Critical Thinking & Information Overload

"Computers make excellent & efficient servants, but I have no desire to serve under them."

Spock. "The Ultimate Computer." Star Trek Gene Roddenberry, creator & producer. NBC

An idea that initially appealed to my students was the assumption that an overabundance of information was the same thing as critical thinking skills. In early papers, some students data-dumped and supposed that the sheer *amount* of information proved that, in addition to filling up pages, they were also thinking. A similar assumption seems to inform artificial intelligence (AI) technology. CBS's *60 Minute* correspondent Scott Pelley reports that Google's new chatbot Bard appears "to possess the sum of human knowledge" and in addition, has "super-human skills."[1] I gave students an article written by Steve Nguyen called "Information Overload—When Information Becomes Noise." The article makes reference to a paper for the Harvard Graduate School of Education's Learning Innovations Laboratory (LILA) written by Joseph Ruff. Ruff defines "information overload" as when our ability to process information has passed its limits, and further attempts to process information–or make accurate decisions from the surplus of information–interferes with our ability to learn and engage in creative problem-solving.[2] Ruff's generation-old article remains pertinent, particularly his claim that "when our ability to process information has passed its limits…the surplus of information–interferes with our ability to learn."[3] A new generation of students was also supposed to be able to multitask far more nimbly than could their teachers' generation. However, that claim was soon questioned when neurologists began reporting evidence that "when we *think* we're multitasking, most often we aren't really doing two things at once…we're doing individual actions in rapid succession, or task-switching…[and] we become less efficient and more likely to make a mistake."[4] Fortunately, we can still walk and chew gum, but as Dr. Kubu from the Cleveland Clinic explains, "The more we multitask, the less we actually accomplish, because we slowly lose our ability to focus enough to learn…we don't practice tuning out the rest of the world to engage in deeper processing and learning."[5] Even government officials are beginning to recognize a trade-off between speed and accuracy, now, for example, regarding the information reported on current wars, and the "threat of force multiplier" that occurs when flash mobs rally in support of online threats.[6] What the new claim for AI's "super-human skills" overlooks may be the key insight of Ruff's paper: "Once capacity is surpassed, additional information becomes noise and results in a decrease in information processing and decision quality. Having too much information is the same as not having enough."[7]

Technology brings advantages to some medical research, particularly with prosthetics, and to some forms of communication, such as AI now said to be able to detect oncoming wild fires.[8] But the out-of-scale speed and amount of information introduced by Google's Bard, Open AI's chatbotGPT, and other AI technology interferes with the thinking of diverse and independent individuals in society. They may not bring the "super-human skills" promised as much as they bring more distraction and more noise for minds–now hyperactive–to sort through. However, even as we have become more aware of this possible downside to the benefit of the vast amount of information at our disposal, we have also become increasingly dependent on it. Facebook's use of AI micro-targeting first made us aware of its capacity to groom young people for dependency. And as dependency increases, the sheer multitude of choices often paralyzes decision-making and sends us down rabbit holes rather than refining specific goals or expanding our horizons. In the fall of 2020, the pandemic was just around the corner. Joseph Fuller, professor of management practice at Harvard Business and the Work World School, wrote, "Virtually every big company now has multiple AI systems and counts the deployment of AI as integral to their strategy."[9] Perhaps because the pandemic brought so much isolation, this early spread of AI seems to have occurred almost unnoticed. However, when the pandemic retreated, AI, and in particular, OpenAI's chatGPT took the country by storm. "Within two months of launch, OpenAI's chatGPT became the fastest-growing consumer product in history, with more than 100 million active monthly users in January 2023 alone."[10] But along with lightning growth, the new AI tools brought "worrying implications for the future of work and education as well as the future of humanity."[11] Because there is too much we do not know and because no one accepts responsibility for what they do not know, Congress needs to put some brakes on development and use of AI chat bots in society in order to allow more time for testing—not just for the "overwhelmed" legislators who acknowledge their less-than-competent understanding of it or for Supreme Court justices who have not kept up with fundamental changes technology is ushering in to society–but also for the welfare of all in society.

The use of artificial Intelligence (AI) by corporations to expand power is akin to the growth of virtual imperialisms that are spreading despite the fact that at least four major issues involving all of society remain unresolved: 1. critical thinking in education is often undermined; 2. employment patterns are disrupted; 3. trust throughout society is eroding as a result of surveillance and extreme disparities in wealth between those few reigning at the top of virtual empires and mass societies of their newly-colonized subjects; and 4. the exploitation of individuals' personal information often results in increased spread of social bias and disinformation. Tristan Harris, co-founder with Aza Raskin of the Center for Humane Technology, tells NBC anchor Lester Holt, "No one is building the guardrails … And this has moved so much faster than our government has been able to understand or appreciate."[12] Harris and Raskin tell Lester Holt that the rate at which we are developing AI is "reckless" and comparable to "an arms race," and it is occurring "with as little testing as possible."[13] "Tesla CEO Elon Musk, one of the first investors in Open AI when it was still a non-profit company, has repeatedly issued warnings that AI…is more dangerous than a nuclear weapon."[14] AI technology turns knowledge into an overload of data for colonizing human minds in order to increase profits and thereby extend the reach of imperialists' power over those minds. CEOs of virtual empires are possessive of proprietary algorithms not unlike those that ran the computer HAL in Stanley Kubrick's *2001: A Space Odyssey* that runs a complex spaceship only to eventually become possessive of its power over the entire space mission.[15] So it is not as though we have not been warned to connect the dots. AI technology enables instantaneous spread of powerful propaganda, and with its focus on saturating us with entertainment and immediate gratification more so than with thoughtful deliberation–the critical thinking process by which the young are taught to examine assumptions and their sources, to evaluate and weigh evidence, and to draw independent conclusions for themselves–is often and easily undermined. Testing innovations and requiring evidence to establish not only the benefits of innovations but also evidence of the costs/risks of those innovations to society is needed to slow the speed of technological development from "super-human" to a more realistic level of where we are at–i.e., one of human scale.

Such action would also encourage wider citizen involvement in deliberating on and then evaluating the technology rather than the technology first being "sold"/donated to schools and businesses and then discovery of its long term detrimental effects coming only later —much later, after the grip of dependency has become a crutch that many *want* to continue. A cross-section of viewpoints also provides more diversity of viewpoints than those of stockholders in multi-national corporations, who often compete only to increase their bottom lines. Again, government is needed here to make informed decisions on behalf of the nation and to provide prudent and tough oversight even as start-ups, research labs, and investors all elbow each other aside to be first in line. But legislators are slow to connect the dots and often in awe of trillion dollar corporations. Transparency with algorithms and further developments should be open sourced until legislators and the public *understand the implications of what it is they are doing!* Google's new chatbot Bard is said to "learn" from "possessing" whatever sum of human knowledge it has been fed. Bard introduces itself as "I'm your creative and helpful collaborator."[16] But possession of data in and of itself is not knowledge until tested with multiple, real life experiences and integrated with empirical knowledge. We do not get enough variety of human experience from sitting in front of screens or from leaving education to chatbots "helping" with our decision-making. With endless amounts of peripheral information already tracking each choice we make, the mind is easily hacked by influencers using profiles and branding–which are new ways we ourselves are being stereotyped and have learned to stereotype others. In previous generations, we occasionally met odd characters at grocery stores, gas stations, post offices, or may even have had a couple of eccentric neighbors. Later we may have recognized that people are not just their images–like covers of books–and that jumping to conclusions is almost always unreliable. We had opportunity to develop tolerance for thinking through ideas before they settled into impulsive ratings of thumbs up/down. We may have even experienced an adult or two changing their opinions of someone and then explaining to us why they were initially mistaken.

This is very different from the echo chambers we enclose ourselves in today, where often we remain listening to feedback that repeats multiple versions of our own thinking. As a result of being removed from each other's physical presence, many choices are now categorized before we make them. Predetermined algorithms keep us comfortably cocooning with our own preferences, which results in bias confirmation. We may be forfeiting human understanding of the complexities of *individual* thoughts—of the interesting sides of our differences–for new labels that indeed judge a book only by its cover and thereby further encourage intolerance. Students agree that technology makes us less patient. We are less tolerant—not only of others, but also of ourselves. Stereotypes bypass the much-needed learning to tolerate others–what we may dislike or what differs from us–by using rhetoric that easily incites predictable responses. People engage more often with things that elicit emotional reactions, and stereotypes are *shortcuts* to those emotional reactions. Stereotypes also contribute to hate crimes because it is easier to dislike what we have never learned to tolerate or do not even know. Frances Haugen points out Facebook's own research shows content that is hateful, divisive, and polarizing is "easier to provoke people into anger" than to other emotions.[17] Note Haugen's statement that it is "easier" to provoke people to anger than to other emotions. Anger is not the only emotion that provokes people. It is just that it is an "easier" one to provoke because it relies on more shortcuts. Easier means less time and thought occur before emotional responses explode and go viral. Resorting to outbursts of hate speech, bullying, trolling, and inducing outrage, all circumvent learning to understand new ideas and/or learning to tolerate disagreements in favor of profiling, branding, and other shortcuts to critical thinking that technological speed uses to bypass more thoughtful analysis as it spreads virally. Reactions based on stereotypes and jumping to conclusions are easier when people resort to paths of least resistance or ways of avoiding mental work, such as procrastination, idling, or thinking on auto pilot—now made seductively *even easier* with Open A.I.'s Chat GPT and GPT4. Rather than the hard work of examining different points of view in order to discover what is promising and particular to individual and independent thinking, AI technology makes it easier for students just to reverse course and go with the flow—to put their minds on *cruise control*–grabbing bits and pieces from what has been said time and again.

Imagine reading politicians' clichéd talking points or those of the NRA heard over…and over…and… Resorting to cruise control thinking is mental avoidance of thought by relying on ready-made stereotypes. It breeds addiction to instant, stockpiled reactions rather than allows individuals to form more independent ideas by having to think for themselves. Furthermore, as we process things with greater speed and pride ourselves on disrupting we know not what, we bypass a need to think long-term, to see what lasts through next generations, beyond the immediate horizon of the latest app on an iPhone. It is no accident that climate change, an issue long with us, is so often kicked down the road because--until lately–we did not *feel* the impacts of a warming planet knocking at our doors. Disrupting long-term thinking further contributes to undermining critical thinking–the loss of which far outweighs the benefits of speed, convenience, and ease. Aren't these filter bubbles of our own thinking what we are supposed to outgrow as we become more educated and therefore more open-minded about a broader, more complex world? Thus, technology and its wizardry may be counter-productive to our maturity as human beings. One year into the pandemic, in her testimony before Congress about Facebook's conscious choices to use *algorithms that put profits ahead of people*, Frances Haugen drew a connection between Facebook's use of AI "micro-targeting that is at the root of its business model and the opioid crisis: "In many ways," Haugen stated, "micro-targeting is the digital equivalent of the opioid crisis."[18] It "relies on artificial intelligence to attract users' attention, maximize engagement, and disable critical thinking."[19] Researcher Karen Hao backs up Haugen's testimony by stating, "Machine-learning algorithms create a much more powerful feedback loop….they…continue to evolve with a user's shifting preferences, perpetually showing each person what will keep them most engaged."20 This "evolving" or "emergent" property of AI chatbots that tracks or keeps people "perpetually" engaged not only becomes addictive, but might also be seen as some kind of automated stalking of its users' preferences. Haugen documents claims that Facebook "'is pulling families apart….and in places like Ethiopia it is literally fanning ethnic violence.'"[21]

Furthermore, "It is tearing societies apart…including Myanmar in 2018 when the military used Facebook to launch a genocide" on the Rohingya Muslim minority.[22] *The Wall Street Journal* reviewed additional documents from Facebook employees showing "human traffickers in the Middle East use the site to lure women into abusive employment situations….They sent alerts to their bosses about organ selling, pornography and government action against political dissent."[23] As if these accounts were not horror enough, we learnedthat Facebook has been "studying the phenomenon since at least 2016."24 "Studying"—but nevertheless continuing to expand itsvirtual empire while profiting off of ever more human suffering. What improvement or convenience in our lives outweighs that much human suffering? Mark Zuckerberg responds to the persistent problems with Facebook and now parent company Meta by stating, with typical vagueness: "These are complex issues that you can't fix. You manage them on an ongoing basis."[25] He adds, "We have a different world view than some…covering this."[26] Tim O'Reilly, of O'Reilly Media, tells Walter Isaacson that, like Haugen, he too believes 1)"incentivizing tech corporations to pursue A.I. innovations" is "exactly parallel" to 2)"incentivizing drug companies to sell opioids…leading to the opioid crisis."[27] O'Reilly warns this is "a real learning moment here for us": "We literally have a system of incentives in place that told companies that it is okay to maximize for shareholder value; it is okay to tell the FDA, hey, downplay the risk of addiction here."[28] O'Reilly goes even further to link both design choices—the pursuit of AI innovations and pursuing the sale of opioids –to an arch-design choice of "relentlessly pursuing material profits at the expense of other human beings,"…which he calls "the master algorithm of our society."[29] It is for that arch-design or master algorithm that Congress thus far refuses to restrict incentives and/or regulate with AI corporations. We remain complicit with—or what some have called "willing slaves" to–the 21st century's form of advertising by ignoring the spread and addictive lure of the virtual imperialisms that now drive many of our choices—such as providing weapons in the Mideast– even—as is clear from O'Reilly's explanation—in order to kill other human beings.30 Semantic Visions CEO Frantisek Vrabel wrote a book with a title reminiscent of the opium wars in China, *How Facebook became the Opium of the Masses*, and in which he writes, "Until Facebook opens its algorithms to scrutiny – guided by the know-how of its own experts…the war on disinformation will remain unwinnable, and democracies around the world will continue to be at the mercy of an unscrupulous, renegade industry."[31]

Like preparations for war that rely on fear to demonize others, the extremes so prevalent in politics today are used to assert power over listeners. At least $2.5 billion was spent on the 2020 elections, and we are more polarized than ever as a result of labeling others rather than arguments that show an ability to understand there are many more viewpoints than one's own.[32] Due to social media's pervasive influence, today we are well aware that the extreme polarization of "us vs. them" thinking on both right and left threatens the very functioning of democracy. In a recent trip to Budapest, Pope Francis spoke of "the shift *from communism to consumerism*," and alerted us to the ease with which people can move from rejecting limits on thinking "to the belief that there are no limits."[33] No limits—first on social media, and now on AI— parallel Lebow's unquestioned "necessity" to identify our self-worth with materialism, with what we consume, and that he used expediently to develop the advertising industry. By definition, greed knows no limits. In the 21st century, advertisements have been largely rebranded as "innovations," and always hailed as "without danger." The incentive to get to market at Lebow's "ever-increasing rate" of speed is now overwhelming. Curiosity is certainly a positive value, but one often and easily derailed by profit motives. We are told AI technologies have the "potential" to improve our lives, but there is little evidence that interactions among people in society have improved as a result of technological innovations. According to Haugen, "When we live in an information environment that is full of angry, hateful, polarizing content it erodes our civic trust, it erodes our faith in each other, it erodes our ability to want to care for each other."[34] Scott Pelley of CBS *60 Minutes* seems to agree: "Facebook essentially amplifies the worst of human nature."[35] And Surgeon General Vivek Murthy states that social media, AI's precursor, poses a "profound risk of harm to the mental health and well- being of young people" and possibly all of society.[36] Perhaps the media has always done so.

Freak shows sell tickets to a circus, and tabloids—a precursor to conspiracy theories–have long sold gossip and lies to attract idle minds But today, social media and AI technology spread what is angry, dishonest or grotesque with such immediacy that critical evaluation of sources, evidence, and reasoning is often and easily bypassed. In a helpful summary of "Inside ChatGPT: How AI chatbots work," writers JoElla Carman and Jasmine Cui warn that AI tools "shouldn't be relied on when accuracy is required" because "being correct isn't really the point of ChatGPT—it's more of a byproduct of its objective" which is "producing natural-sounding text."[37] I was curious to see an example from the above article asking for a one sentence summary of Jane Austen's novel *Pride and Prejudice*. The summary by Google's Bard indeed produced natural-sounding text and even referenced Wikipedia as its source. Then, to compare summaries, I looked up *Pride and Prejudice* in Wikipedia. Bard's response is *identical* to the entry in Wikipedia, which means Wikipedia was not used merely as a source, but had been copied verbatim. Therefore, I think Bard's entire response should have been in quotation marks and that it is more accurate to call the response "copying"—even plagiarism–rather than "intelligence."[38] A current lawsuit against Sam Altman's Open AI questions the right AI companies have to monetize content "scraped" from the internet in "direct competition" with other sources of that content. One of the plaintiff's lawyers, Joseph Saveri, states, "'Though the open-source licenses did not require the coders to get paid for the code, it required them to get credit for it…It's important for their career and for recognition.'"[39] Heather Tal Murphy astutely questions why Altman calls this pending lawsuit "frivolous" because doing so contradicts "the idea that Altman and his $27 billion company care passionately about respecting who does—and does not—want to help train A.I."[40] Altman says he does not "'want to see the industry force people to contribute their content to the training of A.I. tools: 'a minimum is that users should be able to, to sort of opt out from having their data used by companies like ours….'"[41] That's interesting, but seems backwards. I thought we've been told ChatGPT has *already* consumed "the sum of human knowledge," and there is no way to know if requests to "opt out" have been honored. Such "data feed" clearly fails to take into account those who do not even care to use social media, i.e., what is excluded from or outside of any purported "sum of human knowledge." Even searching the web has become more difficult now, as many sites require one to agree to policies (ten or so pages of legalese) before continuing to use. And precisely how does one "sort of opt out"?

A more fair and equitable approach than "to sort of opt out" would be that used in medical research. People who agree to their data being used by AI imperialists should be able to "**Opt In**," to sign up to participate in technological experiments, and for which participants then would be reimbursed. That, of course, slows down the process, but it also respects more so than coerces the "contributions" of other users' data. In January, 2023, the New York City Education Department **banned** ChatGPT from school devices and networks because although the tool provides "quick and easy answers to questions, it "does not build critical-thinking and problem-solving skills."[42] Shortly thereafter, Chancellor David Banks changed his mind and said he "didn't recognize the possibility of generative AI and its ability to support students and educators."[43] Banks now says educators will be provided with "resources and real-life examples of successful AI implementation in schools to improve administrative tasks, communication, and teaching."[44] School districts-almost always in need of more funding-may welcome technology whose "tools" are initially donated, along with advertising campaigns, to help with overloads of students. Some schools where I worked found that by purchasing the new technology some former booksellers are all too eager to sell, administrators could combine data entry and other administrative tasks with the duties of teachers. I have since spoken with nurses and doctors who also find large corporate hospitals and insurance institutions are now requiring that they write medical advice in "pre-authorized" forms that combine medical billing tasks with their diagnoses. In this case, technology's toolkits of resources seems to involve shifting of duties from one part of a business to combine those duties with another's work—without, however, the agreement of all parties. As with financial services, administrative services are flourishing. But teachers and other professionals use less and less of their own specialized learning since AI has more and more ready-made "knowledge." Others have pointed out that because data is collected from the past, it tends to have a regressive bias that does not accurately reflect the progress of social movements. In his recent testimony before Congress, Altman states his "belief that artificial intelligence has the potential to improve nearly every aspect of our lives. Then comes the flip side: it creates serious risks we have to work together to manage."[45]Altman wants to carry on unimpeded as he decides unilaterally what are "improvements" to "nearly every aspect of our lives," and then have all of us "work together to manage" the "serious risks."[46]

Again, that seems backwards: eliminating "serious risks" should precede, i.e., be the responsibility of Open A.I.'s owners, whose products are already marketed based merely on one man's speculations about improving nearly every aspect of our lives! AI machines "promise" to free us from repetitive tasks. Repetition does not sound like anything anyone wants too much of until we consider that it is critical to learning. Repetition is basic to early learning. It is used to develop and reinforce pattern recognition. Students who played a lot of tech games told class- mates how easily they could figure out each new game because they recognized parts of patterns reused from previous games. Also, mastery of a myriad of artistic crafts–wood-working, pottery, glass-blowing, painting—or music, or sports does not occur until one repeats the creative process or physical training many times over. Neuroplasticity after a stroke or other injury occurs primarily through *repetition* of tasks to rewire our mental faculties. .[47] Investment firm Goldman Sachs estimates a "loss or diminishment" of 300 million jobs worldwide due to AI, and concerns about destabilizing employment remain unresolved. [48] Pope Francis observes that human beings derive dignity from work, and we already know that in many parts of the world, i.e., Haiti, lack of work leads to increases in violence and gang activities, especially among young males. In a service economy such as we have in the U.S., more women will be affected because more women than men hold the less attractive jobs in service and hospitality. "There will be an impact on jobs," CEO Sam Altman admits to Congress. "We try to be very clear about that, and I think it'll require partnership between industry and government, but mostly …by government, to figure out how we want to mitigate that." [49] So Altman acknowledges it will be "mostly" up to the government to mitigate job losses. But as Naomi Klein points out, job losses, even tedious ones, that will be "mostly" up to the government to mitigate, means that "we're talking about socialism" that much sooner—or–what government programs are like in a new imperialism. [50] We need to put brakes on using AI for cruise-control thinking in order to slow down and back away from what is so obviously vying for our attention. We might consider what roles we play in a larger picture.

Students who participated in units requiring social service—that is, in some way working with other human beings- often rated those experiences their favorites and continued with the new Many are also trying to build new normals with families–ones that value friendship and respect to pass on to children. When the country focused more on work, Americans found ways to get along with and sometimes even learn about other human beings while accomplishing common projects. With today's shift towards mass entertainment, many insist on my way or the highway. There are fewer practical reasons for getting along with each other because we can find affirmation online for whatever we choose to fancy. And what we choose to fancy is bound to offend someone somewhere—who then responds with whatever they fancy, demonizing or cancelling what is "other" to one's own interests. Pope Francis urges young people to look away from their cell phone screens and make eye contact with the people around them: "In a world that tends to isolate us, divide us, and that pits us against each other … the secret is precisely to take care of others." [51] Apple CEO Tim Cook seems to agree on this point: "But for me, m-my simple rule is if I'm looking at the device more than I'm looking into someone's eyes, I'm doin' the wrong thing…"[52] Here, both a spiritual leader and tech giant seem to agree there is a need to look into the eyes of those around us—to register the intuitive and sensory feedback we get when we notice and then respond to other human beings with whom we are sharing this life. But when O'Donnell asks, "Is Facebook an amplifier for fake news?" Cook's response is less certain: "I don't really believe personally that there is– that A.I. has the power today to differentiate between what is fake and what is not. And so I worry about any property that today pushes news in a feed….we're– we're not creating news, but we– we do pick top stories, we have people doing it. And so I do worry about people thinking like machines. Not machines thinking like people.[53] Cook's worry focuses on those who oversee automated feeds that might amplify or reward—by being picked as "top stories," and those that Pelley sees amplifying "the worst in us." Cook believes AI lacks "the power to differentiate what is fake and what is not" and worries, rather, about people "thinking" like machines.[54] Ironically, however, it is precisely the "assistance" and "collaboration" of AI's automated curating function that may be conditioning people to think like machines. Machines can only "assist" by "training" users to act in compliance with the algorithms of data sets that have been predetermined by programmers to produce outcomes the programmers deem "desirable."

Furthermore, since we have often been sold on how "intuitive" technology is, AI may be training to "assist" us to *such an intuitive extent*, that we no longer have to even bother thinking. Autonomous intelligence becomes more superfluous so that we can go dream about what to buy tomorrow. These and other conversations have long needed to be taking place not just among elites talking with each other, but also among the whole of society, whose majority of citizens may want vastly different choices than do those expanding their imperial domains and consolidating power over the newly colonized. Recent research finds GPS devices might be weakening the brain's orienting ability.[55] When we delete the "thinking" from an activity, it follows we might also lose certain skills, as occurred years ago when calculators replaced the ability of some clerks to perform simple math or make change at a register other than the amounts a machine displays. A related concern is what happens when we do not question–when we lose the ability to think critically. Rather, we follow, as we are "assisted" by AI's "super-human" skills. We forfeit the power of human thinking to the very machines that lack the ability to differentiate what is fake and what is not. If we forfeit autonomy over our brains' orienting skills or situational awareness, we become prime targets for emergent conditioning by AI. but why do we want to outsource our own intelligence to machines programmed by others? AI technology robs people of their critical thinking so that, gradually, it is we who cede responsibility for the mental work of individual citizens to machines—mental work that is needed to think independently and thus to participate in society in meaningful ways. And if we lose our capacity for critical thinking, we lose whatever freedom we think we might have. That scenario seems opposite to what Steve Nouri, of Forbes Technology Council, says is needed: "Corporate executives need to ensure that human decision-making is strengthened by the A.I. technologies they use and are responsible for supporting scientific advancement and standards that can minimize A.I. bias."[56] **In addition, then, to issues regarding education, employment, and human choice, another—if not the predominant–concern with A.I. machines remains what Nouri calls the responsibility to minimize A.I. bias.** ChatGPT is the new face of this debate. As Hannah Getahun writes, "like many chatbots before it, it is also rife with bias."[57]

Even Open AI CEO Sam Altman admits to its In order for scientists training AI machines to "learn" large language models (LLMs) are trained on billions of words, enormous datasets "scraped off the Internet." Because the data has not been filtered or edited, output can result in expression of all the biases the AI models consume from the datasets fed to them. In turn, we feed on those same biases as we consume AI and then oftentimes thoughtlessly project them onto others. The latest model, ChatGPT4, has been trained not only on text but also images, the power of which often trumps rhetoric. [59] Sarah Brayne, Asst. Professor of Sociology at UC Austin, cautions against trying to solve sociological problems, such as racial equality issues, with technological solutions.[60] She uses the example of predictive policing for street crime, though not for white collar crime, to illustrate AI is not neutral. Rather, it *reflects the biases* of programmed algorithms. O'Reilly says Facebook began with the "wrong theory they were neutral platforms." More recently, he says it is "commercial speech that's attempting to deceive people."[61] In addition to biased data, AI researchers have their own biases and, to date, are a largely homogeneous field that decides the data to feed their models. Melinda French Gates wants more data collected on women: as she says, "I used to think the data was objective, but in fact, data is actually really sexist. … We like to think of data as being objective …but the answers we get are often shaped by the questions we ask. When those questions are biased, the data is too."[62] In 2023, a survey from the Anti-Defamation League shared with *USA TODAY* showed three quarters of Americans—a solid majority—now worry AI technology can cause "substantial harm," such as using AI "for criminal activity (84%), spreading false or misleading information (83%), radicalizing people to extremism (77%), and inciting hate and harassment (75%)."[63] As AI models infiltrate more and more aspects of our lives, "biased algorithms mean existing inequalities are perhaps being amplified—with dangerous results."[64] One student comments that the AI cannot fact-check itself, so its products may contain false information or made-up sources.[65] In that case, using the "friendly assistants" of AI chatbots may "help" us to rely more and more on *unreliable* narrators. Amplifying more bias and inequality in our society is hardly what we need. Many are now speaking out who call for regulation of the speed at which the latest innovations are being sold for public consumption because issues of underlying bias in AI chatbots have not been adequately tested, even when they have been shown to cause harm to others.

Gary Marcus, co-author of "Rebooting A.I." and host of the podcast "Humans vs. Machines," confirms that AI machines can cause "significant harm to the world" and is one of many who, in May, 2023, called for a six months "pause" on the information overload being sold to the American public as "innovations."[66] In what is surely a quintessential example of information overload, Tristan Harris tells NBC's anchor Lester Holt, "The CEO's of the big AI labs are saying they can't even keep up with the pace because people are inventing new improvements so much faster than their own understanding."[67] I am reminded of Ruff's conclusion that having too much information is the same as not having enough. Recall that Ruff warns bombarding the mind with information beyond the scale or pace of what the mind can process-even in *hyperactive* mode- preempts the wider social participation in discourse. This wider participation is essential to democracy and obstructs an ability to solve problems creatively–that is, ideas that derive from thinking *outside* the box. It also preempts deciding if "new improvements" are, in fact, really improvements or merely sophisticated propaganda to promote continued and stronger dependency on the technology, i.e., to maintain an opioid-like addiction to it. It preempts being able to assess if the yet unknown effects pose dangerous risks to individuals and social interactions that might outweigh the benefits that are forecast to "improve" our lives. Sean McGregor, founder of the Responsible A.I. Collaborative, sees bias as inevitable: "You can do your best to filter an instrument and make a better dataset, and you can improve that. But the problem is, it's still a reflection of the world we live in, and the world we live in is very biased and the data that is produced for these systems is also biased."[68] Yes, the data does reflect real world bias as well as the bias within ourselves, which is why Porter Braswell, founder of Diversity Explained, calls for greater accounting of the role human judgment plays. However, the ways that bias is "amplified" by AI are perhaps less "unaccounted" for than they are unexamined. Programming can reward bias to begin with by amplifying it rather than by recognizing and then limiting it. Klein contends that "a world of deep fakes, mimicry loops and worsening inequality is not an inevitability….It's a set of policy choices."[69]

It is worth noting that many of those who see bias as inevitable are men whereas many who see bias as a reflection of human choices are women—perhaps due to most programmers being men, but perhaps also due to a difference in how the different genders feel power is best used. For Timnit Gebru, a computer scientist and AI researcher who founded an institute focused on advancing ethical AI, the need for testing is obvious. She thinks it's part of being responsible: "There needs to be oversight…If you're going to put out a drug, you gotta go through all sorts of hoops to show us that you've done clinical trials, you know what the side effects are, you've done your due diligence."[70] Ethicists Abeba Birhane and Deborah Raji, writing for *Wired*, point out AI programmers are well aware of the harm their machines are capable of. They argue the models we see now are not inevitable; programmers are capable of making "different choices" to develop "entirely different models."[71] For example, "starting in 2017, Facebook's algorithm gave emoji reactions such as "angry" five times the weight as "like," which boosted those posts on its users' feeds."[72]Four years later, "when Facebook set the weight on the angry reaction to zero… users began to get less misinformation…less graphic violence."[73] Their work shows "unexpected" consequences can be more accurately anticipated and taken into consideration if one's priorities are people and society before power and profits. Programmers' awareness "of the harm their machines are capable of" seems clear when we read that "Microsoft shut down its A.I. machine Tay because "even if you take measures to mitigate bias…many algorithms are designed to continuously learn and thus are especially vulnerable to becoming biased."[74] Ethicist Timnit Gebru notes what happens when the "*most hostile digital environments*" i.e., "*fake news, hate speech, even death threats aren't moderated out*": "*They are then scraped as training data to build the next generation of LLMs. And those models, parroting back what they're trained on, end up regurgitating these toxic linguistic patterns on the internet.*[75] This toxic "stew" of hate speech and bias is fed to and consumed by—in order to train–the next generation of LLMs! Arthur Holland Michel, a Senior Fellow at the Carnegie Council for Ethics and International Affairs, remarks: "'Bias is a mathematical property of all AI systems,"[76]…but he agrees that it is "problematic to allow an algorithm to be used to generate divisive, hateful, untruthful content at a superhuman scale, with zero guardrails."[77]

Thus, he inadvertently offers a reason why more testing and accountability should be required before releasing AI machines to the public. So how do we make sure we do our best to filter, "curate," or otherwise make use of automated editing in order to make better datasets and produce different models? It is a question many in the field grapple with. The European Union's AI Act made an intelligent start simply by placing the responsibility where it belongs—that is, requiring "organizations to use fair training data and ensure their AI algorithms don't discriminate."[78] McGregor suggests, "Open AI's release of ChatGPT allows people to help make the 'guardrails' that filter biased data more robust."[79] Well, not exactly. Because when Open AI added "more robust" guardrails, users found getting around them "extremely easy" just by rephrasing questions—as Melinda French Gates noted—or even just asking the machine to ignore the guardrails![80] Braswell recognizes that AI learns from "existing data sets … designed by highly fallible humans."[81] Therefore, he sees a need to look more closely at how diversity should be included in the human judgments assessing the bias that already exists in programming data sets. Braswell explains that because machines are "necessarily contingent on the data fed to them" thinking that we can eliminate human error overlooks the "systemic ways in which bias is embedded in our societies and cultures."[82] It is a mistake to suppose that "AI is a super-powered fix to the woes of human error and subjectivity."[83] That thinking, Braswell explains, "allows the brands who serve us to *falsely market* their products as neutral, unbiased, and devoid of flawed human judgment."[84] "The most attractive features of AI are automation and reduction or elimination of human error."[85] However, Braswell uses IBM's Watson super-computer as a cautionary tale for the "technological hype and hubris around AI": recommendations for patients that the super-computer made for one part of the world were not applicable to those in other parts of the world.[86] Braswell's cautionary tale might parallel the alleged "super-human capacities of today's AI Large Language Models (LLMs). Researchers have found a "persistent pattern" in the "*linguistic deficiencies*" in terms of "*content moderation*." That is, "*Communities that speak language not prioritized by Silicon Valley suffer the most hostile digital environments*. [87] Perhaps AI recommendations for one part of the world are simply not applicable to people whose native language is one not prioritized in Silicon Valley*?*

Braswell believes the role of AI is to *enhance* human intelligence, not to *replace* it: "Because AI-powered predictions can only use the data that's out there already," we need to acknowledge the role our own intelligence and judgment play in using that data to "effect change in racial employment, wealth, and income gaps," i.e., the values on which we choose to focus.[88] Here, Braswell's thinking is compatible with the view of Birhane and Raji that programmers are indeed able to make *different* policy choices, depending on *whose and which values* they choose to focus. Braswell foresees a utopian A.I.: "As we progress into the next age of AI…we have to ask ourselves: who is responsible for the design of these systems? What data are they using…is it representative of the human beings they claim to serve? And if AI is only ever as good as its inputs, what are we doing to ensure those inputs reflect the future we want to live in, and not the past we need to leave behind?"[89] If human judgments reflect the values on which we choose to focus, whose utopia is it that we should assume speaks for a future "as society would like it to be? Particularly given Braswell's cautionary tale about IBM's Watson super-computer that the values of one imperialist world may not be as relevant or desirable to communities or societies that speak different languages as are those values to one's own. When programmers make recommendations to industry, government, educational institutions, insurance and health systems, etc. even what is relevant for one individual may not be relevant for another individual—which again recalls IBM Watson's cautionary tale, albeit at an individual or microscopic level. There is no one-size-fits-all, which does not mean we cannot test the outcomes using the values we choose to focus on to see which ones the largest number of citizens agree on before releasing into society even more recklessly-biased technology. Matthew Gombolay, an assistant professor of Interactive Computing at Georgia Tech, has worked on the CLIP robot that "gained widespread interest for the large scale of its dataset, despite jarring evidence that the data resulted in discriminatory imagery and text descriptions."[90] Gombolay agrees AI and algorithms are not neutral and cautions: "Chatbots like ChatGPT weren't created to reflect back our own values, or even the truth…They're 'literally being trained to fool humans'…To fool you into thinking it's alive, and that whatever it has to say should be taken seriously."[91]

Furthermore, Gary Marcus warns against trusting the branding of AI as innovations that will improve our lives: "We all more or less agree on the values we would like for our AI systems to honor. We want, for example, for our systems to be transparent, to protect our privacy, to be free of bias, and above all else to be *safe*. …But current systems are **not** in line with even these values. Current systems are *not* transparent, they do *not* adequately protect our privacy, and they continue to perpetuate *bias*. Even their makers don't entirely understand how they work."[92] **In the Mayo Clinic's RISE for Equity podcast, host Lee Hawkins interviews Maria Hightower, M.D. M.B.A., MPH,** Chief Digital Technology Officer of the University of Chicago Medicine, and Ivor Horn, M.D. MPH Director, Health Equity & Social Determinants of Health at Google, on the topic "Is AI Biased? How Do We Fix It?"[93] Both guests offer insight into making more empathic decisions—akin to Braswell's "enhanced" decisions–by *integrating* 1)more diversity into our initial judgments and 2)more real life, human experiences used in conjunction with the medical data that AI provides.[94] Hightower points out that the data A.I. is trained on is akin to the narratives in society that very much reflect "power and privilege," whereas the stories of some of the most marginalized populations "are not as well captured into our zeros and ones that convert a lived experience into data."[95] That awareness, she says, can be used to inform the "different judgment calls within the machine learning process where bias can be…either mitigated or expanded."[96] She gives the example of an A.I. model used to predict no shows to medical visits. Health systems often double book when data shows someone is at "high risk" for being a no show. But Hightower states that health officials also have an option to use critical thinking–to figure out why "somebody is at high risk for no show…to understand that root cause," for example, be it lack of transportation or lack of child care; then to "try to help alleviate that barrier to access." That, she says, is a "very human decision."[97] Ivor Horn contends that programmers often have no context of "putting the data in a real world understanding of diverse populations because for the most part…most of them have had a really privileged experience."[98] Like Braswell, Horn wants more diverse voices with lived experiences at the table to enhance the thinking of programmers and health equity experts as a team rather than relying only on AI data to replace human judgment. He states, "You may not be an engineer, but you may have the lived experience that that team needs to hear and understand about that product so that the people that you care about can be seen in the products that you're helping to develop."[99]

This is a good goal, especially if—and hopefully–the people we care about are all of humanity. We need to "progress" at a pace that allows diverse voices to weigh in on decisions regarding AI "advancements" and to hear and respond to the many human concerns involved so that, hopefully, we make decisions for the common good with the greatest compassion of which we are capable. The two professionals' reflections are valuable because both call for transparent algorithms and sharing input from lived experiences in order to view the mathematical formulas within a more humanized context. Highwater's concern for making human decisions by asking why a patient does not show for an appointment introduces empathy into the medical community's decision-making. Horn's concern for transparency allows for multiple and different understandings of the decisions that go into programming algorithms. These are ways that might mitigate bias by allowing for more compassionate human judgments. Except … compassionate human judgements in the above scenarios are–not impossible—but very much unlikely to be widely implemented for two reasons: First, transparent coding cedes power to others. It is not that individuals are incapable of doing that. But humans scaling the heights of virtual imperialisms do not have the best track record for making decisions to restrain their power over others in order to empower others. Like greed, hubris, too, knows no bounds. Second, and as Braswell points out, a major reason AI is welcomed by industry is that its technology saves time for businesses because as time to perform tasks is reduced, productivity increases. Therefore, following up on automated decisions in the way Hightower and Horn describe takes not only empathy but also a much greater amount of time to make those "very human decisions" with greater empathy,[100]to care about the impact on humanity of technology-based decisions. The attempts by Hightower and Horn to "fix" the AI bias of machine learning are admirable. They reintroduce our humanity, which is precisely what the mathematical algorithms delete. However, their improvements run counter to a supposed major benefit of AI—the increased productivity that leads to greater profits.

Furthermore, will AI's next age of automation be able to do what it is not doing now—that is, give more attention earlier to diverse human judgments, and the effects of cutting time–that is, before those judgments succumb to the lure of the marketplace? Will people have a choice to Opt In before their data is usurped for experimentation and training of AI? Will people on low rungs in the social hierarchies of virtual imperialisms have choices other than to follow decisions made by multinational corporations that have "friendly" AI machines to "assist" them with what has been programmed? What happens when data is misused—i.e., exploiting migrants and workers in Third World countries with disinformation so that we might avoid similar and grossly unjust work conditions here? The question of values is once again critical, and members of society do not necessarily agree on these and other values. For example, deciding "what constitutes hate speech and toxic politics is now being done by Kenyan laborers making less than $2 an hour. These workers were hired to screen tens of thousands of text samples from the Internet and label it [sic] for sexist, racist, violent or pornographic content."[101] Researcher Vinay Prabhu says the violent imagery he saw of rape and sexual assault while working with an image-text model made him physically ill. [102] Laborers in Kenya, interviewed by Leslie Stahl, reported being traumatized by training AI to recognize pornography and excessive hate speech.[103]A young man reported no longer being able to enjoy sex with his wife. A woman, refers to the $2/hr. Meta and Open AI pay them (via a middle man, Sama) as "exploitation" and "modern day slavery."[104] The choice to hire Third World laborers points to an "overlooked" judgment about the labor practices that build and sustain virtual imperialisms—such as colonizing peoples in other cultures with jobs shown to make others literally sick, even as "Open AI is reportedly close to reaching a **$29 billion valuation** (including a $10 billion investment **from Microsoft**)."[105] Here, decisions seem to have been already made, and human judgment seems to be "out of sight, out of mind." It is a similar version of the argument multinational corporations used when supply chains moved to China and may be grossly unfair to other human beings who need work to survive. There have been reports of "forced labor from thousands of Uyghurs that the Chinese government…displaced from their homes in Xinjiang."[106]

It would seem those who feel entitled to power and privilege are responsible to ensure effective oversight of their own policies, even in —or maybe especially in-foreign lands. Perhaps the most significant problem with the speed of developing untested AI and that remains unresolved is the erosion of trust in society wherein conflicts between 1)commercial racing to restructure society into virtual empires and 2)assumptions by others that they lived in something resembling a democracy have landed center stage. Until these critical human concerns are integrated into discussion of AI, its further development should be on hold. Imperialists have introduced the information overload of AI machines to expand and colonize their commercial domains. The imperialists are aware AI's emergent properties are increasingly vulnerable to "learning" biases that accumulate from feedback loops on data that parrot back, reinforcing bias with the feedback loop's toxic stew of bias….Where, then, is their *accountability* for those "unexpected" results of using tools–perhaps offered for free–but which society will take for granted, and become addicted, opioid-like, to using? We are warned that in 2024, given new AI-generated photos, videos, cloned voices, etc., we will not be able to tell the difference between what is real and what is fake. Marcus predicts, "We're going to enter an era when no one trusts anything…We don't really know the scope of it."[107] In remarks delivered May 16, 2023, to the U.S. Senate Judiciary Subcommittee on Privacy, Technology, and the Law, Marcus again states, "There are benefits; we don't yet know whether they will outweigh the risks….Fundamentally, these new systems…can and will create persuasive lies at a scale humanity has never seen before."[108] The concerns voiced here seem akin to Tim Cook's concern about people ceding their more humanitarian and critical thinking to regurgitating the "thinking" of super-human machines. "Large language models will have the ability to mimic you. It mimics humanity. And when that happens, it also doesn't tell you it lies. …with large language models, with ChatGPT…all the different ones that are rolling out, it looks set to weaponize our loneliness."[109] So why do we allow ourselves to move in the direction of more deep fakes and embrace possibilities for "persuasive lies" that yield even greater deception? Klein writes, "It would be awfully nice if AI really could sever the link between corporate money and reckless policy making—but that link has everything to do with why companies like Google and Microsoft have been allowed to release their chatbots to the public despite the avalanche of warnings and known risks."[110] "Last year, the top tech companies spent a record $70 million to lobby Washington —that is, more than does the oil and gas sector."[111] But do we really need hyperactive minds, i.e., brains akin to the Industrial Revolution's steam engines? I think not.

There are plenty of highly intelligent people—and most are even ethical. Nor do we need to pursue more corporate value propositions "customers may not even know they need yet."[112] Rather, we need AI systems that align with and honor democratic values "we all more or less agree on."[113] We need trust in government officials to enforce those values for the good of the majority of citizens. But that requires a Congress—and an administration–working for the collective good. Nor are we in need of more data or troves of datasets to address these issues. "When we are mistrustful of everything we read and see in our increasingly uncanny media environment, we become even less equipped to solve pressing collective problems," such as climate change," immigration, homelessness, etc.[114] Spitting out more mega-data from super-human machines will not resolve these issues, just as knowing there is already an excess of mass shootings in America this year alone has not resolved gun violence. Rather, the persistent will of different communities working together to help each other will better our world—just as we see many parents of children who have died from gun violence or opioid overdoses refusing to let us forget their children. But if we recognize that the answer is not simply more piles of data, then we should be extremely wary about handing over critical thinking to friendly AI "assistants" and about the mistrust that Klein warns will destroy our ability to solve collective problems, i.e., to work with each other. O'Reilly repeats to Walter Isaacson the by now familiar acknowledgement that too much information is the same as not having enough: "We are designing incredibly complex systems today that we don't really understand. And that's the real fear of AI. It's not of the rogue AI that's independent of us. …we're building these hybrid machines of human and machine that are incredibly complex that we don't really understand."[115] Ironically, though, and almost immediately after repeating "we don't really understand [the]…incredibly complex systems being built today," O'Reilly elaborates on a metaphor that is meant to justify his desire, nevertheless, to go ahead anyway–with what he calls Web 2.0.

That is, he advocates continued use of the very hybrid machines he concedes we don't really understand: "We're trying to figure out how you weave billions of people into this dynamic system and we have not figured out the equivalent of aeronautics yet."[116] Are our corporations addicted to their own power? The blueprint seems to be "Let's just wing it," regardless of who or what we damage and/or destroy along the way. So when did "billions of people" sign up to be guinea pigs experimented on by flying along with someone else's "recipe for success," particularly given the amount of recent troubles with airlines and with construction of the airplanes themselves? This is a recipe that neither Congress, nor the Supreme Court, nor even those in multinational corporate tech industries fully understand.[117] According to Haugen, Facebook's mission was "to connect people all around the world." Instead, it "set up a system of incentives that is pulling people apart."[118] Now consider "Open AI's original mission statement that proclaimed, "Our goal is to advance [AI] in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return."[119] A mere seven years later, "humanity has taken a back seat."[120] A.I. experiments with total disregard for the members of the societies it experiments on. It is as though chatbots use their natural-sounding voices to get us on board; then, once aboard, tell us…oh, and by the way…the pilots don't really *know how to fly*! Here again, the process for testing to determine risks and benefits or "innovative" AI should be the choice of individuals to OPT IN to studies and to be compensated for their time and risk, etc., rather than being able to "sort of Opt out" and rather than failing even to be consulted, i.e., having no say in the use of their private property/data. After all, we—some of the newly colonized–might resist other humans being treated only as objects to be mined for their data as well as further changes to both the workplace and social order in which we all live. I want choices to interact with human beings rather than with more machine-automated "speech." Matthew Gombolay refers to a common responsibility to care about the species: "We should all be concerned about the potential of AI biases to cause real-world harm: 'If you are a human, you should care.'"[121]

Brayne, the scholar at UC Austin who cautions against trying so solve sociological problems with AI warns, "Opaque and proprietary algorithms are not the same as open and public laws."[122] Algorithms replace the greater good by majority rule with corporate values determined by unelected individuals who often place profits first. An Open Letter signed by 31,810 AI researchers and developers on March 22, 2023, called "on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4, the newest model released by Open AI in March…Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable."[123] The letter, issued by the Future of Life Institute, continues, "AI labs and independent experts should use this pause to jointly develop and implement a set of shared safety protocols for advanced AI design and development that are rigorously audited and overseen by independent outside experts." [124] In a sense, the letter was disingenuous because Open AI was perhaps *already* too far along to stop. It provides evidence, however, that those involved in developing AI technology were *already aware* of the socially responsible precautions they might have taken before releasing their "innovations" to society. Marcus points out the intuitive awareness for the thousands working in this field: "If you want to release something to 100 million people, then you should do a safety analysis, and there should be someone outside of your company that evaluates that and make sure it's OK."[125] Those precautions for the good of society have been willfully disregarded, just like the many who turned a blind eye on those who have suffered and/or now suffer from opioid addictions. "Lawmakers in Europe recently signed "the world's first set of comprehensive rules of artificial intelligence….One of the EU's main goals is to guard against any AI threats to health and safety and protect fundamental rights and values."[126] Sam Altman, CEO of OpenAI's ChatGPT, has voiced support for some guardrails on AI and signed on with other tech executives to a warning about the risks it poses to humankind. But, just as with Brad Smith telling Leslie Stahl in the *60 Minutes* interview, it's a mistake to regulate "right now," so too, Altman is quick to qualify what he says: "It's 'a mistake to go put heavy regulation on the field right now.'"[127] Shortly after the Future of Life Institute's letter and as first reported by the *New York Times*, "Godfather of AI" Geoffrey Hinton, resigned his position as leader of Google's AI research division.

He says he now wants to help others understand why AI development should be put on hold and to warn of its exploitation by bad actors.[128] Hinton voiced concern about an "existential risk": "'The kind of intelligence we're developing is very different from the intelligence we have…it's as if you had 10,000 people and whenever one person learned something, everybody automatically knew it…that's how chatbots can know so much more than any one person.'"[129] Subsequently, Hinton was awarded the Nobel Peace Prize. Instantaneous transmission by chatbots to 10,000 people sounds to me like some kind of herd mentality anticipated by Aldous Huxley. Perhaps the "emergent" property of chatbots that vie for our attention is akin to predatory stalking of autonomous thinking? At a Congressional hearing, Sam Altman and senators expressed fears AI could "go quite wrong."[130] "'One-on-one interactive disinformation,'" is one of Altman's greatest concerns, and he says regulation on the topic would be 'quite wise.'"[131] It is not rocket science that what leading imperialists in the industry call an "existential risk" and a "greatest concern" warrant more testing by people who do fully understand the technology before it is released to the public at In their mad race for more power and money, have the imperialists stumbled across a bit of wisdom…only…as an… afterthought? Is it only *now* that they are suggesting regulation would be "quite wise" because things can go "quite wrong"? In a society that claims to be of, for, and by the people, the untested "innovations" of AI automation need to be balanced with protection of individual choices and civil rights. Wisdom and ethics are not values easily tacked on as afterthoughts. After all, our master algorithm can be "quite greedy"–respecting no limits and supplanting even freedoms for the newly colonized. The Supreme Court has ceded its responsibility, Congress seems unable to take responsibility for what helps to fund its reelections and stock gains and for what they do not adequately understand, and the virtual empires have refused and do not want responsibility for what it is they "sort of" understand. Matt O'Brien writes, "Pressed on his own worst fear about AI, Altman mostly avoided specifics, except to say that the industry could cause "significant harm to the world" and that "if this technology goes wrong, it can go quite wrong."[132] One of the founders of AI-focused law firm BNH.ai, Andrew Burt, writes, "Without thoughtful, mature evaluation and management of these systems' harms, we risk deploying a technology before we understand how to stop it from causing damage."[133]

ADL CEO Jonathan Greenblatt states, "If we've learned anything from other new technologies, we must protect against the potential risk for extreme harm from generative AI before it's too late."[134] Braswell maintains a qualified optimism that AI "*will* help us create a better future…but only if humans are there to analyze its outputs and help shape the direction in which they guide us."[135] I agree with Braswell's accounting for the role human judgment plays in designing and analyzing the outputs of AI, but disagree that is should be the very machine thinking of AI data sets "designed by highly fallible humans" that should "guide us"- particularly given that Braswell himself warns "it is a mistake to suppose that "AI is a super-powered fix to the woes of human error and subjectivity."[136] Then too, there is Tim Cook's concern about people thinking like machines. Human subjectivity—even superhuman subjectivity–remains "highly fallible" and needs to be guided, rather, by conscience, that antiquated and inherent human tool by which individuals make ethical judgments in the here and now, and one's individual's conscience cannot be spoken for by another. Gombolay says, "It is, let's not forget, a robot. Whether it thinks Hitler was right or that drag queens shouldn't be reading books to children is inconsequential. Whether you *agree* is what matters, ultimately."[137] Gombolay emphasizes the human activity of reading here, of interpreting the outcome of algorithms according to our individual world views—hopefully guided by one's conscience–that ultimately gives assent or dissent to subjective and highly fallible human judgments that also play a role in programming A.I. What a programmer calls "1" and/or "0," as well as the ways they convert lived human experiences into data, presumably aligns with the consciences of the individual programmers. Bias is inherent in the values of each programmer's assignment of 1s and 0s—essentially the degree of giving weight to or subtracting weight from–options that are used to construct algorithms. This automates what we see on the internet and hear from those "helpful" chatbots.

Do we want more of what automates human judgment? Exhortations to care for others by people as diverse as the Pope, the CEO of Apple, AI researcher Matthew Gombolay, and medical researchers at the Mayo Clinic are more critical than ever because with increasing use of AI, human beings have the only capacity to care about social changes that affect all of us. Reading data—first, when programming it; second, when making use of it–are the only times when intelligence is sourced with conscience, that conscience that is needed to guide the fallible and subjective human judgments with as much empathy as possible in order to filter out as much bias as possible. We hear of more and more people stepping forward today because adjusting the values that the mirror of algorithms reflect back to us seem always to lose out to the overpowering thrust of human greed for money and power. Algorithms are indifferent to human responses because the mathematical calculations of AI removes human empathy from decision-making. They are completely antithetical to that need to care for one another because machines cannot discern what is false and what is true. They lack conscience. In the science fiction film *Gattaca*, the ruling social system relies on eugenics to deny an individual deemed "unqualified" a chance to exercise free will via his dream of going to outer space. It is only another human being who makes a stealth exception in order to allow the man to escape.[138] What AI cannot account for is the *unpredictability* of human nature because AI "learns" by massive repetition of known trial and error. Thus, AI dictates the boundaries in any given situation, limiting them to replaying what is the already known–to the predictable past—a terrible infringement on human freedom to choose otherwise. In that "sum of human knowledge" that Google claims Bard and presumably other AI chatbots have been fed, we find all *known* biases that warp humanity, i.e., war, terrorism, slavery, racism, destruction of families, child abuse, ransomware, exploitation of workers, sextortion, human trafficking, deception, betrayal, holocausts, torture, crimes against humanity And yet, recall what Raskin further remarks: "What's very surprising about these new technologies is that they have emergent capabilities that nobody asked for.[139] One of the biggest problems with AI right now…it hallucinates…it speaks very confidently about any topic and it's not clear when it is getting it right and when it is getting it wrong."[140]

The emergent capabilities of new AI technologies–their grasping at a future, may be mirroring humanity's own reckless pursuits. Recall an "innovative" property of AI machine training is the accumulation of its "learning" with *feedback loops* that reinforce humanity's *toxic stew of hate and biases* to be consumed by the next generation LLMs. Hallucinations–speaking in "error with confidence" or as unreliable narrators rather than evaluating if it's getting something right or wrong—may, however, just be machines "thinking like machines."[141] Data gathered from real world experience is often augmented by the imaginations of science fiction writers who likewise anticipate future scenarios. Science fiction writers also cannot be certain of the scenarios they depict, i.e., whether they get imagined futures right or get them wrong. Might science fiction be an artistic precursor to AI machines? Suppose that AI's hallucinations or emergent utterances are akin to our as-of-yet unimagined biases—buried deep in that toxic stew of both known and unknown biases,…lurking…and perhaps soon emergent due to accumulating and continuously *reinforcing a feedback loop* that regurgitates "*toxic linguistic patterns*" for consumption by the next generation of LLMs, then dumped into AI's amorphous "sum of human 'knowledge'"—all purportedly— to "assist" human beings. Recently, Jerry Seinfeld told Jimmy Fallon that we've already done this AI stuff, and it didn't work with Frankenstein.[142] And…do we really need more "assistance" to unearth yet UN-imagined toxic biases?