**Designing Causal AI-based Coach for the Health care – Pilot Study**

**Designing Causal AI-based Coach for the Health care – Pilot Study**

# Introduction

[**https://www.sciencedirect.com/science/article/pii/S2414644719300508**](https://www.sciencedirect.com/science/article/pii/S2414644719300508)

[**https://app.clinally.com/better-health-outcomes-with-ai-powered-virtual-assistants/**](https://app.clinally.com/better-health-outcomes-with-ai-powered-virtual-assistants/)

**Notes:**

**AI Coach is designed for intensive baby care …. Why focus is baby care?**

There have been tremendous advances in neonatal care over the past several decades, along with significant efforts to improve the quality of care provided to the vulnerable neonatal patients. It is increasingly recognised that, despite best efforts by health care providers, active error detection and an ideal safety culture, errors will inevitably occur in perinatal health care systems. Common patient safety issue may result in injuries to babies in the newborn period. These injuries may lead to short- and long-term consequences, which will have a strong impact for the patient, their families and society. One of the common areas of malpractice risk for neonatologists is the management of the neonates during the transitional period–from antenatal risk management, birth, delivery room management and immediately postnatal phase. A better understanding of the medical malpractice system and common patient safety issues in peri- and (neo)natology can lead to protective strategies to reduce risk for untoward events. By developing and using newer decision support systems and strategies including maintaining competency, following national and

Human-computer interaction (HCI) is a technology which enables users and machines to communicate using natural language [1]. An intelligent communication device (chatbot) is one solution to human-computer conversation that has been developed to persuade people that they are conversing with a human rather than a machine. Chatbots have been commonly utilised in a variety of domains, including customer care, website assistance, and schooling. According to recent reports, 80% of businesses expect to deploy chatbots by 2020 [1].

The biggest advantage of utilising chatbots for businesses is that their customer support systems are streamlined and the chatbot will address queries about goods or services from consumers. Building a smart chatbot, on the other hand, is difficult since it necessitates contextual comprehension, text entailment, and language-understanding technologies. Therefore, artificial intelligence and natural language processing also are needed for a variety of applications [2],[3].

It is still necessary to define AI’s role in providing viable solutions to the defined epidemic. Surprisingly, a global research program must be developed to begin measures against this – but rather potential – epidemic without blaming anyone.

Recently, we have observed an increasing interest in conversational agents, as well as in software that interacts with humans through natural language. Text-based chatbots (or simply chatbots) have proliferated in a wide range of application environments for about the last ten years, allowing humans to communicate with machines using natural written language [4].

During a pandemic, people become unsure about what to do. Taking insufficient precautions (for example, failing to take prophylactic measures) will increase everyone’s risk of infection. Heading to the emergency room for minor symptoms, for example, can overburden the healthcare system, wasting valuable resources. As a result, trustworthy information sources are critical for preventing a “misinfodemic”: propagating a disease aided by viral misinformation.

To manage the large number of requests from citizens during pandemics, such as COVID-19, organizations should have new communication mechanisms. As a result, organizations such as the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) have started to use chatbots to exchange knowledge and suggest behaviours in order to relax the overwhelming majority of nervous individuals [5]. Chatbots are software-assisted intelligent services that converse with people in their native language via voice or text. Such well-known examples include Amazon’s ”Alexa,” Apple’s “Siri,” and Microsoft’s “Cortana.” The Chatbot use for health-related purposes has grown significantly in recent years, from assisting clinicians with clinical interviews and diagnosis to assisting customers in self-managing chronic conditions.

Chatbot systems are the latest digital interface developments, following the growth of the web and smartphone apps [6],[7]. It is well documented for these applications to use automatic conversational agents that operate on software creation or a kind of artificial intelligence (AI) relationship between users and automated systems with the intervention of natural language processing (NLP) [8]. A chatbot, on the other hand, reflects the technical advancement of Question Answering systems that are primarily focused on Natural Language Processing. Among the most popular examples of Natural Language Processing being used in various enterprises’ enduse applications is producing answers to user requests in human-like natural language.

# Related work

In this section, we aim to address the related work for our study. First we will explain the neonatal care and medical errors associated to it. Secondly, we will explain the terms which will be used throughout the study. Lastly, we will explain the constructs which were used to model the AI Coach system.

## Neonatal Care and Medical Errors

(Why neonatal is taken as an initial department)

As mentioned in Section 2.1., neonatology field is a field where (phenomenal) consciousness is the prime phenomenon taken into account.

Team work

Explain forgetfulness here, how can AI help

## Virtual Healthcare Coaches

Pressure on healthcare is increasing due to reduced availability or accessibility of healthcare professionals. To address this challenge, intelligent conversational agents or virtual assistants have proven their potential to serve and reduce the burden of medical health practitioner. For example, Eliza was the first conversational ever designed (Weizenbaum, 1983), or Woebot to help as therapy chatbots (Fitzpatrick et al., 2017) . Ever since, there is a huge trend on improving the capabilities of conversational agents. These capabilities not only focus on human understanding (through natural language processing), but also aim to understand the context of conversation through machine learning and artificial intelligence.

Unlike conventional agents virtual coaching system is used to enhance healthcare interventions by system human interaction. This is an emerging field that not only offers an opportunity for mental wellbeing but it also allows the individuals to self-track and e-coach for their goals (Tsiouris et al., 2020). The system human interaction in healthcare can be possible by conscious choice of users (by clicking certain option) or by monitoring sensors. For instance, in home coaches data maybe gathered by different sensory devices (e.g., Fitbit, Samsung, Polar, etc.), or in hospital where data maybe gathered by Virtual Reality (e.g., Microsoft Kinect, Nintendo, sensing gloves, etc.) (Tsiouris et al., 2020). The main motivation of such virtual assistants or coaches is to provide an interaction to promote changes in behaviors.

Mostly, the designs of such systems are user-centered, so they involve end users throughout development cycle. Thus the ultimate success of such assistants depends whether the end users will find the application useful. Moreover, involving users can reduce the development time, and this can result in more accurate assessments regarding user requirements (Rosson & Carroll, 2002). As mentioned in Section 2.1., neonatology field is a field where (phenomenal) consciousness is the prime phenomenon taken into account. By this we mean that its all about the experience the stakeholders are having at the moment of baby birth. These stakeholders might include health care professionals, parents of the neonate, neonate and more. Therefore, this venture is a joint venture, in which a cautious choice can only result in good health of a baby.

As discussed in Section 2.1, forgetfulness can lead to medical errors and need further assistance to avoid it. There are various conversational agents or coaches designed to serve in healthcare domain, but to our knowledge none of them was designed to assist child birth. Section 2.3. mention the requirements considered to design the prototype of AI Coach.

## Requirement Analysis

interviews  
Cover Social technical Systems in Causal Models

SMILE4Safety -> The concept of a shared mental model has recently received increased attention in medical team performance literature as well as in other domains. Shared mental models are often brought in relation to the quality of team performance and safety.38, 39, 40, 41, 42, 43 A team has a shared mental model when relevant knowledge structures concerning how reality works or should work are held by all team members and when there is sufficient alignment in the internal representations of these knowledge structures. 44, 23, 45

REQUIREMENTS FOR A SOCIO EMOTIONAL

RELATIONSHIP WITH A VIRTUAL SYSTEM from (Roelofsma, 2013)

Knowledgebase (include ML/Script oriented)

Human centered

OL

Modeling different processes is not a novel concept, and is widely used for their comprehension.

It uses different diagrams, mathematical representations, or tools to understand them.

Social factors in social-technical systems refer to the human and interpersonal elements that influence the design, implementation, and functioning of these systems. Social-technical systems are those that combine both technological components and human elements, recognizing that the success and effectiveness of the system depend not only on the technology but also on the people who use and interact with it. These social factors play a crucial role in shaping the outcomes and impact of such systems. Here are some key social factors in social-technical systems:

1. User Behavior and Adoption: The behavior of individuals who use the system and their willingness to adopt and adapt to the technology significantly affect its success. Factors such as user acceptance, resistance to change, and training needs influence how effectively the system is integrated into the organization or society.
2. Organizational Culture: The values, norms, and practices of the organization or community where the social-technical system is implemented can impact how well the system aligns with existing processes and workflows. A supportive and innovative culture can foster successful integration and implementation.
3. Communication and Collaboration: Effective communication channels and collaboration among users, stakeholders, and developers are essential for understanding needs, addressing challenges, and ensuring that the system meets its objectives.
4. Power and Politics: Social-technical systems are not immune to power dynamics and political influences within an organization or community. Decisions about the system's design, deployment, and control can be shaped by different stakeholders vying for influence and control.
5. Trust and Privacy: Users' trust in the system and concerns about privacy and data security are critical factors in determining the level of acceptance and usage. A lack of trust can hinder adoption and lead to resistance.
6. Socioeconomic Factors: The economic conditions and resources available to users and organizations can affect their ability to implement and sustain the social-technical system.
7. Human-Computer Interaction: The usability and user experience of the technology play a significant role in how well the system is embraced and effectively utilized.
8. Ethical Considerations: Ethical concerns surrounding the use of technology in social-technical systems must be addressed, such as issues related to bias, discrimination, transparency, and accountability.
9. Social Impact: Understanding and managing the potential positive and negative impacts of the system on society, including effects on communities, individuals, and the environment, are essential aspects of social-technical system design.
10. Adaptability and Flexibility: Social-technical systems should be designed to adapt to changes in the social and technological landscape over time. Emphasizing flexibility allows for the system to remain relevant and effective in evolving contexts.

Recognizing and addressing these social factors is crucial for the successful implementation and sustainable functioning of social-technical systems, ensuring that technology aligns with human needs and values.

Top of Form

Regenerate response

Socio‐technical system design is based on the premiss that an organization or a work unit is a combination of social and technical parts and that it is open to its environment[[5](https://www.emerald.com/insight/content/doi/10.1108/00251749710173823/full/html?casa_token=SLS40KANcLwAAAAA:n-Hc4c0EVvCaTpQfi-j05eubMLwRNoN4Qo5dsUzxb4_F6fwb6edBLIG7yIRsQzPnDosxLgEWAQMzRSYS8TxX_ZfuJ2d4qk-_EtHq56NpCKFEHZN7n882#b5)]. Because the social and technical elements must work together to accomplish tasks, work systems produce both physical products and social/psychological outcomes. The key issue is to design work so that the two parts yield positive outcomes; this is called joint optimization. This method contrasts with traditional methods that first design the technical component and then fit people to it. The traditional methods often lead to mediocre performance at high social costs.

Organizational development can be defined as a “long‐term, system‐wide application of behavioural science techniques to increase systems” but “neither system should operate at the expense of the other. Co‐ordinated and integrated human and technical activities are possible when one system is supportive of the other”[[2](https://www.emerald.com/insight/content/doi/10.1108/00251749710173823/full/html?casa_token=SLS40KANcLwAAAAA:n-Hc4c0EVvCaTpQfi-j05eubMLwRNoN4Qo5dsUzxb4_F6fwb6edBLIG7yIRsQzPnDosxLgEWAQMzRSYS8TxX_ZfuJ2d4qk-_EtHq56NpCKFEHZN7n882#b2), p. 62]. Because all subsystems are interdependent, changes in one area affect and influence other system elements[[8](https://www.emerald.com/insight/content/doi/10.1108/00251749710173823/full/html?casa_token=SLS40KANcLwAAAAA:n-Hc4c0EVvCaTpQfi-j05eubMLwRNoN4Qo5dsUzxb4_F6fwb6edBLIG7yIRsQzPnDosxLgEWAQMzRSYS8TxX_ZfuJ2d4qk-_EtHq56NpCKFEHZN7n882#b8), p. 49]..

<https://www.emerald.com/insight/content/doi/10.1108/00251749710173823/full/html?casa_token=SLS40KANcLwAAAAA:n-Hc4c0EVvCaTpQfi-j05eubMLwRNoN4Qo5dsUzxb4_F6fwb6edBLIG7yIRsQzPnDosxLgEWAQMzRSYS8TxX_ZfuJ2d4qk-_EtHq56NpCKFEHZN7n882>

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.rama.mahidol.ac.th/ceb/sites/default/files/public/pdf/ACADEMIC/2019/RADS601/Slide/4.2%20IT%20Governance%20%26%20Management%20in%20Healthcare%20Organizations%20-%20Part%202.pdf

The core idea od social technical theory is that the design and performance of any organizationl system can be understood and improved only if both social and technical aspects are brought together and treated as interdependent parts of a complex system.

https://en.wikipedia.org/wiki/Sociotechnical\_system

Social technical means that technology, which by definition, should not be allowed to be the controlling factor when new work systems are implemented. So in order to be classified as 'Sociotechnical', equal attention must be paid to providing a high quality and satisfying work environment for employees.[[21]](https://en.wikipedia.org/wiki/Sociotechnical_system#cite_note-Mumford2001-21)

<https://www.sciencedirect.com/science/article/pii/S0953543810000652#b0300>

The term socio-technical systems is nowadays widely used to describe many complex systems, but there are five key characteristics of open socio-technical systems ([Badham et al., 2000](https://www.sciencedirect.com/science/article/pii/S0953543810000652#b0040)):

* •

Systems should have interdependent parts.

* •

Systems should adapt to and pursue goals in external environments.

* •

Systems have an internal environment comprising separate but interdependent technical and social subsystems.[2](https://www.sciencedirect.com/science/article/pii/S0953543810000652#fn2)

* •

Systems have equifinality. In other words, systems goals can be achieved by more than one means. This implies that there are design choices to be made during system development.

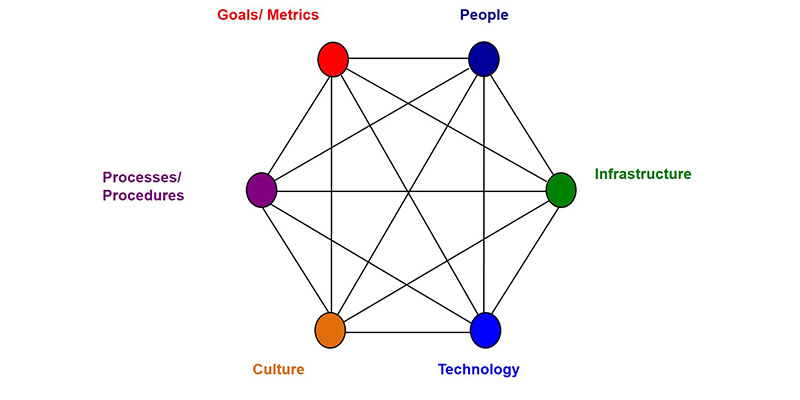
* •

System performance relies on the joint optimisation of the technical and social subsystems. Focusing on one of these systems to the exclusion of the other is likely to lead to degraded system performance and utility.

Sociotechnical refers to the interrelatedness of ‘social’ and ‘technical’. Sociotechnical ‘theory’ is founded on two main principles. One is that the interaction of social and technical factors creates the conditions for successful (or unsuccessful) system performance. These interactions are comprised partly of linear ‘cause and effect’ relationships, the relationships that are normally ‘designed’, and partly from ‘non-linear’, complex, even unpredictable relationships, which are those that are often unexpected. An inevitable consequence of mixing ‘socio’ with ‘technical’ is that the socio does not necessarily behave like the technical, people are not machines, paradoxically, with growing complexity and interdependence even the ‘technical’ can start to exhibit non-linear behaviour. Inevitably, both types of interaction occur when a sociotechnical system is put to work. The corollary of this, and the second of the 4 two main principles, is that optimisation of either socio, or far more commonly the technical, tends to increase not only the quantity of unpredictable, ‘un-designed’, nonlinear relationships, but those relationships that are actually injurious to the system’s performance. Sociotechnical ‘Theory’, therefore, is all about ‘joint optimisation’.

A sociotechnical ‘system’, as well as being the descriptive term given to any practical instantiation of socio and technical elements engaged in purposeful goal directed behaviour, is a particular expression of Sociotechnical Theory. Sociotechnical systems take the concepts and metaphors of general systems theory, in particular the notion of ‘open systems’ (e.g. Bertalanffy, 1950), as a way of describing, analysing and designing systems with joint optimisation in mind, particularly those that embody some degree of non-linearity within themselves as well as the environment they reside in. Sociotechnical systems theory, the term used throughout the current article (and seemingly the term in most widespread use at the present time) reflects certain specific methods of joint optimisation in order to design organisations that exhibit open systems properties and can thus cope better with environmental complexity, dynamism, new technology, and competition.

Within a socio-technical systems perspective, any organisation, or part of it, is made up of a set of interacting sub-systems, as shown in the diagram below. Thus, any organisation employs people with capabilities, who work towards goals, follow processes, use technology, operate within a physical infrastructure, and share certain cultural assumptions and norms.



### 2.3.1. Human Factors and Organizational Learning (to be considered in Healthcare?)

Safety and Just Culture (Take from proposal)

Models

Explain models the reified architecture

Shared Mental Models

Through enhancing contact between clinic–patient and doctor-patient, healthcare chatbots have a lot of potential in medical communication. Remote testing, prescription followup tracking, and telephone appointments will also help meet the high demand for health services. A chatbot will conduct short and easy health surveys, set personal health-related reminders, communicate with clinical teams, plan appointments, and retrieve and review health data [12]. When looking for specific signs or patterns that can be used to diagnose illness, chatbots can react quickly or immediately to patients’ healthcare-related queries.

For example, the Internet-based Doc-Bot communicates via mobile phone or Messenger. The bot can be personalised to suit unique health conditions, demographics, or habits [13]. The bidirectional information exchange between chatbots and patients could be used to test for medication commitment or to gather data. ([LINK](https://ieeexplore.ieee.org/abstract/document/9447652?casa_token=VdLUrtTjdNsAAAAA:y4DtvpmQXacBMfpvD0Gg_sOsv6OQY_rONlTQRucGoun3dH4LKxgbOiaZygZ6jUetABa2EcfbyA))

Concept of

telehealth, [IOT](https://www.silabs.com/applications/connected-health?source=Media&detail=Google-PPC&cid=med-gos-mlt-041923&s_kwcid=AL!16736!3!650955730038!b!!g!!iomt&gad=1&gclid=Cj0KCQjwoeemBhCfARIsADR2QCuRrCoXdCuM3R6QHf_XcgYZChkorllfXmqr6JtvZk79_zcXC017QukaAjTZEALw_wcB)

[pervasive healthcare](https://ieeexplore.ieee.org/abstract/document/1599033)

### 2.3.2. Knowledgebase Design

Why we chose database based on the models? For interaction

### 2.3.2. Decision Support System

Human powered + computerized

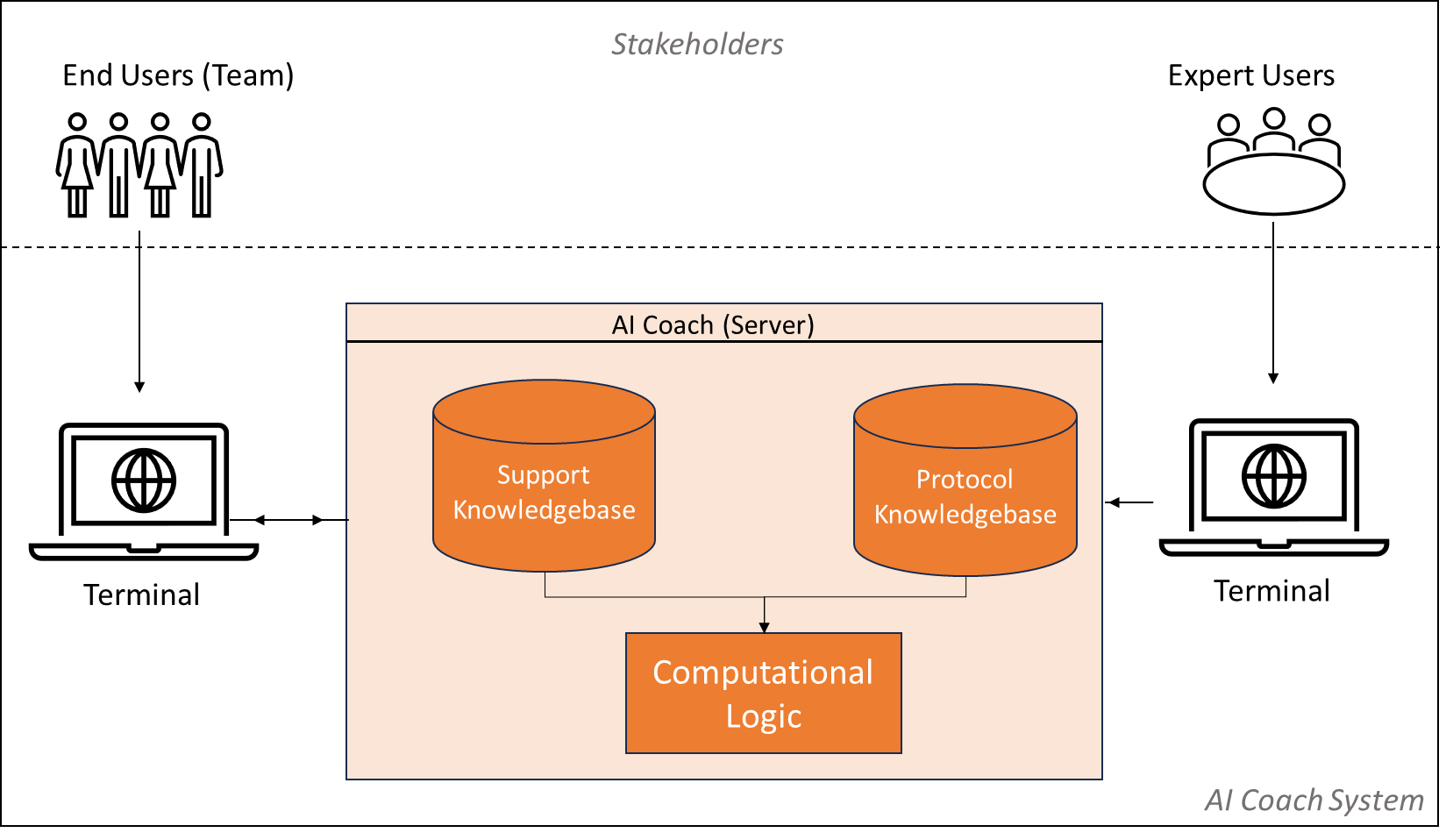
<https://en.wikipedia.org/wiki/Decision_support_system>

# Methods and Methodologies

This section presents the whole process that lays the foundation for the development of AI Coach. Initially, we will discuss the framework design of the AI Coach, then we presents how different models are extracted to provide human-coach interaction, and how does coach can interact with its stakeholders.

## Framework Design of AI Coach

In IT, a framework refers to a structure that consist of many phases or components working together to act as a foundation for the software design and implementation. This article introduces a novel Human-AI based framework to support medical teams in different healthcare settings. It has ability to offer assistance to variety of end users, which can be expanded depending upon culture(s) of an organization like medical healthcare. The framework is mainly composed of two components, i.e., AI Coach and the stakeholders of a system (See Figure XXX).



**Figure** **XXX**: Framework of AI Coach

### Stakeholders

AI Coach system is a software application that aims to interact and serve two type of stakeholders in particular, i.e., end users and experts.

**End users** are the stake holders who interact directly with the system to achieve their goals, learn and have an pleasant experience via usage of technology. AI Coach is designed to serve neonatology in Erasmus MC, for example during baby delivery/care, or intensive care unit. Therefore, it not only has the capability to adhere proper neonatal protocol execution but it can also foster emotional and social wellbeing of the people who are actually on the floor, e.g., patient, their family, or the medical team working together in neonatal care. Understanding the time constraint the coach is designed to fulfil its objective through minimum user engagement. These users are further discussed in Section 3.2.

While **experts** are the stakeholders which work in different domains and interact to provide support in a meaningful activity. They can from different domains like psychology, artificial intelligence, health and safety sciences, who interacted with each other to ensure patient safety and just culture. The main focus of experts is to brainstorm and develop the algorithm of the coach. As a result, the behavior of AI Coach behavior is adaptive by nature. This is accomplished by developing (shared) mental models (van Ments et al., 2021). Further details can be found in Section 3.2.

Current role of AI Coach is to serve neonatology department in Erasmus MC, Rotterdam. So, it can provide support for neonatal care and wellbeing of patient(s) and medical team, where the possible end users are; baby, patient, their family and the staff on the floor.

### AI Coach

AI Coach is an AI based virtual assistant whose design is inspired by the principles of sociotechnical systems (i.e., with the focus on social systems and technical systems in healthcare). Therefore, primarily it is human centered and have a social impact on its users. Still, it is quite flexible to address any changes in the social landscape over the time (within an organization). Therefore, all the stakeholders are considered for human interaction with the AI Coach.

AI Coach is designed with a deep understanding of needs of the medical team who is working on the floor in the neonatology department. This can address different social situations/aspects, capabilities and behaviors in different situations like child birth. For instance, if a baby is not breathing, AI Coach will assist to achieve the particular goal of helping the baby to breath (Xu et al., 2022). Similarly, if a resident wants to speak up during any medical process the coach will help him/her to speak up in a certain situation (Doornkamp et al., 2022). Therefore, end user interaction is adaptive which follows through the series of choices which may have been (un)consciously made by the end users.

Similarly, the design of AI Coach is quite flexible, which means that an organization (like Erasmus MC) can design the usage of AI Coach as per needs (via their expert users). This may address changes in environment, technology or organization, which may evolve over time. In other words, the stakeholders can become an architect of AI Coach. As a result, this may not only effect the assistance process of the coach, but it may also affect the roles of the involved stakeholders and the coach itself. In the section below, we describe the components and functionality of AI Coach.

#### Components of AI Coach System

In this section, we will describe technical aspect of AI Coach system, i.e., how it was designed to offer functionality to its end users. AI Coach is a server based virtual assistant, that is designed based on requirements collected through different interview sessions from the staff of neonatology department in Erasmus MC. Few important considerations were taken into account during the design time, i.e.,:

1. Job is physically demanding, which may lead to forgetfulness, stress or influence on the personal skills of a resident or a doctor
2. Job has a time constraint, they have to respond in a quick and timely manner
3. Baby birth is a team work, and safe execution of protocol ensures no errors in the procedure
4. Different situations may lead to protocol deviations (medical errors) which may lead to poor mental wellbeing, and sufficient support is desired in this.

Key stakeholders were identified and invited to take part in the developmental process of AI Coach for neonatal care. Different stakeholder groups were identified through scoping interviews, therefore, main interacting actors were either the residents or the doctors, while the people who are on floor were the neonate, mother, father (or family member). To offer a deeper analysis, we undertook formal stakeholder analysis, where individual behaviors and related support was modeled along with relevance to the organization and available resources.

Interviews reflected that, shared contribution of all stakeholders ensure better performance and improved safety in neonatology department. We also considered different kind of data (e.g., protocols, user information) to dive into the possible situations that may cause poor adherence of protocol. As a first step, the protocols were investigated along with the interviews, to see the possible deviations in the healthcare protocol of neonatal care (Madar, Roehr, Ainsworth, Ersda, et al., 2021), which helped us to extract use-cases. *Use-case* is a term that describes the functional requirements and explain how stakeholders can interact with the system. Table XXX mention few examples in which each use-case is addressed through possible scenarios. i.e., a success scenario and when AI Coach predicts that support is desired or intervention is necessary.

Table usecases: Scenarios (use cases) in neonatal care

|  |  |  |  |
| --- | --- | --- | --- |
| **Use-case (step)** | **Success Scenario** | **Possible Intervention desired** | |
| **When** | **Possible Causes of Failure** |
| Access neonate breathing and tone | Neonate is breathing adequately and tone is good | Baby is not breathing or needs ventilation support | Mask is not appropriate  Sufficient FiO2 levels are not provided |
| Communication Behaviors | Members of medical team can communicate properly | Poor Communication due to different possible factors | Residents/team members can’t communicate |
| Family of Neonate may get affected | Delivery makes everyone happy | Development of postpartum depression in parents | Adequate support isn’t provided |

Once the requirements and expectations of the stakeholders collected, we designed the basic layout of AI Coach System that can support these use cases. The system layout is divided in two parts, server and a controller. Expert users are allowed to access and design the knowledgebase for AI Coach, that play role in the computational logic of the AI Coach, whereas, end users (i.e., medical team) interact through terminals to monitor and achieve these goals.

##### Knowledgebase for AI Coach

For every conversational agent or virtual assistants a related knowledgebase is vital, as it provides a basis for human computer interaction. Section 2.3 indicate how assistants or conversational agents can use domain knowledge to interact with its users. Here, the knowledgebase of system consist of ‘Support Knowledgebase’ and ‘Protocol Knowledgebase’ (See Figure XXX).

To generate knowledgebase, we modeled the collected requirements which can help the team to follow the medical protocols, through shared mental models (van Ments et al., 2021). The modeled scenarios reflected how different stakeholders can interact with each other to achieve a success scenario. For instance, Xu et al. explains how team should interact to ensure baby can breathe normally (Xu et al., 2022), or how mother can face postpartum depression (Weigl et al., 2022), or how can speaking up behavior of a resident can ensure risk free environment and help in proper execution of protocol (Doornkamp et al., 2022). For this, we also explored when AI Coach needs to intervene and ensure the patient safety and mental wellbeing by different support related messages. The interactions between stakeholders and the support related messages are considered as knowledgebase resource for the human interaction with AI Coach.

The above mentioned studies explore how social elements can be integrated into the design of AI Coach. For example, we studied how a healthcare practitioner may respond in case of development of postpartum depression among parents for a behavioral change. A similar design methodology was considered for the interaction of AI Coach, which aims to reflect the similar behavior towards a family member(s), i.e., to provide support to avoid postpartum depression (Weigl et al., 2022). Thus, the AI Coach interacts to avoid postpartum depression or help if it is developed in father (or a family member). If certain factor is found AI Coach notify the support messages from the support knowledgebase using visual and audio support. The aim of support knowledgebase is to use support messages to make users aware of what is happening or about to happen. Few example some support messages can be seen in Table supmsgs. Therefore, these models not only ensures that system meets its intended goals (adherence of protocol and mental wellbeing by monitoring), but they also manifest the possible behaviors of AI Coach, which may vary as per scenarios. The dynamic behaviors of coach are discussed below, while the related interfaces will be discussed in Section 3.2.

Table supmsgs: Example Support messages that can be used for human coach interaction

|  |  |  |
| --- | --- | --- |
| **State** | **Meaning** | **Example Support Messages** |
| epidural\_used |  | * If epidural is used, then there are more chances of development of postpartum depression |
| birth\_dev |  | * As there is a deviation in birth protocol, an elevated fear or stress levels is expected * you know if there is a deviation in birth, this can increase the fear levels in the patient and can lead to stress too * Support is desired as there is a deviation in normal birth protocol |
| stress | Pppd Father needs support | * In stressful environment, there are more chances to develop postpartum depression. * Father needs support too, as there are high chances of developing postpartum depression. |

##### Computational Logic and Human Coach Interaction

A basic human-coach interaction enables medical professionals to monitor their progress and avoid any possible medical errors. So, AI Coach helps to avoid forgetfulness or ensure mental well-being of the people on the floor. We aimed for a meaningful and limited interaction, so minimal attention and understanding is required from the medical team with a complete functionality.

For communication model, we considered not only the input from a user, but also considered medical protocols and the data related to desired support. Thus the coach analyzes a scenario based on input from the user and then communicates with the user with respect to the different use cases addressed through the shared mental models (e.g., see Table usecases). This communication can benefit the team in different ways, for instance, for protocol monitoring, action impact assessment, future proposals and reminder notification (See Figure clogic).

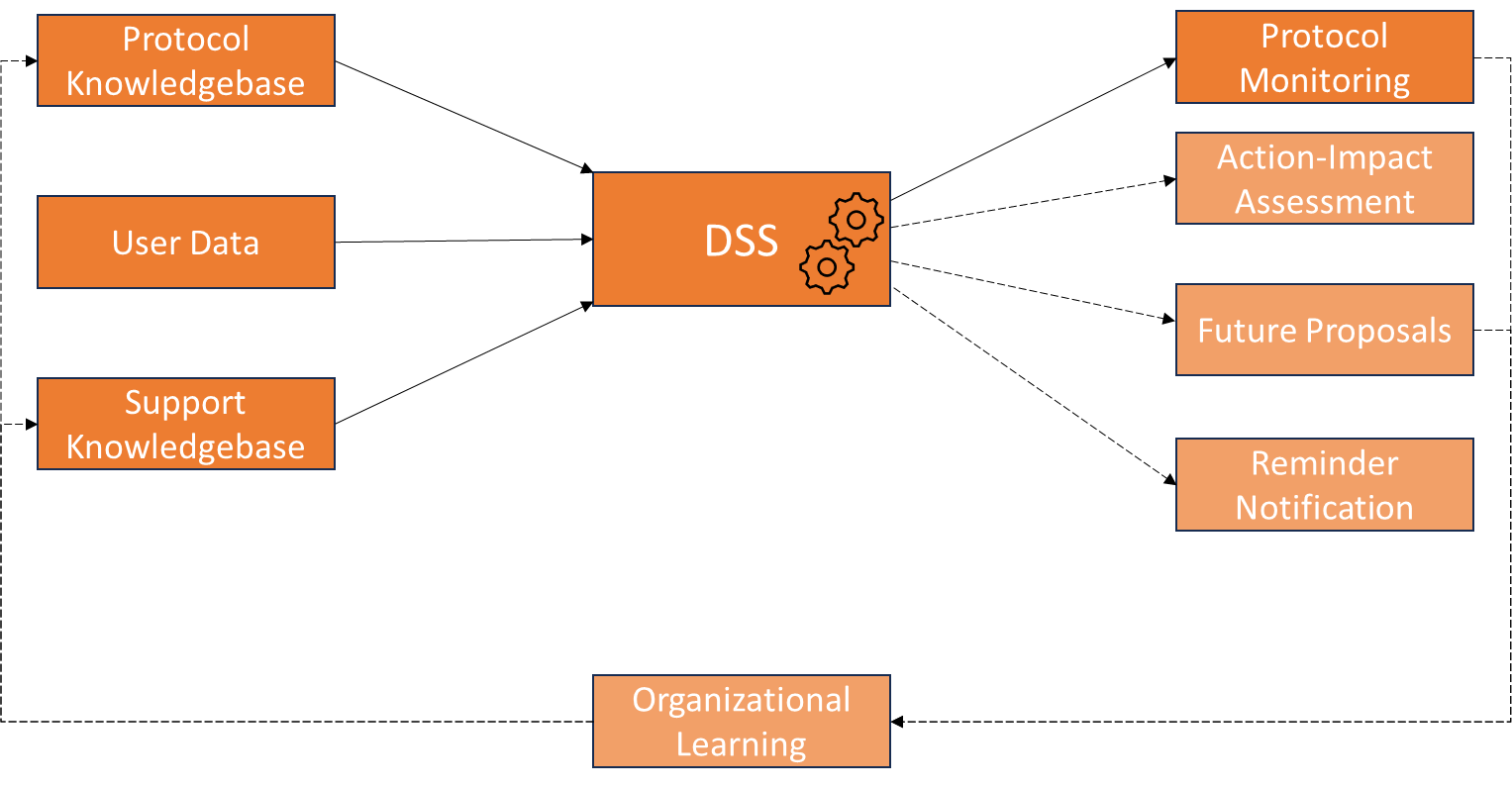


Figure clogic: Computational Logic of AI Coach

As an example, consider the example discussed earlier, i.e., development of postpartum depression (Weigl et al., 2022). For protocol monitoring, there are many factors that can cause depression e.g., deviation in normal delivery (represented by state birth\_dev), or stress (stress). To monitor depression (mppd; pppd), human coach interaction will estimates these factors. If these factors are found (Kowalski, 2011), it will communicate and provide the related support through the designed support knowledgebase. For instance, for state birth deviation (birth\_dev), AI Coach will monitor if mother is in labor (baby\_deliv) and if there was riskiness of giving birth (riskiness). If, both of the factors (states) are observed, AI Coach will enable the monitoring of deviation in birth. Therefore, a related monitoring message can be ‘Please make sure that mother is in labor and there is a risk of giving birth’. Similarly, when mental support is monitored (mental\_supp), then AI Coach ensures that the therapist has been referred (refer\_therapist). So, a related support message can be ‘Please make sure that therapist is referred’.

In other words, human coach interaction acts as a decision support system that manages and help the medical team to make decisions about the current situation(s) in neonatology which cannot be easily specified beforehand. These support messages are designed by the expert users. Here, it is to be noted that we didn’t consider the mental models of the team for communication, for this purpose we rely on the computational logic of the adaptive network oriented models (Treur, 2020). Moreover, we have implemented the functionality of ‘Protocol Monitoring’, however, further research is needed to implement the rest of the functionality of the coach, or extend it as per needs. Open source code is available here.

## User Interaction of the AI Coach

This section presents the user interfaces for the AI Coach system. Preliminary, there are two type of stakeholders who may use interface of AI Coach system, i.e., the expert users who design the functionality of the coach and end users who will interact with the coach. Each user has to login to the AI Coach system to use the related functionality. The considered constraint was that the interaction was intended mainly for the baby birth. It is to be noted that the related requirements were identified by the invited interviews which involved the stakeholders from the organization, AI experts and experts from psychology.

### Expert User Interface

This interface is meant to model the behaviors of the involved stakeholders and the role of AI Coach in assisting them. Therefore, the main aim is to design and reflect the adaptive behavior of AI Coach which varies as per scenarios.

It is to be noted, that all the models are based on the shared mental models which are subjected to preliminary analysis and simulation by AI experts and researchers in the same domain, for example, in neonatal care, the work from Doornkamp et al. which specifies the communication behavior among the medical team to ensure patient safety (Doornkamp et al., 2022), or by Weigl et al. which identifies the factors that can play role in postpartum depression (Weigl et al., 2022).

When an expert user (modeler) is designing the AI Coach communication model, (s)he has to specify a logical name for the model (See Figure statenames - a). The name should be self-explanatory which can be helpful to search the model. Moreover, the modeler has to specify the number of actors who would be interacting as a team. This information helps in selecting the model for the human coach interaction. Once the users information is there, he can design the model using the states information.

|  |
| --- |
| A screenshot of a chat  Description automatically generated |
| (a) |
| A screenshot of a computer  Description automatically generated |
| (b) |

Figure statenames: a) specifies the name of the conceptual model. b) Specifying the related state names

#### Incoming Connections

Figure statenames – b indicates the incoming states at the base level, for example, here it can be seen that Y has two incoming connections that is from states Input X1, and Input X2 (See Section 2.1 for further details). User has to specify different details related to a state, i.e., state name, the user name or the friendly user name, the type of the state (internal or mental state/actionable state) and then the respective incoming state. An ‘internal state’ is a state which does not need user monitoring and will not be shown to the end users, but, they are part of predicting certain behavior of a state. While, an ‘actionable / observation state’ is a state that needs input from the user and thus involved in the monitoring process. First column id is auto generated, it is unique and is used for further specification of any state (see below).

In Figure statenames – b, it can also be seen that X1 has one incoming connection i.e., X1, so if a state (like X1) has an incoming connection from the state itself, the aim is to maintain the state value over the period of time (for further details see Section 2.1). A similar pattern is developed for the adaption layers, i.e., first order level and the second order level. However, the states on these levels can only be the internal states.

#### Connection weight, Speed and Initial Value of a state

Figure specs indicates the connection weights for the incoming states for the complete model specified in Figure statenames. These weights can be either a number ranging between 0 – 1 or a state specified by id. A state having parameter as a state id (e.g., X1, or X2), indicates that the behavior of the a state will vary as per values of the state with that id.

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |

Figure specs: a) Connection weights b) Initial values and Speed factor for the incoming states

#### Impact Computation of a state

Figure combfunc indicates how can a modeler choose the combination function (For theory Section 2.1). By default Euclidean function is selected, however, user can modify the function by choosing it through dropdown list. After choosing the combination function, the modeler has to choose the respective parameters. Modeler can also add or remove another combination function by choosing ‘+’ or ‘-’ respectively. Once the modeler specifies these specifications, (s)he can save them. Now the designed model is ready to be used for a human coach interaction.

|  |
| --- |
|  |
| Figure combfunc: Combination Function Specification of the model states. |

### End User Interface

End users can interact with AI Coach through an interactive interface, with a visual and audio support. AI Coach selects a scenario based interaction, which depends on multiple factors, for example, number of participants, current use case and so on. As a precondition an end user should be aware of how to interact with AI Coach. Figure monitoring shows the interface of AI Coach, responsible for monitoring of neonatal care protocols, where each step is represented by a state (rectangle).

AI Coach uses network models to provide human coach interaction. For example, let’s consider the designed model by Weigel et al. The model reflect different factors, which may lead to postpartum depression (e.g., if mother is nulliparous, epidural is used). If a mother develops postpartum depression, there are higher chances for the other parent (or family member) to develop postpartum depression. Thus AI Coach needs to follow, monitor the development of postpartum depression. The main aim of this interaction is to assist the stakeholders to follow certain steps which might be skipped leading to a medical error. For example, the medical team has to refer a therapist (ref\_therapist) or parent needs to ask for support (ask\_sup) if there is stress (stress). In other words, these models provides a platform to AI Coach interact with the medical team to assist them, leaving no chance of possible error (for modeling details please see (Weigl et al., 2022)).

AI coach is a human based coach, and it interacts as per modeled behaviors of the individuals in the team. These behaviors are dynamic in nature and therefore doesn’t have predefined steps except the initial factors in the main protocol. Figure monitoring illustrates how AI Coach provides interaction based on the causal model designed by Weigel et al. The interaction starts by showing some initial states of the main protocol. Figure monitoring a indicates the initial factors (shown by grey rectangles – e.g., nulliparous mother, epidural used) that might be observed by the medical team. If these states are observed during the process, these states need to be marked as ‘monitored’ or ‘observed’ (marked as green). Marking these states will enable the end users to keep the track. Moreover, when a user clicks any state state, the coach shows further states known as ‘predictable states’ (grey rectangles – e.g., risk and mppd), which need to be monitored to follow the protocol further. If medical team somehow misses these steps, this may cause an error in the protocol. AI Coach plays its role to avoid missing this step by ensuring its monitoring.

|  |  |
| --- | --- |
|  |  |
| Figure monitoring: a) initial states of model by Weigl et al (Weigl et al., 2022). b) Observed states are shown as green, while initial/predicted states are shown as grey states | |

The interface of the coach is kept simple, and is designed to handle time constraint situations and to perform their role with the minimum involvement of AI Coach. An audio support is also added, so that if a user is not able to track a predictable state, the coach assists the user to ensure that the particular state is also monitored. For example, if the user wants to make risk as ‘observed’, AI coach will ensure that riskiness is observed only if age of the mother is between 25 – 29. Therefore, an audio is played (e.g., Mark age as monitored) to ensure is risk is there. We also added a preliminary tone support for audio. To do this, we performed sentiment analysis of the text (e.g., state name, or related message) by huggingface library[[1]](#footnote-1). For instance, if a negative state (e.g., stress) is observed by user, the audio tone of coach is high (pitch = 0.9) and rate = 0.7), while a positive tone is bit low (pitch = 0.8) and rate = 0.85). The implementation details of the coach can be found in appendices.

# System Evaluation and Discussion

Preliminary discussions helped us to clarify the system boundaries and the basic objectives of AI Coach (See Section 3.1.2.1). We considered them as the basis for system evaluation. Therefore, system evaluation was based on, e.g., a) how well AI Coach interaction can be designed by the system, b) how far it can meet the consumers’ needs, and c) what is the usability of the system and AI Coach in terms of its assisting the medical team. Table feedback mentions the feedback and related score of the questions that were asked from 5 participants (applied psychologist = 1; software developers = 2; AI researchers = 2) to evaluate the system.

It is to be noted that, the knowledgebase (models used by AI Coach) is to be designed, simulated and verified mathematically before using AI Coach. However, to give an essence of how it can be created, we explained some test models that can be designed and used dynamically by the AI Coach system. For this we gave them a brief walk-through tutorial which included a) the goals of the AI Coach system, b) the knowledgebase creation of AI Coach, c) interaction with the AI-Coach, d) feedback questions, and e) an informal session about the coach. This session not only helped us to study the look and feel of the users (expert + end users) for human coach interaction, but it also helped the participant to understand how AI Coach will interact with the end users. Following we explain the results of using the AI Coach system.

## Test Model

A sample model was designed to see how far it is easy to design the model using the AI Coach system. This model is inspired from newborn life support (Madar, Roehr, Ainsworth, Ersdal, et al., 2021), and uses adaptive network modelling technique (Treur, 2020). The model designed in this phase is used for protocol monitoring and the results mentioned in Section 4.2 are compiled accordingly.

Therefore, the participants of the study had to play the role of a) an expert users and b) an end users. This is because, our focus is investigate how easy is to use the system to design the coach and using it. Moreover, following this will develop a deeper understanding of the system and algorithm for protocol monitoring. The example model is mentioned in Figure samplemodel, where each state specifies actions to be monitored by AI Coach. These actions are similar to the actions, which are to be performed by medical staff. Here state XXX

Figure samplemodel: Sample model inspired from life support neonatal care (Madar, Roehr, Ainsworth, Ersdal, et al., 2021)

## Results and Discussion

All the participants were explained the motivation of the designed model, and all the participants were successful to design the model (average time = XX mins). Participants found the framework quite easy and useful to develop a human coach interaction, which will serve in a dynamic manner. All users commented that the interactive visualization could help and save time to develop such coaches, which are not only capable of providing human computer interaction to reflect human behaviors, but also can be customized and used for different purposes.

A participant also mentioned that he used to develop rule based chatbots, which rely on structures like decision trees, however, this system can enable users to develop a human-bot interaction to address graph based structures. Moreover, writing sample messages were also found to provide human-bot interaction. Overall, positive user experience was noticed for the coach design.

As a next step, participants were briefed, that the models they designed so far can also be used for monitoring a neonate. For simplicity, only the designed model was considered for monitoring through human coach interaction.

Table feedback: Participants feedback on usefulness and usability of AI Coach (5 point Likert Scale)

|  |  |  |
| --- | --- | --- |
| **Statement** | **Mean** | **S.D** |
| The system is easy to use |  |  |
| I find the system design is helpful to develop the human coach interaction |  |  |
| The system helped me to understand the flow of AI Coach which will interact with the people |  |  |
| AI Coach can help to achieve the goal, i.e., reduction of medical errors |  |  |
| The interaction of AI Coach was useful |  |  |
| The reminder prompts dynamically generated are useful. |  |  |
| Audio utterances were helpful to keep the flow of AI Coach |  |  |
| I enjoyed real-time interaction with AI Coach |  |  |
| I am satisfied with my experience using the system and its coach |  |  |
| I would use the system to design my coach. |  |  |

Table feedback mentions the results of the experience of participants about the human coach interaction.

Overall, all the participants were quite impressed how AI Coach system can be used to generate human coach interaction, however, they also reflected that some conversational elements (like images) can help the monitoring in a better way. Another participant mentioned that it can indeed provide a real time interaction in neonatal care, however, training would be required to interact with the coach.

Our focus was to designing an assistant, which can help to resolve contextual understanding, however, we didn’t use natural language understanding. As a result, participants didn’t felt the warmth of (humanoid) emotions and connectedness with the coach. Lastly, one participant mentioned that on one side bot development is automated, but on the other side it is similar to other chatbots. He further explained that this coach does not have the ability to adapt the new information at runtime, i.e., it only relies on the static knowledge that is preprogramed by the developers. Following section mention further limitations of the study.

**Scrolling is not encouraged. How to map expectations vs real time usage**

# Limitations

In this section, we would like to mention the limitations of the study. Initially, we will explain the limitations in the design of AI Coach, then we will explain the limitation in analysis and results of the study.

While discussing the limitations of the design of the framework, we would like to mention that the basic goals of the framework were to provide a real-time interaction among a team, with the aim to assist healthcare professionals if they forget a step in a medical protocol. The user needs and goals were taken as a primary objective, however, some requirements still need to be incorporated. Few examples are, rapport creation (e.g., a message during a conversation ‘as a coach, I can help you, do you want me to remind you’), or maintaining the relationship (e.g., reminding users if they forgot to monitor a step’), adopting attachment style during human coach interaction (Roelofsma, 2013).

The framework has ability to integrate different features in future. For example, the domain knowledge should be increased by incorporating more mental models. This will help to improve domain knowledge and thus AI Coach can interact in a better way. The proposed framework enables handling of complex tasks based on human behaviors, however, it lacks the human emotions and personalization. Adding an avatar with emotions can be helpful to overcome this (Zmugg et al., 2015).

Unlike many other chatbots, AI Coach takes user input through button click, however, framework also allows to integrate sensors, which can improve the human experience by better understanding of environmental factors to predict an action. For example, one example can be integrating sensors for taking baby vitals’ data, to predict if the baby is breathing normally, and predicting the related outcomes which can be helpful in assisting the healthcare professionals working on the floor. Another example, can be predicting stress from sensors. As a result, stress can be reduced by music, so AI Coach interprets the stress levels through the environmental sensors, and may propose users to play a comforting music (Trappe, 2012). Similarly, adding few question answer sessions via sensors can improve the human-level understanding and intuition, which can improve human coach interaction.

# Conclusion

# Acknowledgements

Doctors

# References

# Appendices

1. **Interface of AI Coach**
2. **Implementing AI Coach**

Conversational agents, also known as chatbots or virtual assistants, have gained significant attention in recent years due to advancements in natural language processing (NLP) and artificial intelligence (AI). These intelligent agents are designed to simulate human-like conversations, enabling users to interact with them through text or voice-based interfaces. The following literature review provides an overview of key research and developments in the field of conversational agents, exploring their applications, challenges, and potential impact on various domains.

1. Applications of Conversational Agents: Conversational agents have found applications in diverse domains, including customer support, healthcare, education, e-commerce, and more. Bickmore et al. (2010) discuss the use of virtual agents in healthcare to provide information, support, and behavior change interventions. They highlight the potential of conversational agents in improving patient outcomes and increasing engagement. Similarly, in the realm of customer support, Li et al. (2019) examine the effectiveness of chatbots in addressing customer queries and enhancing customer experiences. Their study demonstrates the positive impact of conversational agents in reducing response time and improving customer satisfaction.
2. Natural Language Processing and AI Techniques: The development of conversational agents heavily relies on advancements in NLP and AI. These technologies enable chatbots to understand and generate human-like responses. Jurafsky and Martin (2020) provide a comprehensive overview of NLP techniques, including language modeling, dialogue systems, and sentiment analysis, essential for building conversational agents. They discuss the challenges in achieving accurate language understanding and contextualized responses.

Furthermore, deep learning techniques, such as recurrent neural networks (RNNs) and transformer models, have been widely employed for training chatbot models. Vinyals and Le (2015) present Seq2Seq, an encoder-decoder framework based on RNNs, for generating responses in conversational agents. They highlight the importance of large-scale training data and model architecture in improving response quality.

1. User Experience and Personalization: Personalization plays a crucial role in enhancing user experiences with conversational agents. A study by Wang et al. (2020) explores the use of user-specific conversational agents for personalized news recommendations. They propose a method to capture user preferences and generate personalized news summaries. The study showcases the potential of conversational agents in tailoring content to individual users, thereby improving engagement and relevance.
2. Ethical and Social Implications: As conversational agents become more prevalent, ethical and social implications arise. Adversarial attacks on chatbots, where malicious users attempt to deceive or manipulate the system, have been a subject of concern. El Asri et al. (2018) discuss the vulnerabilities of conversational agents to adversarial attacks and propose defense mechanisms. They emphasize the need for robust and secure conversational agents to mitigate potential risks.

Additionally, the impact of conversational agents on employment and human interaction is a topic of debate. Dignum et al. (2020) analyze the ethical challenges of deploying conversational agents in various industries, considering issues related to transparency, accountability, and job displacement. They argue for responsible design and deployment of conversational agents to ensure societal benefits and minimize negative consequences.

Conclusion: The literature on conversational agents highlights their wide-ranging applications, technological advancements, and associated challenges. From healthcare to customer support, conversational agents offer opportunities for improved outcomes and user experiences. However, ethical considerations and the need for robust security remain crucial. As research and development in this field continue, further exploration of conversational agents' capabilities and their impact on society will undoubtedly shape their future applications and implications.

1. <https://huggingface.co/transformers/usage.html> [↑](#footnote-ref-1)