CS 309 (Intro to AI Literacy): Pre-class Reflections

Abdon Morales

October 28, 2024

Week 1

Question: Reflecting on the topics from this module, write a paragraph (5–10 sentences) discussing your current perspective. Choose one of the applications of AI mentioned in this week's course materials (e.g., autonomous vehicles, review of job applications, evaluation of job performance, assistive robots, etc.)

Imagine you are the CEO at a company developing the tool you have chosen. Reflect on some of the potential risks and benefits of this tool. As CEO, what are your top ethical priorities for your team that is developing this tool?

My perspective of the topics discussed over ethics in the development of artificial intelligence combines the idea of human interaction and the need for a safety net to ensure AI is used responsibly. As mentioned in the podcast, "What is AI, Anyway?", the goal of ethics for AI is to ensure AI does more good than harm, which ultimately depends on how we interact with these technologies. The podcast highlights, "AI is not doing anything by itself. It'll be AI technologies and it'll be people who are using AI technologies that are doing things that are either beneficial or harmful." If I were a CEO at a company developing a tool for reviewing job applications, some potential risks include the tool makes biased decisions based on the hard-coded parameters such as age, race, gender, work experience when coming into a conclusion if job applicant is acceptable or not. This can lead to unjust outcomes as seen in the case of a news-bite shown in the lecture video of a possible "racist AI" or beyond. However, some benefits of this tool is its cost-effectiveness and efficiency, where it can process large volumes of applications without the consumption of a lot of resources and money. As CEO, some ethical priorities is for my team for developing this tools is for it to be fair and non-discriminant, transparent and accountable through clear guidelines, human-centered design with proper oversight, and safety and security through fail-safe, regular audits, and monitoring of the tool.

Week 2

Question: Last week, you were asked to write about priorities a CEO might have when developing an AI technology. This week, please write a paragraph (5–10 sentences) to reflect on your current perspective as an AI user. Consider the following prompts to get started:

- What is an example of something you do using AI?
- If you had a choice between two AI technologies that function similarly, what criteria would you use to decide between them?
- Each ethical framework was connected with a guiding principle (e.g. Hindu = duty). Which ethical framework do you think would be most helpful in shaping your perspective on AI and why?

An example of something that I do using AI, is using LLMs such as OpenAI's ChatGPT-4o and Anthropic's Claude Sonnet to help me look up, understand, and study topics from my courses; such as recently with my Calculus II class to study up for Tuesday's quiz. If I had a choice between two AI technologies that function similarly, the criteria that I use to decide between them is: it has to have really advanced mathematical and logical reasoning to understand my questions and the content that I present to it, it must be able to generate comprehensive and understandable images such as graphs and other graphics, and finally [most AI users would agree on this issue] it should be able to have a form of "memory bank" or "brain" that can actively recall things instead of hallucinating and giving gibberish responses. The ethical framework that

I think is the most helpful in shaping my perspective on AI is the consequentialist framework as personally to me, I want to keep myself informed and understand the possible outcomes [or impacts] that AI technology could affect in the long- or short-term run, such as the case of the AI100 study. Furthermore, the consequentialist ethical framework is also helpful in delving into these ethical topics of AI in a more broader sense.

Week 4

Questions: Part 1: Reflecting on the clips from Robot and Frank and your personal values, write a paragraph (5–10 sentences) discussing your current perspective on the following ethical theme. If you had a loved one in need of extra care and an assistive AI tool were available, would you use it? Reflect on requirements/expectations you would have for the tool, and some of the risks and benefits of using an AI tool for healthcare support. Dr. Fleischman suggested that a computing code of ethics might be introduced in the future. What is one principle that, if outlined in a computing code of ethics, would make you feel better about using AI for eldercare?

Part 2: Reflecting on the content of this module (including videos and reading), write a paragraph (5–10 sentences) that includes one or more of the following:

- Insightful questions;
- Clarification questions about ambiguities;
- Comments about the relation of the content to previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the topic's importance; and summaries of the most important things you learned.

Part 2 of this reflection is designed both to encourage you to engage with the videos before Thursday class and also to allow us to incorporate some of your responses into the Thursday class discussions.

Based on my current perspective of both the film and the current issues with care in retirement homes (and extra care home for older adults), I would in any scenario let an assistive AI take extra care of a loved one. However, there should be some form of ethical standards for AI when introduced in a much high-risk area such as healthcare support and similar. A few requirements for AI to be part of a loved one's healthcare or healthcare support is having the intelligence of that of a human, the ability to have some sort of physical form to assist with physical chores or obstacles, and a form of ethics that it itself can interpret what is right and wrong. One principle that if outlined in a computing code of ethics would make me feel better using AI for for healthcare is it duties for it's owner/client/consumer as being the superior directive over all directives as it's the utmost importance for it to help take care of its caretaker/elder.

Week 5

Questions: Part 1: Reflecting on the "Ethics in AI: Guidelines" video and your own personal values, write a paragraph (5–10 sentences) that highlights your thoughts about the challenges of developing AI and ethical AI guidelines in certain parts of the world when the AI technologies and the ethical guidelines will be applied globally. What stood out to you in the video? You may address the questions below or others.

- Which ethical themes would you prioritize and why?
- How well do your priorities line up with the priorities described in the video (which are also those emphasized in published AI ethics guidelines)?

Part 2: Reflecting on the content of this module (including all videos and reading), write a paragraph (5–10 sentences) that includes one or more of the following:

- Insightful questions;
- Clarification questions about ambiguities;
- Comments about the relation of the content to previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the topic's importance; and summaries of the most important things you learned.

Part 2 of this reflection is designed both to encourage you to engage with the videos before Thursday class and also to allow us to incorporate some of your responses into the Thursday class discussions.

A few of the ethical themes that I think myself would prioritize would be the principles of transparency and privacy; as those have been more persistent and brought to light as areas of concern per the public and average consumer. The reason behind this logic and why I chose these themes are because of the recent litigation of how OpenAI uses public data and datasets to feed information that is accessible to chatGPT and the end use of chatGPT. This is something that should be an area of regulation or determine an ethical method deal with intellectual property and how an AI company can use it. In the other hand, there has to be some form of transparency between the AI company that is using public data for its LLMs or etc, its end users who have access to this data, and to those who are within the public data set. As both transparency and privacy have become more of a concern for the end user [consumer] and the general public in the past decade. My priorities do somewhat align or even overlap with some of the priorities described in the video and the published AI ethics guidelines.

Some interesting things, as for someone who is a non-CS major, but have had 5+ years of outside experience in the field; some of the courses that I'm taking for adding CS as my major, is one class that has stuck out to me in this whole entire module: SDS 321. As most of the module covers about probability, the majority of the concepts, texts, and general theme overlaps and even some of the content is exactly the same; it makes me think that in some way as you progress through your CS undergrad, you learn little by little of AI significance in the field, even though it might be a bit more distinct.

Week 6

Questions: Part 1: SCENARIO: You are a serving as a city council member in a city that has seen a recent uptick in crime. The city is considering implementing an advanced AI-powered video surveillance system in some public spaces (e.g., parks, streets, and transportation hubs) with the goals of increasing public safety and responding quickly to emergencies. The AI system can recognize faces, detect suspicious behaviors, and track individuals across multiple cameras.

At a city council meeting, community members express concerns about privacy and the potential for misuse. One resident argues that constant surveillance would make them feel like they are always being watched, and is a violation of their personal freedom. Another expresses concerns that the AI might misidentify individuals, particularly people from marginalized groups, leading to false accusations or discrimination. A third individual is worried the collected data could be hacked or used for purposes beyond preserving well-being, such as for advertising or personal tracking by private companies.

Write a paragraph addressing the ethics of such a surveillance system. You may address the questions below or others.

- What balance should be struck between public safety and individual privacy?
- How should concerns about privacy and data security be addressed?
- What safeguards could be put in place to prevent misuse or discrimination by the AI system?

Part 2: Reflecting on the content of this module (including all videos and reading), write a paragraph (5–10 sentences) that includes one or more of the following:

- Insightful questions;

- Clarification questions about ambiguities;
- Comments about the relation of the content to previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the topic's importance; and summaries of the most important things you learned.

Part 2 of this reflection is designed both to encourage you to engage with the videos before Thursday class and also to allow us to incorporate some of your responses into the Thursday class discussions.

On the question of "What balance should be struck between public safety and individual privacy?", our implementation of this AI-powered surveillance system to address the concerns of privacy is to keep all the data stored in house in the costs of maintaining our own infrastructure; as we understand the concerns of the citizens' right to privacy. Furthermore, the purpose of this surveillance system has no ill-intentions as its main purpose is to maintain public safety. To also address the concerns of data security and integrity, most of this data will be stored at secure locations, as a distributed storage system to maintain the outmost data integrity. Furthermore, with the data being onsite, we also would like to protect it from attacks both within and outside the network, thus for department individuals who want to access the system must have a security key or access the system within the network. The concerns of a biased AI is something that should be addressed; to be direct, this is something that we could try and mitigate, but since our dataset contains everyone general ID/profile, this is something that we can't really control. Therefore, we will also have a human supervisor to have some control and mitigate the discriminational bias of our AI-powered surveillance system. In addition to address the concern of misuse, we will implement form of a security clearance system as a way to also mitigate the possibility of misuse.

My thoughts of this module of Computer Vision are interesting, because of how it applies in our day-to-day such as the OCR technology for analyzing text such as PDFs and images. In addition, AI, computer vision, and LLMs have become powerful for reading documents and images to then store that data for it to respond to the end user. These advances in the field of computer vision could become the "Next Big Thing" that could effect the industries like healthcare, game development, and LLMs such as GE Healthcare, Treyarch, Microsoft, OpenAI, and etc.

Week 7

Part 1: SCENARIO: You are a data scientist working for a healthcare company developing an AI model to predict the likelihood of patients developing a specific disease based on their medical history and lifestyle data. The company has collected a large dataset from a local hospital, consisting primarily of patients from one region with similar demographic backgrounds and lifestyles.

At a team meeting, a colleague expresses excitement over the model's high accuracy when tested on the local hospital data, and the company is eager to deploy the system nationwide. However, another team member raises concerns about the model's ability to generalize to patients from different regions, cultures, and backgrounds. They argue that since the training data is limited to one demographic group, the model might not perform as well on diverse populations, potentially leading to incorrect predictions or even bias against underrepresented groups. Another concern is that the model might be overfitting the local data, meaning it could perform exceptionally well on the training data but fail when exposed to new, unseen medical history of a patient.

Write a paragraph reflecting on the ethical implications of deploying this AI model. You may address questions such as:

- What might be the source of generalization gap when deploying the model nationwide?
- What are the risks of overfitting the model to the training data, and how can they affect the patients' outcomes?
- What steps can be taken to avoid overfitting and ensure the model works across diverse patient groups?

Part 2: Reflecting on the content of this module (including all videos and reading), write a paragraph (5–10 sentences) that includes one or more of the following:

- Insightful questions;
- Clarification questions about ambiguities;
- Comments about the relation of the content to previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the topic's importance; and
- Summaries of the most important things you learned.

Part 2 of this reflection is designed both to encourage you to engage with the videos before Thursday class and also to allow us to incorporate some of your responses into the Thursday class discussions.

Some of the source of the generalization gap when deploying is the consideration of people from different backgrounds; in which the model may somewhat suffer this gap but this can be addressed with redundant testing via validation sets. Furthermore, we can breakdown the dataset to more pieces in order to address the model's inaccuracy, however, this can lead to more erroneous testing that could make the training dataset useless. Then again, there should be more rigorous testing before releasing this model to the public via quality source control through a small internal beta before moving onto a public beta. Some risks of overfitting the model to the training can lead to larger margin of error when it's is deployed nationwide using real-world data versus testing it using training set data; and model biases. This can lead to more erroneous results for many patients.

On this module, for me, this is the very first time for to know what machine learning is conceptually; but in the terms of its application, I have seen it before in my day-to-day life via video games, email spam control, and so much more. With these past explanation in the slides and video, it has made it more clearly me the impacts that machine learning has in our lives. Whether it is apparent via our smartphone and its application; or underneath the hood to power the current day LLMs such as ChatGPT, Copilot, Claude; it has made a great impact in the society around us.

1 Week 8

Part 1: SCENARIO: You are part of a team developing an app that can predict disease outbreaks using data collected from online sources. The team decides to gather large datasets from social media, public websites, and forums to train a machine learning model that will help identify health trends. However, some of this data includes personal information, like location details or personal health stories that people may not have explicitly agreed to share for such purposes.

The app is designed to benefit society by helping predict and prevent the spread of diseases, potentially saving lives. But the data was scraped from the internet without the explicit consent of the individuals involved.

Write a paragraph reflecting on the ethical implications of deploying this app. You may address questions such as:

- What is it ethical to gather this data without permission, even though the goal is to use it for a positive cause?
- If the data is publicly available, does it justify using it without explicit consent?
- What steps could be taken to ensure that data collection respects individual privacy while still enabling beneficial uses of the information?
- What alternative methods could be used to gather the necessary dat while respecting people's privacy and consent?

Part 2: Reflecting on the content of this module (including all videos and reading), write a paragraph (5–10 sentences) that includes one or more of the following:

- Insightful questions;
- Clarification questions about ambiguities;

- Comments about the relation of the content to previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the topic's importance; and
- Summaries of the most important things you learned.

I believe the manner that this data was scrapped is completely unethical; since the user did not give consent for their data to be accessed in a way that is somewhat illegal and a violation of their right to privacy [dependent on the Terms of Service also]. I say again, that even if it is for a positive cause; I think the manner the data was accessed is not ethical or legal in some way or form. If the data is publicly available such as a government website, then yes; but the data should be cleaned for any personal or detailing information that could expose or put the user's privacy in harm; then yes, I believe this is more of an ethical approach than that detailed in the scenario. As I mentioned, one of the ways to respect people's privacy and consent could be through censoring of personal or detailing information; and if the personal information or details are needed, to reach to the individual(s). Furthermore, if the individual(s) does give their consent, then these files or data should be encrypted to protect this valuable, important, and vulnerable data. Some alternative methods that could be used ensure that data collection respects individual privacy and consent is through the public government databases; where they can request such necessary information through legal and respectful means. Another way, is to just gather fresh data via a consent form or an agreement, and so on; thus, creating another alternate way of data collection.

Somethings in this prompt are of concern in the current applications of the real world such as how cookies and other web data store our personal information; and without proper security measures can lead to catastrophic data leak/loss that have occurred in the past 5-10 years.

2 Week 9

Part 1: AI and copyright have been in the news related to the arts. Notably, OpenAI produced a voice for a personal assistant called Sky that sounded strikingly like that of Scarlett Johansson. Artists have likewise been concerned with copyright due to the rise of AI image generators, and recently won a victory against in a landmark case where a judge ruled that that an AI company violated artists' rights by illegally storing work. Likewise, in music, there is a similar case where record companies are suing AI song generators. Consider art and the ethics of copyright for this reflection.

1. https://www.npr.org/2024/05/31/g-s1-2263/voice-lab-analysis-striking-similarity-scarlett-johansson-chatgpt-sky-openai
2. https://news.artnet.com/art-world/artists-vs-stability-ai-lawsuit-moves-ahead-2524849
3. https://www.nbcnews.com/tech/tech-news/us-record-labels-are-suing-ai-music-generators-alleging-copyright-infr-rcna158660

You may answer the following questions or use other ideas in your essay:

- In the case of Scarlett Johansson's voice being imitated by an AI personal assistant, even if the AI was not trained directly on her voice, do you think she has a right to object or be compensated? What factors should influence this decision?
- For visual artists and musicians, how can they protect their creative work from being imitated by AI systems without their permission? Should artists have the right to opt out of having their work used to train AI models, and how should this be enforced?
- Considering the recent legal victory for artists, do you think these lawsuits will have a positive or negative impact on the future of AI development in creative fields? Will it hinder innovation or encourage more ethical AI practices?

- Is it more important to protect the rights of individual creators or to encourage the development of AI tools that can produce creative works efficiently and at scale? Can there be a balance between these goals?
- Where do you think the line should be drawn between inspiration and imitation? In the case of human artists, it's common to be inspired by others, but should AI be held to a different standard? Why or why not?

Part 2: Reflecting on the content of this module (including all videos and reading), write a paragraph (5–10 sentences) that includes one or more of the following:

- Insightful questions;
- Clarification questions about ambiguities;
- Comments about the relation of the content to previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Thoughts on the topic's importance; and
- Summaries of the most important things you learned.

To answer for the case of Scarlett Johansson's voice being imitated by an AI personal assistant, even if the AI was not trained directly on her voice; I do not think she has the legal right to object or be compensated based on the assumption that it sounds like her voice. I can easily say that my friend's voice sounds like Scarlett Johansson's voice, but does not; if OpenAI could have proven this by showing evidence, the case could have been easily closed, but due to OpenAI not being able to, it then rose the suspicion that it had indeed use her voice in the dataset.

For the question of "For visual artists and musicians, how can they protect their creative work from being imitated by AI systems without their permission? Should artists have the right to opt out of having their work used to train in AI models, and how should this be enforced"; one of the ways that artists and musicians can protect their work is via some form of a checksum when downloading music or running via a streaming platform incase, these streaming companies are working with third-parties like AI companies. Beyond that, in my opinion, I would lobby for AI regulation on what it can use on its dataset for THIS INDUSTRY/area; and yes, artists should have the right to opt-out their works from training dataset for AI models along with it being enforced via legal means or possibly some federal intervention.

To answer the third question, some of these recent lawsuits will have a positive impact on AI development in the creative fields as artists and musicians legally have the right to protect their work via the U.S Patent and Copyright Office. Furthermore, I would push for more ethics and regulations of AI in the creative space and push for a more appropriate way to handle copyrighted works in these fields when developing for AI.

I believe that it is important to protect the rights of individual creators, but to also encourage the development and innovation of AI tools that can produce independently creative works efficiently and at scale. There is a possibility for a balance between these two objectives, one of these can be a way on how creative works are handled during AI development and the means to maintain the security and legal protections of those works while AI researchers and developers can use this protected works via legal, secure, and consensual way in their dataset [this can be via citing the works and so on].

I think that AI should be able to create works based on inspirations like humans also base their art, but I'm strongly against that AI imitates on copyrighted human art works and music. But I strongly believe that AI should be able to create its own works in art and music based on inspiration of human works; this leads to the question on why should AI be held to different standards than humans?

Reflecting on the contents of this module; one thing that stuck out to me is the the articles and how outdated some of the examples are such as referencing GPT - 3 and how it's unable to do basic arithmetic. On the contrary, there were some references to probability in the terms of how LLMs respond to user inputs and determine what is the best response.

3 Week 10

Part 1: Reflect on the ethical concerns presented in this week's videos and/or reading in a paragraph (5-10 sentences). In your paragraph, be sure to:

- Clarify the topic/ethical dilemma you are considering;
- Consider the perspective of the different stakeholder and the individuals impacted;
- Connect the topic to any ethical framework you are aware of that may apply in this situation;
- Acknowledge any personal biases you may have with respect to this topic; and
- Suggest some practical and ethical solutions.

Part 2: Reflecting on the content of this module (including all videos and reading), write a paragraph (5-10 sentences) that includes one or more of the following:

- Insightful questions;
- Clarification questions about ambiguities;
- Comments about the relation of the content to the previous content;
- Solutions to problems or exercises posed in the readings or videos;
- Critiques;
- Thoughts on what you would like to learn about in more detail;
- Possible extensions or related studies;
- Summaries of the most important things you learned.

The ethical topic/dilemma that is concerning to me is the determination of a user's privacy when using an LLM or DALL-E like systems; where do the user's right to opt-out from being part of the data set fall on? Via what basis? Nowadays, the modern companies of today have these user's waive their rights via an agreement called the TOS when a user wants to user their service; another form that these modern-day companies try to thwart the efforts of the user from protecting their data is hiding or making it difficult to find the option(s) to opt-out. In addition, companies like OpenAI, Microsoft, Apple, Google, and so many more, are using this practice as a way to "legally" scrape the user's data without the user's knowledge since who nowadays reads the Terms of Service? The topic of determining user privacy in the space of AI, reminds me of the ethical framework of the Rights-based Ethics as the indivudual and really everyone have a natural right(s) that should not be pertrubated by a company or goverment, which fits this topic perfectly. A solution that I suggest for user privacy with an undertone of security is the legality of asking consent of the user, in using their data for training the model and so on in clear terms or a prompt displayed to them so they have to option to deny or agree to these terms. In addition, if the user's data is used via consent; then the company should secure or censor any determining identification of the user when tranining it for the model.

In this module, I found it interesting how the problem of historical and racial bias is currently the problem plaguing AI due to it's diverse dataset. One of the articles that I read was interesting in what are the recommended actions or frameworks that they recommend in their paper in engineering an LLM system; and the possible harms that those engineers should be aware of when designing their system(s).