Large Language Model Policy and Practice

A framework for large language models in an Academic Medical Center

Sean Davis

Invalid Date

Table of contents

Overview **3**
Technical Challenges and Opportunities . . . 3
Ethical Challenges and Opportunities . . . 3
Legal Challenges and Opportunities . . . 4
Social Challenges and Opportunities . . . 4

1 Introduction **6**

2 Guiding principles for AI in healthcare **7**
2.1 Principle 1: AI tools should aim to alleviate existing health disparities . . . 8
2.2 Principle 2: AI tools should produce clinically meaningful outcomes . . . 8
2.3 Principle 3: AI tools should reduce overdiagnosis and overtreatment . . . 9
2.4 Principle 4: AI tools should have high healthcare value and avoid diverting . . . 9
2.5 Principle 5: AI tools should incorporate social, structural, environmental, . . . 9
2.6 Principle 6: AI tools should be easily tailored to the local population . . . 10
2.7 Principle 7: AI tools should promote a learning healthcare system . . . 10
2.8 Principle 8: AI tools should facilitate shared decision-making . . . 10

3 DRAFT policy **11**
3.1 Vision Statement . . . 11
3.2 Stakeholder Considerations . . . 11
3.2.1 Patients . . . 11
3.2.2 Healthcare Providers . . . 12
3.2.3 Researchers . . . 12
3.2.4 Administrators and Support Staff . . . 12
3.3 Monitoring and Compliance . . . 12

References **13**

Appendices **14**
AI principles proposed by select organizations . . . 14

Overview

Large language models, such as ChatGPT, have gained significant attention in recent years due to their ability to generate human-like language and process vast amounts of natural language data. The development of these models presents a significant opportunity to transform healthcare by enhancing clinical decision-making, patient care, and medical research. However, the implementation of these models also presents several challenges, including technical, ethical, legal, and social challenges.

Technical Challenges and Opportunities

One of the significant challenges in developing and implementing large language models is the computational power required for training and inference. These models require a vast amount of data and computational resources to train and fine-tune. However, recent advancements in deep learning frameworks and cloud computing have made it possible to train and deploy large language models on a large scale.

Another technical challenge is the issue of bias in language models. Language models learn from the data they are trained on, which means that if the data is biased, the model will also be biased. This can have significant implications for healthcare, where biased models could lead to incorrect clinical decision-making or reinforce health disparities. To address this challenge, researchers have proposed several techniques to mitigate bias in language models, such as data augmentation, adversarial training, and fairness constraints (1).

Ethical Challenges and Opportunities

The implementation of large language models in healthcare raises several ethical concerns, such as patient privacy, informed consent, and fairness. Language models require vast amounts of data, including personal health information, which raises concerns about patient privacy and data protection. Additionally, patients may not understand how their data is being used or may not have given informed consent for their data to be used in this way.

However, there are also significant ethical opportunities presented by the use of large language models in healthcare. For example, language models can be used to generate natural language explanations for clinical decision-making, which can improve transparency and help build trust

between patients and healthcare providers. Additionally, language models can be used to identify and address health disparities by analyzing large-scale healthcare data and developing targeted interventions.

Legal Challenges and Opportunities

The implementation of large language models in healthcare also presents several legal challenges, such as liability and regulatory compliance. If language models are used to make clinical decisions, healthcare providers could be held liable for any adverse outcomes resulting from the use of these models. Additionally, healthcare providers must ensure that the use of language models complies with existing regulations, such as HIPAA, which govern the use of personal health information.

However, there are also legal opportunities presented by the use of large language models in healthcare. For example, language models can be used to analyze large-scale healthcare data and identify potential areas of fraud or abuse, which can improve the efficiency and effectiveness of healthcare delivery.

Social Challenges and Opportunities

The implementation of large language models in healthcare also presents several social challenges, such as the potential for job displacement and the exacerbation of existing healthcare disparities. The use of language models could lead to the displacement of healthcare workers whose jobs can be automated by these models. Additionally, if language models are biased, they could reinforce existing healthcare disparities, particularly in marginalized communities.

However, there are also social opportunities presented by the use of large language models in healthcare. For example, language models can be used to improve the accessibility or utilization of healthcare services in at-risk or disadvantaged populations. Additionally, language models can be used to improve the quality of healthcare services by providing personalized treatment recommendations and identifying potential areas of improvement in healthcare delivery or clinical operations.

The implementation of large language models in healthcare presents significant challenges and opportunities, including technical, ethical, legal, and social challenges. However, with careful consideration and mitigation of these challenges, the use of language models in healthcare has the potential to transform healthcare delivery and improve patient outcomes. It is essential for healthcare organizations to consider the potential risks and benefits of implementing large language models and to prioritize ethical and responsible use of these models.

As healthcare providers increasingly rely on large language models for clinical decision-making, it is critical to ensure that these models are transparent, explainable, and unbiased. Researchers and developers must work closely with healthcare providers and patients to ensure that the development and implementation of language models are aligned with ethical principles and patient needs.

In summary, the implementation of large language models in healthcare is a complex and rapidly evolving field with both challenges and opportunities. By carefully considering the technical, ethical, legal, and social implications of these models, healthcare organizations can leverage the full potential of large language models to improve patient outcomes and advance medical research.

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

2 Guiding principles for AI in healthcare

Table 2.1: Questions that can be used when considering each principle in the AI development process (Badal, Lee, and Esserman 2023)

Principle	Questions
1. Alleviate healthcare disparities	• What health disparities are reported for the present AI application? • How can the AI tool be designed to be accessible to and improve outcomes for the disadvantaged population? • What clinical interventions are needed to realize the benefit, and are these accessible? • How can data collection be supported in underserved communities for tool retraining over time?
2. Report clinically meaningful outcomes	• How is clinical benefit defined in this domain? • What is the present threshold for the clinical benefit of existing tools, and how can the AI tool improve upon this threshold?
3. Reduce overdiagnosis and overtreatment	• What disease state is an overdiagnosis? • For every case of overdiagnosis, what are the downstream costs to the patient and healthcare system? • How can this AI application reduce the number of overdiagnoses compared to existing approaches?
4. Have high healthcare value	• Is this AI tool addressing a high-priority healthcare need? • What would be the cost to the healthcare system in implementation, maintenance, and update? • What would be the cost to the patient who does and does not benefit from this tool? • Does this tool have high healthcare value, and if not, how can it be improved?
5. Incorporate biography	• What biographical data can be collected or carefully coded for the intended population? • How do these factors vary in the intended population? • How can these factors be included when developing AI tools?
6. Be easily tailored to the local population	• Can the training features be easily collected in different settings? • Are these features reliable for training across different populations? • Will the AI/ML workflow be made open-access?

Principle	Questions
7. Promote a learning health-care system	• How will this AI application be evaluated over time, and at what intervals? • What are acceptable thresholds for performance? • How will the evaluation results contribute to continuous improvement?
8. Facilitate shared decision-making	• Have AI explainability tools been explored and utilized? • Do clinicians and patients find the explainability results helpful? • Have simpler, explainable algorithms been tried and compared to 'black-box' algorithms to determine if a simpler model performs just as well? • How can patient values be easily integrated into the use of the AI tool?

2.1 Principle 1: AI tools should aim to alleviate existing health disparities

Reaching health equity requires eliminating the disparitities in health outcomes that are closely linked with social, economic, and environmental disadvantages. At their very core, AI tools require collection of specialized and high-quality data, advanced computing infrastructure for use, capacity to purchase or partner models from commercial entities, and unique technical expertise, all of which are less likely available to healthcare systems that serve the most disadvantaged populations.

More careful training and model development that accounts for the unique needs of disadvantaged populations is needed to ensure that AI tools do not exacerbate existing health disparities. Creating equitable AI tools may require prioritizing simpler models for deployment, and the trade-off between balancing accuracy and equity can potentially be resolved by designing AI tools that can be easily tailored to the local population. AI tools designed to serve disadvantaged groups must not unnecessarily divert resources from higher priority areas and more effective interventions (Principle 4).

2.2 Principle 2: AI tools should produce clinically meaningful outcomes

AI tools should be evaluated based on their ability to improve clinically meaningful outcomes. The clinical benefit of AI tools should be defined in the context of the existing standard of care, and the AI tool should be evaluated against this standard. If AI practitioners do not define clinical metrics for clinical benefit *a priori*, they risk producing tools that clinicians cannot evaluate or use. Clinician partners of AI researchers should evaluate accuracy, fairness,

and risks of overdiagnosis and overtreatment (Principle 3). They should also evaluate the healthcare value (Principle 4) along with the explainability and auditability of AI tools and models (note principles outlined in Table 2.1.

2.3 Principle 3: AI tools should reduce overdiagnosis and overtreatment

Particularly in the United States, overdiagnosis and overtreatment are major drivers of healthcare costs and patient harm. Overdiagnosis occurs when a disease is diagnosed that would not have caused symptoms or death in a patient's lifetime. Overtreatment occurs when a patient is treated for a disease that would not have caused symptoms or death in a patient's lifetime. AI tools should be carefully constructed with the spectrum of disease and interventions to result in decreased overdiagnosis and overtreatment.

2.4 Principle 4: AI tools should have high healthcare value and avoid diverting

resources from higher-priority areas

AI tools applied in healthcare should result in the same outcomes for reduced cost or better outcomes for costs comparable to current costs. Costs to gather inputs, build, maintain, update, interpret, and deploy in clinical practice must be estimated and included in weighing the decisions around AI tool application. Note that what might be cost-effective, leading to high healthcare value, in one setting might be extremely cost-ineffective in settings where resources are scarce.

2.5 Principle 5: AI tools should incorporate social, structural, environmental,

emotional, and psychological drivers of health

2.6 Principle 6: AI tools should be easily tailored to the local population

2.7 Principle 7: AI tools should promote a learning healthcare system

2.8 Principle 8: AI tools should facilitate shared decision-making

3 DRAFT policy

This is a draft policy for the use of language models in healthcare. It is intended to be a living document that can be updated as needed. Note that it is mostly a placeholder for now.

3.1 Vision Statement

- LLMs must be used in a manner consistent with the mission, vision, and values of the academic hospital system.
- The use of LLMs must align with relevant legal and regulatory requirements, including but not limited to data privacy, security, and intellectual property laws.
- The deployment of LLMs should prioritize patient safety, privacy, and wellbeing.
- LLMs must be used in a transparent manner, with users understanding the capabilities and limitations of the technology.
- Continuous improvement and evaluation of LLM usage should be prioritized to ensure ongoing alignment with organizational goals.

3.2 Stakeholder Considerations

3.2.1 Patients

- LLMs should be used to augment patient care and improve outcomes, without replacing the human touch and empathy of healthcare providers.
- Patients must be informed about the use of LLMs in their care, and they should have the option to opt out if desired.
- Patient data used in LLM applications must be anonymized, encrypted, and securely stored to protect patient privacy.

3.2.2 Healthcare Providers

- LLMs should be deployed to enhance clinical decision-making and efficiency without undermining the autonomy and expertise of healthcare providers.
- Adequate training and support should be provided to healthcare providers to ensure proper use and understanding of LLMs.
- Feedback from healthcare providers must be regularly solicited to improve LLM performance and usability.

3.2.3 Researchers

- The use of LLMs in research must adhere to ethical standards, including obtaining informed consent and minimizing potential harm.
- Collaboration between researchers and LLM developers should be encouraged to drive innovation and address specific research needs.
- Research involving LLMs should be transparent and reproducible, with results and methodologies made available to the wider scientific community.

3.2.4 Administrators and Support Staff

- LLMs should be deployed in administrative and support functions to improve efficiency, reduce costs, and enhance the overall quality of service.
- Staff should receive appropriate training and support to understand and utilize LLMs effectively.
- Employee feedback should be actively sought to identify areas of improvement and potential new applications for LLMs.

3.3 Monitoring and Compliance

- A designated LLM Steering Committee, comprising representatives from various stakeholder groups, will be responsible for monitoring and enforcing compliance with this policy.
- Periodic audits and assessments will be conducted to ensure adherence to this policy and identify areas for improvement.
- Policy violations may result in disciplinary action, up to and including termination of employment or access to LLMs

References

Badal, Kimberly, Carmen M Lee, and Laura J Esserman. 2023. "Guiding Principles for the Responsible Development of Artificial Intelligence Tools for Healthcare." *Communication & Medicine* 3 (1): 47. https://doi.org/10.1038/s43856-023-00279-9.

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

Appendices

AI principles proposed by select organizations

This list is adapted from Badal, Lee, and Esserman (2023), Table 1.

- Ethics and governance of artificial intelligence for health, World Health Organization
  - Human autonomy
  - Human well-being and safety and the public interest
  - Transparency, explainability, and intelligibility
  - Responsibility and accountability
  - Inclusiveness and equity
  - Responsive and sustainable
- Ministries of Health, Medical AI algorithm assessment checklist, FUTURE-AI (an international, multi-stakeholder consortium)
  - Fairness
  - Universality
  - Traceability
  - Usability
  - Robustness
  - Explainability
- Good Machine Learning Practice for Medical Device Development: Guiding Principles, fdFDA, Health Canada, United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA)
  - Leverage multidisciplinary expertise in development
  - Implement good software engineering and security practices
  - Datasets are representative of intended population
  - Training and test sets are independent
  - Reference datasets are well developed
  - Optimize performance of Human-AI Team
  - Thorough clinical testing
  - Information accessible to users
  - Monitor deployed models and mitigate retraining risk

- Defining AMIA's artificial intelligence principles, American Medical Informatics Association (AMIA)
  - Autonomy
  - Beneficence
  - Non-maleficence
  - Justice
  - Explainability
  - Interpretability
  - Fairness
  - Dependability
  - Auditability
  - Knowledge managemen

Index

ChatGPT, 3