxford University, 2023.

Sne Scott Hansen, Ph.D. student, Copenhagen University, 2023.

Søren Brunak, professor and head of research, Novo Nordisk Foundation Center for Protein Research, Copenhagen University, 2023.

Thomas Plough, professor of philosophy, Aalborg University, 2023.

Thomas Trolle, Bioinformatics Specialist, Evaxion Biotech AS, 2023.

Scientific articles and reports

About AI and ethics in healthcare:

'AI in health and medicine,' Nature Medicine, 2022.

'Artificial Intelligence in medicine and the disclosure of risks,' AI & Society, 2021.

'Black box algorithms and the rights of individuals: no easy solution to the "explainability" problem', Internet Policy Review, 2021.

'When Doctors and AI Interact: on Human Responsibility for Artificial Risks,' Philosophy and Technology, 2022.

'Why good work in philosophical bioethics often looks strange,' Theoretical Medicine and Bioethics, 2022.

'The Ethics of Algorithms in Healthcare', Cambridge Quarterly of Healthcare Ethics, 2022.

'The right to a second opinion on Artificial Intelligence diagnosis - Remediating the inadequacy of risk-based regulation', Bioethics, 2023

Analysis of the media coverage of AI:

'AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media,' AI & Society, 2020.

'An Industry-Led Debate: How UK Media Cover Artificial Intelligence', Reuters Institute, 2018.

'Best-practices for science communication,' Rethink Scicomm, EU, 2022.

'Imagining the thinking machine: Technological myths and the rise of artificial intelligence,'
Convergence: The International Journal of Research into New Media Technologies, 2017. "

'Portrayals and perceptions of AI and why they matter', The Royal Society, 2018.

'Public AI imaginaries - How the debate on artificial intelligence was covered in Danish newspapers and magazines 1956-2021,' Nordicom Review, 2021.

'The AI doctor will see you now: assessing the framing of AI in the news coverage,' AI & Society, 2020.

'The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation,' AI & Society, 2022.

'The Role of Media and Deference to Scientific Authority in Cultivating Trust in Sources of Information about Emerging Technologies,' International Journal of Public Opinion Research, 2011.

'What to expect when you're expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news,' Journalism, 2020.

Books

Bioethics and Journalism, Pavel Tishchenko & Boris Yudin, 2011.

Bioethics Committees and Public Engagement, UNESCO, 2019

Deep Medicine, Eric Topol, 2019.

Kunstig intelligens bagfra (Artificial Intelligence in Reverse), Anders Søgaard, 2022.

Power And Prediction: The Disruptive Economics of Artificial Intelligence, Ajay Agrawal, Joshua Gans & Avi Goldfarb, 2022.