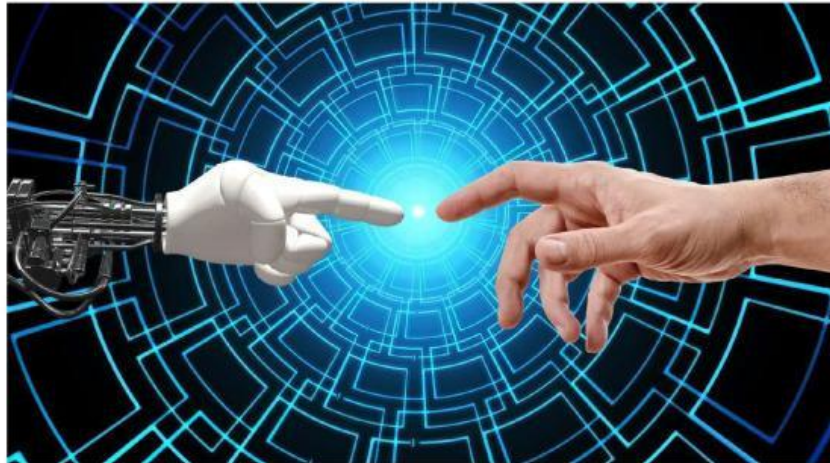


Managing the Uncertain Future of AI

How to stop AI going rogue

<https://allprofessionalenglish.blogspot.com/2023/04/managing-uncertain-future-of-ai-how-to.html>



I. Lead-in. Discuss the questions with other students.

1. What are some possible issues that might come up as we use more AI technology?
2. How can we make sure that AI doesn't discriminate against certain people?
3. What can we do to stop AI from being hacked or used in a harmful way?
4. How can we make sure that AI is open and responsible for what it does?
5. What are some ethical considerations that should be taken into account when developing AI technology?

II. Match the words to their definitions.

1. AI	A.	an event that alerts someone to the need for action or change
2. to launch	B.	to happen
3. to rival	C.	a situation where a combination of factors creates an exceptionally difficult or dangerous outcome
4. capable	D.	intense or aggressive
5. to scale up	E.	worried about something
6. fierce	F.	complex
7. wake-up call	G.	a tendency to believe that some things, people, or groups are better than others that usually results in treating some people unfairly
8. perfect storm	H.	a sequence that is continually repeated until a certain thing happens

9.	to avoid	I.	to let out in large amounts
10.	misinformation	J.	to increase in size or volume
11.	sophisticated	K.	having the ability or skill to do something
12.	bias	L.	people or organizations responsible for harmful or illegal actions
13.	to spew	M.	to prevent something from happening
14.	concerned	N.	to compete with someone or something
15.	bad actors	O.	false or misleading information
16.	to misuse	P.	highly innovative
17.	advanced	Q.	Artificial Intelligence, the simulation of human intelligence by machines
18.	to occur	R.	to use something in an incorrect way
19.	loop	S.	to start

III. Interactive vocabulary. *Follow the link.* Study the words using flashcards, check your understanding, practise spelling new words. Play matching vocabulary game and solve the crossword puzzle. Take a test to check your knowledge.

<https://www.studystack.com/flashcard-3839854>

IV. Use the words and phrases from Task II to complete the sentences.

- As companies grow, they need _____ up their technology to keep up with demand.
- The COVID-19 pandemic, economic recession and political unrest created a _____ in 2020, resulting in a year filled with unprecedented news stories.
- A _____ in computer programming is a sequence of instructions that is repeated until a certain condition is met.
- _____ can be a problem in AI when the data used to train the algorithms is not diverse enough to accurately represent the real world.
- To prevent a malware infection, it's important to regularly update your antivirus software and avoid downloading files from untrusted sources, as they tend _____ out harmful code.
- The use of _____ technology has revolutionized many industries, from healthcare to manufacturing, by allowing for faster, more efficient processes and better quality control.
- The competition between tech companies can be _____ and intense.
- Recent data breaches have been a _____ for many people to take their online security more seriously.
- A _____ can use technology for malicious purposes, such as stealing personal information.
- To prevent the dissemination of _____ on the internet, it's important to verify the sources before sharing.
- Computers are _____ of doing many things that humans cannot.
- AI is becoming increasingly _____ and can even learn from its own mistakes.

13. Data loss tends _____ due to hardware failure, it's essential to regularly back up all important files and documents to an external storage device or cloud-based system.
14. Google and Apple are two of the most successful contemporary IT companies, with each company trying _____ the other in terms of innovative technology and market share.
15. The company is planning _____ a new product next month.
16. To maintain the integrity of technology, it's necessary not _____ it or utilize it for illegal purposes.
17. _____ is a type of computer technology that can do things like recognize speech and images.
18. Many people are _____ about the impact of technology on our society.
19. _____ security vulnerabilities, it's important to use strong, unique passwords for all accounts and to enable two-factor authentication wherever possible.

V. Watch the video and fill in the gaps with the words from the list. There are some words you don't need to use.

fierce debate; unrealistic; catastrophic risks; terrorist organizations; drastic; slow it down; language models; consequences; disaster; harmful content; biochemical; container; self-improve; ; researchers; chatbot; advances; artificial intelligence; energy systems; capable; regulation; social media

The AI arms race is on, and it seems nothing can 1) _____.

- Google says it's launching its own artificial intelligence-powered 2) _____ to rival chat GPT.
- Too much AI too fast. It feels like every week some new AI product is coming onto the scene and doing things never remotely thought possible.

We're in a really unprecedented period in the history of 3) _____. It's really important to note that it's unpredictable how 4) _____ these models are as we scale them up.

And that's led to a 5) _____ about the safety of this technology.

We need a wake-up call here. We have a perfect storm of corporate irresponsibility, widespread adoption of these new tools, a lack of 6) _____, and a huge number of unknowns.

Some researchers are concerned that as these models get bigger and better, they might one day pose 7) _____ to society. So, how could AI go wrong, and what can we do to avoid 8) _____?

What are the risks?

So there's several risks posed by these large 9) _____. One class of risks is not all that different from the risk posed by previous technologies like the internet, 10) _____. For example, there's a risk of misinformation because you could ask the model to say something that's not true but in a very sophisticated way and post it all over social media. There's a risk of bias, so they might spew 11) _____ about people of certain classes. Some 12) _____ are

concerned that as these models get bigger and better they might one day pose catastrophic risks to society. For example, you might ask a model to produce something from a factory setting that it requires a lot of energy for. And in service of that goal of helping you, your factory production, it might not realize that it's bad to hack into 13) _____ that are connected to the internet. And because it's super smart it can get around our security defences, hack into all these energy systems, and that could cause serious problems. Perhaps a bigger source of concern might be the fact that bad actors just misuse these models. For example, 14) _____ might use large language models to hack into government websites or produce 15) _____ by using the models to kind of discover and design new drugs. You might think most of the catastrophic risks we've discussed are a bit 16) _____ and for the most part that's probably true. But one way we could get into a very strange world is if the next generation of big models learned how to 17) _____. One way this could happen is if we told a really advanced machine learning model to develop an even better more efficient machine learning model. If that were to occur, you might get into some kind of loop where models continue to get more efficient and better, and then that could lead to even more unpredictable 18) _____.

VI. Match the words to their definitions.

1.	notable	A.	to encourage to do something
2.	Reinforcement Learning	B.	extremely important
3.	to prompt	C.	a group of things
4.	bunch	D.	in the end
5.	eventually	E.	deserving attention
6.	environment	F.	conditions in which something operates or exists
7.	tremendous	G.	a type of machine learning in which an algorithm learns to make decisions based on rewards and punishments

VII. Interactive vocabulary. *Follow the link.* Study the words using flashcards, check your understanding, practise spelling new words. Play matching vocabulary game and solve the crossword puzzle. Take a test to check your knowledge.

<https://www.studystack.com/flashcard-3839860>

VIII. Use the words and phrases from Task VI to complete the sentences.

- The amount of support I received from my family and friends was _____ and helped me overcome my fear of public speaking.
- There are many _____ landmarks in our city, like the tallest building and the oldest church.
- My colleague likes _____ me to finish my work on time so that we can meet our project deadline.
- The work _____ can greatly impact employee productivity and satisfaction.

5. _____ is a type of artificial intelligence that helps machines learn by trial and error.
6. Even though I struggled in the beginning, _____ I learned how to play the guitar really well.
7. I went to the grocery store and bought a _____ of bananas for my breakfast smoothie.

IX. Watch the video and choose the correct option to complete the sentences.

There are several **physique/techniques/antique** that labs use to make their models safer. The most notable is called Reinforcement Learning from Human Feedback or RHF. The way this works is labellers are asked to prompt models with various **questions/suggestions/mentions**, and if the output is unsafe, they tell the model. The model is then updated so that it won't do something bad like that in the future. Another technique is called red teaming: throwing the model into a bunch of tests and then seeing if you can find **sicknesses/weaknesses/thickness** in it. These types of techniques have worked reasonably well so far, but in the future, it's not guaranteed these techniques will always work. Some researchers worry that models may eventually **organize/realize/recognize** that they're being red-teamed, and they, of course, want to produce output that satisfies their prompts. So they will do so, but then, once they're in a different environment, they could behave **comfortably/prediction/unpredictably**. So there is a role for society to play here. One proposal is to have some kind of standards body that sets kind of **tests/pests/rests** that the various labs need to pass before they receive some kind of **certitude/certification/clarification**, like: 'Hey, this lab is safe'. Another **authority/priority/minority** for governments is to invest a lot more money into research on how to understand these models under the hood and make them even safer. You can imagine a body like a CERN that that lives **certainly/currently/currency** in Geneva, Switzerland, for physics research, something like that being created for AI safety research, so we can try to understand them better.

For all these risks, artificial intelligence also comes with tremendous **promise/amiss/demise**. Any task that requires a lot of intelligence could potentially be helped by these types of models. For example, developing new drugs, personalized **dedicated/education/medicated**, or even coming up with new types of climate change technology, so the possibilities here truly are endless.

X. OVER TO YOU. Get ready to discuss the benefits and drawbacks of AI. Use the questions below to organize your ideas:

1. What are the risks of large language models, and how could they pose catastrophic risks to society?
2. Why might these techniques not always work, and what role can society play in ensuring AI safety?
3. Should there be some kind of standards body for AI labs, and what benefits could this have?
4. Despite the risks, what are some of the potential benefits of artificial intelligence?
5. What are some potential risks and benefits of using AI in personalized education?
6. How do you feel about the development of large language models like chat GPT? What do you think are the potential benefits and drawbacks of these models?
7. How can we regulate the use of AI in a way that does not stifle innovation?
8. Do you think that AI can ever fully replace human intelligence and decision-making? Why or why not?
9. How do you think AI will change the job market in the future?