AI Best Practices in the Wiens Lab

Marissa Miller & Dr. Kirsten Wiens
Department of Epidemiology and Biostatistics
Temple University College of Public Health
September 2023

General Guidelines

The Wiens Lab is supportive of the use of generative AI tools in appropriate applications when used ethically and transparently. Generative AI tools can be used to strengthen one's own work and save time, not to create original ideas or outsource one's thinking.

Limitations

Prior to the usage of any AI, it is necessary to ensure that you are legally and ethically able to share the information that you are copying into the program. Sharing of source code, unpublished work, sensitive information, or data people have not consented to share in this manner is a breach of privacy. Further, using AI to create original work on one's behalf is a form of plagiarism. Finally, AI tools are imperfect. They may generate false or misleading information, are only as exhaustive as their sources, and, intentionally or not, are vulnerable to bias. The process through which AI systems process data and create predictions is often not understood, by those using it and the AI itself, so identifying the source of a problem or how a connection was made can be difficult.

Approved Usages

AI tools may be used to assist in writing, researching, or programming when properly cited. AI may be used to brainstorm, search for relevant sources (though it should not be the only source you use to find literature), summarize published work, create outlines, change language style, and check grammar and spelling. In a coding language that you are familiar with, AI tools can be used to create simple scripts or templates, work through specific bugs or problems, improve style, review code snippets, or test software using AI-generated synthetic datasets. Full project code and draft manuscripts should not be copied into generative AI tools unless they are approved (by your coauthors and the IRB) for public circulation.

Resources

The **Wiens Lab Supplementary Notes and Information Regarding Ethical Artificial Intelligence Use** (appended below) contains further information and freely available resources. These notes and guidelines are subject to change as the field of artificial intelligence is ever-growing and best practices may need to adapt.

Wiens Lab Supplementary Notes and Information Regarding Ethical Artificial Intelligence Use

Authors: Marissa Miller & Dr. Kirsten Wiens

Last Updated: September 2023

Table of Contents

I. Ethical Principles

A. UNESCO Core Principles: A Human Rights Approach

1. Proportionality and Do No Harm
2. Safety and Security
3. Right to Privacy and Data Protection
4. Multi-stakeholder and Adaptive Governance & Collaboration
    a) International law & national sovereignty must be respected in the use of data.
    b) Participation of diverse stakeholders is necessary for inclusive approaches to AI governance.
5. Responsibility and Accountability
    a) AI systems should be auditable and traceable.
    b) There should be oversight, impact assessment, audit, and due diligence mechanisms in place to avoid conflicts with human rights norms and threats to environmental wellbeing.
6. Transparency and Explainability
7. Human Oversight and Determination
8. Sustainability
9. Awareness & Literacy
    a) Public understanding of AI and data should be promoted through open & accessible education, civic engagement, digital skills & AI ethics training, media & information literacy.
10. Fairness and Non-Discrimination

B. Jobin et al. Review of Global AI Ethics: Key Principles
    1. Transparency
    2. Justice, Fairness and Equity
        a) Protection against bias and discrimination
        b) Equity and fairness in accessing AI data, tools, and benefits
    3. Non-Maleficence
    4. Responsibility and accountability
    5. Privacy
    6. Beneficence
    7. Freedom and Autonomy
    8. Trust
        a) Trustworthy AI research and technology
        b) Trustworthy AI developers and organizations
        c) Trustworthy design principles
    9. Sustainability
    10. Dignity
        a) AI should not diminish or destroy, but respect, preserve or increase dignity
    11. Solidarity
        a) In protecting against the implications of AI on the labor market
        b) In ensuring fair distribution of AI benefits to all, especially vulnerable groups

C. Atenas et al. Review of Data Ethics Frameworks: Key Components
    1. Respect Autonomy
        a) Informed consent needs to go beyond data collections, but clearly describe how and by whom data will be used and how the results will be published
    2. Respect Privacy
    3. Promote Fairness
    4. Address Equality
    5. Do No Harm
    6. Promote Sovereignty
        a) This means that subjects, both at individual and collective level, should be in a position to decide when and what data they wish to disclose and to whom, and that a refusal to share data, should not impede or obstruct their access
    7. Address Bias
    8. Challenge Power Structures

D. General Guidelines
    1. Use AI as a stool to strengthen your own work
        a) If it doesn't sound like something you'd say anyway, don't use it
        b) "Outsource your labor, not your thinking"
    2. Be transparent about AI use and cite when used

3. “Ethical people will find a way to be ethical, just like unethical people will find a way to unethical”
4. Always double-check anything- resources, citations, drafts-created by AI to ensure its existence, accuracy, and validity

II. Warnings/Limitations

A. AI Vulnerable to Algorithmic Bias

1. If data put into the program is biased, results will be biased (feedback loop bias)
    a) Intentionally or unintentionally
2. Bias in who is represented
    a) Women and minorities historically absent
    b) When data on which the algorithm is trained are not representative of the target patient population, selection bias occurs.
3. Bias in who benefits, through access, to these new technologies
4. Examples
    a) Amazon attempted to use AI to streamline, and remove bias, from the hiring process. AI used historical hiring data to make decisions, and since historically men were hired, the system was biased against women.
    b) Predictive Policing: using machine learning to ‘predict’ who will commit crimes
        (1) Chicago’s Strategic Subjects List (known as the heat list) was used to intimidate members on the list as the police threatened them with surveillance and/or used a form of confirmation bias, assuming members on the list who were near a crime, were guilty of the crime
        (2) Use of heat list in this way had no statistical impact on the likelihood of individuals on the list being involved in gun violence, nor on the overall gun violence in their communities but radically increased the likelihood of being arrested
        (3) Important to keep in mind the data, historical crime statistics, likely to be biased and often categorize people according to their belonging to a group rather than by more individual characteristics
    c) Use of past data on health care spending led to black patients needing to be much sicker than white patients in order to be recommended the same level of care
        (1) Black patients who were considerably sicker than white patients were given the same risk score
        (2) Risk score guides course of care, which resulted in a detriment to the health of black patients

(3) Past data on health care spending limited by historically biased wealth disparities, opportunities for access to care, trust in medical interventions, past biases of medical providers, etc.

B. Black Box problem
    1. Difficulty in understanding how AI systems and machine learning models process data and generate predictions or decisions
        a) AI does not always track how it comes to conclusions, so it also does not always know
        b) Can be incredibly difficult to backtrack a problem and identity why/how AI was wrong
    2. Program will make unintended connections
        a) Self-driving uber hit pedestrian as the person was jaywalking and the system did not recognize them without a crosswalk
        b) Program designed to identify skin cancer was fed thousands of images to train the program. Program biased towards images with rules, because medical images (actual skin cancer) likely to include ruler to measure size of tumor
    3. Obstacle to 'fixing' issues
        a) Attempts to remove abusive language from ChatGPT lead the software to over remove anything mentioning or produced by marginalized groups

C. False Information
    1. It may make up false information
        a) AI provided a seemingly accurate biography for an entirely fictitious person
    2. The free version of ChatGPT has "Limited knowledge of world and events after 2021"

D. AI Plagiarism & Copyright
    1. Using AI to create 'original work' on one's behalf is a form of plagiarism
        a) Presenting ideas as your own original work, that you did not create
        b) AI does not cite its sources, so neither can you
    2. AI generated images come from other images
        a) AI software is often trained by scraping content from the internet. This means the output could be a copy of or a derivative work of someone else's copyright-protected work.
        b) Getty Images suing over this
    3. AI & Generative Code
        a) To be deployed in a commercial software, the codebase needs to be licensed to the end user
        b) The tool used must have a commercial license

4. ChatGPT does not provide citations, impossible to credit where it's getting its information
    a) Uses web so it had to come from somewhere
    b) If the input is someone else's copyright-protected work or someone else's copyright-protected work was used to create the output, the output is an unauthorized derivative work of that other person's work, aka copyright infringement.
    c) Reinforced by black box issue
5. ChatGPT OpenAI terms states that there's no guarantee that it would not produce the same output for two different users.
6. US Copyright office investigation
    a) The Copyright Office has launched an initiative to examine the copyright law and policy issues raised by artificial intelligence (AI) technology, including the scope of copyright in works generated using AI tools and the use of copyrighted materials in AI training. After convening public listening sessions in the first half of 2023 to gather information about current technologies and their impact, the Office will publish a notice of inquiry in the *Federal Register*.

E. Privacy
1. People do not always know/understand their data is being used and so have not consented for their data to be used in this way
    a) Do people know what data is collected, how it is stored, the security of that storage, how their data is used and how it is linked?
    b) Once linked anonymized data can be de-anonymized and reveal personal information
        (1) An AI tool trained on medical images learned to discern patient's self-reported race, even though this was not its goal
2. Some information (identifying information, health-related data, financial information, etc.) is sensitive and should not be shared to prevent against data from being leaked
    a) Generated code might not ensure/implement security safeguards
    b) Potential that those who use generative AI code less likely to write correct and secure solutions to programming problems
3. Is the information given to the AI, information you *legally* are allowed to share?
    a) National Institutes of Health
        (1) NIH scientific peer reviewers are prohibited from using natural language processors, large language models, or other generative AI technologies for analyzing and formulating peer review critiques for grant applications and R&D contract proposals.

(2) As part of standard pre-meeting certifications, all NIH peer reviewers and members of NIH advisory councils will be required to certify a modified Security, Confidentiality and Nondisclosure Agreement signifying that they fully understand and will comply with the confidential nature of the review process, including complete abstention from Artificial Intelligence tools in analyzing and critiquing NIH grant applications and contract proposals.

(3) Grant application and contract proposal materials and other privileged information cannot be shared or disseminated through any means or entity

b) Source Code

(1) The unauthorized exposure of source code can lead to the loss of intellectual property and competitive advantage.

(2) The upload of proprietary source code to an AI tool, risks exposing valuable trade secrets to third parties.

F. Laws

1. There is no comprehensive federal legislation dedicated solely to AI regulation.
    a) Existing laws and regulations that touch upon certain aspects of AI, such as privacy, security, and anti-discrimination
2. Record of Laws by State
3. White House AI Bill of Rights
4. Many algorithms being used, especially in the clinical space, are under-regulated

G. Human Rights

1. Those working to support AI systems, known as ghost workers, can be woefully underpaid, are often skilled members of vulnerable populations, may be subject to emotionally disturbing work and work in harsh conditions, fearing reprisal
    a) Work in labeling data, content moderation, captioning/translation work, research, beta testing, verification of details (such as addresses/phone numbers), etc.
    b) Often labor is outsourced to other countries to reduce cost to companies
2. Creator of ChatGPT, OpenAI, hired out the task of labeling images that were explicit, abusive or otherwise toxic and meant to be removed from the platform to a firm in Kenya.
    a) Kenyan workers paid between $1.37-$1.42, or $170 a month, whereas foreign workers are paid $2.20 an hour and receive a relocation stipend
    b) Work contracted to Sama for $12.50
3. Amazon's Mechanical Turk

a) Freelance workers, known informally as turkers, complete requests employers post on Amazon's Mechanical Turk website, doing various tasks
b) Most tasks pay a dime or less and many turkers reported earning under $5 an hour
c) Analysis of reported wages found workers earned a median hourly wage of only ~$2/hr, and only 4% earned more than $7.25/hr

4. TikTok Content Moderators
    a) Workers tasked with viewing, and then removing, graphic content
    b) Workers in Morocco complain of 12 hour shifts, little breaks and low pay

III. Approved Usages

A. Personal Correspondence

1. Reply to Emails
    a) For example, crafting a particular tone for an important communication, without putting anything sensitive or identifying into the algorithm
2. **Draft** personal statements
    a) Be careful to not just copy and paste what it gives; make sure that you re-write in your voice, highlighting your unique accomplishments

B. Research Assistance

1. **Keep in mind the limitations of the tool you are using, as the output will have similar limitations, and so should not be solely relied upon**
    a) The free version of ChatGPT does not have access to information past 2021
    b) Many tools use specific databases which are not exhaustive
    c) Use AI tools to build on an idea, not to create one
2. Summarize published work
    a) lengthy passages for main ideas
    b) Search for specific information (definition, method used, etc.)
    c) Put in something you've written and ask for main ideas/most relevant adjectives
        (1) As long as you are OK with sharing this with the AI
        (2) For example, this notes page is probably OK, but you wouldn't want to copy and paste your grant proposal in
3. Brainstorming
    a) Brainstorm research questions that are similar or more specific to your original question
    b) Brainstorm potential consequences/limitations of an idea
    c) Brainstorm potential counter arguments
4. Creating Outlines

5. Generate sub-topics
6. Ask for resources
    a) Find relevant research articles
    b) Identify organizations and experts relevant to a topic
7. Create citations for existing sources

C. Preparation
1. Presentations
    a) Only upload presentations where you can legally and ethically share all the information in it publicly
    b) Simulate audience questions
    c) Review grammar
    d) Prompt talking points
    e) Develop speaker notes
    f) Assess/change readability of information
2. Interviews
    a) Ask for questions, and variations of questions, that an interviewer is likely to ask
    b) Gather information about the company or industry

D. Check grammar and spelling
1. Removing redundant words
2. Add punctuation

E. Change language style
1. Explain complicated stuff in a more simplified way
    a) Explain (this) in a way a (age/grade) would understand
    b) Plain language
    c) “What does this mean”
2. Find synonyms
3. Change communication style
    a) Non-violent
    b) Formal
    c) Informal

F. Coding Assistance
1. **Prior to use of any AI tool, review sections on <u>privacy</u> and <u>plagiarism</u> to ensure it is ethical and legal to use!**
2. It is necessary to have some familiarity with the coding language that you are asking it to write in, so that you can provide supervision and make sure that what it's giving you is appropriate

3. Creating simple script
4. Getting suggestions to complete code
5. Code Review/Bug Fixes
6. Code refactoring
7. Improve style
    a) Does code adhere to style guidelines?
    b) Is it consistent?
    c) Is it readable across a codebase?
8. Test software
    a) Generate test cases/code
    b) Test documentation
    c) Test result analysis

IV. AI Resources

A. ChatGPT

1. Create structure
2. As an editor
3. For personal correspondence
4. To predict, or prepare for, something
5. For code assistance
6. **Not** for finding sources
7. **Not** for conducting literature reviews

B. Paperpal

1. Academic writing assistant
2. Tools
    a) Evaluates text for grammar, spelling and conventions
    b) Rephrase portions of text to add variety, avoid repetitions
    c) Trim text to shorten content to meet word-counts
    d) Find contextual synonyms
3. Free version is limited in number of times a tool can be used

C. SpeakNotes App

1. Transcribe and summarize voice notes
2. Notes can be organized easily into folders
3. Search notes for ideas/terms
4. Editing notes only available on paid version (8/23)

D. Consensus
    1. Ask it a question and it'll give you a summary of top 5-10 (real) published papers, along with links to the article, article abstract, article citations, etc.
    2. It will also give you the consensus of the scientific community about the question.
        a) Feature still in beta phase (8/23) and not completely accurate
        b) Limited amount of consensus reports on free version
    3. Integrates easily with Zotero
    4. Limitations
        a) Can only provide consensus on questions posed as Yes/No
        b) Maximum number of papers analyzed is 20, possible important research is missing
    5. Data source: Semantic Scholar

E. Elicit
    1. Brainstorm research questions
    2. Find relevant sources
        a) Provides links to articles, generated abstract summary and summary of top papers
        b) Allows to include information on intervention, outcomes measured, number of participants
    3. Rephrase sentences
    4. Suggest search terms
    5. Get independent and dependent variables from hypothesis, question, or statement
    6. Elaborate on short statement or clarify a concept
    7. Recommend data processing
    8. Generate ways a plan could fail
    9. Reframe a critical statement more positively
    10. Define terms
    11. Brainstorm possible counterarguments
    12. Bias busters
    13. Create analogous problem describer
    14. Data Source: Semantic Scholar

F. SciSpace
    1. “Read with AI”
        a) Explain portions of text
        b) Find related papers

c) Ask questions regarding main ideas, methods used, data sources, limitations, conclusions, etc. of the paper
d) Summarise entire text or section of text

2. Find resources for a research question, including article insights
3. Help paraphrase
4. Create citations
5. Data source: “a repository of research papers across domains — with metadata of 200 million+ papers and 50 million+ Open Access full-text PDFs”

G. Inciteful

1. Interactive graphs of different papers on a given article to find relevant articles
   a) Creates lists of sources by top authors, most recent papers, most important, review papers
2. Literature Connector tool allows you to enter two papers and it will give you an interactive visualization showing you how they are connected by the literature
3. Data Source: Semantic Scholar, Open Alex, CrossRef, OpenCitations

H. Research Rabbit

1. Assist in literature review process
   a) Creates maps of how a paper is connected to other papers, finding new authors and papers related to original paper
      (1) Can add to this graph
   b) Will create ‘feed’ of studies using past saves it thinks are helpful
   c) Add notes to papers
2. Easily connects with Zotero
3. Data Source: Semantic Scholar

I. Connected Papers

1. Creates maps of how a paper is connected to other papers, finding new authors and papers related to original paper
   a) Cannot add to graph
2. Provides lists of prior and derivative works
3. Limits use by month
4. Data source: Semantic Scholar

V. AI Citation

A. APA

1. As of May 2023, APA does not have official guidelines on citing AI

B. Current Examples

1. Citation:

a) OpenAI. (date). ChatGPT: Optimizing language models for dialogue (Version 3.5) [Web application]. Retrieved op January 26, 2023, from https://openai.com/blog/chatgpt/

b) OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat

2. Parenthetical citation: (OpenAI, 2023)
3. Narrative citation: OpenAI (2023)
4. Putting full response, including prompt, into appendix of paper

C. Links

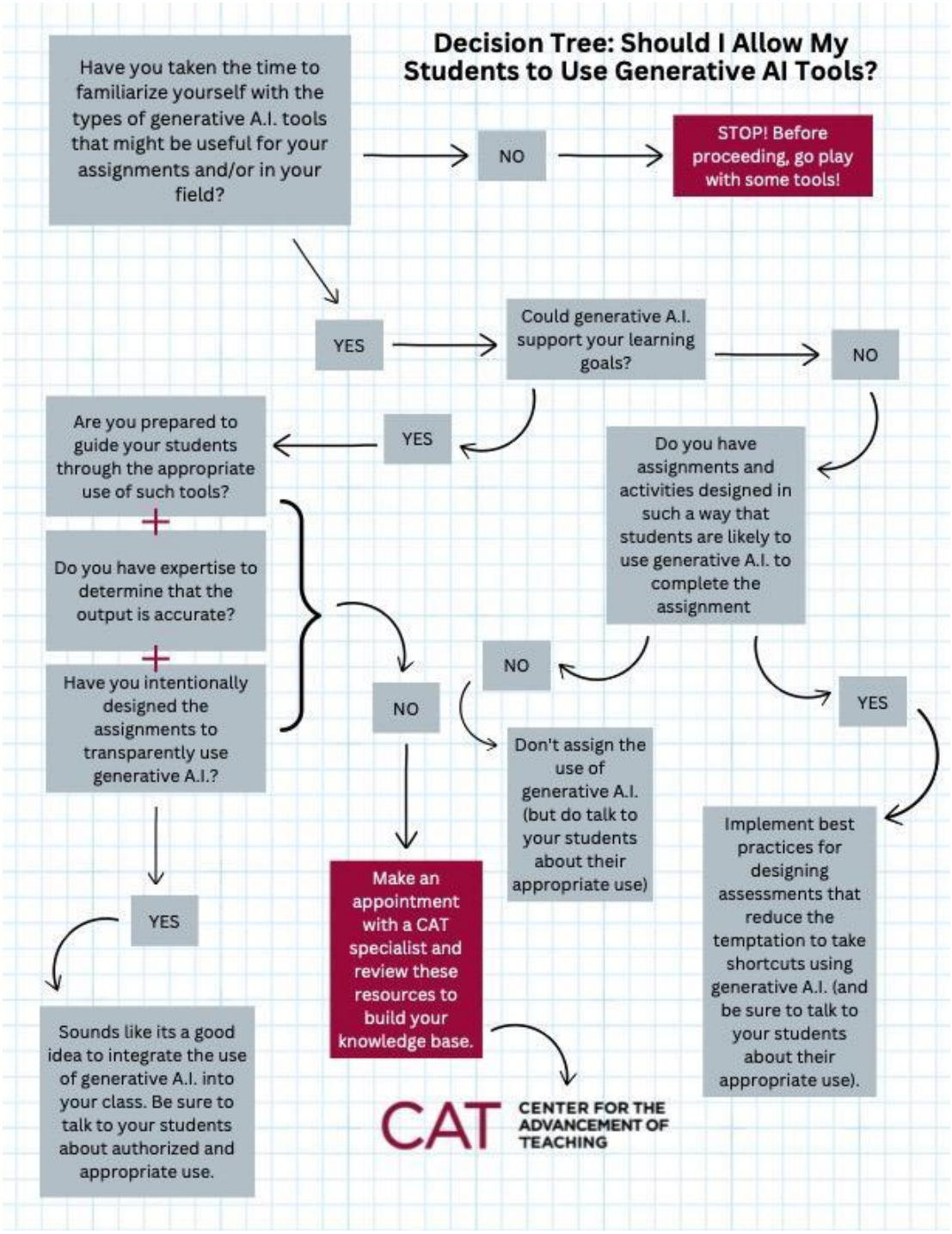
1. https://sdmesa.libguides.com/c.php?g=1296253&p=9521243
2. https://apastyle.apa.org/blog/how-to-cite-chatgpt

VI. Temple Stance

A. Sample Syllabus Statement from Temple's Center for the Advancement of Teaching

B. Faculty Guide

C. Decision Tree for Student Use

1. 

D. AI Detection Apps

1. Writefull
2. Writer

VII. Resources

A. LinkedIn Article
B. Mushtaq Bilal