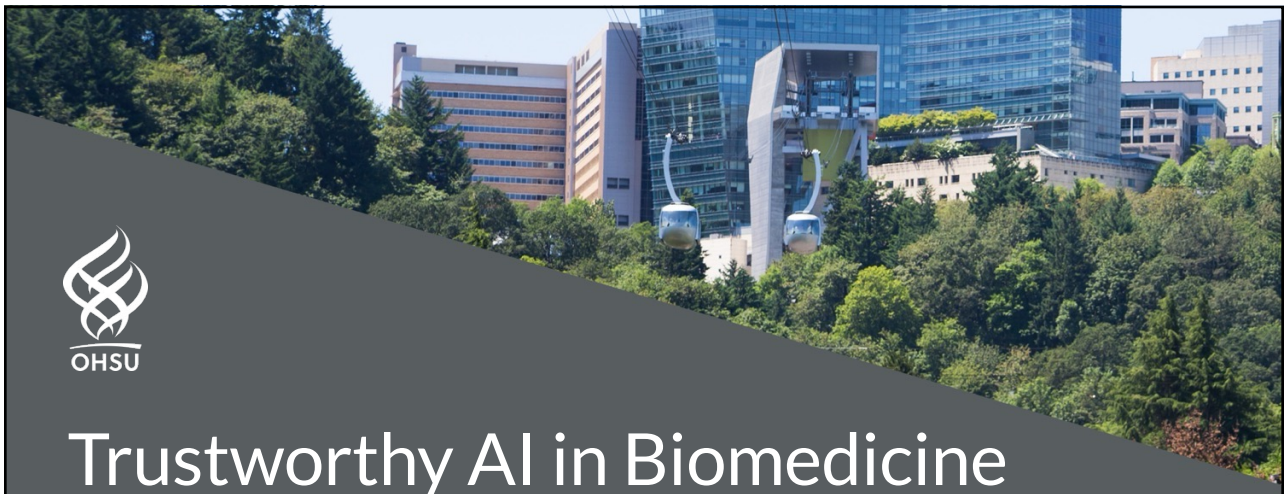


Trustworthy AI in Biomedicine and Health

What is Biomedical and Health Informatics? - http://informatics.health/
William Hersh

Oregon Health & Science University

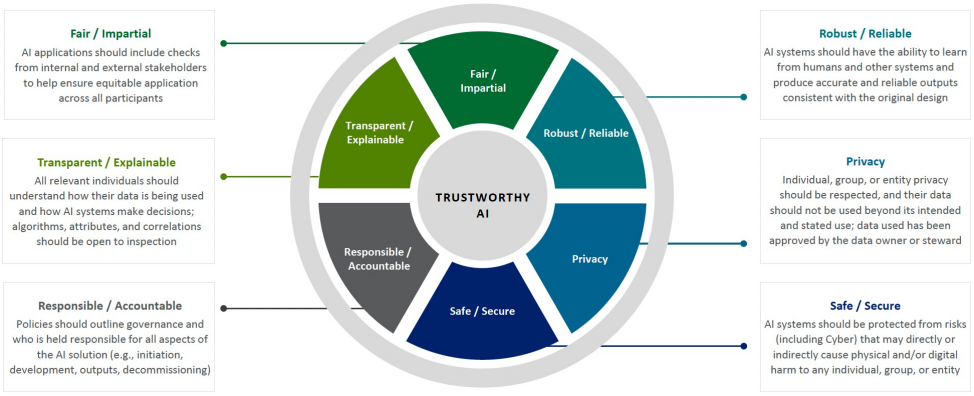
1

Trustworthy AI

- Trust
- Fairness
- Education
- Regulation and liability
- Ethical and equitable AI
- Organizational statements
- Moving forward and future directions

2

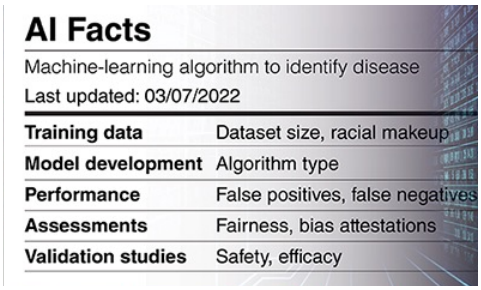
How do we ensure trustworthy AI? (Ammanath, 2022)



HHS Trustworthy AI Playbook (2021)

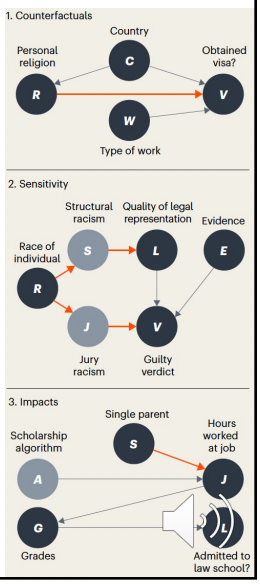
How do we ensure trustworthy AI (cont.)?

- Recommendations from US government
  - Trustworthy AI playbook (HHS, 2021)
  - AI risk management framework (NIST 2021)
- Recommendations in specific uses
  - Clinical (Grote, 2020)
  - Clinical research (Volovici, 2022)
  - Genomics (Whalen, 2022)
- Other identified needs
  - "Nutrition fact sheet" about model and validation (Sendak, 2020; Cohen, 2022)
  - "Medical algorithm audit" (Liu, 2022)
  - Continuous monitoring and improvement (Feng, 2022)



How do we ensure fairness?

- Roadmap to fairer algorithms (Kusner, 2020)
- Framework (Xu, 2022) and toolkit (Bellamy, 2019) to insure fairness and reduce bias for individuals and groups
- Data ownership
  - Who owns patient photographs? (Davis, 2020; Davis, 2022)
  - De-identified data can be re-identified with other data; who should own resulting data? (Ross, 2022)
  - Data should be "democratized" (Allen, 2019), especially for disabled (Newman-Griffis, 2022) and historically underserved groups (Gottschalk, 2022; Hossain, 2023)



5

Accountability

- Even when AI employed fairly and transparently, must also address bias in hands of "imperfect users" (Kostick-Quenet, 2022) and structural challenges in healthcare (London, 2022)
- Concerns over control by tech companies; models should be open-source (Toma, 2023)
- Models should be validated locally and recurrently (Youssef, 2023)
- Safe harbors for red-teaming and evaluation (Longpre, 2024)

6

Next steps for trustworthy AI

- Coalition for Health AI "Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare" (CHAI, 2023) – trustworthy AI should be
  - Valid with respect to accuracy, operability and meeting intended purpose and benefit (clinical validation)
  - Reliable
  - Testable
  - Usable
  - Beneficial
- https://www.coalitionforhealthai.org/
- CHAI creating system to certify and register AI models and publish details about performance before widely adopted (Ross, 2024)

7

Education of clinicians and informaticians

- AI should build on competencies in clinical informatics (Hersh, 2014; Hersh 2020; Hersh, 2023)
- Others note
  - Clinicians must be prepared to practice in a world of AI (James, 2022)
  - Medical schools face dual challenges of needing to teach about AI in practice but also adapt to its use by learners and faculty (Cooper, 2023)
  - Physicians must be prepared for "clinical algorithm era" (Goodman, 2023)
- AI-specific competency frameworks (Russell, 2023; Liaw, 2023)

1. Find, search, and apply knowledge-based information to patient care and other clinical tasks
2. Effectively read from, and write to, the electronic health record (EHR) for patient care and other clinical activities
3. Use and guide implementation of clinical decision support (CDS)
4. Provide care using population health management approaches
5. Protect patient privacy and security
6. Use information technology to improve patient safety
7. Engage in quality measurement selection and improvement
8. Use health information exchange (HIE) to identify and access patient information across clinical settings
9. Engage patients to improve their health and care delivery though personal health records and patient portals
10. Maintain professionalism in use of information technology tools, including social media
11. Provide clinical care via telemedicine and refer patients as indicated
12. Apply personalized/precision medicine
13. Participate in practice-based clinical and translational research
14. Use and critique artificial intelligence (AI) applications in clinical care

8

Competencies for use of AI-based tools by healthcare professionals (Russell, 2023)

Domains	Details
Basic knowledge of AI	Explain what AI is and describe its healthcare applications
Social and ethical implications of AI	Explain how social, economic, and political systems influence AI-based tools and how these relationships impact justice, equity, and ethics
AI-enhanced clinical encounters	Carry out AI-enhanced clinical encounters that integrate diverse sources of information in creating patient-centered care plans
Evidence-based evaluation of AI-based tools	Evaluate the quality, accuracy, safety, contextual appropriateness, and biases of AI-based tools and their underlying datasets in providing care to patients and populations
Workflow analysis for AI-based tools	Analyze and adapt to changes in teams, roles, responsibilities, and workflows resulting from implementation of AI-based tools
Practice-based learning and improvement regarding AI-based tools	Participate in continuing professional development and practice-based improvement activities related to use of AI tools in healthcare

9

Competencies for use of AI in primary care (Liaw, 2023)

Domain	Bottom Line	Competency
Foundational knowledge	What is this tool?	Clinicians will explain the fundamentals of AI, how AI-based tools are created and evaluated, the critical regulatory and socio-legal issues of the AI-based tools, and the current and emerging roles of AI in health care.
Critical appraisal	Should I use this tool?	Clinicians will appraise the evidence behind AI-based tools and assess their appropriate uses via validated evaluation frameworks for health care AI.
Medical decision making	When should I use this tool?	Clinicians will identify the appropriate indications for and incorporate the outputs of AI-based tools into medical decision making such that effectiveness, value, equity, fairness, and justice are enhanced.
Technical use	How do I use this tool?	Clinicians will execute the tasks needed to operate AI-based tools in a manner that supports efficiency and builds mastery.
Patient communication	How should I communicate with patients regarding the use of the tool?	Clinicians will communicate what the tool is and why it is being used, answer questions about privacy and confidentiality, and engage in shared decision making, in a manner that preserves or augments the clinician-patient relationship.
Unintended consequences (cross-cutting)	What are the "side effects" of this tool?	Clinicians will anticipate and recognize the potential adverse effects of AI-based tools and take appropriate actions to mitigate or address unintended consequences.

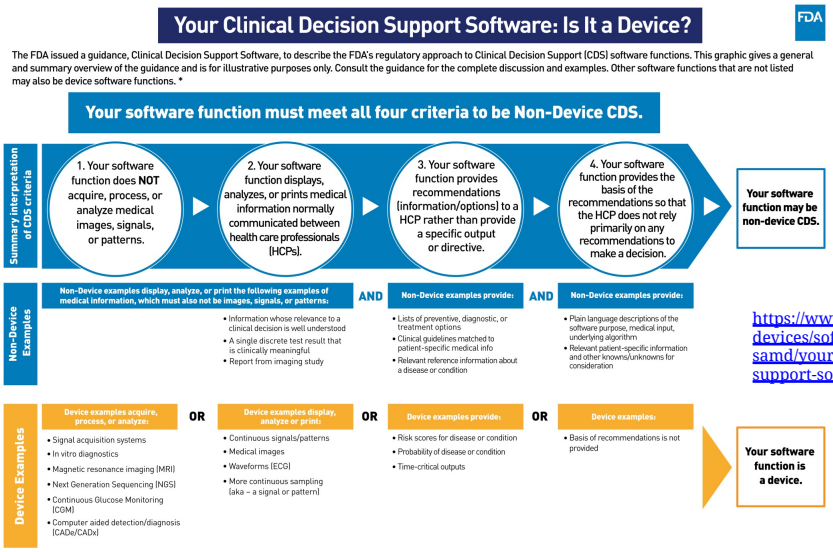
10

Regulation and liability

- What is liability for use or non-use (Price, 2019)?
- Need lifecycle-based regulation – when to allow incremental changes vs. wholesale renewal (Hwang, 2019)
- AMIA position paper on "adaptive clinical decision support" (Petersen, 2021)
- For first 130 medical AI devices approved by US FDA, only 4 had undergone prospective evaluation, with number of evaluation sites and sample sizes often not reported (Wu, 2021)
- About 700 and growing medical AI devices approved as of 2024
  - https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

11

FDA guidelines for when software is or is not medical device



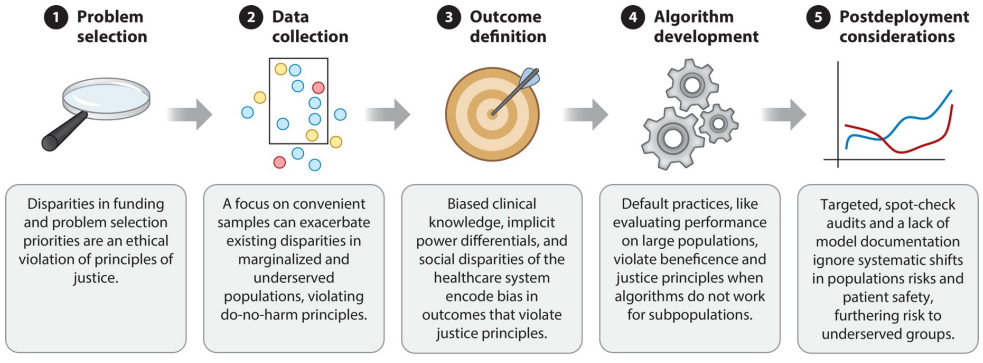
12

Regulation becoming more imperative in era of generative AI?

- Calls for regulation by companies, governments, and others (Park, 2023)
  - Google and Epic resist stronger regulation (Ross, 2023)
- Responsible use – need Code of Conduct for use of AI in healthcare (Dorr, 2023)
- Regulation should be driven by results of patient outcomes (Ayers, 2024)

Ethical AI

- AI and data ethics must be guided by justice and transparency (Basl, 2021)
- Call for responsible AI (Shneiderman, 2021; Shneiderman, 2022)
- Framework (Chen, 2020)



Ethical checklist for generative AI (Ning, 2023)

Ethicist involvement		
☐ Ethical discussions disclosed below involved contributions from philosophers and/or ethicists.	☐ Ethical discussions disclosed below were peer-reviewed by philosophers and/or ethicists.	
☐ Additional involvement of philosophers and/or ethicists from relevant ethics subfield: _____	☐ Peer-reviewed by philosophers and/or ethicists from relevant ethics subfield: _____	
Ethical discussions		
Ethical Principles	Discussions in text (check all that apply)	Position in text
Accountability	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Autonomy	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Beneficence	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Equity	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Integrity	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Morality	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Non-maleficence	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Privacy	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Security	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Transparency	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Trust	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	
	☐ Not applicable	
Others: _____	☐ Discussed relevant issues arising from generative AI	
	☐ Further discussed possible solutions	
	☐ Discussed use of generative AI to resolve relevant issues	
	☐ Further discussed limitations and inadequacies	

Equity in AI

- Need framework to incorporate ethical AI principles into development process in ways that intentionally promote racial health equity and social justice (Dankwa-Mullan, 2021)
- NIH Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD)
  - https://datascience.nih.gov/artificial-intelligence/aim-ahead
  - https://aim-ahead.net/
- Democratizing immunology datasets (Bhattacharya, 2021)
- Algorithmic Bias Playbook – how to define, measure, and mitigate racial bias in live algorithms (Ross, 2021; Obermeyer, 2021)
- Principles to address impact of algorithm bias on racial and ethnic disparities in health and healthcare (Chin, 2023)

Organizational statements about AI

- Professional associations
  - AMIA principles (Solomonides, 2021)
  - American Medical Association (AMA) (2023)
- WHO Ethics and Governance of AI for Health (2021)
  - Follow-on report focused on LLMs (2024)
- US White House
  - Blueprint for an AI Bill of Rights (2022)
  - US Presidential Executive Order on Safe, Secure, and Trustworthy Development and Use of AI (2023)
- VA Trustworthy AI Framework (VA, 2023)
- Responsible AI for Social and Ethical Healthcare (RAISE) (Beth-Goldberg, 2024)

17

AMIA principles for AI (Solomonides, 2022)

Rule	Principle	Definitions
I.	Autonomy	AI systems must protect the autonomy of all people and treat them with courtesy and respect including facilitating informed consent.
II.	Beneficence	AI systems must be helpful to people modeled after compassionate, kind, and considerate human behavior.
III.	Nonmaleficence	AI systems shall "do no harm" by avoiding, preventing, and minimizing harm or damage to any stakeholder.
IV.	Justice	AI systems must include equity for people in representation and access to AI, its data, and its benefits. AI must support social justice.
V.	Explainability	AI developers must describe AI systems in context-appropriate language so that their scope, proper application, and limitations are understandable.
VI.	Interpretability	AI developers must endow their systems with the functionality to provide plausible reasoning for decisions or advice in accessible language.
VII.	Fairness	AI systems must be free of bias and must be nondiscriminatory.
VIII.	Dependability	AI systems must be robust, safe, secure, and resilient. Failure must not leave any system in an unsafe or insecure state.
IX.	Auditability	AI systems must provide and preserve a performance "audit trail" including internal changes, model state, input variables, and output for any system decision or recommendation.
X.	Knowledge management	AI systems must be maintained including retraining of algorithms. AI models need listed creation, revalidation, and expiration dates.
Organizations deploying or developing AI		
XI.	Benevolence	Organizations deploying or developing AI must be committed to use AI systems for positive purposes.
XII.	Transparency	AI must be recognizable as such or must announce its nature. AI systems do not incorporate or conceal any special interests and deal even-handedly and fairly with all good faith actors.
XIII.	Accountability	AI systems must be the subject of active oversight by the organization, and any risk attributed to AI must be reported, assessed, monitored, measured, and mitigated as needed. Complaints and redress must be guaranteed.
Special considerations		
XIV.	Vulnerable populations	AI applied to vulnerable populations requires increased scrutiny to avoid worsening the power differential among groups.
XV.	AI research	Academic and industrial research organizations must continue to research AI to address inherent dangers as well as benefits.
XVI.	User education	AI developers have a responsibility to educate healthcare providers and consumers on machine learning and AI systems.

18

AMA principles

- AMA – prefers term "augmented intelligence" (Crigger, 2022)
  - Trustworthy AI must focus on ethics, evidence, and equity
- Principles for Augmented Intelligence Development, Deployment, and Use (AMA, 2023)
  - "Whole of government" oversight, with collaboration from all other stakeholders
  - Transparency in use
  - How disclosure and documentation should be handled
  - Develop and adopt generative AI policies that anticipate and help manage risks associated with technology prior to use
  - Physician liability for use of AI-enabled technologies should be limited and adhere to current legal approaches to medical liability
  - Patient privacy and data security must be top priority
  - Bias in algorithms be proactively identified and mitigated to promote health equity
  - Payor use of automated decision-making systems should not reduce needed care, nor systematically withhold care from specific groups

How do we move forward?

- Need "algorithmovigilance" (Embi, 2021) and human vigilance (Adler-Milstein, 2024)
- ML in healthcare must have trustworthiness, explainability, usability, and transparency (Cutillo, 2020)
- Moving beyond prediction of diagnostic labels to "wayfinding" (Adler-Milstein, 2021)
- Must shift ML for healthcare from development to deployment and from models to data (Zhang, 2022)
- Also need
  - AI-capable organizations (Novak, 2023)
  - Sustainable deployment – model maintenance (Davis, 2024)
  - Interoperable digital eco-system (Mandl, 2024)
  - Need "translational" AI from basic to clinical science (Hersh, 2024)

Implementation challenges

- As models change healthcare practice, they must be adapted (Lenert, 2019)
- Healthcare organizations may not be able to implement (Panch, 2019)
  - May not be able to re-engineer care
  - Issues of data – who owns, can use, or maintain responsibility?
- Too much emphasis in research on algorithm success ("winner's curse") above all else (Sculley, 2018)?
- ML system more accurate than humans in robustly simulated mammography interpretation (McKinney, 2020), but may be wrong question, since value of screening is uncertain (Aschwanden, 2020)
- AI is more than the algorithm (Kwong, 2024)

Toward the future

- Real-world use still modest
  - As of Sept 2023, only 21% of medical groups using AI applications in practice (MGMA, 2023)
  - EHR usability, patient communications, and billing outrank AI as top tech priorities among medical groups (MGMA, 2023)
  - AI tools used by only 38% of physicians (AMA, 2023)
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
  - True for all physicians, even Dr. McCoy?