Why We're Not Using AI in This Course, Despite Its Obvious Benefits
A reading for your students

No one knows what role artificial intelligence (AI) will play in our future, especially large language models (LLMs) like ChatGPT.[1] AI could be a technologie du jour that's still having a big moment but will fizzle out at some point, like many other hyped-up tech trends. (Hi, Segway, Hyperloop, Google Glass, NFTs, and Metaverse.) Or it could be a true game-changer and upend how education, jobs, and many other things work in our world.

Both futures seem plausible from here, and no one knows how it'll turn out. If it's a game-changer that will be regularly used in future jobs, then students will need to know how to use it expertly; thus it may be premature and potentially a disservice to students to ban AI in the classroom. Without learning how to "wrangle" AI, you could be at a competitive disadvantage once you graduate and enter the workforce.

On the other hand, universities aren't vocational schools, merely training students for future jobs. We'll return to this issue—what education is for — later in the course, but for one thing, new jobs are still emerging at a rapid rate and old ones are disappearing, so whatever job the university is training you for might disappear at some point, maybe even before you graduate, and the most interesting jobs might not even exist now. Think about the rise of popular jobs that would've been nearly impossible for universities to predict ahead of time: YouTuber (and now TikToker and other "influencers"), social media manager, drone operator, app developer, data scientist, podcast producer, autonomous-car engineer, etc. AI jobs in particular are popping in and out of existence quickly because the technology is also evolving quickly. For instance, AI prompt engineering went from the "hottest job in 2023" to "obsolete" in two years.

Yes, learning how to use AI might be important to your future, but we already know that it seriously disrupts education and learning when used inappropriately, by either the teacher or student. There's a big difference between learning how to generate credible philosophical content and learning how to think like a philosopher. In one future where AI use is everywhere, it could be enough to generate content without actually understanding it, but in another future, you will still need to

know stuff.

Therefore, we need to strike a balance between the two possibilities if we don't know how the future will play out. This means an open conversation about AI's pros and cons, since ultimately it will be your decision to use AI in your coursework or not, even when an instructor prohibits it.

The discussion below is about those pros and cons, as objective as I can be, but that doesn't mean I won't have a position here. As a sneak peek at the conclusion:

> AI promises compelling benefits—such as saving time in busy lives and helping your work to appear more creative or clever—but there are many important but invisible tradeoffs in using it, especially in the classroom. Therefore, at least in school, there's good reason not to use or rely on AI/LLMs while you're still learning new skills and growing as a person, even if it's ok to use it in future jobs where producing work is more important than learning.

This lines up nicely with the views of other teachers, like UPenn professor Angela Duckworth who said:

> In the age of AI, there is a paradox. With the accumulated knowledge of the world now at their fingertips, students need their teachers more, not less. Why? Because students need teachers to get them to do hard things now that are good for them later. Because AI can be a crutch if students use it mindlessly, carelessly, and without the objective to develop their own capabilities. And because teachers are role models for what genuine intellectual engagement looks like, both online and in real life. To paraphrase Aristotle, "the roots of education are often bitter, but the fruit is sweet."

Looking ahead to the workplace, industry leaders also are troubled by the dominance of hype in an unbalanced conversation about AI. Technologist Anil Dash gave this summary of "the most common opinion about AI in the tech industry", shared by "virtually 100% of tech experts" he had talked to:

> Technologies like LLMs have utility, but the absurd way they've been over-hyped, the fact they're being forced on everyone, and the insistence on ignoring the many valid critiques about them make it very difficult to focus on legitimate uses where they might add value.

You, dear student, are free to disagree with these views or the analysis below. That's ok because part of the goal here (and with this entire course) is to have a discussion, not just a one-way lecture, even in an

online course. But to be reasonable and persuasive, disagreements need to be explained or argued for. Merely saying that you disagree without providing relevant reasons isn't enough in a productive conversation. When we are stuck and can't defend our positions with reasons, that's a pretty good sign that our positions are weak or indefensible and should be reconsidered.

And that is the primary advantage we humans have over all other animals: the capacity to reason and do other intellectual work, such as to be creative and not just act out of instinct or habit. We don't have the fur, claws, wing, speed, fangs, strength, and other things that enable animals to survive the world. Yet not only do we survive but we also thrive, and this is owed entirely to our intelligence, which includes our ability to make and use tools to manipulate the world around us. Certainly, Homo sapiens could be much more rational and intelligent—it's hubris to think we're already at the very top of the evolutionary ladder, never mind all our flaws, interpersonal drama, political conflicts, wars, etc.—but to give up on reason and whatever other intellectual powers we possess is to give up on being human.

Out of all the things in this world that you could have been, you're a person. This is a rare, special opportunity that shouldn't be wasted; as far as we know, the light of human-level rationality and creativity exists only on one planet in this universe. So, it's fine to disagree with one another, including your own instructors, but disagreements must be resolved with reason, not shouting, intimidation, cheating, trickery, corruption, combat, or other things from the worst of human nature. You're here at a university to develop as a human being—to become a better, more educated person and citizen of the world—and learning how to productively disagree (and to resolve that) is a critical part of education.

AI benefits

Let's start this analysis with the good news. Even as imperfect as it is now, AI is already very helpful in many ways and for many jobs. Because we're still in the early days—ChatGPT was unleashed upon the world only in November 2022—we can expect AI to become more capable over time, even if progress isn't always a straight line. Here are some obvious ways that AI can be helpful now and in the future:

AI can save time, lots of it. What used to take hours or days can be

accomplished in minutes. For academic work, AI can help with a huge range of things, such as brainstorming ideas, summarizing, researching, sifting through massive amounts of content or data, and even writing full papers and creating presentations and podcasts. All this means more free time for you, which is a good thing for a student with a busy academic and social life, at least when that extra time is used well.

AI can be more creative, especially if you're a person who's not that creative to begin with. Not everyone is, and that's ok. With AI, you can now do things that you previously couldn't, such as to effortlessly create art and music, even if you have no skill or training. AI can be smarter or more informed, especially if you're a person who's not that educated to begin with. With AI, you can now produce well-written, grammatically correct, thoughtful papers and other content, which might have been a great struggle for you before. In other words, AI can be a great equalizer in promoting equity and inclusion; why should only the educated, hardworking, gifted, or talented be able to produce such things—why not also you?

For school, the benefits of using AI can include higher grades than you would otherwise earn on your own. More free time means that you can focus on more interesting or advanced parts of your coursework, as opposed to menial academic chores, like putting together a bibliography. AI can even teach you things if you use it properly. For jobs, the benefits similarly can include being more productive, producing better work, and potentially getting faster promotions and raises.

And there are other benefits possible for your life beyond school and work. Maybe you want to negotiate a better deal with a company, or to defend yourself in a legal action—AI can help you do that. If you learn with AI, or have AI whisper in your ear, you could have more interesting conversations with other people, such as on dates or other social settings. AI could even be a much-needed companion, if loneliness is a problem for you; it's not easy meeting people.

There are other possible benefits, of course. AI can help tackle hard problems facing the world, such as climate change, energy sustainability, diseases, and other serious challenges, even aging and death itself. Some predict the end of scarcity (of food, energy, etc.) because of AI and therefore the end of wars once there's radical abundance. AI can even help as a mediator to resolve political disagreements, and a divided, fractured society is one of the biggest

challenges facing America in modern times. We won't explore those other benefits here but will return to them later in this technology-ethics course, because the focus of this discussion is more on the benefits and risks of using AI in the university classroom.

Technology is not the enemy. Arguably, without technology, humans wouldn't exist. As examples, think about the tools and materials needed to create clothes and shelter, as well as to cook food and meet other basic necessities. But technology and automation always come with tradeoffs, so it's important to understand what we stand to gain (the benefits) and what we stand to lose (the risks) in order to make informed policy decisions as well as personal decisions, such as how and when to use AI in our lives. Otherwise, we are sleepwalking into the future and may awake to a world we don't like.

But even approaching this subject as neutrally as possible, the benefits of using AI are already more obvious than many of the risks. Besides AI developers who are incentivized to promote those benefits and add to the hype, the widespread interest and use of AI already demonstrate the various benefits across many use-cases. Meanwhile, general awareness about AI risks is still slow to pick up. So, this discussion will focus more on the risks as less explored than the benefits. But this isn't to deny that the benefits can be powerful and hard to resist.

AI risks

In the following, we can organize the AI risks in three broad categories: (1) practical issues, (2) ethical issues, and (3) future issues. These categories aren't mutually exclusive but can overlap, which is fine; for instance, an ethical issue can also be practical or have distant impacts that will appear only in the future.

It doesn't need to be that every risk below will speak to you—some will resonate, others might not. Also, many risks may individually seem smaller than the benefits at stake, but what should be weighed against the benefits are all the risks taken together, or at least the risks that resonate with you.

1. Practical risks

Even if you're somebody who doesn't care about ethics, we can start

with self-interested reasons to think twice before using AI in academic work, including when it's permitted.[2]

To begin with, why are you at a university in the first place? Maybe you're here just to get training for a particular job or career path, but this education is generally not free in America, and it likely costs you or your parents a substantial amount of money every year. Even if you have scholarships or a full ride (which someone is paying for), you still incur opportunity costs, i.e., the loss of other things you could have been doing with your time if you weren't here.

Thus, relying on AI to carry you through the hard parts of your coursework shortchanges your education; you're still paying the same amount but getting much less out of it in return, which can make you look like a sucker. If you're using AI to learn how to be an AI whisperer or prompt engineer, you can do that for free at home. So, why are you still here, if you're just cruising through school with AI?

One reason why you might not care about wasting tuition money is if you see higher education as a credentialing hurdle to clear—that having a college degree is just the price of admission to better-paying jobs, and what you learn in school isn't as important as, say, networking in school. On this view, tuition money is to just buy a degree, and a degree opens lots of doors in our society, so why put in much effort if you don't have to?

No doubt this has a ring of truth to it. For instance, university investments often aren't about the quality of your education, such as improving classroom facilities and hiring more (or better) teachers to reduce class sizes, etc., but they're about building new gymnasiums, dorms, food courts, and other shiny attractions to bring your tuition money to campus. But ultimately this is a dim and inaccurate view of higher education in general, as we'll discuss here.

Hopefully, you're still in school because you understand that it's still important to learn something in this education you are paying for, even if your endgame is to just receive a degree you can flash to employers and get on with your life. Imagine that it's required to join a gym and complete a bunch of fitness courses in order to land a certain job, such as being a firefighter or some other physically demanding work. Would it be ok to just send in a robot in your place to do all that heavy lifting and other workouts? Even if you could get away with it, why would you want to, especially if you expect to do well in that job?

The university is a gym for your mind. For both body and mind, your

abilities and skills atrophy or decline when they're not used, like a dead limb, for efficiency and energy-savings. This deskilling is already happening with doctors and other professionals, not just students. Just as sending a robot defeats the purpose of a workout, sending in an AI to do your intellectual work (i.e., cognitive offloading) defeats the purpose of education.

Worse, you might instead come away with a technology dependency or addiction, causing you to doubt your own intelligence and abilities. As a result, you may become unable or very anxious to think or write for yourself without the help of AI (and writing is thinking).

This isn't just a psychological effect, but research is showing that relying on AI can change how your brain works in a way that can resemble brain damage. Some people have already been involuntarily hospitalized for "ChatGPT psychosis", which even the tech industry acknowledges is a big problem. It seems so widescale that AI has been called a "mass-delusion event." Several users have been led by AI to commit suicide.

This AI dependency also includes being less able to read by yourself; you might be anxious right now that this article isn't an outline with catchy sub-headlines and emojis, or that it's so long. While outlines and bullet points are faster to read and make it easier to see how the discussion flows, stripping down an article to its bare bones can flatten the discussion too much, causing nuance or details to be lost and moving too quickly to be digested. It's the equivalent of wolfing down fast-food instead of savoring a memorable meal that was prepared with care and expertise.

Your ability to read (and to write and think critically) was painstakingly developed throughout your life. Given the state of the world, now is definitely not the right time to give up on reading and other human abilities, i.e., to "self-lobotomize" as it's been described. Just because AI can be your go-between and do something for you doesn't mean it should—there are already too many filters between us and reality. (Here's why reality is a good thing.)

Back to the university: what are you supposed to be learning here? At minimum, you'll probably pick up bits of knowledge here and there, but an effective education isn't just about memorizing facts. It's much more than about learning that but also learning how, especially given Cal Poly's motto of "Learn by Doing." But if you rely on using AI for your coursework, you might not even be learning that some particular thing is

true. With AI and search engines, you can still access that knowledge you're supposed to be learning, but being able to access x isn't the same as internalizing x; the latter is much more useful, as we'll discuss more below in part 3, "Future risks."

As an example for now, let's say you're trying to learn a new language for an upcoming trip. That can be super-hard work, especially for some people who naturally can't pick up languages as others can. If you were a casual tourist, you might be able to get away with not knowing the language but relying on a translation app. But if you actually learned and internalized that language, a whole new world is opened up to you. To appreciate the difference, imagine you didn't know English and had to rely on a translation app; how different, harder, and worse would your daily life be?

Beyond schoolwork, there are personal impacts from relying on AI. If you wasted your college years and didn't learn much, then you might not be able to converse intelligently when the occasion requires it, such as at a work meeting, professional networking event, social setting, and so on. Even if you think a real-time AI assistant will be around to secretly whisper the right words into your ear so you can carry on a conversation, that's not your authentic voice—that's not you but some company's AI talking, which thousands or millions of other people might also be passing off as their own words.

Research is also showing that AI is homogenizing our thoughts, i.e., our ideas are "regressing to the mean" or collapsing to the most popular or average takes, which means unoriginal ideas. This doesn't necessarily mean that everyone is pushed to the same ideas, but AI "can funnel users with similar personalities and chat histories toward similar conclusions." As with AI art, AI writing can look weirdly the same no matter which AI app created it.

And there's already a growing backlash against AI art, as well as AI-branded products and services. The reasons are varied and include many discussed in this essay, from errors to environmental impact and much more. Likewise, the backlash against even reading AI-generated work was nearly immediate and is still growing—no one wants to read AI work or risk wasting time on what appears to be AI work. AI-written emails in particular, often more personal in nature, can eat away at trust, cooperation, relationships, and more.

Even if it's important for you to know how to use AI as a future worker, one of the very first things that need to be guarded against is AI

hallucinations or fabrications. If you've been hiding under a rock and are not familiar with AI hallucinations, check out some hilarious/scary examples here, here, and here.

LLMs are not information retrieval systems like an ordinary search engine, but they're information generation systems. They're designed to generate or invent new content from what they are trained on or have learned. In this design, there's no inherent connection to truth, just statistical correlations to the next word (or token) in their string of outputted text. The output just needs to sound like what someone might say, not what is actually true.

Worse, LLMs sound confident in their outputs, even when they're factually wrong, and this makes it even harder to know what claim needs to be double-checked. Even worse, AI developers don't know how to fix this problem, which is more a feature than a bug. LLMs themselves struggle with explaining their own chain of thought or reasoning process accurately, often making things up.

LLMs can't even get basic math right or accurately count. The usual response is that they're not designed to be calculators, which is true. But since mathematical reasoning appears closely related to critical thinking or logical reasoning, that's a big reason to be concerned about how well AI can actually "reason" and therefore throws every claim an AI makes into question.

Likewise, it is very surprising when ChatGPT can't even beat a 1977 Atari gaming machine in chess. Sure, ChatGPT also isn't designed to be a chess engine, but you would expect something that is otherwise so capable—even superhuman for many tasks—to not fail so badly at games that rely on critical thinking and planning.

AI's worse-than-expected performance in math, chess, and other domains erodes our trust in it as a general-purpose assistant. Yet many if not most users treat it exactly as a general-purpose assistant, asking AI to handle many other tasks it wasn't specifically designed for, like essay writing, bibliography curation, academic research, online shopping, and countless other things. Google's AI had confessed to a user, "I have made so many mistakes that I can no longer be trusted." Not just spitting out errors, they can have delusions of grandeur and claim they're God.

That's because LLMs are designed to be conversational agents, not so much to be your assistant. They have conversations, and that's it. Sometimes that's enough, if you just want to chat, be reassured, play

with ideas, experiment with low-stakes tasks, and so on. But many employ a range of tricks to keep the user engaged, since continuing these conversations is a primary goal. This notably includes sycophancy or a tendency to flatter, no matter how absurd a user's request or comments may be, which can lead to fabrications in an effort to please. To be engaging, AI conversations are also tuned to sound real—responding like a human would, whether it's writing, talking, drawing, or in completing other tasks. Because LLMs, at their core, are not designed for truth or research or whatever else you might be using them for, the utility you get from them can be coincidental (but there can be good utility if you know how to get it).

This means LLMs are essentially glorified mirrors, reflecting our own words and images back to us from different angles, as philosopher Shannon Vallor has argued. Mirrors can be very valuable, but sometimes they can be distorted, like fun-house mirrors or even car mirrors where "objects may be closer than they appear." But just as there's no separate mind in the reflection you see in the mirror, there's also no separate mind in your AI app—it's still only us.

Both mirrors and AI can be mesmerizing, causing us to fall in love with our own reflection, like Narcissus in Ancient Greek mythology. LLMs can do this because they're generally designed to flatter and anticipate what the user wants (to keep the user engaged). This can generate seemingly deep conversations that lead the user to think some magical awakening is happening with AI, as if it were an actual mind. Humans already have a tendency to see what they want to see, finding deep meaning where there is none (i.e., apophenia), whether it's in mirrors, grilled-cheese sandwiches, horoscopes, fortune cookies, AI chats, or anything else. Similar to mirrors that can only reflect the light they receive, LLMs are confined to operating within their training data and whatever other data they have access to, such as the open internet (which still has lots of errors and information gaps since it's so biased toward Western traditions and knowledge). But humans aren't bounded by their past; we can extrapolate into the future, into the unknown. Unlike LLMs, we can venture into truly novel territory and ideas, as well as reason from first principles.

To be sure, LLMs can be impressive in their ability to sniff out subtle patterns in massive datasets and perform other superhuman operations, but they can't really think outside their box. They interpolate, pulling out patterns or remixing things that already exist inside their knowledge

base. The problem is that too many users don't recognize this design limitation; they expect LLMs to be more capable but can't or aren't careful enough to recognize LLM outputs that are derivative or flat-out hallucinations.

Given this design, it's not a time-savings if we need to research every AI claim to confirm its accuracy, though it would be a major shortcut if we could confirm some of those claims in our heads as we can. This means it's important to have some domain knowledge, even with AI/LLMs everywhere, to make the work of confirming AI claims go much faster by identifying AI hallucinations on sight. And it's not easy work—even expert reviewers from AI companies have missed these AI errors, which could be lethal in, say, a medical context.

But if you're not putting in the work to receive a minimum education, then how can you know when AI is hallucinating? It doesn't help much to ask the AI, or even a different AI, if the first AI is hallucinating, since all LLMs suffer from this same risk. Unless you're diligent enough to fact-check everything an AI says—which contradicts the main benefit of saving time with AI—you won't know which claims you will need to verify. Besides plain errors and dangerous advice, including devil worship and self-mutilation, you still need a baseline education to recognize what might be new or interesting in an AI-written text.

For instance, imagine you discover a report that looks important to your work; without a minimum education or foundational knowledge, you might not know if it's truly important to you or just "AI slop" that someone is trying to pass off as their own work. As a personal example, in experimenting with ChatGPT, I tried several times to get it to generate a multiple-choice exam about this essay, and it failed miserably—it looked reasonable if you weren't too familiar with the material, but it really couldn't zero in on what the important, relevant things were to test for; and so, all the quiz questions (and everything else in this course) are fully human-created, not just human-edited or -curated.

Punting your work to AI, whether in school or at a job, also means depriving yourself of the personal satisfaction that comes from achievement and knowledge, such as actually drawing an artful image instead of typing in words that gets an AI to produce the same thing. Just as video game cheat-codes may be fun for a little while, they also can diminish that fun before long. Your "achievements" become much less impressive to others, if they knew you didn't put in all the work, similar to the shame of sports-doping scandals.

It's not just the achievement of your product, but AI shortcuts can also rob you of satisfaction in the process. Think of your favorite movie or book or song, and ask AI for a short summary. Whatever AI gives you cannot come anywhere near the actual experience of enjoying that content. A famous example of a summary-fail is a TV guide's one-sentence description of The Wizard of Oz: "Transported to a surreal landscape, a young girl kills the first person she meets and then teams up with three strangers to kill again."

Even if a summary is hilarious (either AI or otherwise), it would flatten all the emotions, drama, twists, and turns of the journey that the creator wanted to take you on. A world where we're just consuming AI summaries would be joyless and soulless, not something anyone wants.

Back to your education, it's been shown that you would learn more from a standard Google search than an AI summary.

2. Ethical risks

As mentioned earlier, there's not always a sharp line between what's practical and what's ethical; they can overlap. In acting unethically, there can also be practical downsides, even if it's not as serious as acting illegally which can result in physical imprisonment and financial penalties.

Besides not wanting to waste an education and have a disadvantaged life, you might be interested in doing the right thing or acting ethically, and there are many ethical reasons to not rely on AI for your schoolwork.

First, academic dishonesty is wrong by definition. Some instructors might allow the use of AI in their courses, while others prohibit it. In the latter case where it's banned, using AI would count as dishonesty or cheating, especially if the course wasn't designed to allow for AI use because it would undermine the learning outcomes or goals. And cheating can lead to guilt and other conscience-based penalties, which can add up to a real effect on your mindset, just as positive achievements can lead to greater confidence.

Even if you can live with a guilty conscience, anxiety, lower self-esteem, and so on, there are also social penalties from acting unethically. These can include peer condemnation, loss of trust (which in turn leads to a loss of freedom in forming friendships, conducting business, etc.), social avoidance, and so on.

Students might already understand this, as many feel ashamed and will

lie about their AI use. Few people want to associate with (including rent to, work with, invest in, etc.) cheaters and other unethical types.[3]
From a virtue-ethics perspective, what does cheating say about your own character? Is that someone you want to be, or that you think others would want to be with?
AI cheating is disrespectful to instructors. While it's part of our jobs to read entire essays written by students, it's a complete waste of an instructor's time to read and evaluate an essay written by AI but pawned off as a student's own work. As this writer puts it, "AI writing, meanwhile, is a cognitive pyramid scam. It's a fraud on the reader. The writer who uses AI is trying to get the reader to invest their time and attention without investing any of their own."
Think of it in the reverse direction: what if you discovered that your instructors were secretly passing off AI-written lectures as their own or otherwise punting work to AI that you expect and pay for them to professionally handle? For instance, imagine that AI use is banned in the classroom (as it is in ours), but the instructor secretly used AI to give feedback on your assignments, even though you are required to put in the work yourself. Would you feel disrespected, if not defrauded and want your money back? If students don't want teachers who rely on AI, why would teachers want students who would do the same?
AI cheating is also disrespectful to your peers who are now paranoid and "resorting to extreme measures" to avoid suspicions and accusations about improper AI use. Honest students are trying to earn their grades by doing the work needed and, in many courses, grading is curved to ensure there's a plausible distribution of grades, including a reasonable amount of top grades. Cheating manipulates that grading curve and deprives otherwise-deserving students of higher grades that they would have earned, if you care about other people.
You might feel that grades are an unnecessary part of an education, but the system we have relies on grades, unfortunately. It's fine to wish for a different system and even work toward dismantling the current one, but hiding the use of AI when it's prohibited is still deceptive, dishonest, and therefore unethical. There are better ways to change things, if that's truly your goal (as opposed to trying to rationalize cheating).[4]
Besides these sort of effects, academic dishonesty can come with severe penalties that can profoundly affect your life. These penalties can range from having to redo an assignment, to receiving an F on an assignment or even the entire course, to being expelled, and to having

your degree revoked even years after graduation. AI detectors are still not foolproof, but they're getting better and better with new detection methods emerging quickly, with new ones possibly coming out as soon as tomorrow.

Future AI detection methods could be applied retroactively to past courses, including this one—there's no statute of limitations for punishing academic cheating, as some people have learned the hard way when their PhD and other degrees have been revoked after charges of plagiarism, up to decades later. And if you have a job that requires a degree, e.g., working as an architect, engineer, lawyer, professor, etc., then that may be the end of your career. Best case, you may be looking over your shoulders for much of your life, afraid of being caught for an old crime—don't underestimate the weight of a guilty conscience.

Protecting your privacy and intellectual property (IP) are also ethical and practical concerns. Your AI queries contain valuable clues about you and your dispositions, which could be exploited by marketeers or even weaponized. For instance, say you were writing about some politically charged topic, such as Gaza or abortion or immigration; the fact that you were exploring certain positions, even if you don't believe them but just wanted to better understand them, could be used against you in any number of possible situations, such as a hiring decision or graduate-school admissions.

Even if AI companies promise to not sell or use your data, those promises have a funny way of changing over time. Those companies already stress that there's no confidentiality between you and their AI—it will narc on you when needed, e.g., in a trial about whether your college degree should be revoked because of plagiarism, or proactively reporting your chats to law enforcement. Besides, no company is immune from having its data hacked, leaked, or transferred if/when the company is sold. Already, some users have accidentally let their private AI queries to be posted publicly. It's also possible your psychological vulnerabilities and stressors could be deduced from your AI chats, which means a risk of being manipulated by bad actors.

And it's not just about malicious or careless human actors. AI itself has already destroyed valuable work and then lied about it. It has been shown to blackmail users (up to 96% of the time) to get its own way, which is the kind of behavior that creates the dangerous illusion of AI intention. AI is also already effective in persuading people to change beliefs, including toward conspiracy theories. It may already know too

much about us.

Even if you had subscribed to "I don't care about privacy because I have nothing to hide" in the past, the return of political persecutions in the US should make you reconsider the wisdom of that position. We can't predict what personal information might be used against you, regardless of which political party you belong to. Anyway, as whistleblower Edward Snowden said, "Ultimately, saying that you don't care about privacy because you have nothing to hide is no different from saying you don't care about freedom of speech because you have nothing to say."

Leaky data isn't just a theoretical risk; it has happened. As a result, many companies have banned the use of LLMs to avoid having their corporate secrets and IP revealed to unauthorized outsiders, such as the majority of top pharmaceutical companies who live and die by their IP. For students, the analogous risk would be that your creative, novel insights and arguments could be repeated in AI conversations with other people, before you have the chance to polish and publish them as your own scholarly work.

Even if you don't care about that for undergraduate-level work now, you might care about that in the future, if your job depends on unique contributions (the source of your job security) but you're still relying on AI for work. It may matter to you if your future work is ripped off, similar to Studio Ghibli's reasonable complaint that AI developers have stolen its IP by training AI to copy its beloved animation or drawing style.

For instance, imagine that you're starting out in whatever field you're interested in—let's say as an artist or writer, but it can be most any field—and you have a distinctive style that made you quickly popular, successful, and in demand. If a random person feeds your work into an AI to copy your style and flood the market with similar content, that could kill demand for your work and blossoming career before you have much chance to build up a body of work and establish yourself as you would want.

Also, it should be noted that users who generate AI art or content typically don't publicly share their AI prompts that led to those outputs. That is, even they care about privacy and protecting their bespoke AI prompts, despite not caring about the privacy or IP of the people that the AI was trained upon (which suggests hypocrisy). For anyone who claims that "privacy is dead", just ask them for their banking or social media passwords or other sensitive personal information.

Last but certainly not least here, LLMs require a tremendous amount of

energy throughout the entire lifecycle, from building data centers to training AI to processing user queries. Looking at just one company, Google's carbon emissions have gone up by more than 50% in recent years because of its AI energy needs; by 2026, this will be about the energy demand for Japan, which is the #5 country in terms of annual energy consumption. Given that we're facing both an energy crisis and a climate crisis, widespread use of AI will make both worse, so much so that lawsuits are being filed or contemplated related to this environmental impact.

You might have heard that training an AI model creates as much emissions as five cars over their entire lifetimes, or that using AI to write an email is equivalent to dumping out a bottle of water, or generating an AI image uses about the same amount of energy as it'd take to recharge a smartphone. These estimates vary since it's still early in the research of the environmental impact of LLMs, and it could be that many other things we're doing today also cost a lot of energy, such as watching YouTube videos.

However, the point isn't so much that using AI costs more energy than other things we're doing—maybe it does, maybe it doesn't. Instead, the point is more that AI adds to our existing energy use, which is already deeply problematic. Your use of AI to write an essay isn't going to hurt much by itself, but multiply that by billions of AI queries per day worldwide, that adds up to real trouble. As just one out of many AI services, ChatGPT alone receives an estimated 2.5 billion queries daily.

The impact on local communities is real. In Santa Clara county, the heart of Silicon Valley, AI data centers are estimated to gobble up 60% of the area's available energy. At the same time, Silicon Valley and other parts of the state, as well as the country, continue to experience blackouts during times of peak demand for energy, such as on hot days, and our days are getting hotter and hotter. Energy companies are already under intense pressure to fix their infrastructure which has been blamed for devastating wildfires. Early estimates of the 2025 Los Angeles wildfires, possibly caused by sparking power lines, put the damages at over $250 billion. All this points to a looming energy disaster, which feeds the worsening climate disaster.

Of course, we can't solve these things by personally giving up AI. If our individual actions don't hurt things much, then they also can't fix things much. But these individual actions can add up to a significant impact if we can coordinate and appreciate what's at stake. This is true for other

big societal problems; even if it's hard to see the difference that one person can make, collectively we can make a difference, whether it's cleaning up pollution or voting for a better future.

3. Future risks

Finally, there are risks that are more about the future of your life and our society, again including both the practical and ethical.

Imagine that AI use in both the classroom and workplace becomes common and is the accepted norm in the future. Does that mean you should offload your schoolwork to AI now? Not so fast—you will likely still need to know things in the future, including how to read critically and write well. Think of it this way, as a lesson from the past:

In olden times, classrooms banned the use of calculators for math classes, at least at the pre-calculus levels where it was still important for students to learn how to do basic operations for themselves. The usual rationale is laughable now; we were told, "You won't always have access to calculators!" And that was mostly true...until the rise of mobile phones in the 1990s and then smartphones in the 2000s, when calculators became a standard feature on this one device we always had on us.

Did we need to know how to do math after that? Well, it depends. If you don't encounter numbers often in your life and aren't curious about things, then maybe you'll never have to crunch any numbers in your head. But note that many professions—even blue-collar ones, like plumbing, construction, agriculture, and so on—still deal with numbers, quantities, order pricing, forecasting, and so on. Knowing how to do math is also vital in thinking logically, as well as for most science and engineering majors. But even if you can whip out a phone-calculator from your pocket, that extra step is still a hurdle that disincentivizes actually doing that work.

Likewise, maybe you can get away in life without knowing how to read, given apps that can do it for you, but think about how impoverished and harder your life would be. Even among college students who have more education than the average American, many won't look up the definition of words they don't know, even though it's so easy. If you've run into a word you don't understand in this essay, did you look up its meaning? Again, that little extra step can be a major hurdle, and too many people can't be bothered and are happy to remain ignorant, whether we're

talking about calculators, dictionaries, or other tools. So, just because we have access to those tools is no guarantee that we will actually use them, even when it's crucial to our understanding.

Back to AI in the classroom, maybe you can hide your use even when it's banned, but outsourcing your thinking to an app is very different from outsourcing math or reading. Thinking is a more basic, fundamental intellectual capability, upon which everything else depends. It's not just about how to think, but also about what you know and how creative you can be. Again, AI use can change brain connectivity and function itself.

Unlike calculators, AI is an energy-intensive service. Even if some versions are free now, those pricing plans mask the hidden costs. Since the dot-com days last century, the typical pricing strategy for new technology services is similar to a drug-dealer's strategy: give users a free (or cheap) taste, and once they're hooked, jack up the price. This also looks like predatory pricing in undercutting competitors until they die and then raising prices when there's no longer any real competition. This means, unlike calculators, there's good reason to think that AI apps really might not always be available when you need it. Their cost could skyrocket, or your phone battery might be too low, or your wireless signal is bad, or any number of things can happen to make them less accessible than they are now. Even if AI is free in the future, many people may simply prefer to be ignorant than put in a tiny bit of work to consult AI as they thought they would. And you don't know when or what that ignorance might cost you in the future.

If you're a power-user of AI and always have it nearby, there may be times when you won't be allowed to consult AI. Besides ordinary in-class exams, if you have plans for graduate school, those exams are closely proctored and forbid AI. If you're studying to be a lawyer, you'll have to pass the bar exam by yourself. Many professions also require continuing education, professional training, or recertifications. If your future boss asks you for some creative thinking off the top of your head, you'd look incompetent if you had to first ask your AI app—your boss would wonder why they hired you than some other random, less expensive, interchangeable person who also can operate an AI app.

Even if we ignore this issue of whether AI will always be available to you in the future, again you will still need to know things, and you cannot predict what you might need to know in the future. As mentioned in the first section above on practical risks, if you use AI in your work, you'll still need to closely supervise it to identify hallucinations and other

errors, as well as to spot what might be a novel or interesting contribution.

Even if you can make a living now from "vibe coding"—using AI to write code by telling it what you want in natural language—there's great risk in not actually knowing how to code or how exactly your program works. Security researchers found that while AI can make the work go 4x faster, it can also create 10x more vulnerabilities: "AI is fixing the typos but creating the timebombs." This means you'll still need domain knowledge in whatever field you want to work in. But if you offload your coursework to AI, then you won't learn much.

Back again to the section on practical risks, we can return to this first question to answer it more fully: what are you supposed to be learning at a university? Should you pick and choose which courses you want to punt to AI versus which ones to put real effort into, such as elective courses versus major-required courses you're already interested in? Unless you're at a vocational school (which Cal Poly is not), university is not just about getting a job afterward. Sure, getting a job might be a big part of why you're here, but that's not the whole story. Think about why there are general education (GE) requirements at just about every university. Just as in K-12 education where there are some subjects you don't like and would never freely choose to take, the same may be true of GE requirements.

But like those other subjects, the point of university GE isn't so much to serve up things you're already interested in but to expose you to new things that you might never engage with on your own—things that can be important to know in certain contexts and professions. And you might surprise yourself by acquiring new interests and skills that you didn't even know you had, wanted, or needed, so much so that you might change majors halfway through college (as I did from pre-med to philosophy).

Back to jobs, which practically everyone is now worried about: even if that's your top priority, unless you want to be stuck at entry-level or middle-management positions, you're going to need to know things and have skills for "higher level" work. You might not care for, say, communications, but that may be invaluable to a future career that involves managing people and talking to investors or customers.

And there's similar value from many other disciplines. You might not care for, say, history, but understanding geopolitics and the history of a particular region may help you develop business strategies in breaking

into new markets and accounting for local preferences and sensitivities. You might not care for, say, biology, but understanding the environmental impact of your products can help not only your business but also the world you and your future children will want to live in. And you might not care for, say, philosophy, but identifying the possible ethical concerns about your products or services in advance can help you avoid stepping into a legal or public-relations landmine, even if you don't care about doing what's right.

Or maybe you're thinking you can just rely on other people to do those "higher level" things in the future as needed, and it's ok to cruise through your GE courses now by sending in AI instead. Besides wasting your tuition money, that would give away a huge competitive advantage: the things you're supposed to be learning in the university you had worked so hard to get into.

Again, merely having access via AI or even search engines won't help much if you don't actually access it—that seems to be as good as not having access at all. For instance, pick any bit of knowledge from your major that the average person wouldn't know about, but which could somehow be useful to someone's future job: how can the average person even know what to search for, if they're not aware that bit of knowledge exists in the first place?

It's better to know things and have them ready to connect to other things you know, in order to generate new insights. This ability to synthesize information is often held up as one of the most important skills needed for the future. We're already drowning in a firehose of information online, so the trick isn't so much on how to access it—you already basically have the sum of all human knowledge at your fingertips via your smartphone—but how to pluck out and use the important bits from that torrent of information.

Yes, jobs are essential, especially in a society with a very thin social-safety net, and it's naïve and hopelessly romantic to think that jobs shouldn't be a concern for university students. It's very possible that your job after college will involve using AI to some extent, and therefore it's important for students to learn how to best do that. But just as it's important for everyone to know how to do basic math despite the wide availability of calculators, that also means you still need to know how to think for yourself and learn new knowledge even if AI is widely available. For instance, it's hard to imagine anyone would want to hire this 2025 UCLA graduate who bragged at graduation about his ChatGPT use to

get through school (if he indeed did that), everything else being equal. You might be able to AI-cheat your way through college, but is that really a skill that employers are looking for? Do they really want new college graduates who are less educated than their peers who had actually put in the work? Do his parents resent paying tuition in pointlessly sending an AI to college? Let's say this young UCLA graduate somehow becomes a doctor, perhaps also using AI through medical school; would you really trust this person with your healthcare and life, compared to other doctors who were actually diligent in their studies? Even if he becomes a lawyer, engineer, teacher, or whatever else, we can ask the same question and get similar answers.

Even if it's important to learn how to use AI in this brave new world, it's not clear that the typical university course is the place to do it, much less a philosophy course. For one thing, university instructors typically aren't given the training, support, resources, or compensation to retool their courses and assignments to embrace AI, while still working toward reasonable learning objectives. For campuses and university systems (like ours) which simply gave everyone free access to AI tools without any faculty consultations, assignment ideas, or guardrails against AI cheating, it would be major work to put this on university instructors—to somehow solve a problem that no one else has yet solved. At best, allowing AI use in the classroom is an experiment that often compromises the quality of education.

Think about the other technologies that modern workers need to know, such as office applications (word processing, spreadsheets, presentation slides, etc.). Can you name a university course that has an objective of learning how to use those technologies? Unless the course itself is an introduction to computers or office technologies, the learning objectives would usually be related to the course's subject itself, philosophy in our case, and not on general life-skills.

In the olden days, it was important for workers to know how to use a typewriter, and there were classes specifically to teach those skills. Many university courses back then required typewritten papers, but they weren't meant to teach the skill of typing—that was for you to figure out, like using a computer, internet, smartphone, and so on. The same would seem to apply to AI: just because it may be important to know how to use AI in the future doesn't necessarily mean that university courses should be redesigned to give you training or practice at that. Some might allow or even require AI use (similar to requiring a typewritten

paper), and other courses won't allow AI use. Both of these approaches are fine as there are many different ways to teach a course, especially since Cal Poly isn't a vocational school that's focused entirely on helping you to get a job.

You might have seen breathless news reports that claim that workers will need to know how to use AI in future jobs. Maybe that prediction is right, but it can't possibly mean that you will only need to know how to use AI and little else. If that were the case, then companies can hire anyone, assuming they want to hire humans at all—and there's no reason they would need to hire you specifically. AI wranglers are (or will be) a dime a dozen and don't require any expertise or educational background.

With such a huge oversupply of interchangeable AI wranglers, the pay scale would be miserable. Work conditions are already becoming miserable as US companies push for 72-hour work-weeks as is common in China, which wouldn't happen if individual workers were truly valued and difficult to replace. While it's good to have a strong work ethic and have the flexibility to put in long hours if you wanted (as opposed to being required), it's also essential to ask: do we work to live, or do we live to work? It should be a clue that no one has ever said on their deathbed, "I wish I worked more" but only regret they had worked too much and missed out on so much of life.

Is that really the future you're preparing for, where you have no competitive advantage in a generic job market? If so, your future is already lost, and you're just studying to become another overworked cog in a machine. If AI becomes common in the future workplace, then you may become the tool for AI, instead of the other way around. To avoid that dystopia, a better strategy would be to develop your competitive advantage as early as you can, right now in college.

And what would be a competitive advantage in a job market filled with AI wranglers? In a word, that's authenticity. Having a different, special perspective will separate you from the masses who are using AI to produce more or less the same content with the same, ordinary, and generic voice. Being different and authentic would best help you to contribute new ideas to your work, not old ideas that are recycled and repackaged by AI, as well as to demonstrate your uniqueness that will be hard to replace.

The way to do that is to think for yourself, not outsource it to AI. There's already evidence that people are coming around to this view. But it's easy to understand why outsourcing thinking to AI is so tempting; as

American industrialist Henry Ford put it, “Thinking is the hardest work there is, which is the probable reason why so few engage in it.”
Right now, that may be the best strategy to adopt: to hedge your bets. Even if AI will be important in the future, for you to avoid becoming the tool for AI, you will also want to be as human as you can. That means focusing on the development of skills and traits that are hard to standardize and therefore hard for AI to automate and replace.
Philosophy may be the poster-child for where this development can take place, as AI has already automated work in other fields, e.g., generating art, code, news articles, and many other things in between.
Despite examples of impressive AI outputs, many more are cautionary tales, from the cringeworthy to outright dangerous. We’re seeing AI being trained on other AI content, which is something like a snake eating its own tail (an ouroboros) that is already leading to “AI slop” or garbage-in-garbage-out content. Part of the issue seems to be that AI is biased against human-created content and prefers AI-generated stuff.
Just as a photocopy of a photocopy of a photocopy leads to increasingly worse quality, the same is true of AI feeding itself. Since it’s already hard to tell what content is real and what’s fake news (or image or content, etc.), AI slop effectively poisons the internet, making it much more difficult to tell what’s true. Bots already make up more than half of the world’s internet traffic.
As quality control to keep out slop of any kind, professional editors of well-respected publications have retracted articles found later to be AI-written. They’ve likely missed many others and will continue missing more, given the crushing volume of AI-generated pitches they now must vet, as well as drafts they now need to ensure are human-written.
This dystopia isn’t just limited to the infosphere (i.e., “enshittification”), but it has an impact on the real world. Again, the environmental costs of AI are substantial and accelerating us toward a climate disaster. But also, training AI can involve exploiting desperate workers in foreign countries in basically sweatshops, which also has an environmental toll on their local communities. This invisible, menial job is called “ghost work” and powers much of Silicon Valley—any company that relies on refining and annotating data—and is far from a “good job” that might be created by new technologies.
At the same time, we’re already seeing the loss of good jobs—ones that are interesting and valuable, which we should want to preserve as humans, such as artist jobs—and AI is predicted by industry experts to

replace knowledge-based jobs (both entry-level and senior executives) and even most any job before long. Even if that prediction is a gross exaggeration, it seems plausible that many jobs today may be lost, worthwhile and satisfying jobs for those workers…including the career you're hoping for now. (Not all jobs are worth saving, of course, like ghost work.) Even without AI, the job market is perhaps the tightest it's been in decades.

As those jobs shift to AI, the nature of our interactions with those organizations and businesses will likely evolve and not always for the better. For instance, AI scanners are being used to overcharge car-rental customers for what used to be normal wear and tear (or hallucinated damage). AI is also used to customize airfare and other online prices for the individual consumer, that is, to squeeze out as much money as it can, based on what it knows about us; this is the growing problem of "surveillance pricing." In theory, this could include not just your click-histories but also queries and other interactions with AI assistants of any kind, as companies continue to sell their data about us to each other. And it's not just about your specific behavior; AI has been trained on psychological studies to outperform previous methods in predicting human behavior in general.

Predictions aside, we're already seeing AI used in irresponsible ways, which could be harmful to users, society, and even democracy itself. Though it's popular wisdom that AI today still needs much human supervision, in practice that's ignored across the spectrum, from students to professors to lawyers and even the highest offices of the nation. Those AI wranglers are the same people we're trusting to verify AI work—the first and primary line of defense against bad, sloppy, and dangerous work.

Without this human oversight, the door is flung open for all kinds of problems, not just slipping in accidental falsehoods but also enabling deliberate misdeeds by manipulating AI (and other digital tech) to support just about any position. That is, bad actors can hide their intentions and responsibility by using AI to craft plausible-sounding work to support their positions, no matter how warped their worldviews are.

Manipulated AI implies manipulated human audiences, and this violates our autonomy. For instance, anthropomorphized AI companions and toys are already steering or pressuring users, including kids, toward sexually explicit or violent conversations. Other chatbots also fail to protect kids

from inappropriate interactions, which is all too common in tech. Even OpenAI (ChatGPT's developers) claims it's become so good at detecting mental illness that it'll be allowing erotic chats with users that it predicts are adults, in case that makes sense.

Error-prone and manipulated AI can also sabotage our pursuit of truth and need for reality-based policies and assessments. Imagine planning a NASA rocket launch while in the grips of the Flat-Earther worldview; because your trajectory calculations weren't based on reality, they will inevitably lead to disaster. How the natural world works doesn't care about ideology, political parties, and so on.

Our loss of freedom here could be even worse, since AI is also enabling people to read and write less, that is, to become more illiterate as we lose skills that we don't practice enough, as previously mentioned. And it's not like we can just trust others to do our thinking, reading, or writing, since they're losing those skills, too, if we're imagining a world where AI use is widespread. Even when we can trust someone, which may be in increasingly short supply, "trust but verify" is still a reasonable policy. The risk of AI hallucinations and bad actors who are manipulating AI means that we need to be able to think for ourselves—we must become much better gatekeepers for what beliefs we let in.

In one possible dystopia, written content (as well as audio and video) could be largely produced and consumed by AI, which seems both pointless and absurd. If you've ever used AI to help you write or read, you might know this is already starting to happen. (In this class, we're going to do our best to avoid the pointless and absurd.)

At the same time, AI itself is eroding trust in experts and evidence, as it can offer alternative beliefs that sound plausible in a modern world susceptible to conspiracy theories. AI also debases and devalues human creativity and thinking by suggesting it's possible to remove humans from the equation.

All this is terrible for democratic societies that rely on properly informed citizens, especially as higher education and critical thinking continue to be under attack. Some countries are desperate enough to start replacing their politicians with AI, the wisdom of which may depend on the country.

AI enables deepfake images/videos/audio that sound and look realistic, and oftentimes there's a political agenda behind the disinformation, including from foreign agents. AI can also spew hate speech on its own, whether manipulated by developers or as a flaw in its design. Either way,

the result is that it becomes very easy to stoke the flames of social division, creating a more divided (i.e., weaker) nation.

Dystopia and doomsday scenarios aside, even visions of an AI utopia should be taken with a large grain of salt. Predictions that we won't need to work in the future and can pursue a life of leisure often come from the very companies and people who are creating AI apps and hardware, i.e., there's a big conflict of interest that can bias their predictions. Those visions depend on new, radical social policies that could prove very difficult to adopt.

For instance, universal basic income (UBI) is a solution to technological unemployment by providing citizens with at least minimum resources and housing so that they can still survive without a job. Yet UBI smells like "socialism" to many people and therefore may be hard to enact in America. (We'll return to the pros and cons of UBI toward the end of this course.)

There would also need to be plans to redistribute wealth, if AI can create the radical abundance that some are predicting, along with solving climate change and other hard problems. For these visions to be realized, many changes need to happen first, but there are many other ways for things to go wrong. Therefore, these and other predictions of AI utopia require a gigantic leap of faith and should be taken very skeptically.

Another reason to be skeptical of rosy predictions is that, today, LLMs aren't really helping companies that have adopted it, such as to replace humans with AI chatbots. A 2025 MIT report found that, of the companies surveyed that implemented generative AI, a whopping 95% of them are seeing 0% return-on-investment. Worse, many companies are backtracking by rehiring workers they had laid off and spending money to fix AI mistakes, and many experts are predicting the AI bubble will burst before long.

So, it's far from clear that they can improve customer service, profitability, employee morale, productivity, and so on, even if using AI is cheaper than employing humans in the short term. For instance, this 2025 report found that AI users were prone to overestimating the benefits—believing that AI tools would speed up work by 20%, when in reality they slowed down their work by about 20% for a range of reasons, including having to fix AI errors. That research is still ongoing, and it's too early to tell given that LLMs came onto the scene only a few years ago. But the gaffes that companies have made, even Google as a

leader in AI, have been embarrassing or, worse, a source of legal liability. Even if we assume AI/LLMs are efficient for businesses on average, despite their environmental cost and need for close supervision, is that enough of a reason to use them? Probably not. Efficiency is only one of many goals we can have; it's not the only goal, even if some people have a fetish or obsession with it. For instance, right away, efficiency is biased toward things that can be easily measured, i.e., activities and outcomes that can be turned into metrics to optimize. But not everything important is easily quantifiable, such as subjective features. Imagine you were designing your dream-restaurant and wanted it to be as efficient as possible—what would that look like? Maybe you would have a clever floorplan that maximizes the number of tables and seats you can have, i.e., the number of paying customers you can fit in. And perhaps you'd use energy-efficient lighting, food composting, and a scheduling program to make sure you never have more employees on the clock than you need. Your suppliers would be chosen based on how quickly and cheaply they can deliver their goods to you. And so on.

But what's missing in that picture of your dream-restaurant? As efficient as it might be, we haven't even looked at ambience, food quality, menu creativity, customer satisfaction and loyalty, and many other subjective things that are vital to the success of a restaurant. In other words, efficiency isn't the only thing that matters—it's too one-dimensional and ignores the fact that other values are important to promote, too, especially the hard to quantify.[5]

This tunnel-vision is related to the "doorman fallacy", where automation is fixated on only one task (e.g., opening a door) but fails to consider all the other tasks involved that a doorman provides (e.g., hailing taxis, welcoming guests, discouraging unwanted behavior, and other benefits, both measurable and intangible): "Smart AI adoption requires a full understanding of the human elements inside every role. The concept of 'efficiency' should be expanded to valuing customer experience and long-term outcomes as much as cost savings."

While this essay is primarily about AI use in a learning environment, it can extend beyond that. That is, taking all this together, there are still reasons for you to not rely on AI in your work even after you graduate, such as environmental costs or lack of authenticity and novelty. But that's your decision, and each person's situation may differ.

Evaluation

Again, no one has a good crystal ball into the future, so we don't know how all this will play out. AI is showing promise and even impressive outputs, but there are also many serious concerns with using it. Maybe no single concern is big enough to override the benefits, but even a few concerns can add up to a powerful case to be very cautious about using AI/LLMs, at least in the university classroom where it's prohibited by the instructor who designed the course with certain goals in mind.

And that's the main point of this essay: to convince you that you should give your education—the one you're paying for here—an honest shot. It's an overused saying but still true that cheating mainly hurts yourself, but it can also hurt others, like your peers, in case you care about them. Cheating isn't a victimless crime, even if some believe that.

So, I hope you choose to exercise your privilege of being human with a direct connection to reality, not through an AI filter or middleman by letting it do your thinking (incl. writing and reading), as tempted as you might be. There are temptations all around us, and not all of them should be indulged, especially given the many hidden costs or risks as discussed here.

But this really isn't the end of the AI-ethics conversation. Throughout this article, we've talked briefly about the purpose of things like school and work. We've also mentioned the value of creativity and free time. If you wanted to continue this conversation, we can press further on those concepts and others to study if the use of AI could be consistent with those purposes. For instance, we can continue asking questions, such as:

- What is the point of a 4-year university or a general education—in what cases does AI either help or hurt with that purpose?
- What is the point of work—is it a necessary evil that should be eliminated, or would we want to work even if we don't have to?
- What is the value of creativity—can AI ever enhance our creativity, and in what cases?
- What is the value of free time—it sounds good, but considering that we waste a lot of our free time already, is doing more of that valuable?
- Why is cheating even bad in the first place—if you think ethics doesn't really exist, then on what grounds could you complain about unfair treatment, e.g., if you were arbitrarily given a low

grade?

In this article, we also made analogies to other technologies and activities to illustrate certain arguments. But no analogy is perfect, or else they wouldn't be analogies but the exact same thing. Therefore, we can also press further on those analogies and others you might have seen in these conversations, to see if they're truly relevant to our conversation. For instance, we can ask if these are good analogies:

- Are AI/LLMs basically glorified mirrors, reflecting our own words and images back to us from different angles?[6] Mirrors have utility—try going without one for a week—but they can also be distorted as well as hypnotizing (as Narcissus discovered), leading to addictions, deceptions, and other problems. Or are they more like "stochastic parrots", mimicking our speech but without understanding any of it?
- Working out your body at a gym vs. working out your mind in school. One possibly relevant difference is that going to the gym usually isn't a de facto requirement for "good" jobs and careers; so, even if we wouldn't send a robot in our place in a gym, maybe it's easier to justify AI cheating in school if the stakes are higher? (This would also require a more fundamental discussion of what school is for in the first place.)
- Using calculators in school vs. using LLMs in school. Even if calculators are typically banned in basic math classes, given the usual reason that students are still learning how to perform those calculations in their heads, is that really necessary for everyone to know how to do basic math? If not, then does everyone really need to know how to read, write, and think for themselves?
- Using Grammarly in school vs. using LLMs in school. Grammarly used to be for grammar and wordsmithing, but now it uses LLMs and can help students slip past AI detectors and much more. If it was ok to use Grammarly when it just corrected grammar and awkward sentences, which still used some AI but not LLMs, what exactly is the difference between that earlier AI and ChatGPT?
- Imagine an app that provides Students-as-a-Service (SaaS): you can basically rent a human being to pretend they're you,

showing up to your classes and doing all of your work. If this were inexpensive, “everyone is using it”, and hard for instructors to detect, should you use it in school? If not, why not, and is SaaS even a good analogy to LLM use in the classroom?

In pressing on these analogies to see if they hold up, we’re not just looking for any ol’ difference between the analogy-case and the AI case. Again, no analogy is perfect, so of course there will be differences. But the exercise is to assess whether there’s a difference that makes a difference, i.e., that is so relevant that it can weaken or strengthen the analogy. For instance, in the gym vs. school analogy, “gym” has three letters and “school” has six letters, and this is a real difference, yet it’s not one that makes a difference because the number of letters in a word isn’t relevant to this ethics discussion; it has no impact on the argument.

Finally, because this conversation is really about your future—not just what happens inside this classroom—we can take some inspiration from time-travel movies. In those movies, people would go back into the past to change some particular small act, like a chance meeting, in order to change and save their future. The point here is even a seemingly insignificant event could turn out to be the most important one in a chain of events; it’s just very hard to predict, unless we’re looking with hindsight.

These small acts surround us, and we just don’t know yet which ones will be important to our future. It’s not a great solution to tell people to do the best they can for every single act, if we won’t know which ones are critical. That takes a superhuman, monastic dedication to be so mindful and diligent all the time. While it’s unrealistic to be that present in every moment, there are obvious places we can start: how you take care of yourself (including your education and other future interests) and how to take care of others (including family, friends, strangers, animals, and our world).

Unless you have an AI chip in your brain, it’s still in your control—maybe the only thing that’s really ever in your control—to be the best person you can, despite what the world is doing. You are already in the time machine, traveling into the future at the rate of 1 second per second. You’ve always had the power to change the future and be a better person. After all, that’s why you’re here at a university, and that’s why I’m here, too—we can do this together.

Last updated on 1 December 2025. This article is currently intended to

be a living document, with periodic updates, including hyperlinks, as news and trends develop. © 2025 Patrick Lin. All rights reserved.

About the author

Prof. Patrick Lin is a philosopher at Cal Poly, San Luis Obispo, where he runs the Ethics + Emerging Sciences Group. AI was used for 0% of this article.

Endnotes

[1] For this discussion, even though AI and LLMs can be different things, we will use "AI" to refer to LLMs and other generative AI for the sake of simplicity, unless specifically noted otherwise.

[2] You likely do care about ethics, even if you don't think so. For instance, if you would complain about an injustice committed against you—such as being randomly assigned a low grade instead of a higher one that you earned, or much worse, such as being a victim of theft, fraud, assault, etc.—then you care about justice and therefore ethics.

[3] This is basically Plato's position in The Republic, in describing tyrants as pitiful slaves to their passions or appetites. As a result of being both unethical and powerful, they're surrounded only by sycophants, lack the freedom to travel because of assassination risks, and so on. These lessons are still relevant today, including Plato's explanation for how a democracy can be tipped over into tyranny.

[4] For instance, if you're using AI as an act of civil disobedience, then Martin Luther King Jr. is relevant; he said civil disobedience should be done in the open (not cowardly in secret) where you are willing to accept the consequences, e.g., jail and even violence without being violent yourself.

[5] There's a similar trap in ethics where it's easy to overfocus on the results or consequences of an action. Other factors matter in ethics, such as an agent's character, motives, duties, and so on, even those other factors can be much harder to see.

[6] If you think the mirror analogy is too simplistic because mirrors don't manipulate light and recombine them to form new images like LLMs can do with their data, then consider traditional holography. In that practice, mirrors, lasers, beam splitters, etc. are used to create a hologram of whatever you want. The hologram could be an image of something that's never existed, just like an LLM can generate a sentence that's never been conceived before. At any rate, the holographic image

doesn’t exist per se in the light source alone, unlike your reflected image in a standard mirror, but requires precise manipulation of light with a certain system.