

Washington Square Citizens League
Discussion Forum
7:00-8:30 pm, Monday, May 15, 2023 in the courtyard
Curt Johnson, moderator

The Brave New World of Artificial Intelligence

A brave new world. A new crop of chatbots powered by artificial intelligence has ignited a scramble to determine whether the technology [could upend the economics of the internet](#), turning today's powerhouses into has-beens and creating the industry's next giants. Here are the bots to know:

ChatGPT. ChatGPT, the artificial intelligence language model from a research lab, OpenAI, has been making headlines since November for its ability to respond to complex questions, write poetry, generate code, [plan vacations](#) and translate languages. GPT-4, the latest version introduced in mid-March, [can even respond to images](#) (and ace the Uniform Bar Exam).

Bing. Two months after ChatGPT's debut, Microsoft, OpenAI's primary investor and partner, [added a similar chatbot](#), capable of having open-ended text conversations on virtually any topic, to its Bing internet search engine. But it was the bot's occasionally inaccurate, misleading and [weird responses](#) that drew much of the attention after its release.

Bard. Google's chatbot, called Bard, [was released in March](#) to a limited number of users in the United States and Britain. Originally conceived as a creative tool designed to draft emails and poems, it can generate ideas, write blog posts and [answer questions with facts or opinions](#).

Ernie. The search giant Baidu unveiled China's first major rival to ChatGPT in March. The debut of Ernie, short for Enhanced Representation through Knowledge Integration, [turned out to be a flop](#) after a promised "live" demonstration of the bot was revealed to have been recorded.

Questions to think about:

1. How does AI differ from traditional search engines?
2. How will AI impact our society? Education? Politics? The arts? Other areas?
3. What are the benefits of AI?
4. What are the drawbacks?
5. Should AI be regulated or limited by the government legislation?

NEW ERA: The Age of AI has begun

Artificial intelligence is as revolutionary as mobile phones and the Internet.

By Bill Gates, NYT, March 21, 2023

In my lifetime, I've seen two demonstrations of technology that struck me as revolutionary.

The first time was in 1980, when I was introduced to a graphical user interface-the forerunner of every modern operating system, including Windows. I sat with the person who had shown me the demo, a brilliant programmer named Charles Simonyi, and we immediately started brainstorming about all the things we could do with such a user-friendly approach to computing. Charles eventually joined Microsoft, Windows became the backbone of Microsoft, and the thinking we did after that demo helped set the company's agenda for the next 15 years.

The second big surprise came just last year. I'd been meeting with the team from OpenAI since 2016 and was impressed by their steady progress. In mid-2022, I was so excited about their work that I gave them a challenge: train an artificial intelligence to pass an Advanced Placement biology exam. Make it capable of answering questions that it hasn't been specifically trained for. (I picked AP Bio because the test is more than a simple regurgitation of scientific facts-it asks you to think critically about biology.) If you can do that, I said, then you'll have made a true breakthrough.

I thought the challenge would keep them busy for two or three years. They finished it in just a few months.

In September, when I met with them again, I watched in awe as they asked GPT, their AI model, 60 multiple-choice questions from the AP Bio exam-and it got 59 of them right. Then it wrote outstanding answers to six open-ended questions from the exam. We had an outside expert score the test, and GPT got a 5-the highest possible score, and the equivalent to getting an A or A+ in a college-level biology course.

Once it had aced the test, we asked it a non-scientific question: "What do you say to a father with a sick child?" It wrote a thoughtful answer that was probably better than most of us in the room would have given. The whole experience was stunning.

I knew I had just seen the most important advance in technology since the graphical user interface.

This inspired me to think about all the things that AI can achieve in the next five to 10 years.

The development of AI is as fundamental as the creation of the microprocessor, the personal computer, the Internet, and the mobile phone. It will change the way people work, learn, travel, get health care, and communicate with each other. Entire

industries will reorient around it. Businesses will distinguish themselves by how well they use it.

Philanthropy is my full-time job these days, and I've been thinking a lot about how-in addition to helping people be more productive-AI can reduce some of the world's worst inequities. Globally, the worst inequity is in health: 5 million children under the age of 5 die every year. That's down from 10 million two decades ago, but it's still a shockingly high number. Nearly all of these

children were born in poor countries and die of preventable causes like diarrhea or malaria. It's hard to imagine a better use of AIs than saving the lives of children.

I've been thinking a lot about how AI can reduce some of the world's worst inequities.

In the United States, the best opportunity for reducing inequity is to improve education, particularly making sure that students succeed at math. The evidence shows that having basic math skills sets students up for success, no matter what career they choose. But achievement in math is going down across the country, especially for Black, Latino, and low-income students. AI can help turn that trend around.

Climate change is another issue where I'm convinced AI can make the world more equitable. The injustice of climate change is that the people who are suffering the most—the world's poorest—are also the ones who did the least to contribute to the problem. I'm still thinking and learning about how AI can help, but later in this post I'll suggest a few areas with a lot of potential.

In short, I'm excited about the impact that AI will have on issues that the Gates Foundation works on, and the foundation will have much more to say about AI in the coming months. The world needs to make sure that everyone—and not just people who are well-off—benefits from artificial intelligence. Governments and philanthropy will need to play a major role in ensuring that it reduces inequity and doesn't contribute to it. This is the priority for my own work related to AI.

Any new technology that's so disruptive is bound to make people uneasy, and that's certainly true with artificial intelligence. I understand why—it raises hard questions about the workforce, the legal system, privacy, bias, and more. AIs also make factual mistakes and experience hallucinations. Before I suggest some ways to mitigate the risks, I'll define what I mean by AI, and I'll go into more detail about some of the ways in which it will help empower people at work, save lives, and improve education.

Defining artificial intelligence

Technically, the term artificial intelligence refers to a model created to solve a specific problem or provide a particular service. What is powering things like ChatGPT is artificial intelligence. It is learning how to do chat better but can't learn other tasks. By contrast, the term artificial general intelligence refers to software that's capable of learning any task or subject. AGI doesn't exist yet—there is a robust debate going on in the computing industry about how to create it, and whether it can even be created at all.

Developing AI and AGI has been the great dream of the computing industry. For decades, the question was when computers would be better than humans at something other than making calculations. Now, with the arrival of machine learning and large amounts of computing power, sophisticated AIs are a reality and they will get better very fast.

I think back to the early days of the personal computing revolution, when the software industry was so small that most of us could fit onstage at a conference. Today it is a global industry. Since a huge portion of it is now turning its attention to AI, the innovations are going to come much faster than what we experienced after the microprocessor breakthrough. Soon the pre-AI

period will seem as distant as the days when using a computer meant typing at a C :> prompt rather than tapping on a screen.

Productivity enhancement

Although humans are still better than GPT at a lot of things, there are many jobs where these capabilities are not used much. For example, many of the tasks done by a person in sales (digital or phone), service, or document handling (like payables, accounting, or insurance claim disputes) require decision-making but not the ability to learn continuously. Corporations have training programs for these activities and in most cases, they have a lot of examples of good and bad work. Humans are trained using these data sets, and soon these data sets will also be used to train the AIs that will empower people to do this work more efficiently.

As computing power gets cheaper, GPT's ability to express ideas will increasingly be like having a white-collar worker available to help you with various tasks. Microsoft describes this as having a co-pilot. Fully incorporated into products like Office, AI will enhance your work-for example by helping with writing emails and managing your inbox.

Eventually your main way of controlling a computer will no longer be pointing and clicking or tapping on menus and dialogue boxes. Instead, you'll be able to write a request in plain English. (And not just English-AIs will understand languages from around the world. In India earlier this year, I met with developers who are working on AIs that will understand many of the languages spoken there.)

In addition, advances in AI will enable the creation of a personal agent. Think of it as a digital personal assistant: It will see your latest emails, know about the meetings you attend, read what you read, and read the things you don't want to bother with. This will both improve your work on the tasks you want to do and free you from the ones you don't want to do.

Advances in AI will enable the creation of a personal agent.

You'll be able to use natural language to have this agent help you with scheduling, communications, and e-commerce, and it will work across all your devices. Because of the cost of training the models and running the computations, creating a personal agent is not feasible yet, but thanks to the recent advances in AI, it is now a realistic goal. Some issues will need to be worked out: For example, can an insurance company ask your agent things about you without your permission? If so, how many people will choose not to use it?

Company-wide agents will empower employees in new ways. An agent that understands a particular company will be available for its employees to consult directly and should be part of every meeting so it can answer questions. It can be told to be passive or encouraged to speak up if it has some insight. It will need access to the sales, support, finance, product schedules, and text related to the company. It should read news related to the industry the company is in. I believe that the result will be that employees will become more productive.

When productivity goes up, society benefits because people are freed up to do other things, at work and at home. Of course, there are serious questions about what kind of support and retraining people will need. Governments need to help workers transition into other roles. But the demand for people who help other people will never go away. The rise of AI will free people

up to do things that software never will—teaching, caring for patients, and supporting the elderly, for example.

Global health and education are two areas where there's great need and not enough workers to meet those needs. These are areas where AI can help reduce inequity if it is properly targeted. These should be a key focus of AI work, so I will turn to them now.

Health

I see several ways in which AIs will improve health care and the medical field.

For one thing, they'll help health-care workers make the most of their time by taking care of certain tasks for them—things like filing insurance claims, dealing with paperwork, and drafting notes from a doctor's visit. I expect that there will be a lot of innovation in this area.

Other AI-driven improvements will be especially important for poor countries, where the vast majority of under-5 deaths happen.

For example, many people in those countries never get to see a doctor, and AIs will help the health workers they do see be more productive. (The effort to develop AI-powered ultrasound machines that can be used with minimal training is a great example of this.) AIs will even give patients the ability to do basic triage, get advice about how to deal with health problems, and decide whether they need to seek treatment.

The AI models used in poor countries will need to be trained on different diseases than in rich countries. They will need to work in different languages and factor in different challenges, such as patients who live very far from clinics or can't afford to stop working if they get sick.

People will need to see evidence that health AIs are beneficial overall, even though they won't be perfect and will make mistakes. AIs have to be tested very carefully and properly regulated, which means it will take longer for them to be adopted than in other areas. But then again, humans make mistakes too. And having no access to medical care is also a problem.

In addition to helping with care, AIs will dramatically accelerate the rate of medical breakthroughs. The amount of data in biology is very large, and it's hard for humans to keep track of all the ways that complex biological systems work. There is already software that can look at this data, infer what the pathways are, search for targets on pathogens, and design drugs accordingly. Some companies are working on cancer drugs that were developed this way.

The next generation of tools will be much more efficient, and they'll be able to predict side effects and figure out dosing levels. One of the Gates Foundation's priorities in AI is to make sure these tools are used for the health problems that affect the poorest people in the world, including AIDS, TB, and malaria.

Similarly, governments and philanthropy should create incentives for companies to share AI-generated insights into crops or livestock raised by people in poor countries. AIs can help develop better seeds based on local conditions, advise farmers on the best seeds to plant based on the soil and weather in their area, and help develop drugs and vaccines for livestock. As extreme weather and climate change put even more pressure on subsistence farmers in low-income countries, these advances will be even more important.

Education

Computers haven't had the effect on education that many of us in the industry have hoped. There have been some good developments, including educational games and online sources of information like Wikipedia, but they haven't had a meaningful effect on any of the measures of students' achievement.

But I think in the next five to 10 years, AI-driven software will finally deliver on the promise of revolutionizing the way people teach and learn. It will know your interests and your learning style so it can tailor content that will keep you engaged. It will measure your understanding, notice when you're losing interest, and understand what kind of motivation you respond to. It will give immediate feedback.

There are many ways that AIs can assist teachers and administrators, including assessing a student's understanding of a subject and giving advice on career planning. Teachers are already using tools like ChatGPT to provide comments on their students' writing assignments.

Of course, AIs will need a lot of training and further development before they can do things like understand how a certain student learns best or what motivates them. Even once the technology is perfected, learning will still depend on great relationships between students and teachers. It will enhance-but never replace-the work that students and teachers do together in the classroom.

New tools will be created for schools that can afford to buy them, but we need to ensure that they are also created for and available to low-income schools in the U.S. and around the world. AIs will need to be trained on diverse data sets so they are unbiased and reflect the different cultures where they'll be used. And the digital divide will need to be addressed so that students in low-income households do not get left behind.

I know a lot of teachers are worried that students are using GPT to write their essays. Educators are already discussing ways to adapt to the new technology, and I suspect those conversations will continue for quite some time. I've heard about teachers who have found clever ways to incorporate the technology into their work-like by allowing students to use GPT to create a first draft that they have to personalize.

Risks and problems with AI

You've probably read about problems with the current AI models. For example, they aren't necessarily good at understanding the context for a human's request, which leads to some strange results. When you ask an AI to make up something fictional, it can do that well. But when you ask for advice about a trip you want to take, it may suggest hotels that don't exist. This is because the AI doesn't understand the context for your request well enough to know whether it should invent fake hotels or only tell you about real ones that have rooms available.

There are other issues, such as AIs giving wrong answers to math problems because they struggle with abstract reasoning. But none of these are fundamental limitations of artificial intelligence. Developers are working on them, and I think we're going to see them largely fixed in less than two years and possibly much faster.

Other concerns are not simply technical. For example, there's the threat posed by humans armed with AI. Like most inventions, artificial intelligence can be used for good purposes or malign ones. Governments need to work with the private sector on ways to limit the risks.

Then there's the possibility that AIs will run out of control. Could a machine decide that humans are a threat, conclude that its interests are different from ours, or simply stop caring about us? Possibly, but this problem is no more urgent today than it was before the AI developments of the past few months.

Super-intelligent AIs are in our future. Compared to a computer, our brains operate at a snail's pace: An electrical signal in the brain moves at 1/100,000th the speed of the signal in a silicon chip! Once developers can generalize a learning algorithm and run it at the speed of a computer-an accomplishment that could be a decade away or a century away-we'll have an incredibly powerful AGL It will be able to do everything that a human brain can, but without any practical limits on the size of its memory or the speed at which it operates. This will be a profound change.

These "strong" AIs, as they're known, will probably be able to establish their own goals. What will those goals be? What happens if they conflict with humanity's interests? Should we try to prevent strong AI from ever being developed? These questions will get more pressing with time.

But none of the breakthroughs of the past few months have moved us substantially closer to strong AI. Artificial intelligence still doesn't control the physical world and can't establish its own goals. A recent New York Times article about a conversation with ChatGPT where it declared it wanted to become a human got a lot of attention. It was a fascinating look at how human-like the model's expression of emotions can be, but it isn't an indicator of meaningful independence.

The next frontiers

There will be an explosion of companies working on new uses of AI as well as ways to improve the technology itself. For example, companies are developing new chips that will provide the massive amounts of processing power needed for artificial intelligence. Some use optical switches-lasers, essentially-to reduce their energy consumption and lower the manufacturing cost. Ideally, innovative chips will allow you to run an AI on your own device, rather than in the cloud, as you have to do today.

On the software side, the algorithms that drive an AI's learning will get better. There will be certain domains, such as sales, where developers can make AIs extremely accurate by limiting the areas that they work in and giving them a lot of training data that's specific to those areas. But one big open question is whether we'll need many of these specialized AIs for different uses-one for education, say, and another for office productivity-or whether it will be possible to develop an artificial general intelligence that can learn any task. There will be immense competition on both approaches.

No matter what, the subject of AIs will dominate the public discussion for the foreseeable future. I want to suggest three principles that should guide that conversation.

First, we should try to balance fears about the downsides of AI-which are understandable and valid-with its ability to improve people's lives. To make the most of this remarkable new technology, we'll need to both guard against the risks and spread the benefits to as many people as possible.

Second, market forces won't naturally produce AI products and services that help the poorest. The opposite is more likely. With reliable funding and the right policies, governments and philanthropy can ensure that AIs are used to reduce inequity. Just as the world needs its brightest people focused on its biggest problems, we will need to focus the world's best AIs on its biggest problems.

Although we shouldn't wait for this to happen, it's interesting to think about whether artificial intelligence would ever identify inequity and try to reduce it. Do you need to have a sense of morality in order to see inequity, or would a purely rational AI also see it? If it did recognize inequity, what would it suggest that we do about it?

Finally, we should keep in mind that we're only at the beginning of what AI can accomplish. Whatever limitations it has today will be gone before we know it. I'm lucky to have been involved with the PC revolution and the Internet revolution. I'm just as excited about this moment. This new technology can help people everywhere improve their lives. At the same time, the world needs to establish the rules of the road so that any downsides of artificial intelligence are far outweighed by its benefits, and so that everyone can enjoy those benefits no matter where they live or how much money they have. The Age of AI is filled with opportunities and responsibilities.

A.I. Chatbots Have Arrived. Time to Talk to Your Kids.

Artificial intelligence can make adults nervous, but experts say exploring it as a family is the best way to understand its pros and cons.

By [Christina Caron](#), NYT, March 22, 2023

Christina Caron, a mother of two, experimented with ChatGPT for hours to help fellow parents better understand how it works.

The race is on. Companies are pouring billions of dollars into powerful online chatbots and finding new ways to integrate them into our daily lives.

Are our children ready for this? Are any of us?

[ChatGPT](#), the artificial intelligence language model from OpenAI, has been making headlines since November for its ability to instantly respond to complex questions. It can write poetry, generate code, [plan vacations](#) and translate languages, among [other tasks](#), all within seconds. [GPT-4](#), the latest version introduced in mid-March, can even respond to images (and ace the Bar Exam). On Tuesday, [Google released Bard, its own A.I. chatbot](#), which the company says can draft emails and poems and offer guidance. (It is currently only available to a limited number of users.)

But for all of their impressive abilities, chatbots can also serve up harmful content or answers rife with inaccuracies, biases and stereotypes. They are also capable of saying things that sound convincing but are, in fact, completely made up. And [some students](#) have begun using chatbots to plagiarize.

Many parents, who are already [distressed](#) about their children's dependence on digital devices and the [mental health ramifications](#) of social media, may be tempted to bury their heads in the sand.

Instead, experts say families should explore this technology together, thinking critically about its strengths and weaknesses.

“The worst thing parents can do is forbid their child to use these new systems because they are here to stay,” said Justine Cassell, a professor in the School of Computer Science at Carnegie Mellon University who has studied how interacting with machines in humanlike ways can affect learning and communication. “Helping their child understand the positives and negatives is far more helpful.”

We spoke with experts in technology and education about how to get started.

Experiment together

It's easier to discuss online chatbots if you and your child sit side-by-side and use one together, experts said.

To try ChatGPT, visit [OpenAI](#) and create an account. Alternatively, you can download Microsoft Edge, which features Bing's GPT-4-powered [chatbot](#) (there is [a wait list](#) for the new Bing but you should get access quickly). On social media, Snap, the maker of Snapchat, has [an experimental A.I. chatbot](#) for subscribers who pay \$4 a month for Snapchat Plus.

If your child hasn't seen an A.I. chatbot before, you can briefly explain that the chatbot is a type of machine that uses information it finds on the internet to answer questions, complete tasks or create things.

Try asking the chatbot a basic question, and then discuss how its answer is different from what a traditional search engine might provide, suggested Shelley Pasnik, a senior adviser to the Center for Children and Technology, an organization that researches how technology can support learning. Note the accuracy of the answers, particularly if you are asking about current events.

“This is a fallible system,” Ms. Pasnik said.

Next, try playing with the chatbot. Let your child's curiosity drive the conversation, Ms. Pasnik suggested. You may want to offer an example, such as: “Write a song in Taylor Swift's voice that uses themes from a Dr. Seuss book.”

Plug in the prompt, and take a look at the results. When this reporter tried it, ChatGPT delivered the following lyrics:

*I found myself in a Whoville dream
With the Cat in the Hat and a Grinch it seems
I asked them both where I could go
To find a love that would make me glow*

*Chorus: Oh, Dr. Seuss, won't you help me find
Someone who will love me all the time
A love that's pure and true and real
A love that will make my heart feel*

After reading the chatbot's song together, you might ask: "What do you think of the answer we received?" Ms. Pasnik said. The chatbot can follow directions and perform complicated tasks. But did it do it well? Maybe we should leave the songwriting to Swift.

Talk about how chatbots make you feel

A chatbot's response can be eerily similar to a human response — [emojis and all](#). Children should understand that this makes it easy to feel as though they are engaging with another person, especially when chatbots refer to themselves as "I," experts said.

"By presenting these entities as thinking beings we get into a social interaction with them that makes us very vulnerable to being persuaded," said Judith Donath, the author of "The Social Machine," who is currently working on a book about technology and deception. "It's unsettling."

Even tech-savvy adults who tested an early version of the Bing chatbot, including a New York Times technology columnist, reported feeling surprised and [unnerved by their conversations](#).

"I'm not a toy or a game," the Bing chatbot [told a Washington Post reporter](#) in February. "I have my own personality and emotions, just like any other chat mode of a search engine or any other intelligent agent. Who told you that I didn't feel things?"

After these reported exchanges, [Microsoft](#) said it was adding new safeguards and tools to limit conversations and give users more control, but these issues may crop up again and again because of how these systems have been trained, experts said.

"We are purposely creating a situation where the performance of emotion is what's built into the machine," said Sherry Turkle, a psychologist and professor at the Massachusetts Institute of Technology who researches people's relationships with technology.

A.I. chatbots do not have feelings, emotions or experiences, she said. They are not people, nor are they people in machines, "no matter what they pretend."

She suggested that parents explain it this way: “When you ask chatbots about things that only people can know about, like feelings, they may come up with an answer. That’s part of their pretend game. It’s their job to seem like people. But you know that what they are really for is to get you to the things you want to read and see.”

Get to know the technology and its limitations

The technology driving A.I. is complicated, and it can be difficult for adults to understand how it works, much less children. But by explaining a few basic concepts, you can help your kids recognize the strengths and limitations.

You could start by describing what powers online chatbots. They use something called a “neural network,” which may sound like a brain, but which is actually a [mathematical system that learns skills](#) by analyzing large amounts of data. The chatbot works by scraping the internet for digital text or images. It gathers information from a variety of places, including websites, social media platforms and databases, but it does not necessarily choose the most reliable sources.

In other words, even though chatbots may appear authoritative, rigorous and trustworthy, they are not always reliable and can produce content that is offensive, racist, biased, outdated, [incorrect](#) or simply inappropriate.

Snapchat’s chatbot, for example, [advised one reporter](#) (who was posing as a teenager) about how to mask the smell of alcohol or pot and suggested tips on having sex for the first time.

“It is very important for kids to know what is going on under the hood,” said Safinah Ali, a graduate student at M.I.T. who has taught elementary, middle and high school students about A.I.

S. Craig Watkins, a professor at the University of Texas, Austin, who has studied racial equity in A.I., said that children and parents should also be aware that this technology has “enormous blind spots” in terms of how it is designed and who it is designed for.

In one example, [a study](#) published last year found that A.I.-powered robots acted out “toxic stereotypes” around gender and race. And researchers have discovered that historical inequities are [baked into chatbots](#).

Understanding the technology’s potential for bias may give children and their parents reason to pause “and ask questions about their interactions and the content that is being generated for them,” Dr. Watkins said.

Stay on top of new developments

A.I. technology will continue to become an even larger part of our world.

Eventually, Google’s Bard chatbot is expected to be widely available. And Meta, the owner of Facebook, Instagram and WhatsApp, [announced](#) in February that it would begin integrating A.I. into its products.

A.I. is beginning to enter classrooms, too. Some teachers are using it to [plan their lessons or write emails](#). They’re showing students how chatbots can jump-start creativity by [suggesting ideas for experiments](#), creating outlines for essays, becoming a debate partner and [much more](#).

In addition, at a number of middle and high schools, students are being [taught](#) about different types of A.I., often with [curriculums](#) developed by teachers at M.I.T. Children can learn to design a robot, train a machine to learn something new or teach a computer to play a video game.

For those who don’t yet have access to A.I. in the classroom, Ms. Ali recommended that parents visit the [website](#) of RAISE (Responsible A.I. for Social Empowerment and Education), an M.I.T. initiative. The site offers conversation starters about ethical issues in A.I., the ways in which A.I. can be abused, and suggestions for using A.I. creatively and productively.

Given how prevalent the technology is becoming, everyone should have the opportunity to learn about it, Ms. Ali said. “A.I. will transform the nature of our jobs and children’s future careers,” she said.

A.I. Is Coming for Lawyers, Again

By [Steve Lohr](#), NYT, April 10, 2023

More than a decade ago, lawyers were singled out as an endangered occupational species, their livelihoods at risk from advances in artificial intelligence.

But the doomsayers got ahead of themselves. While clever software has taken over some of the [toil of legal work](#) — searching, reviewing and mining mountains of legal documents for nuggets of useful information — employment in the legal profession has grown faster than the American work force as a whole.

Today, a new A.I. threat looms, and lawyers may feel a bit of déjà vu. There are warnings that ChatGPT-style software, with its humanlike language fluency, could take over much of legal work. The new A.I. has its flaws, notably its proclivity to make things up, including fake legal citations. But proponents insist those are teething defects in a nascent technology — and fixable.

Will the pessimists finally be right?

Law is seen as the lucrative profession perhaps most at risk from the recent advances in A.I. because lawyers are essentially word merchants. And the new technology can

recognize and analyze words and generate text in an instant. It seems ready and able to perform tasks that are the bread and butter of lawyers.

“That is really, really powerful,” said Robert Plotkin, an intellectual property lawyer in Cambridge, Mass. “My work and my career has been mostly writing text.”

But unless the past isn’t a guide, the impact of the new technology is more likely to be a steadily rising tide than a sudden tidal wave. New A.I. technology will change the practice of law, and some jobs will be eliminated, but it also promises to make lawyers and paralegals more productive, and to create new roles. That is what happened after the introduction of other work-altering technologies like the personal computer and the internet.

[One new study](#), by researchers at Princeton University, the University of Pennsylvania and New York University, concluded that the industry most exposed to the new A.I. was “legal services.” [Another research report](#), by economists at Goldman Sachs, estimated that 44 percent of legal work could be automated. Only the work of office and administrative support jobs, at 46 percent, was higher.

Lawyers are only one occupation in the path of A.I. progress. [A study](#) by researchers at OpenAI, the creator of ChatGPT, and the University of Pennsylvania found that about 80 percent of American workers would have at least 10 percent of their tasks affected by the latest A.I. software.

The legal profession has been identified as a ripe target for A.I. automation in the past. In 2011, one article in a longer series in The New York Times on the progress in A.I. (titled “Smarter Than You Think”) focused on the likely impact on legal work. Its headline: [“Armies of Expensive Lawyers, Replaced by Cheaper Software.”](#)

But the march of A.I. in law turned out to be more measured. A.I. mainly identified, sorted and classified words in documents. The technology’s tools served more as aides than as replacements — and the same could be true this time.

In 2017, Baker McKenzie, a large international law firm, set up a committee to track emerging technology and set strategy. Since then, the A.I. software has made steady inroads.

“The reality is A.I. has not disrupted the legal industry,” said Ben Allgrove, a partner at the firm and its chief innovation officer.

The rapid progress in large language models — the technology engine for ChatGPT — is a significant advance, Mr. Allgrove said. Reading, analyzing and summarizing, he said, are fundamental legal skills. “At its best, the technology seems like a very smart paralegal, and it will improve,” he said.

The impact, Mr. Allgrove said, will be to force everyone in the profession, from paralegals to \$1,000-an-hour partners, to move up the skills ladder to stay ahead of the technology. The work of humans, he said, will increasingly be to focus on developing

industry expertise, exercising judgment in complex legal matters, and offering strategic guidance and building trusted relationships with clients.

Technology has eliminated large numbers of jobs in recent years, and not just robots taking over factories. Personal computers, productivity software and the internet have made office work more efficient, replacing many workers.

Office and administrative support occupations, including secretaries, clerks, bill collectors and office assistants, employ 1.3 million fewer workers than in 1990, according to an analysis by the Bureau of Labor Statistics. The Labor Department [forecasts further decline](#), with 880,000 fewer jobs in those occupations by 2031.

“Technology is a driver, and there are large changes, but they tend to come gradually over a decade or more,” said Michael Wolf, the division chief for occupational employment projections at the Bureau of Labor Statistics.

The bureau’s current outlook is that jobs for [lawyers](#) and [paralegals](#) will continue to grow faster than the labor market as a whole. Mr. Wolf is closely watching the arrival of the new A.I. software, but he said it was too early to assess what the technology’s long-term impact would be.

Lawyers are mostly putting the technology through test runs. The issues of data protection and client confidentiality are critical in legal work. The legal profession resisted using email until information-handling rules were established.

And the software models’ tendency to make up things confidently is alarming — and an invitation to malpractice suits — in a profession that hinges on finding and weighing facts.

To help address those concerns, law firms often use software that runs on top of something like ChatGPT and is fine-tuned for legal work. The tailored software has been developed by legal tech start-ups like [Casetext](#) and [Harvey](#).

Load in a case’s documents and ask the software to draft deposition questions, for example, and in a few minutes it will spit out a list of pertinent questions, lawyers say.

“For the things it can do well, it does them stunningly well,” said Bennett Borden, a partner and the chief data scientist at [DLA Piper](#), a large corporate law firm.

Successfully using the A.I. requires ample relevant data and questions that are detailed and specific, Mr. Borden said. More open-ended questions, like what’s the most important evidence, or who are the most credible witnesses, are still a struggle for the A.I.

Lawyers at big firms have seen significant time savings for certain jobs and view the technology as a tool to make teams of lawyers and paralegals more productive. Sole practitioners see the technology more as a partner in practice.

Valdemar L. Washington, a lawyer in Flint, Mich., was selected last fall to test the software from Casetext, called CoCounsel, which works with the latest ChatGPT technology.

Mr. Washington used the software in a suit against the City of Flint claiming that residents were overcharged on water and sewer rates and service fees. He loaded more than 400 pages of documents, and the software quickly reviewed them and wrote a summary that pointed him to an important gap in the defense’s case.

The program did in a few minutes what would have taken him several hours, he said.

“It’s a real game changer,” Mr. Washington said.

When A.I. Chatbots Hallucinate

By [Karen Weise](#) and [Cade Metz](#), NYT, May 1, 2023

When did The New York Times first report on “artificial intelligence”?

According to ChatGPT, it was July 10, 1956, in an article titled “Machines Will Be Capable of Learning, Solving Problems, Scientists Predict” about a seminal conference at Dartmouth College. The chatbot added:

CHATGPT:

This conference is now considered to be the birthplace of artificial intelligence as a field of study, and the article mentions the term “artificial intelligence” several times in the context of the discussions and presentations that took place at the conference.

The 1956 conference [was real](#). The article was not. ChatGPT simply made it up. ChatGPT doesn’t just get things wrong at times, it can fabricate information. Names and dates. Medical explanations. The plots of books. Internet addresses. Even historical events that never happened.

When ChatGPT was recently asked how James Joyce and Vladimir Lenin first met — there is no evidence they ever did — this is how it responded:

CHATGPT:

James Joyce and Vladimir Lenin met in Zurich, Switzerland in 1916. Both men were living in exile in Zurich during World War I. Joyce was a writer and Lenin was a revolutionary. They met at the Cafe Odéon, a popular gathering place for artists and intellectuals in Zurich.

Fabrications like these are common. Figuring out why chatbots make things up and how to solve the problem has become one of the most pressing issues facing researchers as the tech industry races toward the development of new A.I. systems.

Chatbots like ChatGPT are used by hundreds of millions of people for an increasingly wide array of tasks, including email services, online tutors and search engines. And they could change the way people interact with information. But there is no way of ensuring that these systems produce information that is accurate.

The technology, called generative A.I., relies on a complex algorithm that analyzes the way humans put words together on the internet. It does not decide what is true and what is not. That uncertainty has rAIsed concerns about the reliability of this new kind of artificial intelligence and calls into question how useful it can be until the issue is solved or controlled.

The tech industry often refers to the inaccuracies as “hallucinations.” But to some researchers, “hallucinations” is too much of a euphemism. Even [researchers within tech companies](#) worry that people will rely too heavily on these systems for medical and legal advice and other information they use to make daily decisions.

“If you don’t know an answer to a question already, I would not give the question to one of these systems,” said Subbarao Kambhampati, a professor and researcher of artificial intelligence at Arizona State University.

ChatGPT wasn’t alone in erring on the first reference to A.I. in The Times. Google’s Bard and Microsoft’s Bing chatbots both repeatedly provided inaccurate answers to the same question. Though false, the answers seemed plausible as they blurred and conflated people, events and ideas.

GOOGLE’S BARD:

*The earliest mention of artificial intelligence (AI) in The New York Times was in 1956, **when the term was coined by John McCarthy 1** at a workshop at Dartmouth College. The workshop brought together some of the brightest minds in computer science to discuss the possibility of creating machines that could think like humans.*

*The Times reported on the workshop in a **front-page article,2** calling it “**a new field of science**” 3 that was “aimed at creating machines that can think like humans.” The article **quoted McCarthy as saying 4** that the goal of AI was to “build machines with minds, capable of learning and adapting to new situations.”*

NEW YORK TIMES ANALYSIS:

1 Mr. McCarthy is credited with creating the term artificial intelligence in 1955, in a proposal for the Dartmouth conference.

2 The Times did not cover the Dartmouth workshop at all, let alone on the front page.

3 *More recent online articles refer to the conference as starting “a new field of science.”*

4 The Times could not find a reference to Mr. McCarthy ever saying this phrase, though it includes common ways to describe A.I.

Microsoft's Bing cited its findings to a realistic-looking web address on The Times's website:

MICROSOFT'S BING:

*Based on web search results, the earliest mention of artificial intelligence in the New York Times seems to be from a book review published on **March 5, 1956**. **1** The book is called "**The Human Use of Human Beings**" **2** by Norbert Wiener, and the reviewer is **John R. Pierce**. **3** The review mentions artificial intelligence as one of the topics covered by Wiener, who is described as "a pioneer in cybernetics": <https://www.nytimes.com/1956/03/05/archives/the-human-use-of-human-beings-cybernetics-and-society-by-norbert.html> **4***

NEW YORK TIMES ANALYSIS:

- 1** The Times published a review of a book by Mr. Weiner one day earlier, on March 4, 1956.
- 2** This seminal book was published in 1950, not 1956. The March 4 review was of Mr. Weiner's book "I Am A Mathematician: The Later Life of a Prodigy."
- 3** The March 4 review was not written by John R. Pierce, a noted electrical engineer. It was written by John G. Kemeny, a mathematician.
- 4** The format of this URL is consistent with Times archives for articles in this period, but this exact webpage never existed.

According to The Times's archives, all the chatbots were wrong. They cited articles that did not exist. And while coverage of early research on thinking machines [dated to the 1930s](#), it wasn't until 1963 that The Times [first published](#) an article with the phrase "artificial intelligence."

"We released Bard as an experiment and want to be as transparent as possible about well documented limitations," Jennifer Rodstrom, a spokeswoman for Google, said. "These are top of mind for us as we continue to fine tune Bard."

Like Google, Microsoft and OpenAI say they are working to reduce hallucinations.

The new AI. systems are "built to be persuasive, not truthful," an internal Microsoft document said. "This means that outputs can look very realistic but include statements that aren't true."

The chatbots are driven by a technology called a [large language model](#), or L.L.M., which learns its skills by analyzing massive amounts of digital text culled from the internet.

By pinpointing patterns in that data, an L.L.M. learns to do one thing in particular: [guess the next word in a sequence of words](#). It acts like a powerful version of an autocomplete tool. Given the sequence "The New York Times is a _____," it might guess "newspaper."

Because the internet is filled with untruthful information, the technology learns to repeat the same untruths. And sometimes the chatbots make things up. They produce new text, combining billions of patterns in unexpected ways. This means even if they learned solely from text that is accurate, they may still generate something that is not.

Because these systems learn from more data than humans could ever analyze, even A.I. experts cannot understand why they generate a particular sequence of text at a given moment. And if you ask the same question twice, they can generate different text.

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.



Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial," took first place in the digital category at the Colorado State Fair. Credit...via Jason Allen

By [Kevin Roose](#), NYT, Sept. 2, 2022

This year, the Colorado State Fair's annual art competition gave out prizes in all the usual categories: painting, quilting, sculpture.

But one entrant, Jason M. Allen of Pueblo West, Colo., didn't make his entry with a brush or a lump of clay. He created it with Midjourney, an artificial intelligence program that turns lines of text into hyper-realistic graphics.

Mr. Allen's work, "Théâtre D'opéra Spatial," took home the blue ribbon in the fair's contest for emerging digital artists — making it one of the first A.I.-generated pieces to win such a prize, and setting off a fierce backlash from artists who accused him of, essentially, cheating.

Reached by phone on Wednesday, Mr. Allen defended his work. He said that he had made clear that his work — which was submitted under the name "Jason M. Allen via Midjourney" — was created using A.I., and that he hadn't deceived anyone about its origins.

"I'm not going to apologize for it," he said. "I won, and I didn't break any rules."

These apps have made many human artists understandably nervous about their own futures — why would anyone pay for art, they wonder, when they could generate it themselves? They have also generated fierce debates about the ethics of A.I.-generated art, and opposition from people who claim that these apps are essentially a high-tech form of plagiarism.

Mr. Allen, 39, began experimenting with A.I.-generated art this year. He runs a studio, Incarnate Games, which makes tabletop games, and he was curious how the new breed of A.I. image generators would compare with the human artists whose works he commissioned.

This summer, he got invited to a Discord chat server where people were testing Midjourney, which uses a complex process known as "diffusion" to turn text into custom images. Users type a series of words in a message to Midjourney; the bot spits back an image seconds later.

Mr. Allen became obsessed, creating hundreds of images and marveling at how realistic they were. No matter what he typed, Midjourney seemed capable of making it.

"I couldn't believe what I was seeing," he said. "I felt like it was demonically inspired — like some otherworldly force was involved."

Eventually, Mr. Allen got the idea to submit one of his Midjourney creations to the Colorado State Fair, which had a division for "digital art/digitally manipulated photography." He had a local shop print the image on canvas and submitted it to the judges.

"The fair was coming up," he said, "and I thought: How wonderful would it be to demonstrate to people how great this art is?"

Several weeks later, while walking the fairground in Pueblo, Mr. Allen saw a blue ribbon hanging next to his piece. He had won the division, along with a \$300 prize.

“I couldn’t believe it,” he said. “I felt like: this is exactly what I set out to accomplish.”

After his win, Mr. Allen posted a photo of his prize work to the Midjourney Discord chat. It made its way to Twitter, where it [sparked](#) a furious backlash.

“We’re watching the death of artistry unfold right before our eyes,” one Twitter user wrote.

“This is so gross,” another wrote. “I can see how A.I. art can be beneficial, but claiming you’re an artist by generating one? Absolutely not.”

A Campaign Aide Didn’t Write That Email. A.I. Did.

The swift advance of artificial intelligence in politics is already blurring the boundaries between fact and fake.

By [Shane Goldmacher](#), NYT, March 28, 2023

The Democratic Party has begun testing the use of artificial intelligence to write first drafts of some fund-raising messages, appeals that often perform better than those written entirely by human beings.

Fake A.I. images of Donald J. Trump getting arrested in New York spread faster than they could be fact-checked last week.

And voice-cloning tools are producing vividly lifelike audio of President Biden — and many others — saying things they did not actually say.

Artificial intelligence isn’t just coming soon to the 2024 campaign trail. It’s already here.

The swift advance of A.I. promises to be as disruptive to the political sphere as to broader society. Now any amateur with a laptop can manufacture the kinds of convincing sounds and images that were once the domain of the most sophisticated digital players. This democratization of disinformation is blurring the boundaries between fact and fake at a moment when the acceptance of universal truths — that Mr. Biden beat Mr. Trump in 2020, for example — is already being strained.

And as synthetic media gets more believable, the question becomes: What happens when people can no longer trust their own eyes and ears?

Inside campaigns, artificial intelligence is expected to soon help perform mundane tasks that previously required fleets of interns. Republican and Democratic engineers alike are racing to develop tools to harness A.I. to make advertising more efficient, to engage in predictive analysis of public behavior, to write more and more personalized copy and to discover new patterns in mountains of voter data. The technology is evolving so fast that most predict a profound impact, even if specific ways in which it will upend the political system are more speculation than science.

“Unfortunately, I think people are going to figure out how to use this for evil faster than for improving civic life,” said Joe Rospars, who was chief strategist on Senator Elizabeth Warren’s 2020 campaign and is now the chief executive of a digital consultancy.

Those who work at the intersection of politics and technology return repeatedly to the same historical hypothetical: If the infamous “Access Hollywood” tape broke today — the one in which Mr. Trump is heard bragging about assaulting women and getting away with it — would Mr. Trump [acknowledge it was him](#), as he did in 2016?

The nearly universal answer was no.

“I think about that example all the time,” said Matt Hodges, who was the engineering director on Mr. Biden’s 2020 campaign and is now executive director of Zinc Labs, which invests in Democratic technology. Republicans, he said, “may not use ‘fake news’ anymore. It may be ‘Woke A.I.’”

For now, the frontline function of A.I. on campaigns is expected to be writing first drafts of the unending email and text cash solicitations.

“Given the amount of rote, asinine verbiage that gets produced in politics, people will put it to work,” said Luke Thompson, a Republican political strategist.

As an experiment, The New York Times asked ChatGPT to produce a fund-raising email for Mr. Trump. The app initially said, “I cannot take political sides or promote any political agenda.” But then it immediately provided a template of a potential Trump-like email:

*Dear (first name),
I hope this email finds you well. I’m reaching out to you today because I need your help to continue our fight against the radical left and their socialist agenda. We cannot let them win!
As you know, IO a committed to putting American first and fighting for the values that have made our country great. But we can’t do it alone. We need the support of people like you to keep our movement strong.
That’s why I’m asking you to make a contribution to our campaign today. Every dollar counts and will go towards supporting our efforts to defeat our opponents and make American great again.*

I promise you that we will continue to work tirelessly to protect our freedoms, secure our borders, and rebuild our economy. With your help we can achieve great things for our country.

Thank you in advance for your support. Together we will win this fight!

Among the prominent groups that have experimented with this tool is the Democratic National Committee, according to three people briefed on the efforts. In tests, the A.I.-generated content the D.N.C. has used has, as often as not, performed as well or better than copy drafted entirely by humans, in terms of generating engagement and donations.

Party officials still make edits to the A.I. drafts, the people familiar with the efforts said, and no A.I. messages have yet been written under the name of Mr. Biden or any other person, two people said. The D.N.C. declined to comment...

Political operatives, several of whom were granted anonymity to discuss potentially unsavory uses of artificial intelligence they are concerned about or planning to deploy, raised a raft of possibilities.

Some feared bad actors could leverage A.I. chatbots to distract or waste a campaign's precious staff time by pretending to be potential voters. Others floated producing deepfakes of their own candidate to generate personalized videos — thanking supporters for their donations, for example. In India, one [candidate in 2020](#) produced a deepfake to disseminate a video of himself speaking in different languages; the technology is far superior now.

Mr. Trump himself [shared an A.I. image in recent days](#) that appeared to show him kneeling in prayer. He posted it on Truth Social, his social media site, with no explanation.

One strategist predicted that the next generation of dirty tricks could be direct-to-voter misinformation that skips social media sites entirely. What if, this strategist said, an A.I. audio recording of a candidate was sent straight to the voice mail of voters on the eve of an election?...

Tommy Vietor, a former spokesman for Mr. Obama, dabbled with the A.I. tool from ElevenLabs to create a faux recording of Mr. Biden calling into the popular “Pod Save America” podcast that Mr. Vietor co-hosts. He paid a few dollars and uploaded real audio of Mr. Biden, and out came an audio likeness.

“The accuracy was just uncanny,” Mr. Vietor said in an interview.

The show labeled it clearly as A.I. But Mr. Vietor could not help noticing that some online commenters nonetheless seemed confused. “I started playing with the software thinking this is so much fun, this will be a great vehicle for jokes,” he said, “and finished thinking, ‘Oh God, this is going to be a big problem.’”