UDC: 61:004.8:001.891 https://doi.org/10.15407/jai2024.01.010

O. Mintser
Shupyk National Healthcare University of Ukraine, Ukraine
9, Dorohozhytska St., Kyiv, 04112
o.mintser@gmail.com
https://orcid.org/0000-0002-7224-4886

FORMULATING TASKS, INTERPRETATION, AND PLANNING THE IMPLEMENTATION OF RESEARCH RESULTS USING ARTIFICIAL INTELLIGENCE IN MEDICINE

Abstract. Strategic issues of artificial intelligence use in medicine are considered. Summarizing, as of today, AI supports doctors but does not replace them. It is emphasized that AI in healthcare typically solves important, but rather limited in scope, tasks. Difficulties in further implementation of AI are analyzed. **The aim of the study** was to address the analytical generalization of AI capabilities in healthcare, analyze the problems of using the Universum of medical-biological knowledge as a global unified resource, and conceptually justify the need to structure medical-biological knowledge, introducing fundamentally new forms of knowledge transfer in healthcare.

Conclusions made: 1. The goal of AI implementation should be to find a delicate, mutually beneficial balance between its effective use and the judgments of trained doctors. This is extremely important, as artificial intelligence, which may practically fully replace the labour of doctors in the near future, today is an issue that might otherwise hinder obtaining benefits from it. 2. AI will become an integral part of future medicine. Therefore, it is important to teach the new generation of medical interns the concepts and principles of AI application, to function effectively in the workplace. It is extremely important to develop skills such as empathy in AI. 3. A systematic approach to the continuous improvement of diagnostic and treatment processes and systems for patients, first and foremost, requires bridging the gap between accumulated medical knowledge and the logic and results of AI use.

Keywords: artificial intelligence, machine learning, deep learning, knowledge Universum, structured and unstructured data, coreference of medical information, knowledge mobilization, coding of medical information.

Introduction

Artificial Intelligence (AI) already allows for new discoveries and the refinement of processes across the entire healthcare continuum. It has become possible to utilize unstructured data to uncover countless interrelations among data elements, enabling the use of dynamic and context-rich data. This unconventional approach and analysis yield valuable information about human behaviour [1].

Neural networks have become more complex as computational power has allowed for real-time functional inference of queries. Transformers (i.e., deep learning models that differently weigh the importance of each part of the input data) have enabled natural language processing. This approach has increased the complexity of foundational computer models and the data arrays from which these models could be built, making them more powerful [2].

AI excels in well-defined tasks with clearly defined input data and a binary output signal that is easily verifiable. In these cases, AI can demonstrate its efficiency relative to a physician [3]. For example, in classifying suspicious skin lesions, where the input is a digital photograph and the output is a simple binary classification: benign or malignant neoplasms [4].

However, medicine differs significantly from other fields where AI is applied. Ethical, management, and regulatory considerations are crucial in designing, implementing, and integrating each component of AI applications and systems. Due to concerns about utility and safety, new programs generally must meet the same standards as other medical technologies. This will require a level of testing rigor similar to that used in other areas of medicine but may also create problems such as "data set shift," which can occur when there is a mismatch between realities and the data set an AI system operates with [1].

Research Objective

Analytical generalization of AI capabilities in healthcare. Examination of the challenges in using the medical-biological knowledge Universum as a global unified resource. Conceptual justification for the need to structure medical-biological knowledge and introduce fundamentally new forms of knowledge transfer in healthcare.

Obtained Results

The most significant progress in AI use is seen in economic sectors with easy access to structured data sets. This is understandable, as machine and deep learning models require large data sets for accurate situation identification, classification of possible event developments, and provision of potential forecasts.

In medicine, the situation is opposite with the availability of information. Firstly, since patient records are often considered confidential, there is a natural reluctance among institutions to share medical data. Ideally, machine learning-based systems should continually improve as new data is added to their training set, but internal corporate resistance can hinder achieving this goal.

There is a strong dependence of AI results on the quality of data obtained during the examination and treatment of patients. The development of machine learning models requires well-structured training data about a phenomenon that remains relatively stable over time. Deviating from this leads to "overfitting," where AI gives excessive importance to false correlations in past data.

In 2008, Google attempted to predict the seasonal prevalence of influenza using only search queries entered into its search engine. The necessary data was anonymized and digitized, as this facilitated research and development [5]. However, as people's search habits drastically change every year, the model so poorly predicted the future that the developers quickly abandoned it.

Now, consider other drawbacks of artificial intelligence apart from the unavailability of appropriate data. Biased data collection processes used for model development can lead to potentially distorted processing results.

A significant problem can be the development of algorithms that are often used to study insignificant connections between patient characteristics and treatment outcomes (overtraining). This happens when many variables influence the results. It leads to algorithms that may work well in a training data set but give inaccurate results when predicting future events in real cases [4].

Data leakage is another concerning issue. The ability of a method to predict events beyond the training data set diminishes if the algorithm requires an extremely high predictive accuracy, as covariates within the data set can contribute to a false outcome.

One of the typical critical remarks about AI systems is the so-called "black box" problem. Deep learning algorithms generally lack the ability to provide convincing explanations for their forecasts. This could lead to people losing faith in the medical system altogether. Although this discussion continues, it should be noted that the mechanism of action of many drugs is poorly understood, and most doctors have only a basic understanding of diagnostic imaging tools such as MRI and CT. As a result, clinical recommendations and their application can be accompanied by erroneous decisions. In such cases, AI systems do not have the capability to legally protect themselves.

AI lacks human qualities such as compassion, and thus patients must understand that consultations are conducted by human physicians. Moreover, patients cannot be expected to immediately trust AI, a technology shrouded in distrust.

Among other problems associated with AI use, interpretational issues hold a special place. A key question in practical medicine becomes, what constitutes a patient's normal state of health. This question aptly illustrates one of the weaknesses in using artificial intelligence and machine learning in medicine in the form it is widely applied today [3].

There is also no answer to how bias in the "learning" method of AI and machine learning algorithms affects their functioning in a real clinical setting, and how to incorporate human values into these algorithms so that the results reflect the real problems faced by healthcare professionals.

The use of AI for automated analysis of pathomorphological data is of great interest. A major problem in computational pathology is that images of tumor tissue often vary in colour and scale in different research laboratories and medical institutions due to differences in tissue preparation methods and imaging tools. Additionally, the false evaluation of histopathological images and decision-making using tissue slides containing millions of cells can be time-consuming and subjective [6]. Automated image analysis has shown that spatial heterogeneity can make its quantitative assessment of pathological changes unreliable. Analyzing "hotspots" is the most common approach to quantitatively assessing unevenly distributed tissue parameters. The idea is to consider only areas with particularly high or abnormal values, which are considered characteristic of the parameter distribution. Understandably, its results critically depend on both the location and the size of the areas considered. When performed manually, the selection of both the location and size of hotspots tends to be highly subjective. This issue is known as data discrimination (data censorship), also referred to as algorithm discrimination. Essentially, data discrimination is a bias that arises when certain types of data or data sources are intentionally or unintentionally treated differently than others [7].

Many methods have been proposed to combat data discrimination. Recently in computer graphics, tiles - images used to create textures by laying copies of this image like tiles, such that the joints are unnoticeable, have become widespread. In tile graphics, a set of tiles (tileset) and a matrix of cells determining which tile will occupy each cell are used [8]. A pathology visualization method has been proposed for the quantitative analysis of images in full-slide histological section processing [9], using ensemble deep learning methods to employ modern algorithms to improve approaches to diagnostic training and detection of various subtypes of cancer [10]. There is well-known experience in using computational methods based on Convolutional Neural Networks (CNN) which led to the creation of an autonomous pipeline for the efficient classification of different histopathological images for various types of cancer [11]. We use the 4S algorithm (systematization, structuring, stability of states), related to the technology of creating stable morphological, histological, or other structures. The experience of its use provides a basis for cautious optimism. Thus, deep learning approaches are becoming leading tools in medical imaging machine learning, where they have been proven to provide valuable results in various tasks such as segmentation, classification, and prediction [12].

The issues of ensuring the correct functioning of artificial intelligence and machine learning programs in conditions of multiple uses must be resolved, along with the questions of changing classical approaches to statistical conclusions for interventions based on artificial intelligence and machine learning, and the issue of integrating findings from different research methods, including clinical, genomic, metabolomic, and environmental studies. The latter problem is particularly significant in terms of justifying intermediate generalizations regarding the state of individual body systems, identifying contradictions in the obtained diagnostic or therapeutic information, and formulating plans for further diagnostic strategy and prevention of possible complications.

Important issues remain in the sufficiency of diagnostic (prognostic) information for creating convincing and valid clinical conclusions, the necessity of using specialized high-tech procedures, and identifying systematic errors in the process of patient examination.

Finally, attention should be focused on the need for correct data usage. Today, data comes from many additional sources, including social networks, blogs, chats, product review sites, communities, web pages, emails, documents, images, videos, music, and environmental sensors. The capacity of various storage facilities is so large that it allows for the storage and easy access to vast portions of the recorded human knowledge and activity. Many people make parts of their medical documentation and personal genetic data available online for everyone. Denser information storage and faster computations have enabled practically real-time mathematical calculations, which can be used to search for data correlations that were previously considered in isolation.

But progress in data science is not just a matter of increasing productivity, speed, and storage volume. In addition to the type of information stored in libraries, data generated in organizations, and established systems designed for data collection and systematization, the use of fundamentally new information processing technologies generated by both humans and machines is absolutely necessary. The relevance of the problem is underscored by the fact that data is often chaotic and unstructured.

Significant difficulties are posed by the issues of coreference or referential identity—the relationship between names—components of an utterance, in which the names refer to the same object (situation) of extralinguistic reality.

In different databases, semantic compatibility of information can be ensured through the use of formalized attributes—codes. The coding of clinical information is an important element of its formalization and standardization. The main purpose of information coding, apart from ensuring the identification of the information object, is the implementation of the principle of data normalization. The principle of data normalization essentially involves the elimination of repetition and duplication of information in the subject area.

It is clear that the coding of clinical information must comply with industry standards. It involves creating codes according to specific rules and assigning them to an object or group of objects, allowing their names to be replaced with a few signs (symbols). Codes enable the identification of objects in the shortest possible way, i.e., using the minimum number of signs. Minimizing the number of signs identifying objects enhances the efficiency of collection, accounting, storage, processing (analysis), and information retrieval.

The basis of the information coding algorithm is its classification according to a hierarchical, faceted, or mixed principle: defining a set of objects subject to classification for solving specific tasks; identifying characteristics for subdivision into subsets; and choosing the optimal number of subset division levels, ensuring the convenience of practical use of the subject area.

The specific algorithm for coding clinical informational objects depends on the purpose of the classifier, the specificity of the information, and the nature of the applied tasks of the subject area. For example, databases of disease symptoms are primarily intended for automating the formation of electronic records of clinical protocols and creating diagnostic decision support systems. Syndromes (nosological forms), as informational objects, provide the subject scope for forming an expanded clinical diagnosis, solving medical statistics tasks, and clinical management.

An example of this approach's implementation is the International Classification of Diseases. However, it is primarily intended for statistical purposes and does not address various clinical tasks. To date, there are no universally accepted nomenclatures of symptoms, syndromes, and nosological forms. It should be noted that classifiers of complications and consequences of diseases have been proposed, as well as a classifier of pharmacotherapeutic drug groups for "establishing uniform requirements for the diagnosis and treatment of patients with various diseases." Obviously, the further creation of nomenclatures of medical terms, syndromes, and symptoms, and the development of algorithms for coding medical information are *necessary conditions* for the further implementation of AI and information technologies in general in clinical practice. Justifying the principles of formalization of clinical information is one of the current tasks in improving the search for necessary information and the treatment-diagnostic process in general.

It is observed that in medicine, the creation of global systems for storing medical data and the necessity of their computer analysis are particularly urgent.

Today, we urgently need methods for structuring, searching, identifying, and processing data. Google has become a leader in online search by using search queries to determine what people want to know. This required a second revolution — mathematical algorithms that quickly and reliably track search logic and assist the end user in finding specific information. As a result, a considerable level of success in finding necessary information has been achieved. Recent studies show that 25.6% of desktop search queries and 17.3% of mobile device search queries end without transitioning to another Google resource. Moreover, considering the various features of Google's search results today, less than half of both types of searches end with an organic click [13].

For a fairly large number of tasks, such efficiency is insufficient. Other mechanisms and resources are necessary. One such mechanism is associated with the use of accumulated knowledge. Processing the knowledge resource should primarily be based on the use of artificial intelligence [14]. However, the presence of a semantic gap does not allow for the widespread use of AI.

In summary, as of today, AI supports doctors but does not replace them. AI typically solves important, but rather limited in scope, tasks, so the primary responsibility for patient care lies with the human physician. For example, clinical trials are conducted using AI for more accurate and much faster calculation of target zones for head and neck radiation therapy than by humans. Accordingly, AI plays an important role in protecting the patient from harmful radiation. At the same time, the interventional radiologist still bears full responsibility for conducting the therapy [15]. Thus, the goal of implementing AI should be to find a delicate, mutually beneficial balance between its effective use and the judgments of trained doctors.

The Logic of the Universum of Knowledge as a Global Unified Knowledge Resource

The Universum of knowledge is the aggregate of knowledge that has been preserved to date (B K Sen, 2009). A segment is a part of the Universum of knowledge, which contains a subject, and sometimes acts as a component of the subject system. The concept of the Universum correlates with the concept of the Absolute (the Absolute of the world, the absolute of life, the Absolute of expectations). Nominally, the Universum (Latin: Universum, summa rerum) means "all that is present," "everything that exists." Robert S. Solomon defines absolute knowledge as "knowledge that is unbiased, undistorted, unconditional, all-encompassing, free from counterexamples and internal contradictions. The concept of the Universum is encountered in many directions today.

In philosophy, the Universum is the aggregate of objects and phenomena as a whole, considered as a single system that unites time and space. In mathematics, the universum is the class of all elements considered in a given mathematical context.

In informatics, the Universum is most often considered alongside the subject plane (the subject area - part of the real world considered within a certain context. The context can be understood as a field of study or an area that is the object of a certain activity.

The definition "Universum of knowledge - the aggregate of knowledge that has been preserved to date. Knowledge that is being and will be created in the future and will become part of the universe of knowledge" is often used. However, the concept of the Universum of knowledge is a certain metaphor that has great significance in the classification theory of knowledge organization.

The general concern of researchers of the Universum is the necessity to reflect or transform knowledge into specific and logically developed classes that make up the classification system. Thus, the Universum is one of the main aspects of the organization and development of knowledge, which has extraordinary importance in the modern education system and can play a decisive role in organizing the processes of using AI in healthcare.

The development of intelligent services for workplace learning represents a special challenge for researchers and developers of learning technologies. One reason is that considering learning as a situational and social practice is nowhere as important as in cases where learning is closely integrated with workplace practices. During workplace learning, learning is closely related to work practices and tasks, and the simplified transfer of technologies used in the formal educational environment seems to ignore some opportunities in integrating work and learning.

Therefore, it is important that the development of technology for workplace learning is based on a social and situational perspective of learning, focused on issues such as the context in which learning takes place, or the social and artifact-mediated nature of learning.

However, the desire to increase the efficiency of healthcare and use the obtained information to improve diagnostic and treatment methods is often complicated by the low quality of medical data collection, as a high percentage of medical data is unstructured and stored in various systems and formats. Processes of information asymmetry are not explored. Furthermore, there is not always a consensus on which artificial intelligence and machine learning methods work better in different problem areas and what computer tools could make their application more convenient and flexible. Therefore, there is an obvious interest in using knowledge mobilization systems. Recent research shows a growing interest in the topics of "knowledge mobilization" (KMb) and "knowledge mobilizer" (KM) in the context of professional learning networks and their significance for institutional changes. The KMb process consists of two distinct stages: activation of knowledge resources and mobility of knowledge by transferring it to other contexts using various mechanisms [16]. However, all researchers emphasize the need to consolidate existing knowledge to develop the potential of KMb, including key concepts and practices, and to identify significant gaps in the current literature. Such a process is also of exceptional importance for the development of strategies for using AI. KMb can be understood as situational organizational changes that require the creation of an organizational culture.

For this purpose, we have developed a structural scheme of the availability of knowledge (Figure 1).

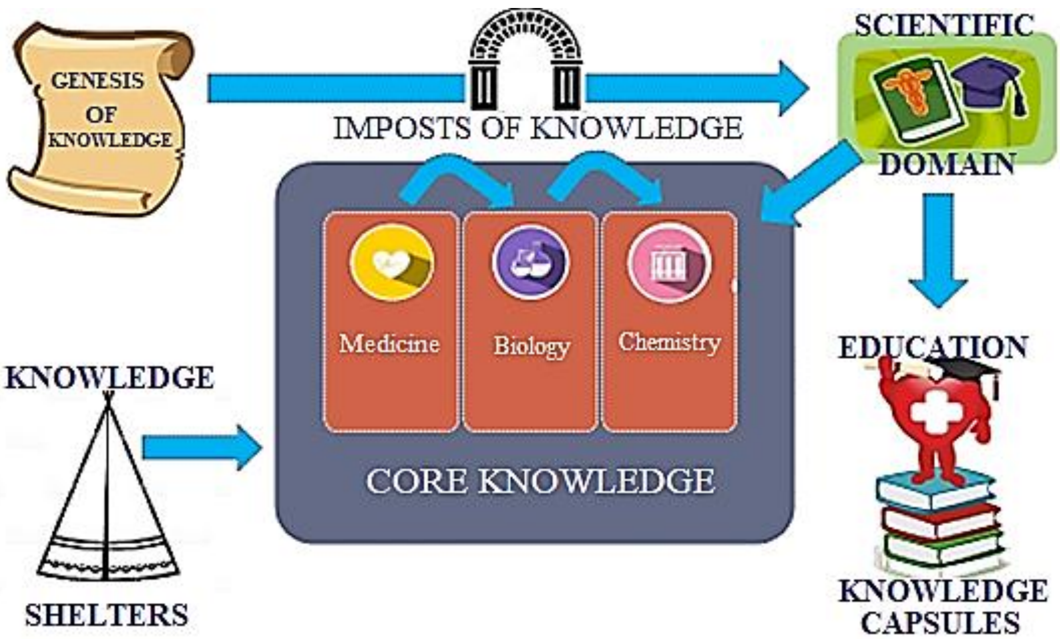

Fig. 1. Structural scheme of existing knowledge in the Universum format

New knowledge must be compatible with existing belief structures to become legitimized and useful in local conditions. This is where AI can play a significant role. In healthcare, bridging the gap between medical knowledge and health outcomes requires a systematic approach to continuous process and system improvement. The importance of this is emphasized by the competence requirements set by accrediting, certifying, and other bodies in the field of medical education and practice.

The success of the process depends on how useful new knowledge appears to recipients or how well it can be linked to existing concepts. Here, it is useful to rely on the work of Nonaka and Takeuchi (1995), related to the difference between explicit and tacit knowledge. Although explicit knowledge is more accessible, its complexity and/or tastiness determine how often it is used.

A fundamental question raised in health care research is how knowledge is created, organized, and used in the minds of physicians. Most efforts made to answer this question have so far focused on teaching methods and techniques, approaches to basic and clinical education, or the integration of disease scenarios into encapsulated knowledge. That's why knowledge management processes are gaining global importance. Knowledge management (KM) is understood as a set of principles, tools, and practices that allow people to share, translate, and apply what they know to create value and increase efficiency. In other words, organizational KM includes "regular, targeted, and systematic socio-technical functions designed to facilitate the assimilation, acquisition, production, organization, storage, retrieval, sharing, dissemination, transmission, development, and evaluation of experience and knowledge assets (tacit and explicit) to create competitive advantages and added value by improving the quantity and quality of organizational decisions and actions (overall levels), by making changes to technical, administrative, and structural strategies and implementing intelligent organization."

Conclusions

1. The goal of implementing AI should be to find a delicate, mutually beneficial balance between its effective use and the judgments of trained physicians. This is extremely important, as artificial intelligence, which may in the near future almost entirely replace the labour of physicians, is currently a problem that might otherwise hinder the benefits derived from it.
2. AI will become an integral part of future medicine. Therefore, it is crucial to educate the new generation of medical interns in the concepts and principles of AI application, to function effectively in the workplace. It is extremely important to develop skills such as empathy in AI.
3. A systematic approach to the continuous improvement of diagnostic and treatment processes and systems for patients, primarily, requires bridging the gap between accumulated medical knowledge and the logic and outcomes of AI use.

References

1. Beam, A.L., Drazen, J.M., Kohane, I.S., Leong, T.Y., Manrai, A.K., Rubin, E.J. (2023) Artificial Intelligence in Medicine. *N Engl J Med*. 388(13):1220-1221. https://doi.org/10.1056/NEJMe2206291.
2. Haug, C.J., Drazen, J.M. (2023) Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N Engl J Med*, 388(13):1201-1208. https://doi.org/10.1056/NEJMra2302038.
3. Buch, V.H., Ahmed, I., Maruthappu, M. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract*. 2018 Mar;68(668):143-144. https://doi.org/10.3399/bjgp18X695213.
4. Khan, B., Fatima, H., Qureshi, A., Kumar, S., Hanan, A., Hussain, J., Abdullah, S. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. (2023) *Biomed Mater Devices*. 8:1-8. https://doi.org/10.1007/s44174-023-00063-2
5. Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science*. 343(6176):1203-5. https://doi.org/10.1126/science.1248506.
6. Mintser, O., Sinyenko, N. O. (2023). Data discrimination in pathomorphology. ways of coping. Medical Informatics and Engineering, (3), 7–10. https://doi.org/10.11603/mie.1996-1960.2022.3.13359
7. Alamsyah, Riza., Luhur, Budi., Wiranata, A. D., Luhur, B. (2019) Defect Detection of Ceramic Tiles using Median Filtering, Morphological Techniques, Gray Level Cooccurrence Matrix and K-Nearest Neighbor Method. *Scientific Research Journal*, 4(6):2201-2796 https://doi.org/10.31364/SCIRJ/v7.i4.2019.P0419632
8. Ammar, M., Rania, K. (2023) A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*. 35(2):757-774 https://doi.org/10.1016/j.jksuci.2023.01.014
9. Kothari, S., Phan, J.H., Stokes, T.H., Wang, M.D. (2013) Pathology imaging informatics for quantitative analysis of whole-slide images. J. Am. Med. Inform. Assoc., 20 ,1099-1108 https://doi.org/10.1136/amiajnl-2012-001540
10. Khosravi, P., Kazemi, E., Imielinski, M., Elemento O., Hajirasouliha, I. (2018) Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images. EBioMedicine. 27:317-328. https://doi.org/10.1016/j.ebiom.2017.12.026
11. Greenspan H., van Ginneken B., Summers R.M. (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique IEEE Trans. Med. Imaging, 35:1153-1159 https://doi.org/10.1109/TMI.2016.2553401
12. Chaffey, D. (2023) Search engine marketing statistics 2023. The Search Engine Optimisation (SEO)

toolkit contains 16 Feb, 2023 https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/

13. Golhasany, H., Harvey, B. (2023) Capacity development for knowledge mobilization: a scoping review of the concepts and practices. Humanit Soc Sci Commun 10: 235 https://doi.org/10.1057/s41599-023-01733-8

14. Khakpour, A. (2020) Effectiveness of Knowledge Acquisition in Medical Education: An argumentative literature review of the resources's requirements. Future of Medical Education Journal. 10(3): 56-63. https://doi.org/10.22038/FMEJ.2020.40998.1271

15. Ley T. (2020) Knowledge structures for integrating working and learning: A reflection on a decade of learning technology research for workplace learning. *Br J Educ Technol*. 51(2): 331-346. https://doi.org/10.1111/bjet.12835

16. Jesacher-Roessler, Livia. (2021) The travel of ideas: the dual structure of mobilized knowledge in the context of professional learning networks. Journal of Professional Capital and Community. ahead-of-print. https://doi.org/10.1108/JPCC-06-2020-0048

The article has been sent to the editors 20.12.23.
After processing 16.01.24.
Submitted for printing 20.03.24