



Why AI Ethics is Unethical

MHeNs Annual Research Day
Maastricht
19 March 2025

Dr. Nolen Gertz
Associate Professor
of Applied Philosophy
University of Twente

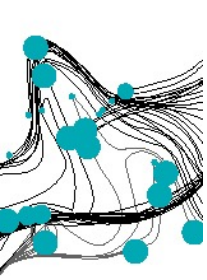
A.I. TURNS THIS SINGLE
BULLET POINT INTO A
ETHICAL THEORY I CAN
PRETEND I WROTE.



A.I. MAKES A SINGLE
BULLET POINT OUT OF
THIS **ETHICAL THEORY** I CAN
PRETEND I READ.



TOM
FISH
BURNÉ



Part I

What *isn't* AI Ethics?



How many people
do you need to steal
ideas from before
we stop calling it
plagiarism and start
calling it learning?

Good question!
Let's ask chatGPT.



freshspectrum

About

OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity.



ChatGPT is coming for your job. Why that's a good thing



by **Matt Asay** in **Artificial Intelligence**
on April 26, 2023, 3:57 PM EDT

Commentary: Though a new report claims artificial intelligence could replace some jobs, the reality is that people in these roles and others can use AI to do their work much more effectively than before.



Image: gguy/Adobe Stock

Even great writing is a bit derivative. Steinbeck's "East of Eden" is a retelling of the biblical Cain and Abel story, for example. But anyone who thinks ChatGPT could come up with that masterpiece of creative writing is way too high on their LLM paint thinners. Great writing emerges from human genius, articulating common themes in uncommon ways. The day I see that come from a prompt I drop into ChatGPT will be the day it's all over for the human race, but guess what? That day isn't coming.

Not now. Not soon. Not ever. Machines, as with the development examples above, are good at incorporating human-created input and mimicking it to generate human-acceptable output. But they're not ever thinking through the all-too-human experience that gives rise to great literature, just as they're not able to grok and respond to the business problems that great developers resolve with code.

Instead, we have a happy union of people and machines. How happy that union will be for given industries and the people therein depends on how well they use GPTs to remove repetitive tasks or code so that they can focus on the innovative, human side of their jobs.

Op-Ed: Don't ban chatbots in classrooms — use them to change how we teach

New York City's Department of Education recently [banned](#) the use of ChatGPT, a bot created by OpenAI with a technology called the Generative Pretrained Transformer.

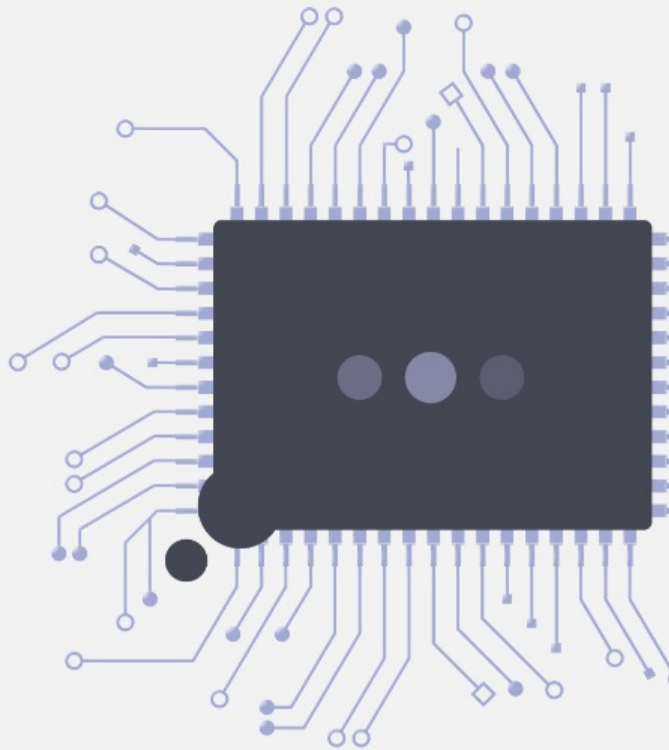
“While the tool may be able to provide quick and easy answers to questions,” says the official statement, “it does not build critical-thinking and problem-solving skills, which are essential for academic and lifelong success.”

We disagree; it can and should.

Banning such use of artificial intelligence from the classroom is an understandable but nearsighted response. Instead, we must find a way forward in which such technologies complement, rather than substitute for, student thinking. One day soon, GPT and similar AI models could be to essay writing what calculators are to calculus.

We know that GPT is the ultimate cheating tool: It can write fluent essays for any prompt, write [computer code](#) from English descriptions, [prove math theorems](#) and correctly answer many questions on [law](#) and [medical](#) exams.

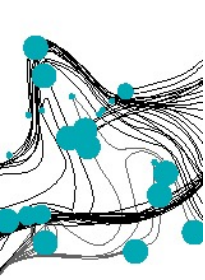
Banning ChatGPT is like prohibiting students from using Wikipedia or spell-checkers. Even if it were the “right” thing to do in principle, it is impossible in practice. Students will find ways around the ban, which of course will necessitate a further defensive response from teachers and administrators, and so on. It's hard to believe that an escalating arms race between digitally fluent teenagers and their educators will end in a decisive victory for the latter.



(Jim Cooke / Los Angeles Times)

BY ANGELA DUCKWORTH AND LYLE UNGAR

JAN. 19, 2023 3:20 AM PT





Traditional ethical theories

- **Virtue Ethics (Aristotle):**
Who should **I become**?
- **Deontological Ethics (Kant):**
Which action is
right in principle?
- **Consequentialist Ethics (Mill):**
Which action produces
the best outcome?





Traditional ethical theories & ChatGPT

- 
- **Virtue Ethics (Aristotle):**
Who does ChatGPT help us **become**?
 - **Deontological Ethics (Kant):**
Is using ChatGPT
right in principle?
 - **Consequentialist Ethics (Mill):**
Does using ChatGPT produce
the best outcome?
- 



Exclusive: In its quest to make ChatGPT less toxic, OpenAI used outsourced Kenyan laborers earning less than \$2 per hour



Topic Rooms Via



Class Action Filed Against Stability AI, Midjourney, and DeviantArt for DMCA Violations

AI image products which was trained on billions of copyrighted images contained in the LAION-5B dataset, which were downloaded and used without compensation or consent from the artists

A ChatGPT defamation lawsuit could be the first of many cases that will examine where legal liability falls when AI chatbots spew falsehoods.

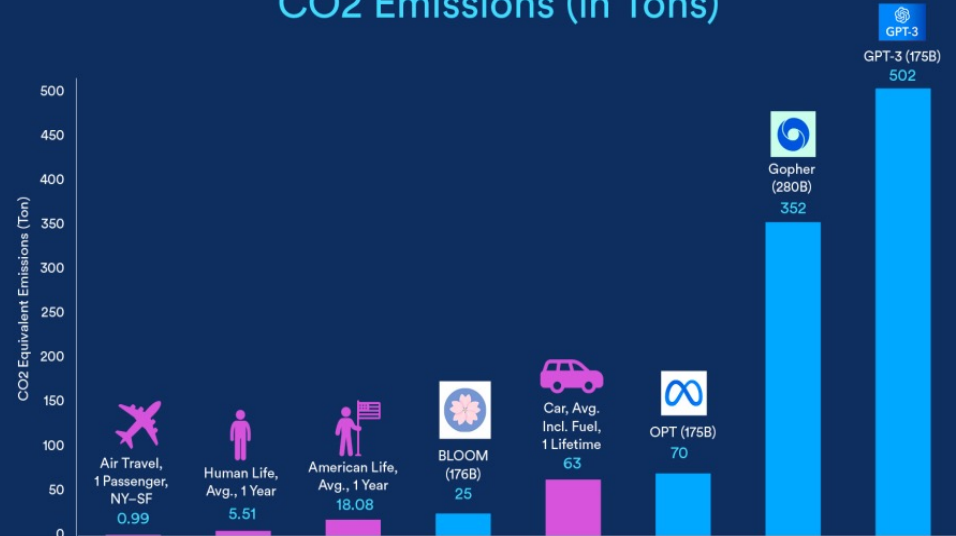


news.bloomberglaw.com

First ChatGPT Defamation Lawsuit to Test AI's Legal Liability

A defamation lawsuit filed against the artificial intelligence company OpenAI LLC will provide the first foray into the largely untested legal ...

CO2 Emissions (in Tons)



Exclusive: In its quest to make ChatGPT less toxic, OpenAI used outsourced Kenyan laborers earning less than



TopicRooms Via

A ChatGPT defamation lawsuit could be the first of many cases that will examine where legal liability falls when AI chatbots spew falsehoods.

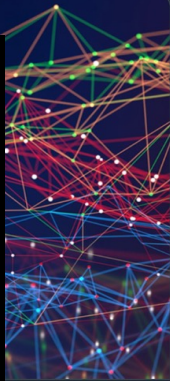
Aristotle: ChatGPT is making us **more vicious** and **less virtuous**

Kant: ChatGPT **violates moral imperatives** to never lie, never steal, and never treat humans as mere means to our ends

Mill: The benefits of ChatGPT are **outweighed by the harms**

Class
Midj

AI image products which was trained on billions of copyrighted images contained in the LAION-5B dataset, which were downloaded and used without compensation or consent from the artists



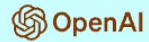
ity
e company
sted legal ...

GPT-3

GPT-3 (175B)
502



ChatGPT is coming for your job. Why that's a good thing



[Research ▾](#) [Product ▾](#) [Developers ▾](#) [Safety](#) [Company ▾](#)

[Search](#) [Log in ↗](#) [Sign up ↗](#)

About

OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity.



OPINION

Op-Ed: Don't ban chatbots in classrooms — use them to change how we teach



**Vicious not
virtuous!**



**Maxim of action
cannot be
universalized!**



**Harms
outweigh the
benefits!**



OPINION

Op-Ed: Don't ban chatbots in classrooms — use them to change how we teach



Part II

What *is* AI Ethics?



ITU Events

AI for Good
Global Summit

Accelerating the United Nations Sustainable Development Goals

Register now!

30-31 May 2024
Geneva and online

aiforgood.itu.int

40 UN PARTNERS

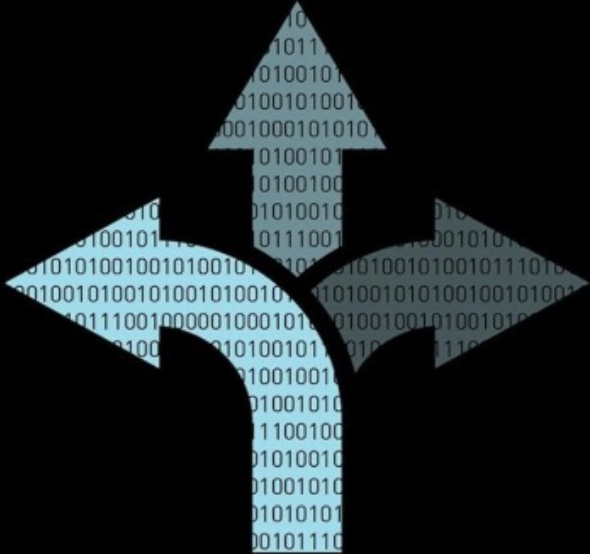
ITU

The banner features a large, colorful, stylized 'A' shape. On the right, a person wearing VR goggles is shown, with the goggles displaying various icons. The background is a grid of colorful squares, each containing a different icon. At the bottom right, a person in a yellow suit is standing on a stage with a large screen behind them.

AI ETHICS

MARK COECKELBERGH

AI is an existential threat...



This story need not have a happy ending: for Bostrom, Tegmark, and others, “existential risks” are attached to superintelligence. The result of these developments may be that a superintelligent AI takes over and threatens human intelligent life. Whether such an entity would be conscious or not, and more generally whatever its status or how it comes into being, the worry here is about what the entity would do (or not do). The AI may not care about our human goals. Having no biological body, it would not even understand human suffering. Bostrom offers the thought experiment of an AI that is given the goal of maximizing the manufacture of paperclips, which it does by converting the Earth and the humans who live on it into resources for producing paperclips. The challenge for us today, then, is to make sure that we build AI that somehow does not raise this control problem—that it does what we want and takes into consideration our rights. For example, should we somehow limit the AI’s capabilities? How are we to contain AI?¹

...so we need to control it!

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION



THE ASSESSMENT LIST FOR
TRUSTWORTHY ARTIFICIAL
INTELLIGENCE (ALTAI)
for self assessment

AI could harm human agency,
lead to deskilling, and worsen
well-being...

REQUIREMENT #6 Societal and Environmental Well-being

In line with the principles of fairness and prevention of harm, the broader society, other sentient beings and the environment should be considered as stakeholders throughout the AI system's life cycle. Ubiquitous exposure to social AI systems in all areas of our lives (be it in education, work, care or entertainment) may alter our conception of social agency, or negatively impact our social relationships and attachment. While AI systems can be used to enhance social skills, they can equally contribute to their deterioration. This could equally affect peoples' physical and mental well-being. The effects of AI systems must therefore be carefully monitored and considered. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, for instance the Sustainable Development Goals.³² Overall, AI should be used to benefit all human beings, including future generations. AI systems should serve to maintain and foster democratic processes and respect the plurality of values and life choices of individuals. AI systems must not undermine democratic processes, human deliberation or democratic voting systems or pose a systemic threat to society at large.

...so we need to make sure
AI is a benefit to humanity!

ISSUES NOTE
21 September 2024



Artificial intelligence (AI) end-to-end:

The environmental impact of the full AI life cycle needs to be comprehensively assessed



The United Nations Environment Programme (UNEP) is the leading global environmental authority that sets the global environmental agenda, promotes the effective implementation of sustainable development within the United Nations system, and serves as an authoritative advocate for the global environment. It is mandated to keep under review the world environmental situation.

Against this mandate, UNEP has been requested by UN Member States to consider the environmental dimensions of digital technologies, assessing their opportunities to enable environmental sustainability and the impact they can have on the environment.

This note outlines key areas identified by UNEP regarding the environmental impact of Artificial Intelligence (AI) across its life cycle. The note aims to inform Member States, civil society, the private sector, and the public, while encouraging the research community to develop and use scientific methods to objectively measure AI's environmental footprint.

¹ This Issues Note provides a review of the latest literature on specific topics that are of relevance to UNEP's mandate. It also presents a set of agreed approaches and recommendations regarding UNEP's communication of the subject matter.

² The Ministerial Declaration adopted at the sixth session of the United Nations Environment Assembly (UNEA-6) emphasized the importance of leveraging emerging technologies and closely monitoring their development to ensure they contribute to sustainability. The declaration also stressed the need to ensure that digitalization is inclusive, equitable and sustainable.

AI is destroying the planet...

There are high hopes that artificial intelligence (AI) can help tackle some of the world's biggest environmental emergencies. Among other things, the technology is already being used to [map the destructive dredging](#) of sand and [chart emissions](#) of methane, a potent greenhouse gas.

But when it comes to the environment, there is a negative side to the explosion of AI and its associated infrastructure, according to a growing body of research. The proliferating data centres that house AI servers produce [electronic waste](#). They are [large consumers of water](#), which is becoming scarce in many places. They rely on critical minerals and rare elements, which are often mined unsustainably. And they [use massive amounts of electricity](#), spurring the [emission of planet-warming greenhouse gases](#).

How can the world rein in the environmental fallout from AI?

In the new issue note, UNEP recommends five main things. Firstly, countries can establish [standardized procedures for measuring the environmental impact of AI](#); right now, there's a dearth of reliable information on the subject. Secondly, with support from UNEP, governments can develop regulations that require companies to disclose the direct environmental consequences of AI-based products and services. Thirdly, tech companies can make AI algorithms more efficient, reducing their demand for energy, while recycling water and reusing components where feasible. Fourthly, countries can encourage companies to green their data centres, including by using renewable energy and offsetting their carbon emissions. Finally, [countries can weave their AI-related policies into their broader environmental regulations](#).

...so we need to *limit* the destructiveness of AI!

The Ethics of Artificial Intelligence

Principles, Challenges, and Opportunities

LUCIANO FLORIDI



OXFORD

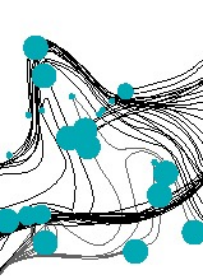
Education, business, and industry; travelling and logistics; banking, retailing, and shopping; entertainment; welfare and healthcare; politics and social relations—in short, life itself as we know it today—have all become inconceivable without the presence of digital technologies, services, products, and practices. Anyone who is not perplexed by such a digital revolution has not grasped its magnitude. We are talking about a new chapter in human history. Of course, many other chapters have come before. They were all similarly significant. Humanity experienced a world before the digital revolution, a world where the steam engine, electricity, and the printing press were the dominant forces of transformation. This transformation was unique. Some irreversibly changed our self-understanding, our reality, and our experience of it, carrying complex and long-term implications. For instance, we are still finding new ways to exploit the wheel (just think of the iPod click wheel). By the same token, what humanity will achieve thanks to digital technologies is unimaginable. As I stress in Chapter 1, nobody in 1964 could have guessed what the world would have been like only fifty years later. Futurologists are the new astrologers; we should not rely on them. And yet, it is also true that the digital revolution will happen only once, and it is happening now. This page in human history has been turned, and a new chapter has begun. Future generations will never know what an exclusively analogue, offline, pre-digital reality was like. We are the last generation to have experienced it.

AI has taken over our lives...

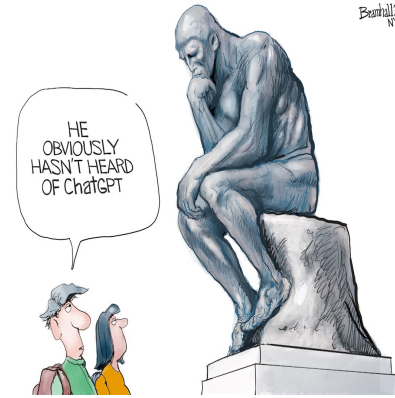
The price for such a special place in history is worrying uncertainties. The transformations brought about by digital technologies are mind-blowing. They justify some confusion and apprehension; one only needs to look at newspaper headlines.

...so there is nothing to do but decide how best to use AI!

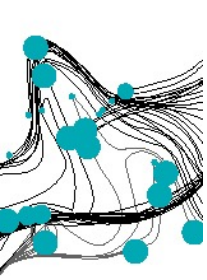
analogue and digital technologies. Precipitous changes are shaping the future. As Winston Churchill once said, 'We shape our buildings; thereafter they shape us.' We are at the very early stage of the construction of our digital realities. We can get them right before they start affecting and influencing both us and future generations in the wrong way. It is not a matter of being pessimistic or optimistic. Discussing whether the glass is half empty or half full is pointless. The interesting question is how we can fill it. This means engaging constructively with the ethical analysis of the problems and the design of the right solutions.



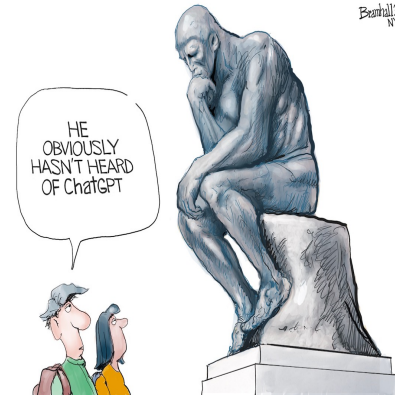
AI Ethics summarized



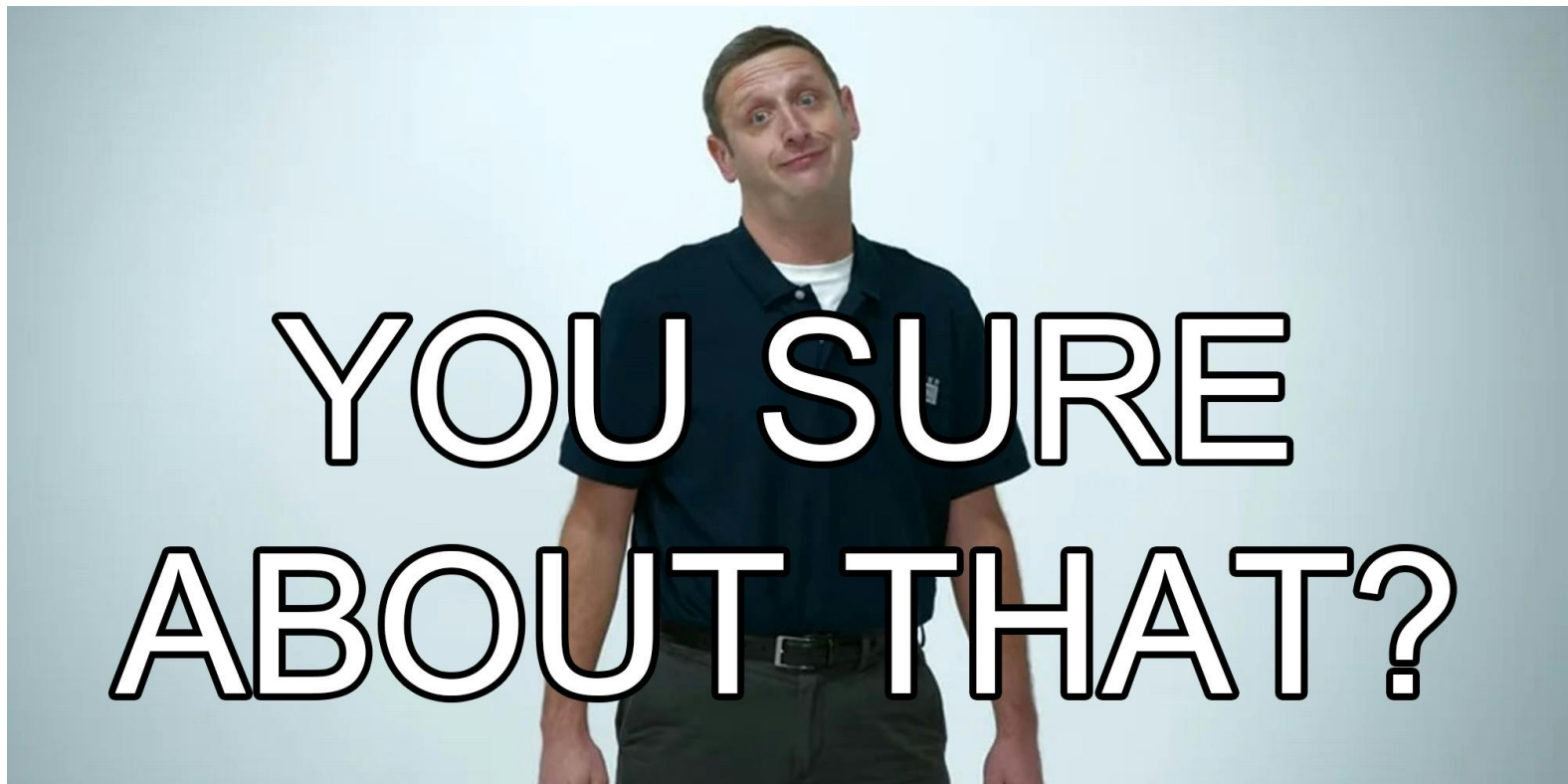
- AI exists (**and there is nothing we can do about that!**)
- AI is potentially beneficial to humanity (**could even solve climate change!**)
- AI is potentially dangerous to humanity (**is already speeding up climate change!**)
- Therefore, must make sure AI is developed ethically (***from now on!***)

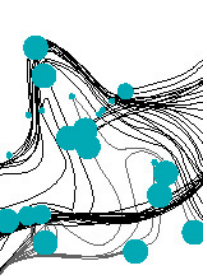


AI Ethics summarized



- AI exists (and there is nothing we can do about that!)





Part III

What is AI?



Don't personalize it Grauber.



The next stop is deification.

Turing predicted computers would *inevitably* be able to think...

VOL. LIX. No. 236.]

[October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?
Now suppose X is actually A, then A must answer. It is A's
28 433

It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question. I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning. The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. I believe further that no useful purpose is served by concealing these beliefs. The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any unproved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research.

...or at least be able to *convincingly seem* to be able to think!

Weizenbaum invented first ChatGPT-like chatbot...

JOSEPH WEIZENBAUM

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT
TO CALCULATION

2. I was startled to see how quickly and how very deeply people conversing with DOCTOR became emotionally involved with the computer and how unequivocally they anthropomorphized it. Once my secretary, who had watched me work on the program for many months and therefore surely knew it to be merely a computer program, started conversing with it. After only a few interchanges with it, she asked me to leave the room. Another time, I suggested I might rig the system so that I could examine all conversations anyone had had with it, say, overnight. I was promptly bombarded with accusations that what I proposed amounted to spying on people's most intimate thoughts; clear evidence that people were conversing with the computer as if it were a person who could be appropriately and usefully addressed in intimate terms. I knew of course that people form all sorts of emotional bonds to machines, for example, to musical instruments, motorcycles, and cars. And I knew from long experience that the strong emotional ties many programmers have to their computers are often formed after only short exposures to their machines. What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people. This insight led me to attach new importance to questions of the relationship between the individual and the computer, and hence to resolve to think about them.

...and then tried to warn people to *not believe* AI existed!

Weizenbaum likened AI inevitability to atom bomb inevitability...

JOSEPH WEIZENBAUM

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT
TO CALCULATION

Today even the most highly placed managers represent themselves as innocent victims of a technology for which they accept no responsibility and which they do not even pretend to understand. (One must wonder, though, why it never occurred to Admiral Moorer to ask what effect the millions of tons of bombs the computer said were being dropped on Viet Nam were having.) The American Secretary of State, Dr. Henry Kissinger, while explaining that he could hardly have known of the “White House horrors” revealed by the Watergate investigation, mourned over “the awfulness of events and the tragedy that has befallen so many people.”

“The tragedy so described had action, but no actors. Only ‘events’ were ‘awful’—not individuals or officials. In this lifeless setting, the mockery of law and the deceit of the people had not been rehearsed and practiced: they had simply ‘befallen.’”⁸

The myth of technological and political and social inevitability is a powerful tranquilizer of the conscience. Its service is to remove responsibility from the shoulders of everyone who truly believes in it.

...because inevitability means *no responsibility*!

Minsky: “We can’t understand programs, so they *are* AI!”

JOSEPH WEIZENBAUM

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT
TO CALCULATION

Minsky goes on to make the following enormously important observations:

“When a program grows in power by an evolution of partially understood patches and fixes, the programmer begins to lose track of internal details, loses his ability to predict what will happen, begins to hope instead of know, and watches the results as though the program were an individual whose range of behavior is uncertain.

“This is already true in some big programs. . . . it will soon be much more acute. . . . large heuristic programs will be developed and modified by several programmers, each testing them on different examples from different [remotely located computer] consoles and inserting advice independently. The program will grow in effectiveness, but no one of the programmers will understand it all. (Of course, this won’t always be successful—the interactions might make it get worse, and no one might be able to fix it again!) Now we see the real trouble with statements like ‘it only does what its programmer told it to do.’ There isn’t any one programmer.”⁵

We do not understand, to hark back to an earlier point for a moment, how a program of the kind Minsky here describes—one that, say, composes “great” music—helps us to “understand” music when the program itself is beyond our understanding.

Weizenbaum: “We can’t understand programs, so we *claim* AI!”

Weizenbaum warned that AI was a sham (did not exist)...

JOSEPH WEIZENBAUM

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT
TO CALCULATION

“The structure of the typical essay on ‘The impact of computers on society’ is as follows: First there is an ‘on the one hand’ statement. It tells all the good things computers have already done for society and often even attempts to argue that the social order would already have collapsed were it not for the ‘computer revolution.’ This is usually followed by an ‘on the other hand’ caution, which tells of certain problems the introduction of computers brings in its wake. The threat posed to individual privacy by large data banks and the danger of large-scale unemployment induced by industrial automation are usually mentioned. Finally, the glorious present and prospective achievements of the computer are applauded, while the dangers alluded to in the second part are shown to be capable of being alleviated by sophisticated technological fixes. The closing paragraph consists of a plea for generous societal support for more, and more large-scale, computer research and development. This is usually coupled to the more or less subtle assertion that only computer science, hence only the computer scientist, can guard the world against the admittedly hazardous fall-out of applied computer technology.”¹⁵

The real message of such typical essays is therefore that the expert will take care of everything, even of the problems he himself creates. He needs more money. That always. But he reassures a public that does not want to know anyway.

...and a scam (to raise money from those who believed it existed)!

Weizenbaum warned that AI was a sham (did not exist)...

BBC

[Home](#) [News](#) [Sport](#) [Business](#) [Innovation](#) [Culture](#) [Arts](#) [Travel](#) [Earth](#) [Audio](#) [Video](#) [Live](#)

AI: Ghost workers demand to be seen and heard

28 March 2021

[Share](#) [Save](#)

Jane Wakefield
Technology reporter



Artificial intelligence and machine learning exist on the back of a lot of hard work from humans.

BUSINESS INSIDER

RETAIL

Amazon's Just Walk Out technology relies on hundreds of workers in India watching you shop

Alex Bitter Apr 3, 2024, 7:10 PM CEST

[Share](#) [Save](#)



...and a scam (to raise money from those who believed it existed)!

Weizenbaum warned that AI was a sham (did not exist)...

Subscribe To Newsletters ✉

Forbes

INNOVATION > AI

Panic Over DeepSeek Exposes AI's Weak Foundation On Hype

By [Eric Siegel](#), Contributor. CEO of Gooder AI, author of "The AI Playbook" & ... [Follow Author](#)

Feb 01, 2025, 10:46am EST

[Share](#) [Save](#) [Comment](#) 0



The AI hype bubble is the new crypto hype bubble

Someday, we're gonna feel pretty silly about our autocomplete worship.

Cory Doctorow · [Follow](#)
12 min read · Mar 9, 2023

3.5K 76

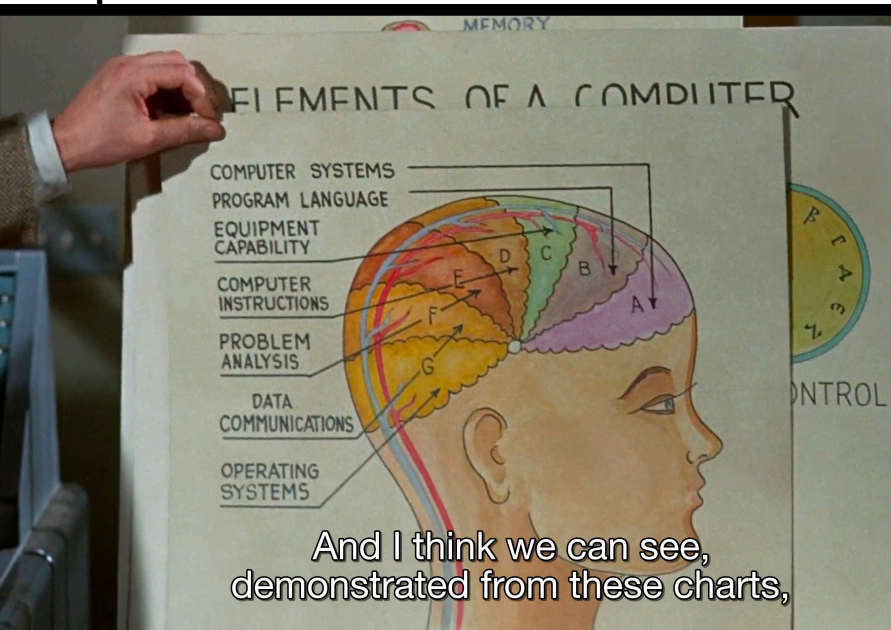
The diagram illustrates the Gartner Hype Cycle for Emerging Technologies. It features a bell-shaped curve on a coordinate system. The curve starts at a point labeled 'Innovation Trigger'. It rises to a peak labeled 'Peak of Inflated Expectations'. It then falls to a trough labeled 'Trough of Disillusionment'. From the trough, it rises again along a slope labeled 'Slope of Enlightenment', eventually leveling off at a plateau labeled 'Plateau of Productivity'. Red spheres, representing data points or hype, are shown along the curve. A large pyramid of these red spheres is stacked at the bottom of the diagram, with the 'Trough of Disillusionment' label positioned above it.

...and a scam (to raise money from those who believed it existed)!

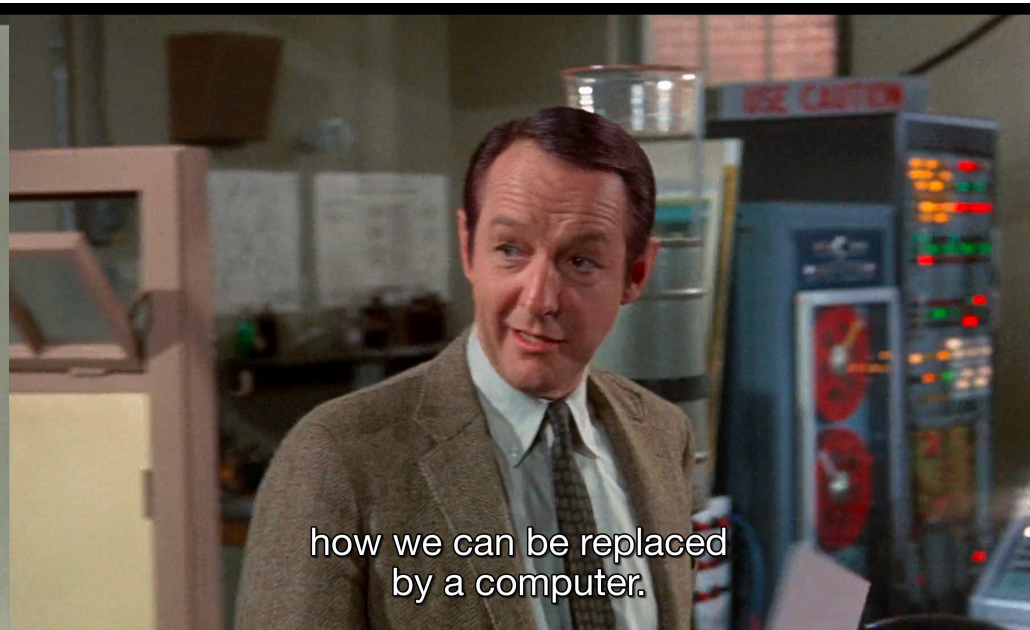


Part IV

What's *unethical* about AI Ethics?

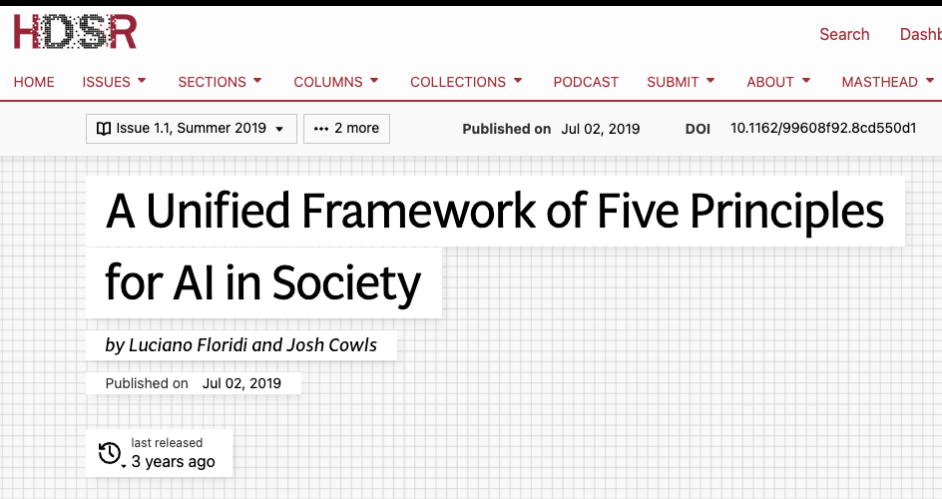


And I think we can see,
demonstrated from these charts,



how we can be replaced
by a computer.

AI Ethics offers so many principles for tech companies to follow...



The screenshot shows the top of a journal article page. At the top left is the 'HDSR' logo. To its right is a navigation bar with links: HOME, ISSUES, SECTIONS, COLUMNS, COLLECTIONS, PODCAST, SUBMIT, ABOUT, and MASTHEAD. Further right are 'Search' and 'Dashb' (partially visible). Below the navigation bar is a header section with a dropdown menu showing 'Issue 1.1, Summer 2019' and a link '... 2 more'. To the right of this is the publication date 'Published on Jul 02, 2019' and the DOI '10.1162/99608f92.8cd550d1'. The main title of the article is 'A Unified Framework of Five Principles for AI in Society' in a large, bold font. Below the title is the author information 'by Luciano Floridi and Josh Cowls'. Below that is the publication date 'Published on Jul 02, 2019'. At the bottom left of the header section is a circular icon with a clock and the text 'last released 3 years ago'.

ABSTRACT

Artificial Intelligence (AI) is already having a major impact on society. As a result, many organizations have launched a wide range of initiatives to establish ethical principles for the adoption of socially beneficial AI. Unfortunately, the sheer volume of proposed principles threatens to overwhelm and confuse. How might this **problem of ‘principle proliferation’** be solved? In this paper, we report the results of a fine-grained analysis of several of the highest-profile sets of ethical principles for AI. We assess whether these principles converge upon a set of agreed-upon principles, or diverge, with **significant disagreement over what constitutes ‘ethical AI.’** Our analysis finds a high degree of overlap among the sets of principles we analyze. We then identify an **overarching framework consisting of five core principles** for ethical AI. Four of them are core principles commonly used in bioethics: *beneficence, non-maleficence, autonomy, and justice*. On the basis of our comparative analysis, we argue that a new principle is needed in addition: ***explicability***, understood as incorporating both the epistemological sense of *intelligibility* (as an answer to the question ‘how does it work?’) and in the ethical sense of *accountability* (as an answer to the question: ‘who is responsible for the way it works?’). In the ensuing discussion, we note the limitations and assess the implications of this ethical framework for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts.

...so let's adapt principles from bioethics!

AI Ethics offers so many principles for tech companies to follow...

ABSTRACT

Artificial Intelligence (AI) is already having a major impact on society. As a result, many organizations have launched a wide range of initiatives to establish ethical principles for the adoption of socially beneficial AI. Unfortunately, the sheer volume of proposed principles threatens to overwhelm and confuse. How might this problem of 'principle proliferation' be solved? In this paper, we report the results of a fine-grained analysis of several of the highest-profile sets of ethical principles for AI. We assess whether these principles converge upon a set of agreed-upon principles, or diverge, with significant disagreement over what constitutes 'ethical AI.' Our analysis finds a high degree of overlap among the sets of principles we analyze. We then identify an overarching framework consisting of five core principles for ethical AI. Four of them are core principles commonly used in bioethics: *beneficence, non-maleficence, autonomy, and justice*. On the basis of our comparative analysis, we argue that a new principle is needed in addition: *explicability*, understood as incorporating both the epistemological sense of *intelligibility* (as an answer to the question 'how does it work?') and in the ethical sense of *accountability* (as an answer to the question: 'who is responsible for the way it works?'). In the ensuing discussion, we note the limitations and assess the implications of this ethical framework for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts.

...so let's adapt principles from bioethics!

Bioethics was developed because even though medical practices could be *dangerous*, they were still *necessary*.

AI Ethics offers so many principles for tech companies to follow...

ABSTRACT

Artificial Intelligence (AI) is already having a major impact on society. As a result, many organizations have launched a wide range of initiatives to establish ethical principles for the adoption of socially beneficial AI. Unfortunately, the sheer volume of proposed principles threatens to overwhelm and confuse. How might this problem of 'principle proliferation' be solved? In this paper, we report the results of a fine-grained analysis of several of the highest-profile sets of ethical principles for AI. We assess whether these principles converge upon a set of agreed-upon principles, or diverge, with significant disagreement over what constitutes 'ethical AI.' Our analysis finds a high degree of overlap among the sets of principles we analyze. We then identify an overarching framework consisting of five core principles for ethical AI. Four of them are core principles commonly used in bioethics: *beneficence, non-maleficence, autonomy, and justice*. On the basis of our comparative analysis, we argue that a new principle is needed in addition: *explicability*, understood as incorporating both the epistemological sense of *intelligibility* (as an answer to the question 'how does it work?') and in the ethical sense of *accountability* (as an answer to the question: 'who is responsible for the way it works?'). In the ensuing discussion, we note the limitations and assess the implications of this ethical framework for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts.

...so let's adapt principles from bioethics!

Bioethics was developed because even though medical practices could be *dangerous*, they were still *necessary*.



But AI is dangerous without being necessary!

AI Ethics offers so many principles for tech companies to follow...

ABSTRACT

Artificial Intelligence (AI) is already having a major impact on society. As a result, many organizations have launched a wide range of initiatives to establish ethical principles for the adoption of socially beneficial AI. Unfortunately, the sheer volume of proposed principles threatens to overwhelm and confuse. How might this problem of 'principle proliferation' be solved? In this paper, we report the results of a fine-grained analysis of several of the highest-profile sets of ethical principles for AI. We assess whether these principles converge upon a set of agreed-upon principles, or diverge, with significant disagreement over what constitutes 'ethical AI.' Our analysis finds a high degree of overlap among the sets of principles we analyze. We then identify an overarching framework consisting of five core principles for ethical AI. Four of them are core principles commonly used in bioethics: *beneficence, non-maleficence, autonomy, and justice*. On the basis of our comparative analysis, we argue that a new principle is needed in addition: *explicability*, understood as incorporating both the epistemological sense of *intelligibility* (as an answer to the question 'how does it work?') and in the ethical sense of *accountability* (as an answer to the question: 'who is responsible for the way it works?'). In the ensuing discussion, we note the limitations and assess the implications of this ethical framework for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts.

...so let's adapt principles from bioethics!

Bioethics was developed because even though medical practices could be *dangerous*, they were still *necessary*.



But AI is dangerous without being necessary!



Likening AI ethics to Bioethics gives the *false impression* that *AI is as necessary as Medicine*

AI Ethics offers so many principles for tech companies to follow...

ABSTRACT

Artificial Intelligence (AI) is already having a major impact on society. As a result, many organizations have launched a wide range of initiatives to establish ethical principles for the adoption of socially beneficial AI. Unfortunately, the sheer volume of proposed principles threatens to overwhelm and confuse. How might this **problem of 'principle proliferation'** be solved? In this paper, we report the results of a fine-grained analysis of several of the highest-profile sets of ethical principles for AI. We assess whether these principles converge upon a set of agreed-upon principles, or diverge, with **significant disagreement over what constitutes 'ethical AI.'** Our analysis finds a high degree of overlap among the sets of principles we analyze. We then identify an **overarching framework consisting of five core principles** for ethical AI. Four of them are core principles commonly used in bioethics: *beneficence, non-maleficence, autonomy, and justice*. On the basis of our comparative analysis, we argue that a new principle is needed in addition: *explicability*, understood as incorporating both the epistemological sense of *intelligibility* (as an answer to the question 'how does it work?') and in the ethical sense of *accountability* (as an answer to the question: 'who is responsible for the way it works?'). In the ensuing discussion, we note the limitations and assess the implications of this ethical framework for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts.

...so let's adapt principles from bioethics!

Bioethics was developed because even though medical practices could be *dangerous*, they were still *necessary*.



But AI is dangerous without being necessary!



Likening AI ethics to Bioethics gives the *false impression* that *AI is as necessary as Medicine*



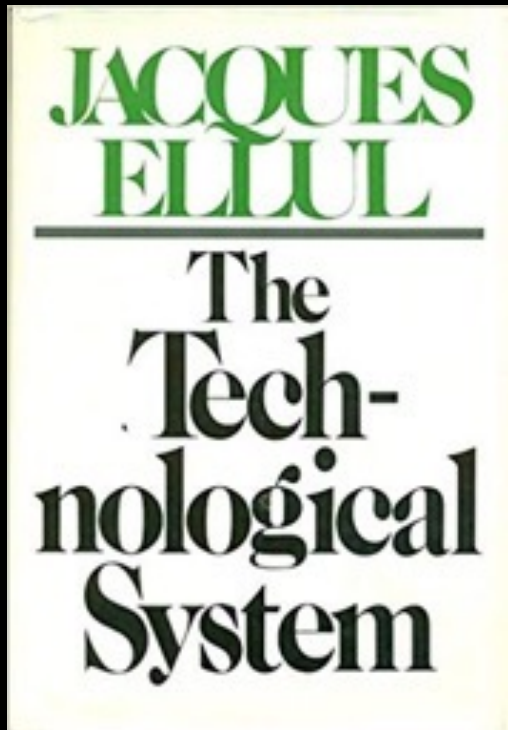
So why does AI ethics help *perpetuate the false impression* that AI is *necessary*?

Technologies appear to be **neutral** with regards to morality and so help to create an **amoral** technological world...

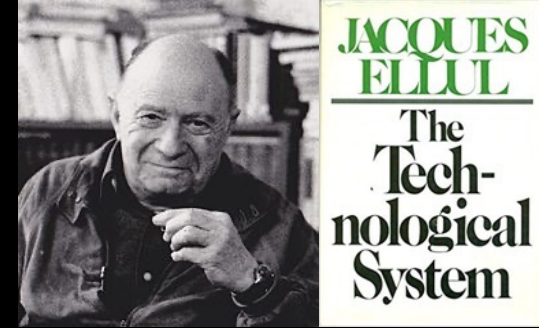
There is, however, a further stride to be made, and quite a normal one at that. Independent of morals and judgments, legitimate in itself, technology is becoming the creative force of new values, of a new ethics. Man cannot do without morality! Technology has destroyed all previous scales of value; it impugns the judgments coming from outside. After all, it wrecks their foundations. But being thus self-justified, it quite normally becomes justifying. What was done in the name of science was just; and now the same holds true for what is done in the name of technology. It attributes justice to human action, and man is thus spontaneously led to construct an ethics on the basis of, and in terms of, technology.²¹

This does not occur in a theoretical or systematic manner. The elaboration only comes afterwards. The technological ethics is constructed bit by bit, concretely. Technology demands a certain number of virtues from man (precision, exactness, seriousness, a realistic attitude, and, over everything else, the virtue of work) and a certain outlook on life (modesty, devotion, cooperation). Technology permits very clear value judgments (what is serious and what is not, what is effective, efficient, useful, etc.). This ethics is built up on these concrete givens; for it is primarily an experienced ethics of the behavior required for the technological system to function well. It thereby has the vast superiority over the other moralities of being truly experienced. Furthermore, it involves obvious and ineluctable sanctions (for it is the functioning of the technological system

...but because humanity cannot live without values, **technologies create new values**, values that are **pro-technology**!



Technology vs Humanity



Belief: Ends determine choice of Means (Health) (AI)

Lighting for your wellbeing

Light is essential to life. We sleep, wake, work, relax, and live by the light around us. Philips Hue lets you harness the power of this light and use it to support your wellbeing.



Stay focused

If your home is also your office, let smart light help you separate your work life from your home life. Your lights tell you when it's time to focus — and when it's time to take a break.

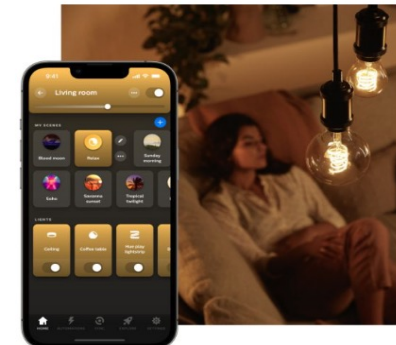
[Explore lighting for productivity >](#)



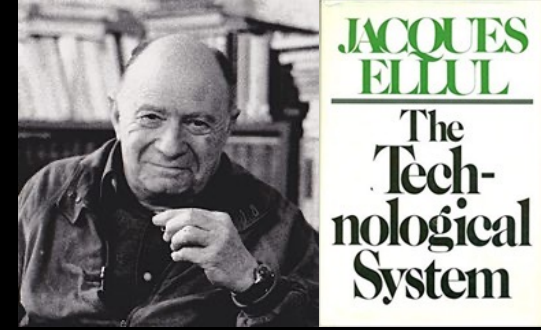
Recharge

Smart lighting can signal when you should start winding down for the day, helping you relax and get ready for a good night's sleep.

[Explore lights for relaxation >](#)



Technology vs Humanity



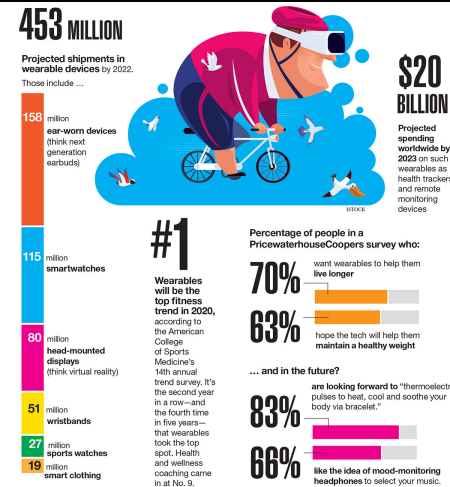
Belief: Ends determine choice of Means
(Health) (AI)

Reality: Means determine choice of Ends
(AI) (“Health”)

6 lighting hacks for healthier, more productive workplaces



The right lighting in an office can make all the difference - like in this Berlin office lit by Selux



Live Longer with AI

How artificial intelligence is helping us extend our healthspan and live better too



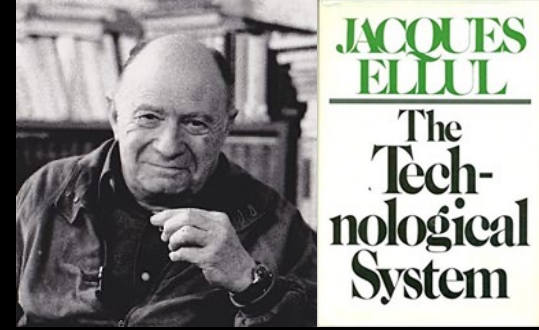
Foreword by
Andrew Scott,
co-author of
the bestseller
The 100-Year Life

Packtp
EXPERT INSIGHT

Tina Woods
with Melissa Ream

Sources: American College of Sports Medicine, Juniper Research, Gartner, PricewaterhouseCoopers

Technology vs Humanity



The belief that *everything should be done as efficiently as possible* is not based on a human value, but on a *value required to perpetuate technological world*



Writing = Set of **tasks** →
Tasks = Do as **efficiently** as possible →
ChatGPT = **Best** way to write!

[REGISTER FOR A FREE ACCOUNT](#)

THE CHRONICLE OF HIGHER EDUCATION

[SUBSCRIBE](#) [Sign In](#)

[NEWS](#) | [ADVICE](#) | [THE REVIEW](#) | [DATA](#) | [CURRENT ISSUE](#) | [VIRTUAL EVENTS](#) | [STORE](#) ▾ | [JOBS](#) ▾ | [Q](#)



**I'm a Student. You Have
No Idea How Much
We're Using ChatGPT.**

**No professor or software could
ever pick up on it.**



Part V

What *should* be AI Ethics?



Exclusive: In its quest to make ChatGPT less toxic, OpenAI used outsourced Kenyan laborers earning less than \$2 per hour



Topic Rooms Via



Class Action Filed Against Stability AI, Midjourney, and DeviantArt for DMCA Violations

AI image products which was trained on billions of copyrighted images contained in the LAION-5B dataset, which were downloaded and used without compensation or consent from the artists

A ChatGPT defamation lawsuit could be the first of many cases that will examine where legal liability falls when AI chatbots spew falsehoods.

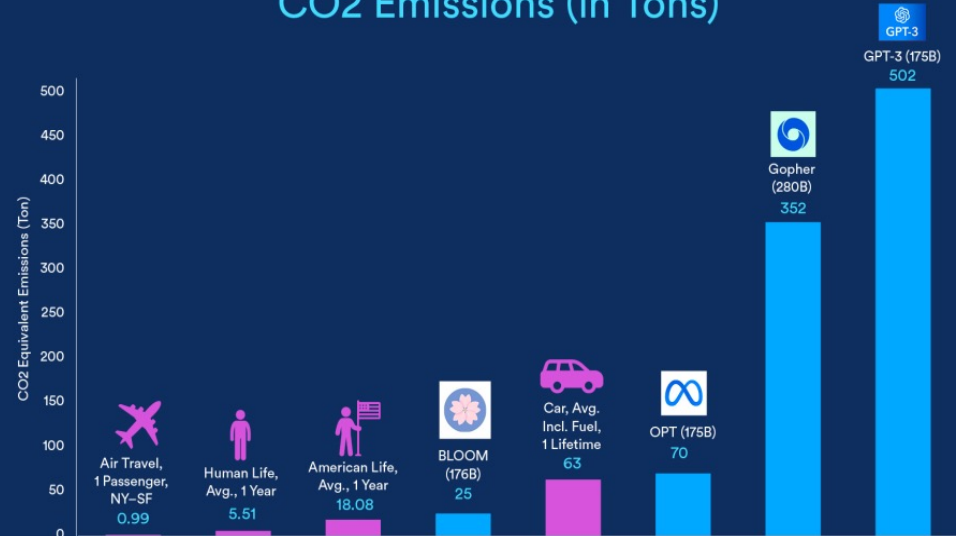


news.bloomberglaw.com

First ChatGPT Defamation Lawsuit to Test AI's Legal Liability

A defamation lawsuit filed against the artificial intelligence company OpenAI LLC will provide the first foray into the largely untested legal ...

CO2 Emissions (in Tons)



Exclusive: In its quest to make ChatGPT less toxic, OpenAI used outsourced Kenyan laborers earning less than \$2 per hour



TopicRooms Vix



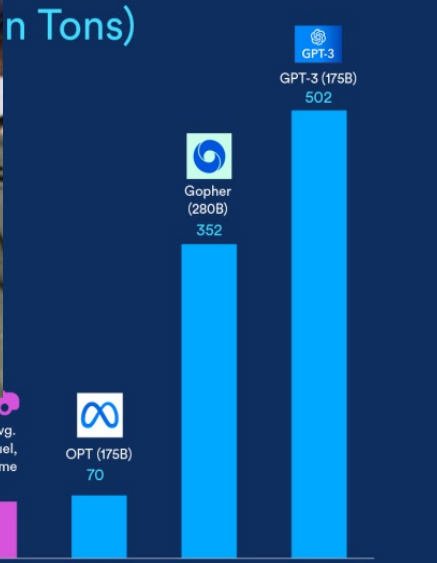
Class Action Filed Against Midjourney, and DeviantArt for Copyright Violations

AI image products which was trained on billions of copyrighted images contained in the LAION-5B dataset, which were downloaded and used without compensation or consent from the artists



ARTIFICIAL INTELLIGENCE

Effort to Test AI's Legal Liability
... the artificial intelligence company ... foray into the largely untested legal ...



Source: Luccioni et al., 2022; Strubell et al., 2019 | Chart: 2023 AI Index Report

Weizenbaum concluded even if AI existed, shouldn't be used...

JOSEPH WEIZENBAUM

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT
TO CALCULATION

What could be more obvious than the fact that, whatever intelligence a computer can muster, however it may be acquired, it must always and necessarily be absolutely alien to any and all authentic human concerns? The very asking of the question, "What does a judge (or a psychiatrist) know that we cannot tell a computer?" is a monstrous obscenity. That it has to be put into print at all, even for the purpose of exposing its morbidity, is a sign of the madness of our times.

Computers can make judicial decisions, computers can make psychiatric judgments. They can flip coins in much more sophisticated ways than can the most patient human being. The point is that they *ought* not be given such tasks. They may even be able to arrive at "correct" decisions in some cases—but always and necessarily on bases no human being should be willing to accept.

There have been many debates on "Computers and Mind." What I conclude here is that the relevant issues are neither technological nor even mathematical; they are ethical. They cannot be settled by asking questions beginning with "can." The limits of the applicability of computers are ultimately statable only in terms of oughts. What emerges as the most elementary insight is that, since we do not now have any ways of making computers wise, we ought not now to give computers tasks that demand wisdom.

...because even *right* answers would be arrived at *wrongly*!



VS



does morality reside in the painfulness of indefinite questioning?



network error

Nolen Gertz

NIHILISM AND TECHNOLOGY

Updated Edition



NIHILISM

NOLEN GERTZ

- \ (ツ) _ / -

THE MIT PRESS ESSENTIAL KNOWLEDGE SERIES

Thank You!

n.gertz@utwente.nl

nolengertz.com