EDITORIAL

Design criteria for AI-based IT systems

Heinz U. Lemke[1] · **Franziska Mathis-Ullrich**[2]

Abstract

Purpose This editorial relates to a panel discussion during the CARS 2023 congress that addressed the question on how AI-based IT systems should be designed that record and (transparently) display a reproducible path on clinical decision making. Even though the software engineering approach suggested for this endeavor is of a generic nature, it is assumed that the listed design criteria are applicable to IT system development also for the domain of radiology and surgery.

Methods An example of a possible design approach is outlined by illustrating on how to move from data, information, knowledge and models to wisdom-based decision making in the context of a conceptual GPT system design. In all these design steps, the essential requirements for system quality, information quality, and service quality may be realized by following the design cycle as suggested by A.R. Hevner, appropriately applied to AI-based IT systems design.

Results It can be observed that certain state-of-the-art AI algorithms and systems, such as large language models or generative pre-trained transformers (GPTs), are becoming increasingly complex and, therefore, need to be rigorously examined to render them transparent and comprehensible in their usage for all stakeholders involved in health care. Further critical questions that need to be addressed are outlined and complemented with some suggestions, that a possible design framework for a stakeholder specific AI system could be a (modest) GPT based on a small language model.

Discussion A fundamental question for the future remains whether society wants a quasi-wisdom-oriented healthcare system, based on data-driven intelligence with AI, or a human curated wisdom based on model-driven intelligence (with and without AI). Special CARS workshops and think tanks are planned to address this challenging question and possible new direction for assisting selected medical disciplines, e.g., radiology and surgery.

Introduction

One major question that was addressed in panel discussions at CARS 2023 related to transparency (or lack thereof) with respect to truthfulness/trustworthiness, complexity and incomprehensibility when employing AI assistance in decision making in health care, specifically:

How should AI-based IT systems be designed that record and (transparently) display (incl. the machine learning part of the AI system) a reproducible path on clinical decision making?

✉ Heinz U. Lemke
hulemke@cars-int.org

Franziska Mathis-Ullrich
franziska.mathis-ullrich@fau.de

1 International Foundation for Computer Assisted Radiology and Surgery – IFCARS, Küssaberg, Germany

2 Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen-Nuremberg, Germany

Answers to this question focused on outlining specific design criteria for AI-based IT systems and a suggestion for a DIKMoWi AI-related process transparency. DIKMoWi stays for moving from Data, Information, Knowledge, and Models to Wisdom-based decision making, and is an appeal for a human model-driven intelligence with, but possibly also without AI [1]:

Historically, a related set of questions and answers have been discussed in a CARS supported panel some 20 years ago in Dresden, Germany, on the topics of telemedicine, robotics and AI [2]. An important signal from this panel consisting of physicians, (computer) scientists, engineers, healthcare providers, philosophers, and theologists was, that the different professions involved in health care have to:

1. work closely together,
2. find respect for each other's point of view, and to
3. balance view point summaries reflecting society as a whole.

Because of the expected transformational impact of AI systems and translating this signal into our times, the design and employment of AI-based IT systems should be a multidisciplinary undertaking and not left to one particular stakeholder or interest group alone.

It was also stated in that panel in Dresden, that ethical questions on research with respect to human genetic technologies are more important than artificial intelligence and telemedicine. Now, 20 years later, it can be observed that research and development about AI-based IT systems have at least the same significance in health care as human genetic questions have had in the past. Recent and expected advances in the field of SLMs, LLMs and GPT like systems make it, therefore, obligatory to address how AI-based IT systems should be designed.

Methods

In order to design deep learning models (generative models) that can generate "realistic" or "truthful" text, audio, images or videos in the domain of health care, a 10-step procedure [1] has been suggested (Fig. 1), and if appropriately implemented, can be considered to be part of a modest GPT-like system for radiology and surgery.

These 10 steps may also be divided into 5 phases as described in reference [1]. Steps 1 and 2 correspond to the data-driven phase, step 3 to the information-driven phase, step 4 to the knowledge-driven phase, steps 5–8 to the model-driven phase and steps 9–10 to the wisdom-driven phase. Availability and quality of domain specific data, information, knowledge, and models are key to achieving a high truth value or even wisdom-related output from such a GPT-like system.

Of particular importance are the quality of the models derived from the selected/relevant corpus of knowledge on which the subsequent model-driven steps are based. These models should be made transparent and verifiable in order to enable a Model Guided Medicine (MGM), as being pursued within the CARS and other communities. In principle, the focus on situational and process awareness is given in all phases, when moving from data, information, knowledge, and models to wisdom-based decision making [1]. This somewhat expands the classic DIKW (Data, Information, Knowledge, Wisdom) hierarchy as proposed in the past and augmented for the present by Liew [3], with the concept of DIKIW. In the context as given by the needs for the CARS community, the "I" standing for Intelligence in DIKIW is replaced here by "Mo" for models, thereby creating the acronym "DIKMoWi". This is in line with the observation, that intelligence/awareness is needed in all phases as given in the DIKW hierarchical structure and that the creation of models represents a special level of intelligence, filling the

link between knowledge and wisdom, thereby justifying an explicit position in the DIKW structure.

The 10-step procedure is defined as follows:

1. The first step of this procedure is the encoding step, which in principle transforms user requests and prompts given as text into numbers or, if given as an input image, into a low-dimensional latent space representation.
2. The second step implies token generation. A token can be considered as a content/feature vector containing a specific characteristic digital pattern, e.g., for a word in a sentence or a patch in an image.
3. The third step implies embedding the tokens into a sequence in which the tokens are weighted according to their importance in the given context.
4. The fourth step is the attention processing step that focusses on the essentials of the question to be answered by referring to the relevant corpus of (non-) curated training data such as books, articles, images, online encyclopedias, letters, knowledge graphs and other sources.
5. The fifth step deals with model generation and updating such as parameter tuning in ANN (Artificial Neural Network) models and tests whether the model is complete and the query is answerable.
6. The sixth step implies temperature and sentiment sampling giving a measure of probabilistic appearances of words, patterns and text.
7. The seventh step deals with token synthesis for corresponding text or image generation.
8. The eighth step secures that the best text or image for the given query has been generated.
9. The ninth step performs the text or image integration and statistical estimation of goodness by means of a beam search as part of a quality assessment of different textual or image candidates.
10. The tenth step is the decoding step, which generates a machine formulated text, audio, images or videos for human consumption or machine learning. Some LLM criteria such as the use of knowledge models, for example, the use of mathematical representations of domain knowledge may assist in this step.

Following the given 10 steps and 5 phases, some human controlled verification and validation should take place, in particular, a procedure for testing and approving GPT model correctness and applicability (incl. biases detection). For example, in order to secure appropriate transparency for bias detection, steps 4 and 5 need to be made explainable in the sense of the principles of explainable AI (XAI).

These and other systems design criteria [4] need to be observed for the 10-step procedure discussed above or any
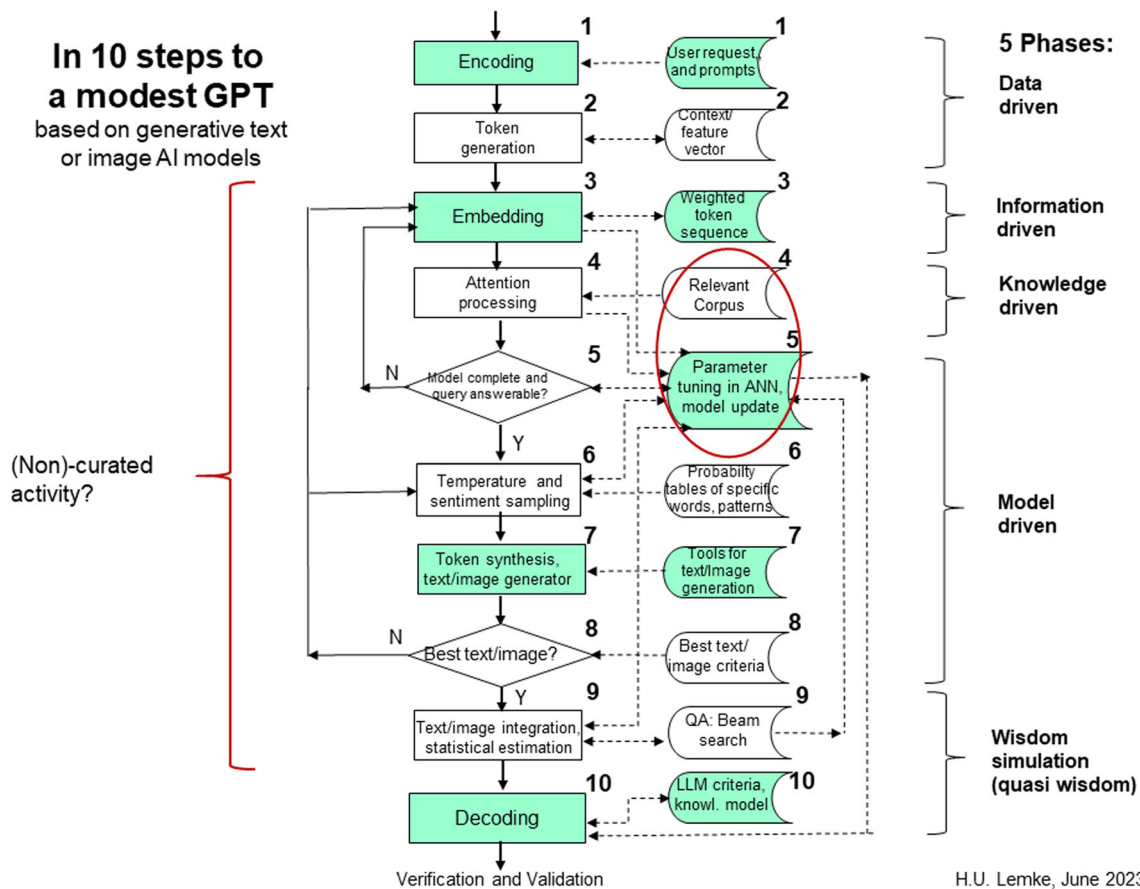
Fig. 1 In 10 steps to a modest GPT

other related algorithmic structure, that may be used for an AI-based IT system design.

With the goal of human intelligence augmentation through natural partnership with AI, when utilizing AI-based IT systems in healthcare, we need to define certain design criteria to provide functionality, performance, safety, and usability. In this scenario, clinicians or medical staff are supported and not replaced by AI systems that provide active support through cooperation. Natural and active support in the healthcare system can only be provided sufficiently by AI-based systems through fulfilling the requirements of system quality, information quality, and service quality, which are paramount during the development process.

System quality encompasses critical aspects such as safety, security, and privacy. From both regulatory and ethical standpoints, it is imperative to maintain rigorous control over personal data, ensuring transparent disclosure of its usage. The utilization of data should be restricted, only permitting access as required. Furthermore, while the availability of vast quantities of high-quality data can augment the efficiency and precision of systems, it is equally important to exert control over which specific data subsets are employed to safeguard user interests and maintain system integrity.

Information quality is paramount in establishing trustworthiness of a system. Particularly in the medical field, where IT systems offer recommendations and decisions to healthcare professionals, the trust placed in these systems is crucial. Trust may be achieved through understanding and explainability, such that a recommendation is clear and comprehensible to the human user. A transparent decision-making process ensures that users can understand, verify, and consequently use the outcomes presented to them.

Service quality, particularly in terms of a system's performance, hinges on the availability of high-quality data and the mutual learning between humans and systems, requiring effective communication and interaction. On one hand, humans possess non-formalizable, implicit knowledge and deep domain expertize that needs to be accurately captured. On the other hand, the system excels in processing and analyzing vast datasets.

Here, we loosely base the definition of design criteria for AI-based IT systems in healthcare on the *Three Cycle View on Design Science Research* by Hevner [4]. Figure 2 shows an overview of the proposed design process. In the center of the design process is the design research. The design research is mostly performed to research and generate new AI models
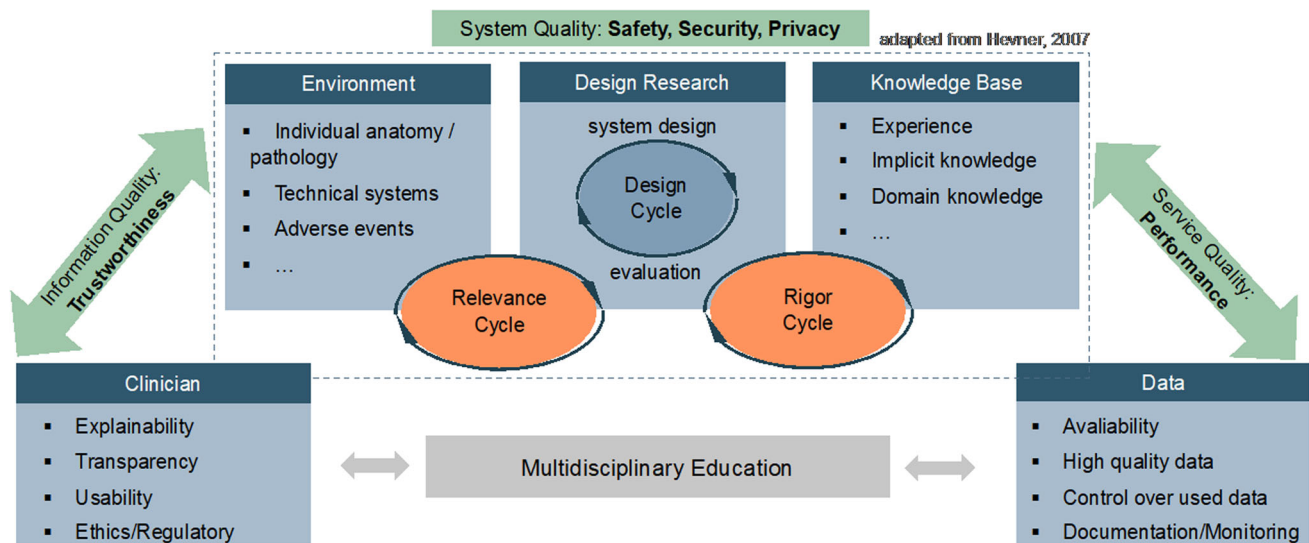
Fig. 2 The design cycle of AI-based IT systems in healthcare. The design cycle is connected by the relevance cycle and the rigor cycle. System quality, information quality, and service quality are crucial aspects and strongly depend on involved clinicians and high-quality data. Common understanding is supported by multidisciplinary education and communication between the scientific disciplines (e.g., medicine and technology), adapted from Hevner [4]

with novel tasks in healthcare applications. After implementation and training of a new design, the model is evaluated through extensive ablation tests. During this phase, an AI-based system must undergo several iterations that evaluate the performance of the algorithms and system architecture, including its accuracy, inference speed and other.

Design research is closely connected to the knowledge base through the rigor cycle as well as to the envisioned working environment of the AI-based IT system through the relevance cycle.

The rigor cycle connects the design phase—often performed by computer scientists and engineers—to the medical knowledge base. This includes non-formalizable or implicit medical knowledge and domain specific knowledge, such as experience of a certain surgical discipline.

The relevance cycle, on the other hand, connects the design research to the environment where an AI-based system is to be deployed to demonstrate performance under clinically relevant conditions. In healthcare, this may include patient-specific properties, such as individual medical records or patient-specific anatomy and pathologies, as well as risk factors that may lead to adverse events during treatment. In addition, the environment includes factors such as other technical systems and medical devices in the operation room as well as the medical staff and clinical workflows. It is particularly important to take the AI system's environment and circumstances into account during design. Both, the rigor cycle and the relevance cycle, must be iteratively advanced to render a system completely suitable and supportive for implementation into the clinics.

The design and development of functioning and suitable AI-based assistance systems for healthcare are highly dependent on the availability of large amounts of high-quality data. These data may include annotated radiological or surgical images, health reports and disease progressions, omics data and many others. However, data from several sources often differ in quality during acquisition, analysis, and modeling, such that certain regulations and rules become paramount, which must be agreed upon by the international research community. Researchers must agree on how to identify good data and who can do this. An example of such an agreement is the Surgical Data Science Initiative [5]. In addition to acquisition of high-quality data, control over the data used by an AI-based IT system must be given. According to FDA guiding principles, models that are deployed in the clinics must be monitored and well documented regarding usage of data, maintenance, safety, and performance. Monitoring also includes management of risks when re-training a network with new data after deployment, including overfitting, unintended bias, or a drift in the dataset.

In addition to carefully handling of medical data, clinicians and medical staff must always be consulted during the complete design and development process of AI-based IT systems in healthcare. With the goal to allow for augmentation of the clinical user through the AI system, these assistive systems must allow for usability and transparency during usage. In many cases, explainable AI has been deployed to communicate certain functionalities and reasoning of an AI-based system to the user. Clinicians and medical staff, but also regulators, lawyers, and further guidelines (e.g., FDA,

MDR) are to be consulted to adhere to ethics and regulations of AI-based systems in healthcare.

The challenge of design, development, and deployment of assistive systems in the clinical workflow, require multidisciplinary research and close collaboration between medical and technical experts. Nonetheless, clear understanding of the clinical requirements on one hand (technical experts) and a basic knowledge about functioning and feasibilities of AI methodologies on the other hand (medical experts) is often lacking. Thus, it is crucial to educate the emerging scientists in both, the clinical and technical fields, beyond their own disciplinary boundaries to foster knowledge and understanding of the partner discipline in AI research for healthcare.

Results

Assuming that many engineers and scientists consider certain state-of-the-art AI algorithms and systems (such as LLMs or even GPTs) to be incomprehensibly complex, how can we expect patients, physicians and healthcare providers be well advised to actually using them? In line with past discussions [2], some critical questions need to be addressed such as:

1. What basic value system, if any, should be reflected in AI-based IT systems that are designed to assist in clinical decision making?
2. Why do we need to re-examine communication behavior of humans with intelligent and networked machines?
3. How should IT systems be designed that record and (transparently) display a reproducible path on clinical decision making?
4. How can possible negative side effects in the use of AI-based IT systems be minimized?
5. Who assumes responsibility for damages incurred through the use of AI systems in health care?
6. Where and when can different concepts and models relating to AI-based IT systems be realized in a controlled (certified?) and verifiable manner?
7. How should AI-based IT systems be employed as an empowering tool for all stakeholders involved in the domain of radiology and surgery, in order to enable a wisdom-oriented healthcare system?

As a generic answer to the concerns reflected in these questions would be a concept that each of the prime stakeholders in a patient-oriented health care, e.g., patients, physicians/medical staff, healthcare providers, researchers, will eventually have access to an AI system that has been designed to reflect their specific value system (bias). In the respective AI system, this will primarily be reflected in the algorithmic steps for selecting the relevant corpus of knowledge and the step of parameter tuning of the corresponding machine learning models. In some cases, also the subgroups of the different stakeholder categories will have their specific AI system, for example for different clinical disciplines of physicians such as radiologists and surgeons.

A possible design framework for a stakeholder specific AI system would be a (modest) GPT based on an SLM [6], into which the 10 procedural steps as indicated in Fig. 1 are appropriately mapped. The relevance and the rigor cycle as suggested by Heyner et al. [4] augmented with situational details, for example, described in appropriate domain specific knowledge graphs, could serve as fundamental tools in the design of user specific AI systems. Provided with sufficient training data, the tools and code already publicly available for implementing LLMs can make them a powerful AI assistant in specific areas of radiology and surgery.

Discussion

Questions which surface from the theme of AI and MGM relate to why, how, where and when these methods and tools will impact an increasingly AI-based (biased!) decision-making process in health care.

For example, how can MGM become an enabler for moving from a data-driven machine learning/AI to a model-driven machine learning/AI in Medicine? In particular, how can certain AI concepts such as transparency, predictability, cause-effect reasoning, cooperativeness, agent and safety driven, data and model interoperability be promoted with MGM? Should model-driven machine learning be the basis on a transparent machine intelligence and replace a current rather black box-based artificial intelligence? Finally, what role will a model-based domain evidence play when it comes to verifying, validating and to evaluating AI algorithms?

Market forces will likely determine the granularity level of stakeholder/user specific AI systems as well as the usage of SLM methods and tools in the domain of health care. Special CARS workshops and think tanks are planned as enablers for this new direction for assisting selected parts of medicine, e.g., radiology and surgery. A long term aim of CARS for professionals in these disciplines is a personalized AI system. A fundamental question for the future remains whether society wants a quasi-wisdom-oriented healthcare system based on data-driven intelligence (with AI) or a human curated wisdom, based on model-driven intelligence (with and without AI)?

References

1. Lemke HU (2022) Moving from data, information, knowledge and models to wisdom-based decision making in the domain of Computer Assisted Radiology and Surgery (CARS). Int J Comput Assist Radiol Surg 17:1513–1517

2. Weizenbaum J (2003) …der Mensch, nicht die Maschine ist das Maß. In: Niederlag W, Lemke HU, Bondolfi A, Rienhoff O (eds) Ethik und Informationstechnik am Beispiel der Telemedizin, Health Academy, 02/2003, Dresden, pp 58–63

3. Liew A (2023) DIKIW: Data, Information, Knowledge, Intelligence, Wisdom and their interrelationships. Bus Manag Dyn 2(10):49–62

4. Hevner AR (2007) A three cycle view of design science research. Scand J Inf Syst 19(2):87–92

5. Maier-Hein L, Vedula SS, Speidel S, Navab N, Kikinis R, Park A et al (2017) Surgical data science for next-generation interventions. Nat Biomed Eng 1(9):691–696

6. Guo Z, Wang P, Wang Y, Yu S (2023) Improving small language models on PubMedQA via Generative Data Augmentation. https://doi.org/10.48550/arXiv.2305.07804

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.