Large Language Model Policy and Practice

A framework for large language models in an Academic Medical Center

Sean Davis

2023-05-13

Table of contents

1 Overview

Large language models, like ChatGPT, have garnered significant interest due to their human-like language generation and immense natural language processing capabilities. These models offer opportunities to revolutionize healthcare by enhancing clinical decision-making, patient care, and medical research. However, implementing them also poses technical, ethical, legal, and social challenges.

1.1 Technical Challenges and Opportunities

Developing and implementing large language models entail considerable computational power for training and inference. These models demand extensive data and computational resources, but recent advancements in deep learning frameworks and cloud computing have facilitated their large-scale deployment.

Bias in language models is another technical challenge. Models trained on biased data can produce biased outcomes, potentially leading to incorrect clinical decisions or reinforcing health disparities. Researchers have proposed various techniques to mitigate bias, such as data augmentation, adversarial training, and fairness constraints.

1.2 Ethical Challenges and Opportunities

Implementing large language models in healthcare raises ethical concerns like patient privacy, informed consent, and fairness. Models require vast amounts of data, including personal health information, which can compromise patient privacy and data protection. Patients may also be unaware of how their data is used or may not have provided informed consent.

Conversely, large language models present ethical opportunities. They can generate natural language explanations for clinical decisions, improving transparency and trust between patients and providers. Furthermore, these models can identify and address health disparities by analyzing large-scale data and developing targeted interventions.

1.3 Legal Challenges and Opportunities

Legal challenges include liability and regulatory compliance. If language models contribute to clinical decisions, providers may be held liable for adverse outcomes. Compliance with existing regulations, such as HIPAA, is also crucial.

On the other hand, legal opportunities arise from using large language models to analyze extensive healthcare data, identifying potential fraud or abuse, and enhancing healthcare delivery efficiency and effectiveness.

1.4 Social Challenges and Opportunities

Social challenges involve potential job displacement and exacerbation of healthcare disparities. Language models could automate healthcare jobs and, if biased, reinforce existing disparities, particularly in marginalized communities.

However, social opportunities also emerge, such as improving healthcare accessibility for at-risk or disadvantaged populations and enhancing healthcare service quality through personalized treatment recommendations and identifying areas for improvement in healthcare delivery or clinical operations.

1.5 Use in Medical Education

Large language models can also play a significant role in medical education settings. They can assist in developing personalized learning pathways, providing instant feedback on complex clinical scenarios, and facilitating access to a wealth of medical knowledge. By incorporating these models into medical curricula, educators can enhance the learning experience and better prepare future healthcare professionals.

2 Conclusion

Implementing large language models in healthcare presents various challenges and opportunities. With careful consideration and mitigation of these challenges, these models have the potential to transform healthcare delivery and improve patient outcomes. It is crucial for healthcare organizations to weigh the risks and benefits of implementation and prioritize ethical and responsible use.

As healthcare providers increasingly depend on large language models, ensuring transparency, explainability, and unbiased models is critical. Researchers and developers must collaborate with healthcare providers and patients to align language model development and implementation with ethical principles and patient needs.

In summary, the use large language models in healthcare is a complex and rapidly evolving landscape. By addressing the technical, ethical, legal, and social implications, healthcare organizations can harness their full potential to enhance patient outcomes, medical research, business operations, and education.

Preface

This book is a work in progress. It develops a framework for adopting ChatGPT and related tools for the Campus. It is not a complete strategic plan and is not meant to be proscriptive. It is meant to help organize and communicate efforts.

The key component of the framework, in my opinion, is the recognition that, while the campus is a single entity, it is composed of four relatively distinct "Domains" that each have their own needs, requirements, resources, and priorities. The four domains that I have identified are:

- Education
- Research
- Business operations
- Clinical

The framework is based on the idea that each of these domains has a different set of needs and requirements, and that the campus should develop plans and implementations that are tailored to each domain. The framework includes a set of "principles" that should be applied to each domain, and a set of "strategies" that should be applied to each domain. The principles and strategies are not meant to be proscriptive, but rather to provide guidance and a framework for thinking about how to approach each domain.

The framework is also based on the idea that the campus should adopt a "platform" approach to the development of tools and services. I have included the "platform" concept as a set of identifiable cross-domain workstreams.

I also note that this framework is not meant to be a "top-down" approach. Rather, it is meant to be a "bottom-up" approach that is driven by the needs of the individual domains. The framework is meant to provide a common language and common set of tools that can be used to develop solutions that meet the needs of the individual domains and the campus as a whole.

The framework is also meant to be a living document. It is meant to be updated and revised as new information becomes available and as new needs and requirements are identified. Progress on the framework should be tracked and reported on a regular basis.

Finally, this framework can form the basis for adoption of any AI (or even other technology) on campus. It is not specific to ChatGPT, though I have found that approaching frameworks with concrete examples is key to success. In that regard, time invested in creating robust processes for ChatGPT will pay dividends in the future as other AI technologies are adopted.

3 Values and Principles

While there are many potential applications for LLMs in healthcare, the following guiding principles should be considered when developing and deploying LLMs in the academic hospital system. Note that these principles largely apply to any Artificial Intelligence or Machine Learning applications in use on the campus.

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

Part I

Implementation

4 The Framework

A Framework for Implementing AI and Large Language Models across an Academic Medical System

In order to successfully integrate AI and Large Language Models into an academic medical system, it is essential to adopt a flexible and agile approach that accounts for the varying pacing, priorities, and levels of risk associated with different aspects of the institution. By organizing the implementation plan into distinct domains and workstreams (see Figure 4.1), we can address the unique requirements of each area, ensuring that resources are allocated effectively and progress is made at an appropriate pace. This structure also enables rapid adaptation to changing circumstances, allowing for seamless collaboration between various teams and promoting a proactive response to any challenges that may arise. Ultimately, the use of domains and workstreams fosters a comprehensive and resilient approach to AI integration, maximizing the potential benefits while minimizing potential risks across the entire academic medical system. Note that the framework is not intended to be prescriptive or exhaustive; rather, it is meant to serve as a starting point for discussion, planning, and implementation. A top-level coordination unit (e.g., a steering committee) will work with the domain areas as a resource and to ensure that the overall implementation plan is aligned with the institution's strategic goals and priorities.

4.1 Domains

The implementation plan for integrating AI and Large Language Models into an academic medical system consists of four main domains:

- Education
- Research
- Clinical
- Business Operations

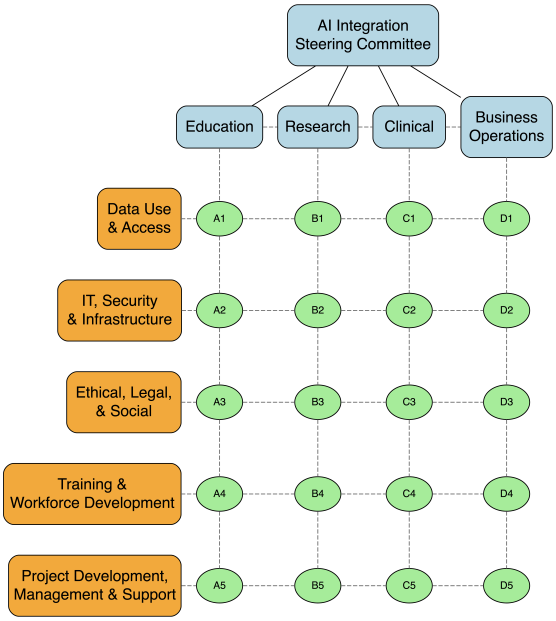

Figure 4.1: A schematic framework for organizing workstreams (orange boxes), domains (blue boxes), and work products and tasks (green ovals). Domains (vertical dimension) capture semi-independent organizations, each with largely independent use cases, budgets and business plans, priorities, and leadership. The workstreams (horizontal dimension) will often require similar or overlapping expertise, and can serve as knowledge resources to provide synergy and uniformity in implementation across domains.

4.1.1 Education

This domain includes all activities related to teaching, learning, and evaluation within the institution. It also encompasses the development of new educational programs and the management of existing ones.

4.1.2 Research

This domain focuses on the practice of the basic, clinical, and translational research programs within the institution. In addition, it includes the management of research grants, the development of new research programs, and the dissemination of research findings.

4.1.3 Clinical

This domain encompasses all activities related to patient care, including the management and implementation of clinical services, decision support and clinical decision-making, automation, and point-of-care or electronic patient support.

4.1.4 Business Operations

This domain focuses on the management of the institution's business operations, including finance, human resources, information technology, and facilities management. It also includes the development of new business processes and the management of existing ones.

4.2 Workstreams

Within each domain, we have identified five workstreams that are critical for the successful implementation of AI and Large Language Models. These workstreams are:

- Data Access & Use
- IT, Security, & Infrastructure
- Ethical, Legal, & Social
- Training & Workforce Development
- Project Management & Support Personnel

4.2.1 Data Access & Use

This workstream focuses on managing and optimizing data access, use, and sharing within the academic medical system. It ensures that data is available, reliable, and secure for AI integration and that the necessary infrastructure is in place to support data-driven activities.

4.2.2 IT, Security, & Infrastructure

This workstream addresses the technical aspects of AI integration, including the development and maintenance of IT systems, ensuring data security, and providing the necessary hardware and software infrastructure to support AI and Large Language Models.

4.2.3 Ethical, Legal, & Social

This workstream focuses on the ethical, legal, and social implications of AI integration in the academic medical system with domain-specific focus as appropriate. It aims to ensure that AI is used responsibly and ethically and that any legal and social concerns are addressed proactively.

4.2.4 Training & Workforce Development

This workstream is dedicated to developing the skills and knowledge of domain community members (including staff *and* leadership) within the domain to understand and, where appropriate, to effectively use and manage AI and Large Language Models. It includes training programs, workshops, and other educational opportunities to build competency in AI-related technologies.

4.2.5 Project Management & Support Personnel

This workstream is responsible for ensuring that the project management of AI and Large Language Models across the four domains. Among its roles are to provide project management, helping to align resource requests, support services around the usage of LLMs. This group will also cooordinate support staff who work collaboratively within and across domains to to ensure that AI integration occurs smoothly and efficiently.

The implementation plan is structured in a way that allows for cross-functional collaboration between the domains and workstreams. This ensures that AI and Large Language Models are integrated cohesively across the entire academic medical system, maximizing the benefits and minimizing potential risks.

5 Domain Implementation Guide

Each Domain will work to meet the needs of the campus that align with its community membership and mission. This section provides a rough guide (adapted from Lindegaard (2023)) to implementing a LLM plan within the Domains. Each Domain can work through this guide to develop policies, plans, resources, and deliverables. The combination of cross-domain Workstreams and reporting to the AISC will help to align resource requests and facilitate knowledge sharing, management, and transfer.

5.1 Suggested Tasks

5.1.1 Establish Ownership Leadership

Assign a Domain leader and convene a Domain Working Group (represented in blue boxes in Figure 4.1) as the driving force behind ChatGPT implementation. This leader and Working Group members need to become proficient in the technology and subsequently guide the organization through the process of integration. This includes setting the vision, aligning key stakeholders, and ensuring that the implementation aligns with the organization's strategic goals.

In practice, this Domain leader will be the primary point of contact for the AISC and will be responsible for reporting on the Domain's progress and deliverables. The Domain leader will also be responsible for convening the Domain Working Group and ensuring that the Domain's Workstreams are progressing.

5.1.2 Lean Into Cross-functional Collaboration

Form a cross-functional team with representatives from HR, IT, Legal, and other relevant departments to ensure that various perspectives are considered, and organizational needs are met during ChatGPT implementation. This collaboration will help address potential challenges, optimize resources, and facilitate effective knowledge transfer across the organization.

5.1.3 Educate Leadership Teams

Keep leadership teams informed about ChatGPT to enable swift, informed decision-making. Offer workshops and seminars to provide an in-depth understanding of the technology, its potential benefits, and its limitations. This empowers them to make quick, informed decisions that will shape the organization's adoption and use of the technology.

5.1.4 Identify Potential Use Cases

Map potential applications across all functions and establish a cohesive implementation plan. Conduct thorough analyses of business processes and functions to identify areas where ChatGPT can bring significant value, prioritize these use cases, and create a detailed roadmap for implementation, including timelines and milestones.

5.1.5 Craft a Communication Strategy

Develop a comprehensive communication strategy that addresses employee concerns and questions about the implementation of ChatGPT. This approach should be transparent, informative, and reassuring to ensure a smooth transition. It should also highlight the benefits of the technology and address potential misconceptions.

5.1.6 Prepare and Integrate High-Value Proprietary Datasets

Build APIs and interfaces to combine ChatGPT with organization-specific data for improved innovation and efficiency. Invest in the development of custom solutions that seamlessly integrate ChatGPT with existing systems, databases, and workflows to fully harness the potential of the technology.

5.1.7 Create Employee Training and Support

Provide comprehensive training programs that laverage publicly available content where possible for employees to effectively use ChatGPT in their daily work. Offer hands-on workshops, e-learning modules, and on-the-job training to equip employees with the skills and knowledge necessary to use ChatGPT effectively. Provide ongoing support and resources to help employees adapt to the new technology and address any challenges they may face.

5.1.8 Monitor Progress and Impact

Regularly assess the performance of ChatGPT within the organization and evaluate its impact on specific use cases. Develop key performance indicators (KPIs) and metrics to track the technology's effectiveness, and use this data to inform future improvements and adaptations.

5.1.9 Develop Continuous Improvement and Adaptation

Stay current with AI advancements and adapt ChatGPT implementation to maximize benefits. The AI landscape is constantly evolving, and it's crucial to stay up-to-date with advancements in the field. Continuously evaluate the performance of ChatGPT and be prepared to adapt its implementation to maximize its benefits and stay ahead of the competition.

5.1.10 Recognize, Document, and Build Processes to Address Ethical Considerations

Develop guidelines for responsible use and educate employees on potential risks and challenges of AI technologies like ChatGPT. Be mindful of the ethical implications of using AI-powered technologies. Develop guidelines and best practices for responsible use, and ensure that employees understand the potential risks and challenges associated with AI, such as algorithmic bias or unintended consequences.

6 Barriers and Obstacles and How to Overcome Them

Inevitably, you will encounter barriers and obstacles during your ChatGPT implementation journey. Recognizing these challenges and understanding how to overcome them is essential to ensure a successful adoption.

The Domains will need to identify their own barriers and adapt solutions to their specific circumstances. The goal is to empower each Domain to lead their organization through the complexities of AI integration.

This is a living document and will be updated as we learn more. Please feel free to contribute to this document.

6.1 Barriers and Obstacles

6.1.1 Resistance to Change

: Employees and leaders may be resistant to adopting new technologies due to fear of job loss or discomfort with the unknown. To overcome this, emphasize the benefits of ChatGPT, such as increased efficiency and improved decision-making, and provide ample training and support. Encourage open discussions and showcase successful examples of AI adoption.

6.1.2 Lack of Technical Expertise

Limited knowledge of AI and ChatGPT may hinder successful implementation. Address this by investing in training programs, partnering with AI experts, or hiring professionals with relevant experience. Create an internal AI community for knowledge sharing and peer support.

6.1.3 Insufficient Collaboration

Inadequate communication and collaboration between departments can impede progress. Foster cross-functional teamwork through regular meetings, workshops, and collaborative platforms. Encourage leaders to champion the initiative and create a culture of cooperation.

6.1.4 Resource Constraints

Limited budget, time, or personnel can pose challenges. To overcome this, prioritize use cases based on potential impact and feasibility, and secure buy-in from top management for necessary resources. Consider leveraging external partnerships or outsourcing certain tasks to reduce internal workload.

6.1.5 Data Privacy and Security Concerns

Handling sensitive proprietary data may raise concerns. Collaborate closely with IT and Legal departments to establish robust data security protocols and comply with regulations. Communicate these measures transparently to build trust among employees and stakeholders.

6.1.6 Ethical Concerns

The potential for biased or unethical AI outcomes may create apprehension. Develop guidelines for responsible AI usage, create an ethics review board, and offer training on potential risks and challenges. Emphasize the importance of ethical AI practices throughout the organization.

As we reflect on the potential barriers and obstacles to ChatGPT implementation, remember that overcoming these challenges is an integral part of the journey towards AI-driven success. By anticipating and addressing these issues proactively, we can foster a resilient and adaptable organization that is well-prepared to navigate the ever-evolving landscape of artificial intelligence.

6.2 Overcoming Barriers and Obstacles

As noted above, there are many barriers and obstacles to ChatGPT implementation and overcoming these is crucial for realizing its full potential within the organization.

In this section, we provide a combination of conventional and less traditional strategies to address the challenges you might face during the adoption process. By embracing these adaptive approaches, we can foster a culture of adaptability and resilience, enabling your organization to successfully harness the power of AI-driven solutions like ChatGPT.

6.2.1 Conventional Strategies for Overcoming Barriers and Obstacles

1. Resistance to Change : To overcome resistance to change, emphasize the benefits of ChatGPT, provide ample training and support, encourage open discussions, and showcase successful examples of AI adoption.
2. Lack of Technical Expertise : Address this by investing in training programs, partnering with AI experts, hiring professionals with relevant experience, and creating an internal AI community for knowledge sharing and peer support.
3. Insufficient Collaboration : Foster cross-functional teamwork through regular meetings, workshops, and collaborative platforms, and encourage leaders to champion the initiative and create a culture of cooperation.
4. Resource Constraints : Prioritize use cases based on potential impact and feasibility, secure buy-in from top management for necessary resources, and consider leveraging external partnerships or outsourcing certain tasks to reduce internal workload.
5. Data Privacy and Security Concerns : Collaborate closely with IT and Legal departments to establish robust data security protocols and comply with regulations, and communicate these measures transparently to build trust among employees and stakeholders.
6. Ethical Concerns : Develop guidelines for responsible AI usage, create an ethics review board, and offer training on potential risks and challenges, emphasizing the importance of ethical AI practices throughout the organization.

6.2.2 Unconventional Strategies for Overcoming Barriers and Obstacles

I often find that we need to find "back-doors" and just different approaches in the context of change and transformation projects. Thus, here are some less conventional approaches to the barriers.

1. Gamification : Introduce gamification elements to the training and adoption process, incentivizing employees to engage with ChatGPT and learn its capabilities. Offer rewards or recognition for participation and achievements.
2. Reverse Mentoring : Encourage younger or more tech-savvy employees to mentor older or less experienced colleagues, facilitating knowledge sharing and promoting a more inclusive approach to technology adoption.
3. Innovation Contests : Organize internal contests or hackathons for employees to develop creative ChatGPT use cases or solutions, fostering a sense of ownership and excitement around the technology.
4. External Showcasing : Publicly share successful ChatGPT implementation stories or use cases to build a positive reputation, attract talent, and create a culture of innovation within the organization.
5. AI Sabbaticals : Offer employees the opportunity to take short-term sabbaticals to focus on AI-related projects or training, providing dedicated time for learning and exploration. This can help develop in-house expertise and promote a culture of continuous learning.

Part II

Domain Resources

This section contains resources that the domain teams may find useful for their specific domains.

- Education
- Clinical
- Research
- Business Operations

7 Clinical Domain

Table 7.1: 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?

7.1 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).

7.2 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 7.1.

7.3 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.

7.4 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.

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

emotional, and psychological drivers of health

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

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

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

8 Research

9 Education

10 Business Operations

Part III

Workstream Resources

11 Training and Workforce Development Resources

12 Training and Workforce Development Resources

13 Training and Workforce Development Resources

14 Training and Workforce Development Resources

15 Project Management and Support Resources

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.

Lindegaard, Stefan. 2023. "LinkedIn." https://www.linkedin.com/pulse/chatgpt-implementation-scaling-organization-your-guide-lindegaard/. https://www.linkedin.com/pulse/chatgpt-implementation-scaling-organization-your-guide-lindegaard/.

A 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

B Whitehouse AI Bill(s) of rights

The Whitehouse OSTP has developed an extensively documented Blueprint for an AI Bill of Rights. The document is a comprehensive overview of the current state of AI and the challenges it poses to society. The Blueprint for an AI Bill of Rights is a set of five principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of artificial intelligence. Developed through extensive consultation with the American public, these principles are a blueprint for building and deploying automated systems that are aligned with democratic values and protect civil rights, civil liberties, and privacy.

It is a good starting point for understanding the issues and the current state of the art.

- **Safe and Effective Systems** You should be protected from unsafe or ineffective systems
- **Algorithmic Discrimination Protections** You should not face discrimination by algorithms and systems should be used and designed in an equitable way.
- **Data Privacy** You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used.
- **Notice and Explanation** You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you.
- **Human Alternatives, Consideration, and Fallback** You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter.

B.1 Commentary and references

- Opportunities and blind spots in the White House's blueprint for an AI Bill of Rights
- 6 Reactions to the White House's AI Bill of Rights The nonbinding principles are being both celebrated and vilified
- Applying the Blueprint for an AI Bill of Rights
- **How Does the White House AI Bill of Rights Apply to Healthcare?** Experts from Mayo Clinic Platform and DLA Piper weigh in on how the White House's Blueprint for an AI Bill of Rights may impact healthcare and health AI regulation.

- The US AI Bill Of Rights Should Kickstart The Debate On Bias In Artificial Intelligence