

Conversational AI in Healthcare:

Risks, Challenges and Solutions

AdalanAI

Executive Summary

Artificial Intelligence technologies have enabled many industries to transform itself and healthcare industry is not an exception. One of the most widely used solutions in healthcare is Conversational AI delivered via chatbots or virtual assistants. Conversational AI is revolutionising patient care making it easier for healthcare providers to provide quality services to patients. In the past few years the industry has been witnessing proliferation of Conversational AI systems as benefits are becoming evident.

At the same time, using conversational AI in the healthcare involves much bigger risks than in any other industry, as it deals with high-impact scenarios and may directly affect human life. We mapped these risks with our AI GovBox - AI Governance Approach.

However, despite the challenges that are unique to the healthcare industry, Conversational AI can be highly beneficial if the right strategies and approaches are applied by the healthcare institutions. We discuss this through 4 key stages of designing and building Conversational AI agents - proof of concept, development, deployment, scaling & feedback - which involves two major stakeholders: developers and healthcare providers.

Conversational AI in Healthcare

The first well-established chatbot, ELIZA, was created in 1966 to simulate a Rogerian psychotherapist. Since then, conversational agents have become more sophisticated and widely used, with many innovations to look forward to.

```
Welcome to
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II    ZZ    AA  AA
EEEEEE LL      II    ZZZ   AAAAAA
EE      LL      II    ZZ    AA  AA
EEEEEE LLLLLL  IIII  ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Depending on their type, conversational agents can fulfil various tasks: rule-based conversational agents work based on predetermined keywords and commands and are used for simpler tasks (e.g. ordering food, booking appointments, purchasing merchandise, etc.), while conversational AI agents use Artificial Intelligence (NLP, ML) to simulate conversation with users and can continually learn from queries, undertake more complex tasks and perform human-like conversations with a user.

The conversational AI is used in finance, retail, commerce, marketing, education, healthcare and other industries. They are increasingly applied in healthcare due to the rise and complexity of data and its potential benefits to resolve these complexities. The Covid-19 pandemic has further advanced the adoption of Conversational AI agents. The conversational agents in healthcare fulfil administrative functions connected to patient engagement but are also involved in diagnosis and treatment recommendations.

There are three types of most commonly adopted conversational agents in the healthcare industry: the first category assists doctors and clinics, the second category assists patients, and the third category is designed to boost and simplify a doctor-patient relationship.

1

To assist doctors and clinics, conversational AI reduces the administrative burden of scheduling appointments, making additional spare time for doctors by navigating frequently asked questions and facilitating symptom assessments to decide whether the patient needs to go to the hospital. CAs can also monitor patients in real-time and provide updates to doctors. Another example is the virtual patients, designed and actively used by the students of medical schools to train them in recording the patient's medical history or deciding a treatment option. It is seen as a risk-free learning platform, as no real patients are involved. NLP And ML hold the possibility for more individual conversations between students and virtual patients.

2

The conversational agents that fall into the second category are designed to give patients practical or medical information relevant to their medical history or answer simple patient queries. For example, during the Covid-19 pandemic, different organizations and governments created chatbots for spreading coronavirus information and symptom checking. Another example of using CA to assist patients is when patients, especially the elderly, require special care during and after hospitalisation. Conversational agents can delay some Alzheimer's disease symptoms through cognitive exercises or stimulate the skills of the elderly with normal cognitive abilities. Other functions of the conversational agents can be using personalised data to support the elderly with treatment recommendations, diet or medications.

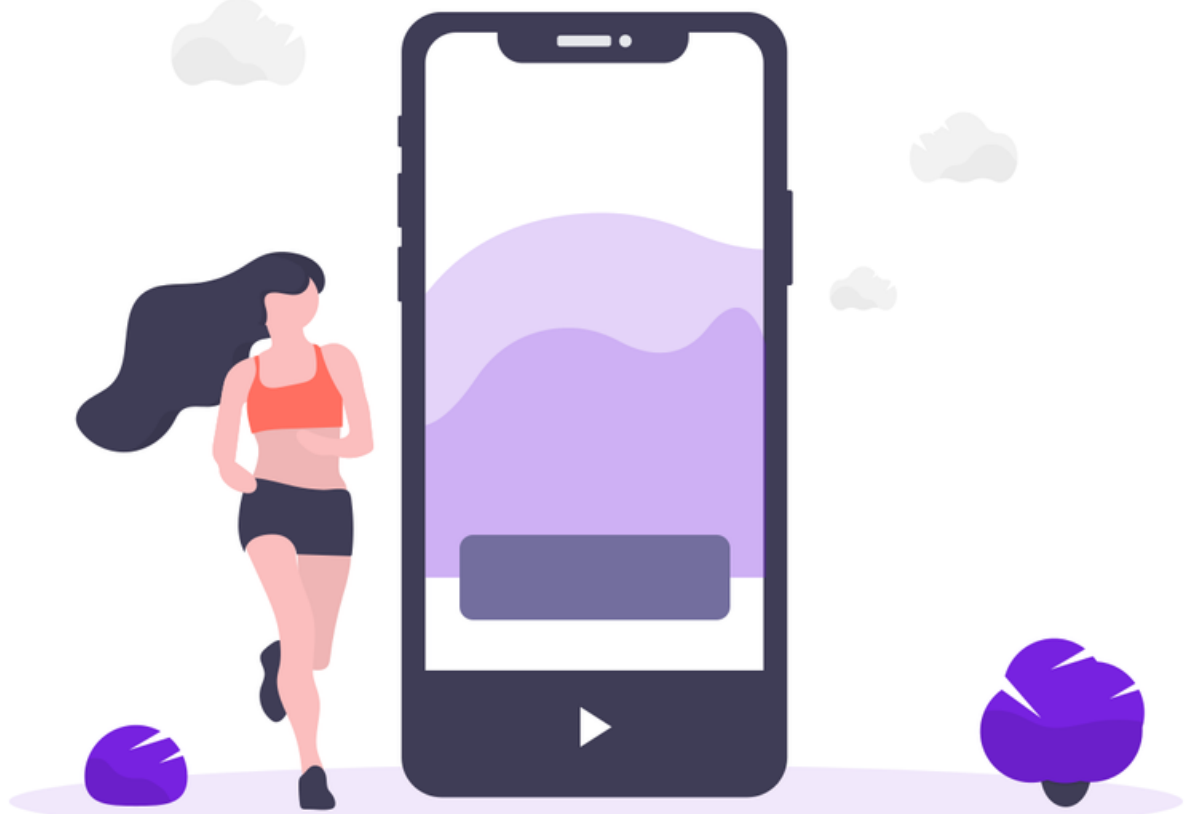
3

The third category includes the use of CA agents, which take the doctor-patient relationship to the next level when chatbots offer therapeutic solutions based on the symptoms provided by the user. One of the most promising areas is mental healthcare. CA can gather diagnostic information and symptom monitoring and provide the users with psychological support. Countries, for instance, the US, experience a decrease in the number of clinicians, with a rise in the number of mental illness cases. Conversational AI can make up for this decrease since it has no time or attention limit, unlike a human clinician and has a limitless capacity to hear and analyse conversations. Moreover, the psychotherapists usually rotate out of the treatment centres, while the conversational agents can have a more long-term relationship with the patients. On the other hand, the psychotherapists and the clinic administration will be free from some mundane tasks (e.g., giving prescriptions, making new appointments, managing patient intake, etc.).

Conversational AI agents enable better access and affordability in healthcare, facilitating adherence to UN Sustainable Development Goal 3 – Good Health and Well-being, which aims to ensure healthy lives and promote well-being for all ages.

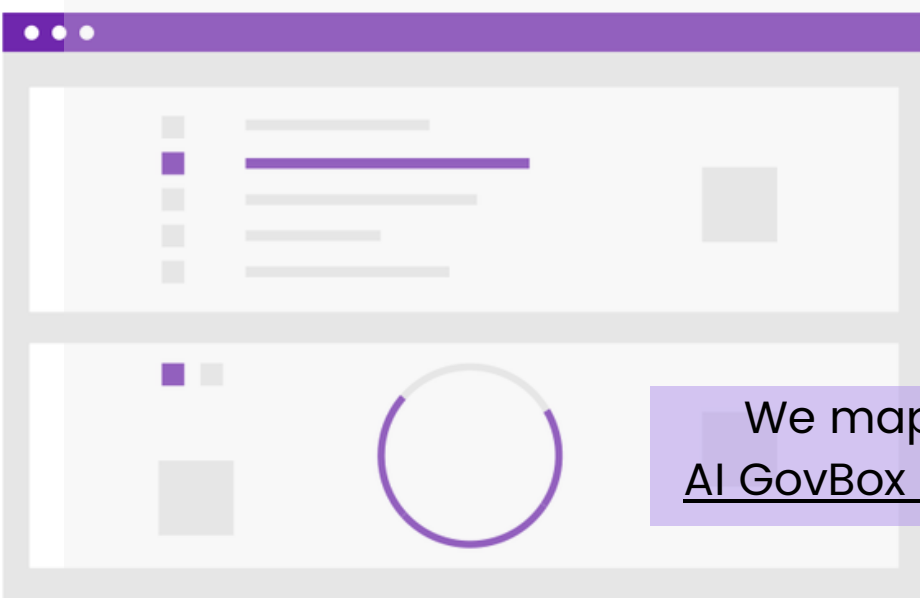
One of the long-standing challenges of health access is to accommodate rural populations and improve the involvement of the people who find traditional therapy stigmatising. On the one hand, conversational AI promises to overcome the challenge and reduce inequality. Still, on the other hand, it raises new risk factors, such as uneven accessibility to qualified physicians. For example, when human intervention is needed in a decision, there are more qualified physicians in developed countries than in developing ones, or there may be a problem with physicians' digital skills in developing countries regarding technology adoption.

Using conversational AI in healthcare involves much bigger risks than in any other industry, as it deals with high-impact scenarios and may directly affect human life.



Risks and Challenges

While there are a lot of benefits conversational AI agents have in healthcare industries, they also pose several risks and challenges. Understanding what could be the possible misuses or risk factors of conversational AI agents in healthcare and defining the ways to avoid them can unlock the vast potential. We mapped these risks with our AI GovBox - AI Governance Approach.



We mapped these risks with our AI GovBox - AI Governance Approach

1 Fit for Purpose

2 Data Quality and Privacy

3 Responsibility and Accountability

4 Fairness

5 Continuous human oversight

6 Transparency and Explainability

7 Safety and Robustness

Fit for Purpose

Research shows that a higher degree of personification of Conversational AI agents and human-like characteristics correlates with a higher level of satisfaction for the users. The delivery of accurate algorithmic predictions alone is insufficient for effective human-AI collaboration. Conversational agents need human-like characteristics to ease tension during communication with users and make them feel free to share their thoughts and feelings. However, with the human-like characteristics of CA, questions connected to anthropomorphism and the uncanny valley* arise. There's no agreement on why uncanny valley occurs: some regard it as a natural reaction, stating that one can't distinguish between a human and non-human. Others focus on the contrast between the extremely human-like design of a robot and its jerky motions.

*Defined as an unsettling feeling while seeing a human-like robot by the roboticist Masahiro Mori in 1970, the Uncanny Valley phenomenon has been debunked by some studies while proved to be true by others.

Those who treat the Uncanny Valley phenomenon as a real problem state that we can never overcome it through engineering – a person, for instance, might be in a relationship with a robot without knowing it. Still, if there's an accident when human damages the robot and its "robotness" is exposed, the person will feel uncanny. But these are all suppositions.

What we must consider instead of theorising on possibly unrealistic things is whether conversational agents today really are human-like? In fact, they are not. Conversational AI today doesn't have fully human-like language abilities, which often leads to misunderstandings and user dissatisfaction. Chatbots have been surveyed in the US with 500 users, and they have reported, among other concerns, the poor usability of the conversational agents and the difficulty of discussing more complicated health questions with them. Other problems that the studies have revealed include poor understanding (due to the limited vocabulary), poor management of word inputs, the lack of voice recognition accuracy, and repetitive content (frequently reported by the users). Some users have also pointed out that the aim of the chatbot has remained unclear to them.

Therefore, the concerns with the Uncanny Valley or anthropomorphism are not the most relevant for today's context, i.e., for today's technological abilities of the conversational agents and the engineering abilities of humans. The challenge is how companies define the purpose of their system, whether it's better for CA to have more human-like characters or more technical to avoid misunderstanding. Yet another challenge is how they communicate with the users about what to expect from the CA.

Responsibility and Accountability

Different AI ethics frameworks have recently been adopted to address ethical challenges raised by AI systems. However, at this point, AI-specific ethics frameworks in healthcare have limited adoption, and they have mostly been used simultaneously with bioethics, which has long been practised in healthcare.

Clear responsibility and accountability are considered fundamental factors in building trust in AI according to the 2019 Code of conduct for data-driven health and care technology by the UK Department of Health and Social Care. Whether there are implemented mechanisms to ensure responsibility and accountability for AI systems, such as auditability, ability to redress, or clear liability, has an impact on how conversational AI agents can foster responsible innovation in the healthcare industry.

It's still unclear to stakeholders who is responsible if the recommendation given by a conversational agent is not accurate, misleading or not delivered on time.

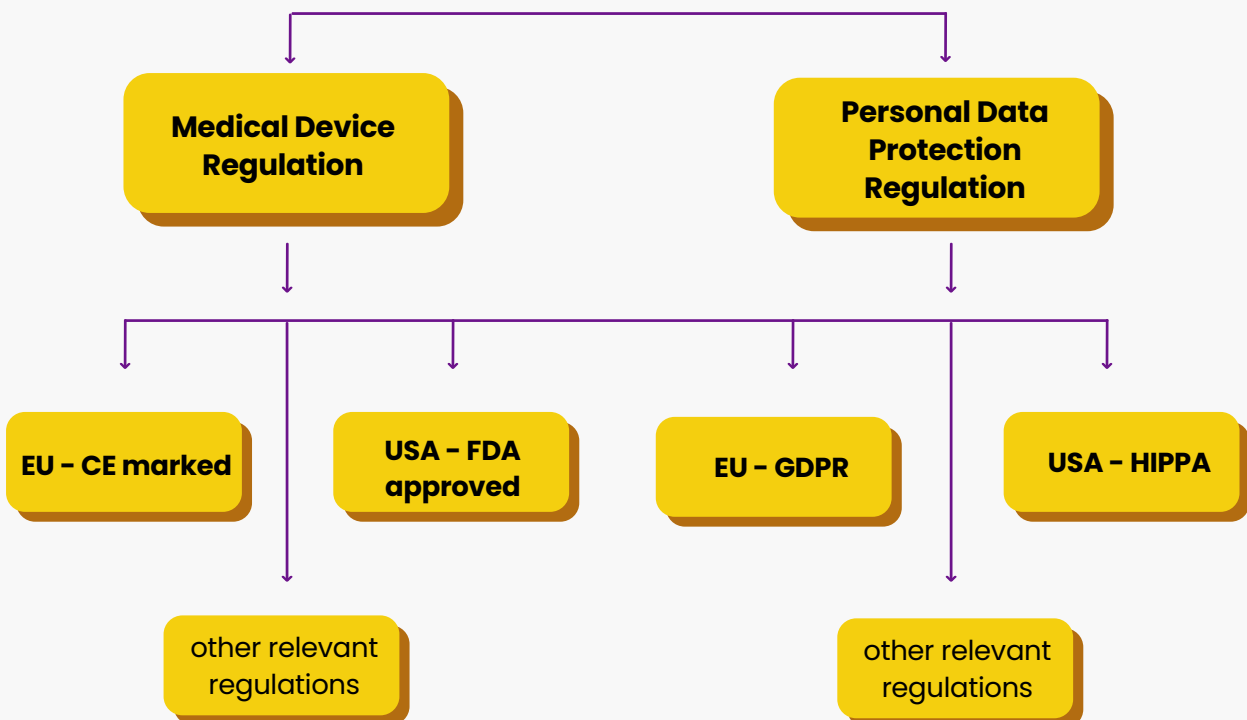


While there is still a regulatory grey area regarding the issue, mechanisms for minimising negative impact, documenting trade-offs, whether the chatbot conversations are human audited and if so, how often, remain one of the challenges for the responsible use of CA.

On the regulatory level, currently, existing data protection regulations apply as well as medical device standards and regulations (Figure 1). For the healthcare industry, medical device regulation from the European Commission is an important document (2017). It defines a medical device as a product or equipment intended for a medical purpose. A conversational AI agent could be considered a medical device if designed to diagnose diseases or cure, mitigate, treat, or prevent disease. However, if conversational AI is intended for only providing information or as guidelines, it may fall in the well-being product category, not the medical device. If the product is labelled as a medical device, the manufacturer has a higher product liability. The manufacturers, on the other hand, should clearly state whether their product is intended for medical purposes, i.e. whether it is a medical device or a well-being product (i.e. whether it is focused on diagnosis or gives some medical guidelines for being healthy).

The new European Union AI regulatory proposal will apply when it enters into force. Even though at this moment, it might have overlapped with medical device regulations, they need to be addressed and resolved in the coming years. The lack of clear regulations on AI technology and conflicting national and international legislation can pose a challenge.

Figure 1: Existing Regulations



Data Quality and Privacy

The accessibility of CAs with mobile apps will lead to another concern – privacy, as the healthcare industry is a high-risk and sensitive area, especially concerning personal data. CAs with free text and voice input raise concerns especially for vulnerable populations, such as the elderly and children, as voice commands are a much simpler format, and are more likely to share sensitive and private information without much prior thinking.

AI is defined as “systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals” by the European Commission. In healthcare, the environment ranges from clinical points of care in nursing homes to private homes. Patients use chatbots and voice assistants mostly at home, where the collection and usage of the information are not protected by the culture of ethics in the medical field. This raises questions about how and when data is collected, who owns it, and how it will be used? Does the developer have a record of data governance? Can a patient request to delete it? – On the other hand, health data is the most valuable resource in defeating diseases, its quality has a crucial role as it is used for diagnosis and treatment.

Currently, roughly 80% of the data collected in the healthcare industry is unstructured or conversational. Moreover, the existing data in the clinics may not be machine-readable or use different medical terminology for symptom descriptions.

Patients use chatbots and voice assistants mostly at home, where the collection and usage of the information are not protected by the culture of ethics in the medical field.



Currently, roughly 80% of the data collected in the healthcare industry is unstructured or conversational. Moreover, the existing data in the clinics may not be machine-readable or use different medical terminology for symptom descriptions. This poses a challenge to train the data, ensure the data quality, and have clear responsibilities between stakeholders regarding data quality. Moreover, while Machine Learning promises to make sense of the previously unused and unstructured information, it is still worth asking how conversational AI with free voice or text commands helps in finding business value and health impact. Making the best out of CA as a solution requires integrating it into hospitals' internal information systems, enabling different stakeholders to access medical records. This raises a lot of concerns about the privacy of the patient's personal information, for instance, whether third parties will have access to it, and if so, what will the patient's health data be used for.

Transparency and Explainability

As the technology develops, it is becoming more and more difficult to keep track of the reasoning behind the responses of CA (the so-called black box effect). This factor is important not only for patients but also for doctors and clinics, as doctors need to be aware of the reasoning behind the model decision, which input data was gathered, and how it was selected and connected. Moreover, clinicians desire information about the model's strengths and limitations. If communication between the user and CA fails, the user might lose trust not only in the CA but also in the therapy. The users might be reluctant to trust the system if the solutions provided by it are unclear or its capabilities are vague, or its intended use seems ambiguous.

On the other hand, sometimes it's difficult for the users to distinguish whether they are communicating with humans or AI, especially if there is a human-like conversation flow. While some chatbots clearly emphasise to the user that they are interacting with the AI, some may not focus much, which leads to confusion, disappointment or even anger.

Here it is worth noting that oftentimes terms transparency and explainability are used interchangeably while these two terms can be very distinct in what they mean to different stakeholders. Transparency is important for the users in order to understand what technology they are interacting with, how and why this technology is processing the user data and what inferences are being made. On the other hand, explainability makes sense when it comes to understanding inside working of the algorithm itself that is useful for the auditors, developers and policy makers.

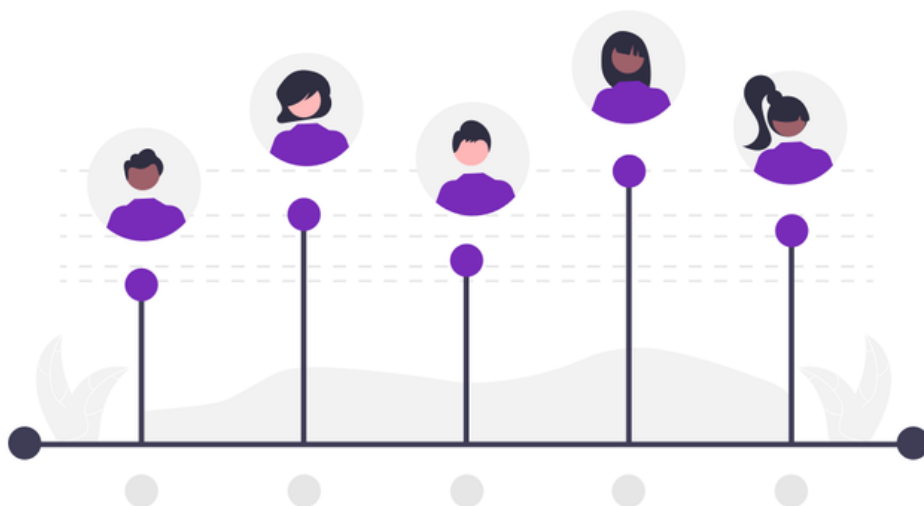
Fairness

While we are discussing the anthropomorphic character of conversational AI and its impact on human autonomy, they still pose the most discussed ethical challenge in AI - bias.

In voice command conversational AI agents, language biases concern diversity and representation of different language and ethnic groups. Voice assistants have trouble understanding different dialects, which can be an issue for patients and doctors, especially in mental healthcare, as verbal communication is an essential part of the therapy. The success of the therapy depends on good communication, so language bias is not about fairness only but influences the treatment and its potential outcome too.

Researchers from University of Massachusetts found that accuracy of several common Natural Language Understanding tools was far lower for speakers of certain varieties of English, such as African American Vernacular English, slang or those with strong accents.

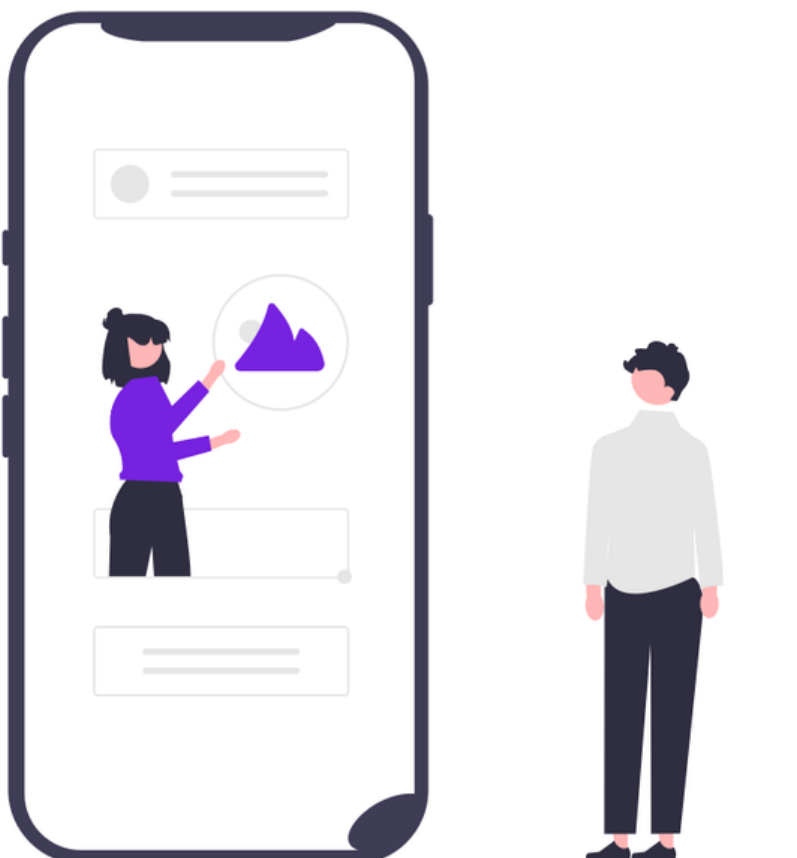
Although all the voice assistants claim they are genderless when asked or give users the option to choose the voice, academic studies show that users are opting more for female voice. On the other hand, the prominence of female-voiced virtual assistants strengthens the stereotypical perception of women as submissive and compliant. Even more crucial is whether conversational AI agents are trained on the dataset, leading to gender-biased



Continuous Human Oversight

The usage of Conversational agents in healthcare is still less explored, its efficacy or potential harm is untested. There are no studies on cost-effectiveness or improved productivity either. Using a conversational agent in therapy means diverting it from its traditional path, i.e. if before the clinician was responsible for sensitive information, now the data is recorded and analysed by the conversational agent too. In some cases, it facilitates patient disclosure, but in others, it can be an impediment.

Even if the users communicate freely with the conversational agent, there is still a potential for harm – for instance, the conversational agents might not be able to respond to suicide and homicide ideations. Certain types of disclosure carry legal mandates, and homicide is one of them – the clinician has to report it if the patient admits such inclinations during the therapy. While conversational agents might not be able to handle the situation and identify potential homicide or suicide cases, it leaves the user or their family under a fatal threat.



Even if the users communicate freely with the conversational agent, there is still a potential for harm – for instance, the conversational agents might not be able to respond to suicide and homicide ideations

Currently, the most successful way of integrating AI in healthcare is a hybrid system in which humans constantly supervise conversational agents. In other words, conversational AI is best seen as an assistant to doctors rather than a substitution for them. For several reasons, even the fields in which automation is easier, such as pathology or radiology, show very slow progress in AI usage.

First, radiologists do not merely interpret images. They also consult with their colleagues, perform medical interventions (e.g. biopsy), are involved in the treatment process, compare their findings with the results of other medical tests, and discuss the whole process with the patient. Conversational agents, on the other hand, cannot perform so many of these tasks yet. AI needs a significant amount of the “labelled data ” (the images of the pathologies such as cancer, a broken bone, etc.), while there is no system of radiology images, neither labelled nor otherwise.



How can providers and developers ensure to train healthcare professionals about the supervision, how can they be sure when is the right time to trust the algorithmic decision and when not remains one of the most important challenges. Yet another issue is what can be a human's role in the AI integration process, other than supervision. The active feedback – the human-in-the-loop approach has proved to be quite successful in the healthcare system. Studies have shown that the human-in-the-loop AI models work better than AI or human doctors alone. What could be the practical implementation of this approach in case of a conversational agent needs to be actively discussed, for instance, whether companies can involve end-users directly in improving the quality of communication.

Safety and Robustness

The main issue with the healthcare sphere is that there's no exact recipe for how to treat a patient. In fact, the treatment is a whole process of weighing risks and benefits and ensuring that the patient understands them and agrees to the suggested solution. Integrating this principle, i.e., the principle of non-maleficence in the conversational agents, is a challenge that needs to be overcome for virtual assistants, doctors and patients alike. While ML succeeds in finding proxies that can predict what we are looking for, they may struggle to understand when and how a proxy is no longer a good predictor. Hence, sudden changes in patient behaviour when interacting with the CA can debilitate models' performance.

As with every AI system, conversational AI systems could also face adversarial attacks where incentivised agents attempt to subvert the model. This could apply to illegal behaviours such as stealing patient data.

The main issue with the healthcare sphere is that there's no exact recipe for how to treat a patient. Sudden changes in patient behaviour when interacting with the CA can debilitate models' performance.



Another challenge is how providers and developers assess potential attacks, such as damage to physical infrastructure, cyber-attacks or data pollution, and what resilience measures are needed while considering the high-risk character of the healthcare industry. The damage or harm that CA may cause in the healthcare industry includes miscommunication with users, too much trust in CA and self-diagnosis without going to a doctor or even wrong diagnosis and following the wrong treatment.

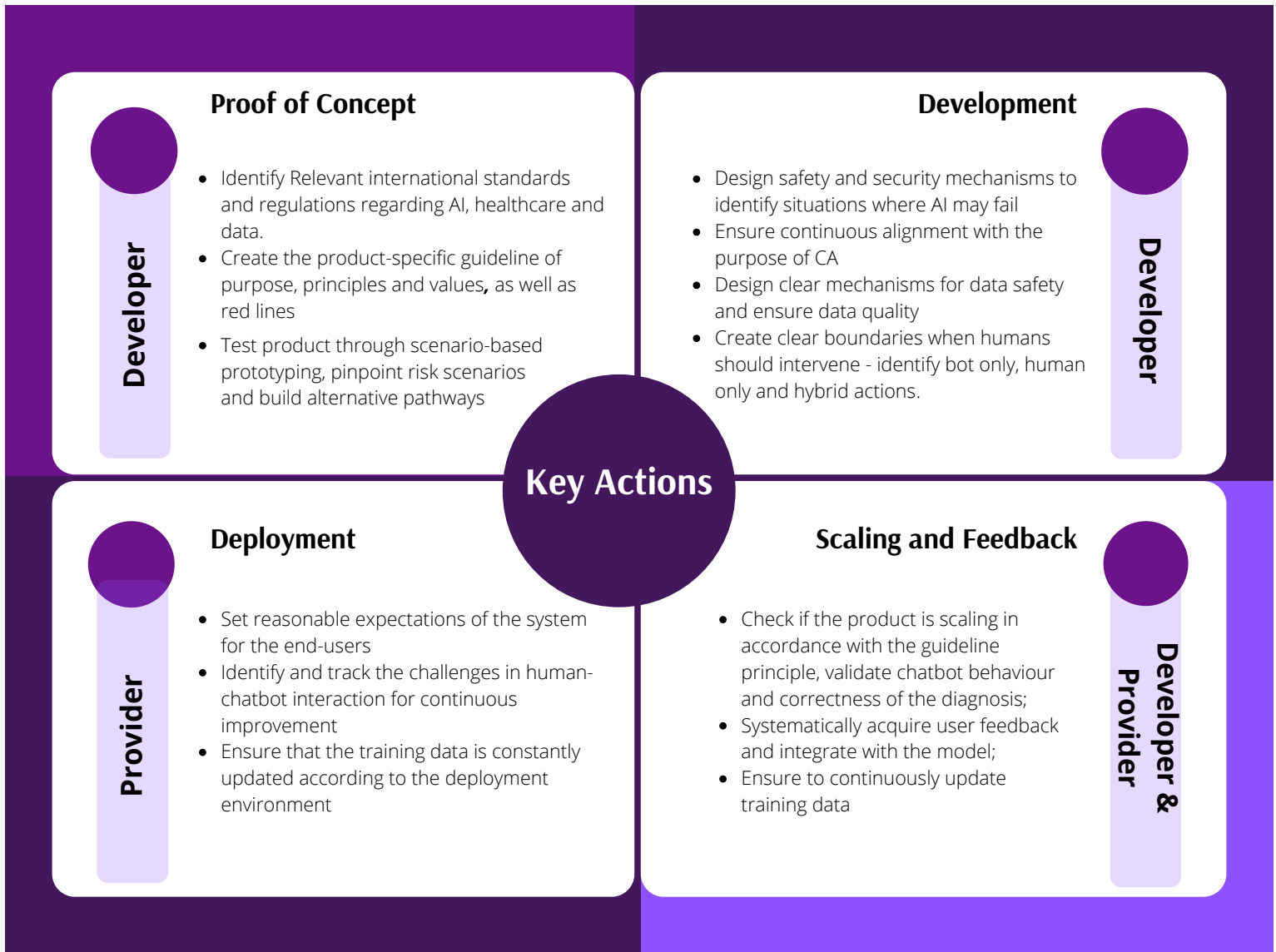
Recommendations and solutions

To ensure the best possible outcome of using conversational AI in the healthcare industry, it requires (1) combining both AI ethics and healthcare ethics and (2) involving different stakeholders, such as healthcare professionals, chatbot developers, AI ethicists and human-AI interaction specialists in all stages of chatbot development.

We propose high-level key actions in the following 4 stages of CA development that need to be taken in order to drive operational excellence, achieve best business value and greater user experience.



The first stage for designing and building the conversational AI is Proof of Concept where you define the purpose and limits for the CA, ethical principles, values, and relevant internal policies. In the next, the development stage, technology developers build AI systems that meet the requirements set in the proof of concept stage and are in alignment with the purpose. The next, the deployment stage, includes the process when CA is piloted in small-scale environments with continuous oversight of involved stakeholders. And in the last stage, CA is scaling while developers and providers systematically collect user feedback in order to adapt CA accordingly.



Conclusion

While Conversational AI agents in healthcare industry are highly promising to improve patient care, certain risks have to be considered at all stages of CA development. If healthcare institutions implement right strategies and processes in place, the benefits of these technologies can be realized at its fullest.

About AdalanAI

The logo for AdalanAI, featuring the text "AdalanAI" in a bold, black, sans-serif font. The text is centered within a bright yellow rectangular background. Below the text, there is a thin, horizontal line with a small, stylized graphic element resembling a double-headed arrow or a bridge structure in the center.

Adalan AI is management consulting & SaaS platform for AI Governance, Policy and Ethics. AdalanAI is about driving operational excellence, achieving greater customer satisfaction and gaining competitive edge on AI market. We help corporates, VCs, investors and policy-makers in Artificial Intelligence product risk assessment and management, impact assessment, policy-tracking and policy formulation by improving/building internal governance strategies, structures, processes and people skills across organizational functions.

Explore how AdalanAI can help in identifying and mitigating Conversational AI risks to achieve best business value and greater customer satisfaction

 We are global

 www.adalanai.com

 ana.chubinidze@adalanai.com