Australian
National
University

COMP8410-Data Mining

Semester 1, 2024

AiDoc: What we need to know, what we need to care about and what we need to do?

Course Code: COMP8410

Course Name: Data Mining

Student number	Student name	Contribution	Signature
u7564091	Yue Zheng	100%	Yue Zheng

AiDoc: What we need to know, what we need to care about and what we need to do?

1. What is AiDoc?

1.1 The Project AiDoc

With the advancement of deep learning, artificial intelligence (AI) is now acquiring the ability to make judgments based on previous experiences. Consequently, such capabilities allow AI to be potentially effective in diagnostics by analyzing prior diagnoses made by doctors.

In light of the scenario mentioned above, the AiDoc project has been initiated. This project aims to develop a machine learning model that equips AI with the ability to conduct corresponding tests based on patients' symptoms and to diagnose based on the symptoms, test results, and medical history.

Beyond diagnosing, AiDoc is also endowed with the capacity for self-learning. AiDoc will revise its model based on feedback from patients and co-diagnoses from doctors. It is anticipated that AiDoc will continuously improve its diagnostic accuracy as it accumulates more diagnostic experiences.
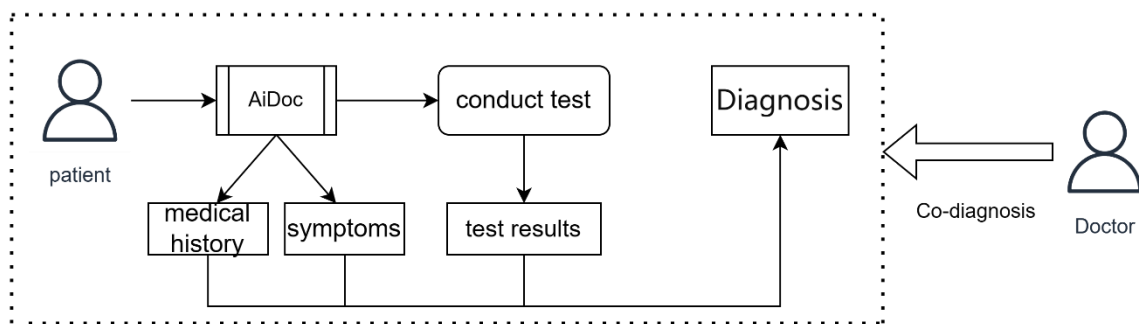

Figure 1. How AiDoc works.

1.2 The Idea of AiDoc

AiDoc is designed to harness the power of AI to mimic and enhance the diagnostic process traditionally carried out by human doctors. By integrating deep learning algorithms, AiDoc can analyze vast amounts of medical data, learn from previous diagnostic cases, and make informed decisions about a patient's health. By this kind of approaches and combination with machine learning, AiDoc is expected to be advanced comparing with traditional diagnostic in fields below:

Speed and Efficiency: AiDoc can process and analyze vast amounts of data at a speed unattainable by human doctors. This rapid data analysis enables quicker diagnoses, reducing the time patients wait for results and potentially increasing the throughput of diagnostic processes in medical facilities. Also, the ability of dealing multiple requests at the same time will alleviate the shortage of medical resources.

Consistency: Human diagnosis can sometimes vary due to subjective interpretations, fatigue, or differing levels of expertise among practitioners. AiDoc, by contrast, applies the same rigorous analysis consistently across all cases it examines, minimizing the variability in diagnostics.

Accessibility: AiDoc can support remote diagnostics, reaching the area where doctor is unable to come, making specialist knowledge more accessible to underserved areas. Patients in remote or rural locations can receive expert-level evaluations without the need for travel, broadening access to quality healthcare.

Besides these primary benefits, AiDoc might offers several additional advantages:

Personalization: AiDoc can integrate and analyze data from a patient's genetic information, lifestyle, and

environment, along with their medical history, to provide more personalized and effective treatment plans. This holistic approach can improve outcomes by tailoring interventions to the individual's unique health profile. With this functionality, AiDoc will lower the costs of having a personal physician in a great degree.

Predictive Analytics: Driving by machine learning model, AiDoc can use historical data to predict potential health risks and diseases before they manifest, offering a chance for preventive measures to be taken, which can lead to better patient outcomes and reduced healthcare costs.

Resource Optimization: By automating routine diagnostics, AiDoc can free up healthcare professionals to focus on more complex cases and patient care, optimizing the use of human resources in healthcare settings. This might no be significant since the doctor needs to be co-diagnosis. However, if there is one day that AiDoc can be responsible and diagnosis solely, it will help a lot to resource optimization.

1.3 The Technology Behind AiDoc

To delve into the technology behind AiDoc, we can categorize it into two principal components: the model and the data.

To craft an appropriate model for AiDoc, it's crucial to identify the nature of the input and the desired output. Since the output entails providing a diagnosis, it should manifest as a few sentences pertinent to the patient's condition. The input is tripartite: symptoms, test results, and medical history. Symptoms primarily consist of the patient's descriptions, often in sentences, and may also include images of the patient, along with other possible inputs like sounds or smells. Both test results and medical history will predominantly be in the form of sentences and images.

Aiming to develop AiDoc into a robust AI diagnosis model necessitates the application of machine learning technology, particularly deep learning models. Given the nature of AiDoc's tasks and the type of input data it processes, the deep learning model inspired by ChatGPT will be adapted for AiDoc. This model will leverage the transformer architecture to enhance its capability in generating coherent responses and organizing sentences effectively. Furthermore, to accommodate the diversity of input data—such as images and videos—specific layers designed to process these varied inputs will be integrated, enhancing AiDoc's proficiency in its designated tasks.

After determining the model and data type, understanding how AiDoc is trained is crucial to comprehending its functionality.

Initially, AiDoc's model will be pre-trained, providing each instance of AiDoc with a foundational understanding of diagnostics. With consent obtained from patients, doctors, and hospitals, AiDoc gathers a vast array of diagnostic data to serve as prior experience for model training. Following this training phase, AiDoc becomes operational.

During its diagnostic processes, AiDoc will inquire whether patients consent to their data being collected. If they agree, AiDoc will utilize this newly acquired data to further train and refine its model.
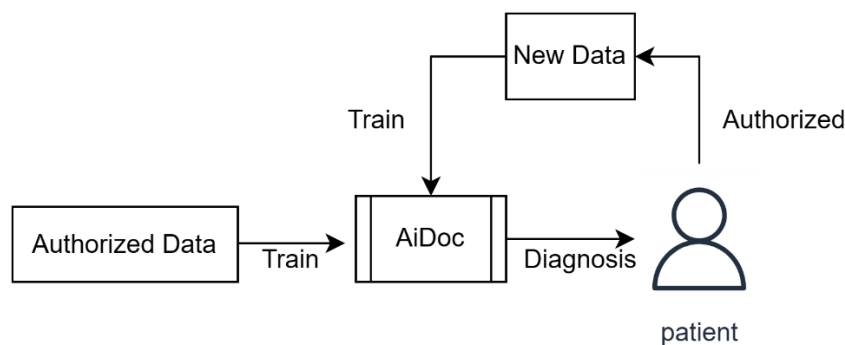

Figure 2. How AiDoc trained.

1.4 Further Application of AiDoc

For instance, its ability to learn from new data could make personalization a viable function for AiDoc. With this functionality, AiDoc could serve as a personal physician for individuals, tailoring diagnoses and health advice to their unique medical histories and current conditions. This personalized approach could significantly enhance patient care by offering more accurate and relevant health recommendations.

Moreover, AiDoc could function as a training tool for medical students. By comparing the diagnoses made by AiDoc with those made by students, it can provide immediate feedback on the accuracy of their assessments. This application not only helps in validating the students' diagnostic skills but also exposes them to a wide range of medical conditions and scenarios, enhancing their learning experience.

Additionally, AiDoc's comprehensive database and learning capabilities could be utilized in epidemiological research and public health monitoring. By analyzing patterns in the data it collects, AiDoc could identify trends in disease outbreaks, effectiveness of treatments, and other valuable insights that could inform public health policies and strategies.

Overall, the potential applications of AiDoc extend well beyond its initial scope, offering opportunities to enhance medical education, research, public health, and accessibility of care, thereby contributing significantly to the advancement of global health initiatives.

2. What we need to care about AiDoc?

As an AI diagnostic model, the implementation of AiDoc will raise some ethical concerns, as listed below.

2.1 Individual Privacy

AiDoc will utilize individuals' data to train its model; since it is used for diagnosis, certain personal information, including living area, gender, and income, may be associated with the diagnosis. Including this type of data can aid in diagnosis; however, it may not fully align with UNESCO's recommendations on the "Right to Privacy and Data Protection"[1]. Therefore, a trade-off between sufficient data collection and individual privacy will be necessary during AiDoc's data collection process.

2.2 Transparency and Explainability

Given AiDoc's complexity as a deep learning model, it may be challenging to explain its internal workings. However, transparency and explainability are crucial as they enable the examination of whether a machine learning model relies on actual evidence rather than biases present in the training data [2]. Therefore, regardless of AiDoc's eventual complexity, maintaining transparency and explainability is essential to ensure the model's meaningfulness.

2.3 Potentially Harmful Decision Making

Guidelines for the responsible application of data analytics [3] state, 'Don't delegate to a device any decision that has potentially harmful effects without ensuring that it is subject to specific human approval prior to implementation, by a person who is acting as an agent for the accountable organization.' However, AiDoc may not fully adhere to these guidelines since diagnostics are crucial to patient health, and a misdiagnosis could cause harm. Given this concern, AiDoc requires additional regulation and responsibility allocation to ensure the project is accountable.

2.4 Fairness and non-discrimination

According to UNESCO's recommendations, a model should be fair and serve everyone equally [1]. However, when dealing with diseases that only have a small number of samples, AiDoc might not perform as well as it does with more common diseases. This discrepancy highlights the importance of ensuring that AiDoc incorporates mechanisms to address and mitigate bias, ensuring its diagnostic capabilities are equitable across a wide range of conditions. Enhancing the model's ability to learn from limited data and

actively seeking diverse datasets can help improve its performance on rare diseases, promoting fairness and non-discrimination in healthcare access and outcomes.

Moreover, while AiDoc might ensure that individuals who have access to the model receive equal medical expertise for the same disease, it inadvertently widens the medical gap for those unable to access AiDoc. This disparity needs careful consideration. Efforts should be made to bridge this gap by expanding access to AiDoc, potentially through partnerships with public health organizations, governments, and NGOs. Ensuring wider availability and addressing digital divide issues are crucial steps toward mitigating healthcare inequalities and fostering a more inclusive approach to medical diagnostics.

2.5 Honesty by User

In the spirit of honesty as outlined in the ACS Code of Professional Conduct [4], it's crucial to address a potential ethical issue: the risk of some doctors and hospitals misrepresenting AiDoc as their own creation. Such deceptive practices not only breach the trust placed in healthcare professionals and institutions but also undermine the integrity of the profession as a whole. Upholding honesty means ensuring that credit for the development and achievements of AiDoc is accurately attributed, preventing any attempts to falsely enhance reputations through the misappropriation of credit. This commitment to transparency and integrity is fundamental in maintaining public trust and confidence in the healthcare system's use of advanced technologies like AiDoc.

2.6 Cybersecurity Threats

As previously mentioned [5], cybercriminals are exploring AI text generators to create sophisticated phishing attacks, harassment, fake news, and software code for ransomware and other malware. A similar scenario could occur with AiDoc. Criminals might use AiDoc to generate fake diagnoses to threaten the public, which will reduce the public's belief in AiDoc and causing misdiagnosis.

3. What can we need to do?

3.1 Set Up Government Guideline

To address these ethical issues, establishing guidelines for governance is both crucial and a straightforward approach. Once such guidelines are in place, they delineate what is permissible and what is not, thereby establishing a clear boundary between lawful and criminal behavior. Consequently, individuals with no intention of causing harm will likely avoid harmful actions if the guidelines are appropriately designed.

Furthermore, these guidelines should also encompass mechanisms for enforcement and accountability, ensuring that any violations are promptly identified and addressed. This includes the establishment of monitoring systems and regulatory bodies dedicated to overseeing AI applications like AiDoc, as well as clear procedures for reporting and investigating potential abuses. By fostering a regulatory environment that promotes ethical use while penalizing misconduct, the government can create a safer landscape for AI innovations to benefit society without compromising individual rights or public safety.

3.2 Educate Public about AiDoc

In many cases, ethical problems arise from misunderstandings. Educating the public about new technology can help decrease their fear of it.

Continuing this approach, comprehensive public education campaigns about AiDoc should focus on explaining how the technology works, its potential benefits, and the ethical considerations that have been addressed during its development. This could include community workshops, informational websites, and collaboration with media outlets to spread awareness. Additionally, providing clear examples of how AiDoc improves healthcare outcomes can demystify the technology, making it more accessible and less

intimidating to the general public. By fostering a well-informed community, individuals can make more educated decisions about their healthcare and engage with AiDoc and similar technologies more confidently and safely. Besides, individuals will also know that AiDoc might not be a hundred percent correct so they should not be fully relying on it.

3.3 Develop Industry Standard Procedural

When addressing aspects such as transparency, explainability, and fairness, it is crucial to establish an industry-standard procedure to assess whether a project meets these standards. By setting these benchmarks, the compliance of a project can be evaluated, enabling project developers to identify areas for improvement in their work.

The development of such standards requires collaboration among industry leaders, regulatory bodies, and stakeholders to ensure that the criteria are comprehensive, practical, and reflective of ethical considerations. These standards should not only cover technical specifications but also ethical guidelines that address privacy, data protection, and user consent. Furthermore, implementing a certification or accreditation process can help in publicly recognizing projects that adhere to these standards, thereby promoting trust and confidence among users and the broader community. Regular updates to these standards are also necessary to adapt to technological advancements and emerging ethical challenges, ensuring that the industry remains responsible and responsive to societal needs.

3.4 Establishing Punitive Measures

For individuals intending to threaten society with models like AiDoc, it is essential to establish punitive measures. These measures should be clearly defined and implemented to deter malicious use of AI technologies. Legal frameworks must be adapted to address the unique challenges posed by AI, ensuring that penalties are not only applicable but also proportionate to the harm caused.

Furthermore, these punitive measures should include fines, sanctions, and, in severe cases, criminal charges, depending on the nature and extent of the violation. Collaboration with international bodies can also ensure that these measures have a global reach, preventing offenders from evading consequences by operating across borders.

Additionally, a system for reporting violations should be made accessible to the public, empowering users to play an active role in monitoring and safeguarding the ethical use of AI. By establishing robust punitive measures, society can better protect itself from the risks associated with the misuse of technologies like AiDoc, ensuring that AI serves the common good.

3.5 Determining responsibility allocation

Before AiDoc is deployed, it's critical to establish clear guidelines for responsibility allocation for any harm it might cause. This involves defining who will be held accountable in various scenarios, whether it's the developers, the users, or the overseeing organizations.

This process requires a comprehensive framework that considers the different stages of AiDoc's lifecycle, from development and testing to deployment and everyday use. For instance, developers could be responsible for ensuring the model's accuracy and fairness, while users, such as healthcare providers, might be responsible for appropriately interpreting and applying AiDoc's recommendations. Regulatory bodies could oversee the entire ecosystem, ensuring compliance with established ethical and operational standards. Moreover, establishing a clear responsibility allocation will necessitate legal and ethical considerations to be integrated into AiDoc's development process. It will also require ongoing monitoring and assessment to adapt to new challenges and technological advancements. This proactive approach will help mitigate risks and ensure that AiDoc remains a beneficial tool for society.

Citations:

[1] United Nations Educational, Scientific and Cultural Organization. (2022). Recommendation on the ethics of artificial intelligence.

[2] Du, M., Liu, N., & Hu, X. (2020). Techniques for interpretable machine learning. Communications of the ACM, 63(1), 68-77.

[3] Clarke, R. (2017). Guidelines for the responsible application of data analytics. Computer Law & Security Review: The International Journal of Technology Law and Practice.

[4] Australian Computer Society. (2014). ACS Code of Professional Conduct (V2.1) [PDF]. Professional Standards Board.

[5] Savage, N. (2023). The rise of the chatbots. Communications of the ACM, 66(7).