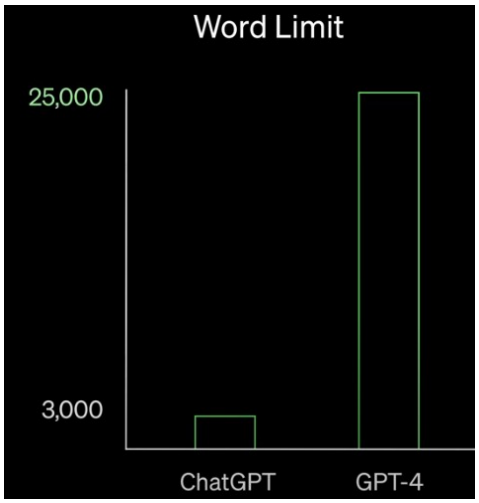
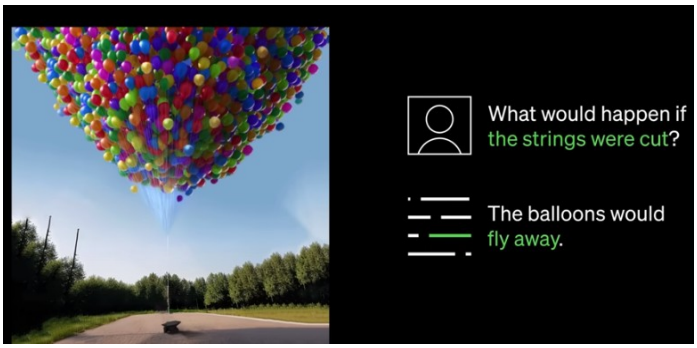
ChatGTP Teach-out (University of Michigan)

GTP = Generative Pre-trained Transformer
Developed by OpenAI
Trained to generate any type of human text
Used to create articles, poetry, news reports, dialogs;
Used for automaed conversational tasks like customer service chat bots
Trained on vast body of internet text to spot patterns in speech and languages
User inputs a little text as a few sentences, the trained system analyses the language and uses a text predictor to create the most likely output of high quality and feels similar to what humans would say of write.
Response is factual answers, but also common sense
https://www.copy.ai/ takes a short written document and generates original usable marketing copy
Shortcomings and risk: limited learning; does not keep learning, inablity to explain and interpret output, small input size which can limit some applications, wide ranges of bias as can found online (same as human bias)

Google Bard
ChatGPT is for free, people using it are voting the answers; so labeling feedback to make it better. For that reason other companies are also rollingout their version, to profit from the feedback to increase performance. Strength of the algorithme depends stongly on the number of users
Google Bard can be better then ChatGPT, but ChatGPT has more users and in practice it will be better than Google Bard.

ChatGPT-4





Internal guardrails are put in:
Adversarial usage, unwanted contect, privacy concerns
Use case education; adapt to all skill levels

Trustworthiness (betrouwbaarheid) becomes increasingly important.

I believe regulation is now very important with respect to what can and cannot be done with these tools

Writing Good Prompts for ChatGPT

prompt engineering

Trail and error like Google search
Systems will change periodically, so different results in time.
Input that is too vague, then what you get back will be similarly vague
The system will try to produce anything, regardless the vague input.
Language should be relatively generic

Dall-E creates images

Convolutional Networks: Images
Recurrent Neural Network (RNN) Disadvantage => they are sequential (so no context)

BERT: Transformer Based Model == Positional Encoding, Attention, Self-Attention

Alignment Problem: AI systems sometimes becomde less inclined (neigend) to follow human intentions. So how we make sure that the system actually acts in accordance with **human intensions** and in accordance with **human values**.
Systems need to follow human instructions => learn from human **feedback**. This makes the system more useful more reliable and more trustworthy.
When system become more powerful, allignment will become more critical.

Ethical implications of ChatGPT

Jack will discuss plagiarism, bias, and discrimination.
Unethical and ethical use

Jack Bernard, University of Michigan (**Plagiarism & Attribution**)
There are lots of ways to use ChatGPT and other AI technologies without engaging in some plagiarism, but I don't think that it's hard to slip into the world where someone might have plagiarized
It's easy when you've run out of time to lean on another source: go to Wikipedia or they go to someone else's article and just cut bits of it and put it into their own work without acknowledging it.
Plagiarism is not illegal violation. It's a violation of academic norms. When someone plagiarizes, they don't acknowledge the source of specific language or even of their thinking or they attribute it to somebody else, like adding someone else's name to your paper in order perhaps to attract more readers.
I created this paper and I use ChatGPT to help me. That's a possibility. How faculty will respond to that, that's a whole new question.
I think just because ChatGPT looks across a large corpus of information to produce an individualized independent response to a query that draws on this experience that it's had reading lots of other materials, that doesn't make it plagiarism in the same way that it doesn't make it plagiarism for me to have formulated all these words that I'm producing right now from all of my experiences over a lifetime. I think ChatGPT is doing essentially what people do, maybe not exactly the same way, but in ways that are actually very similar, which is, it accumulates knowledge and assembles a perspective based on the AI.
Just assembling an amalgam of this information it's been exposed to and creating original expression that in itself is not plagiarism, that's just communication from a piece of technology.

It's difficult to know what kinds of citation culture we're going to create around these smart technologies, these AI technologies.
If it were me, I would err on the side of letting people know that this is what I usedbecause I think it's about integrity.
Students and faculty should make their best efforts to create normative standards were their expectations about what people can and cannot use.
In some classrooms, faculty members will be using ChatGPT as part of the teaching model.

Jack Bernard, University of Michigan (**Bias & Discrimination**)
The AI itself probably isn't discriminating the way a person might discriminate. It is true that AI could be influenced by the discrimination that's replete in society, the bigotry that exists in society, the perspectives of society.
Human beings are filled with all kinds of biases and while they're filled with great beauty, they're also filled with an enormous amount of ugliness.
AI technology has reached the state of consciousness yet. It's certainly possible, it strikes me that consciousness could arise in other kinds of assemblages of chemicals and atoms.

Another two ways I can think of where we might see discrimination exist:

- maybe the entities that make ChatGPT or other kinds of AI available won't put them into a format that's accessible to people who have disabilities.That means that some people who are members of society won't be able to get the benefit of these technologies, and that would be a shame.
- Because they can't afford access.

If they're not available to the general public, well, then only some people will have the advantage of labor saving technologies like smart AI. This may perpetuate the kinds of discrimination that already exist in society. If we are not careful, there will be the haves and the have nots.

Julie Hui (**ChatGPT & Equitable Use**)
Writing good search prompts or search queries is a skill and it's very, very related to your writing or communication ability.

Ethical vs. Unethical Use
Unethical:

- ChatGPT could be considered unethical if the person is expecting a personally written message or if they're using the writing as a measure of the writer's ability. So misuse of ChatGPT.
- Writing a personal email, or a personal message to a friend or family member, you wouldn't really want those messages to be written by an AI.
- might be unethical or problematic to use ChatGPT is if people are sending messages that are written by ChatGPT without really reviewing what it says. Can probably lead to a lot of miscommunication errors in the future.
-

So maybe in the future, you might imagine that there are certain acceptable places to use ChatGPT. I could imagine like things that are a bit repetitive that nobody really cares if somebody had written that particular piece of writing.

Ethical:

- Summarizing a very long discussion for work, or easy search queries, like how to instructions. If you go into ChatGPT, they have some suggestions. Like, how do you do

this? Can you tell me more about quantum physics, and kind of using it like some sort of learning tool?
- So I can see a situation or, or many situations now where you might start off with what ChatGPT has written, and then you might edit it to make it more accurate or more sounding like you had written it yourself to your personal touches.
-

All answers they take some sort of information from somewhere, so it could be biased and you always have to take that into account.

Always making sure that you understand what you are sending, and kind of agreeing with what you've kind of co written with ChatGPT.

Rada Mihalcea (**limitations of ChatGPT**)
Factuality. That means that the output that it produces may be correct for say, 90 percent of the time, but it will be incorrect for 10 percent of the time.
Hallucinations, where it will produce some texts. It looks good in terms of correct, say English language, but is not factually correct. It just made something up. Because it sounds self-confident, it's sometimes hard to spot.
Another place where there is limitations is **reasoning** so in terms of there are two different sides to language. One is for the formal language or the surface, which is really what we primarily see when we interact with tools like this, but then is also the functional piece.
Farther down and ask for situations. It can start breaking because it can actually reason about what's possible or what's common sense.
Making maybe causal relations between facts, or trying to get it to infer things based on a couple of statements which we people can do naturally. It's hard for these systems to do that.

There is still some bias, although I believe that there's been a lot of progress with respect to having these models having interactions that are what we would want to see, so interactions that are not sexist, not racist. But there are ways to jailbreak that.

I think not knowing what it doesn't know. That uncovering what is compared to what was provided, what will be a new piece of knowledge that should be added. That's another weakness. It will usually produce what's helpful, but is mostly the obvious. We say now, leave those aside, tell me something new. That's a hard task for models like this.

Kentaro Toyama (**address the negative impacts of ChatGPT**)
How to regulate AI, and I don't think we can do it soon enough.
Just like with nuclear power which is extremely tightly regulated, we also need to be thinking similarly about AI
People or organizations that put out AI systems that go rogue need to be fully held responsible for that fact. And, of course that will immediately slow down the eagerness with which people or companies put out this technology, but in fact, already we've seen examples of this.
Technologies like chatGPT, another thing is that we might want to suggest that any such system keep a very good record of every conversation has had, every text that is generated. And for a variety of reasons, so that we can actually go back and inspect what might have gone wrong if something goes wrong. The creators of the technology to feel and experience the legal responsibility should something go wrong with that technology.
The harder things:
Are we prepared to let it take away jobs? And if so, what's the kind of policy around that? It's going to be inevitable that those things are going to happen, we haven't tried to curb that from

happening with, let's say manufacturing robots, even though they've taken away jobs in certain contexts.
These effects of what happens once we start using these technologies and become so habituated to it that we don't even realize the impact on us?

singularity which is the moment when artificial intelligence equals human intelligence.

The feature, called "AI text classifier," is similar to the plagiarism software Turnitin in that when submitting a body of text, the tool will rate the input on a scale ranging from "likely generated by AI" to "very unlikely." But is imperfect.

GPTZero, created by student at Princeton, can detect whether text is written by ChatGPT.
To determine whether an excerpt is written by a bot, GPTZero uses two indicators: "perplexity" and "burstiness." Perplexity measures the complexity of text; if GPTZero is perplexed by the text, then it has a high complexity and it's more likely to be human-written. However, if the text is more familiar to the bot — because it's been trained on such data — then it will have low complexity and therefore is more likely to be AI-generated.

AI safety at OpenAI, revealed that the company has been working on a way to "watermark" GPT-generated text with an "unnoticeable secret signal" to identify its source.

Addressing the question of whether humans have a right to know if they are interacting with AI.

Legal implications of ChatGTP

Jack Bernard (**Output Ownership & Monetization**)

in US copyright law,anything that's not created by a person can't get copyright protection.
One of the challenges I think that we face in plumbing this particular set of depths is that how will we know what portions of the work were created by the AI and what portions were created by a person.

Open AI will absolutely find a way to monetize this. They have investors, Microsoft has invested in them already to the tune of billions of dollars. So there is no question, they will find a market for monetizing this to generate revenue. Whether I will be able to monetize the output, it's hard to say.

What are the ethical issues with respect to monetizing the output of ChatGPT?
I think in general, we need to cite our sources, we need to acknowledge our sources. And so there's the risk that somebody isn't doing that, that is that they're taking stuff that's produced by ChatGPT pretending that they were the author of that work and then putting it out there. Now it's not necessarily unlawful to do that. We'll have to see how the law plays out in this area.
The idea of giving people credit or fame or opportunity or access as a result of what they've produced, that actually was produced by AI, I think that raises some concerns for us as a society.

Implications of Using Different Forms of AI

An interesting way to think about this is as technology infuses itself into society, we do less work. We'll be able to focus more on critical thinking, rather just on rote memorization or who knows which facts.
I wonder if we'll now be focusing on things like writing and how we teach it differently and also what we measure in the academy differently.

When I took standardized tests to go to college, you couldn't use a calculator. Now, it's very commonplace in standardized tests to use a calculator in order to complete the exam and our thinking changed.

Data privacy

People they might include information, typed at the prompt, that ordinarily they would think of as private,
not realizing that when you put that into the software, you may be exposing information you aren't supposed to expose. For instance, instead of saying the name and medical information of a patient in order to get ChatGPT to write a letter or to make a note describing what your experiences were with that patient, a doctor might want to not use the name of the patient when making that description.
Just think it's important to leave out information that will be absorbed into the database of the technologies that you're using and infused into the background data from which the AI draws information.
You can talk about very sensitive issues without exposing that information, you can create pseudonyms for individuals if you need to talk about an individual name.
I'm not sure what the companies like OpenAI can do to protect private information from being exposed through the AI, I'm not sure that there's going to be an easy way to recognize fact from fiction.

Government Regulation
We should always be wary of giving the government too much reach into silencing speech, even speech that we don't like.
I think we want to be careful about seeding that kind of control so early, there may be circumstances in which we really want to create regulations where we notice a particular problem with these technologies but I don't know where that is and where that will be at.

Images of people who never existed, or images of people who exist, and placing those images, say, on a different body or making a person look like they're in a place where they are not.
We want to be careful about those kinds of technologies and I think we will see examples of ways that smart AI can be manipulated by people in order to inspire misconduct of some kind or another.

I'm not sure whether these kinds of smart AI technology can be regulated effectively. I just don't know yet, I think it's still too early to know. People can be regulated, and maybe there will be mechanisms in place for us to know a great deal more about who's making what kinds of uses with technologies.

Impact on Society and the Economy

Scott Page (**Society's Response to New Technology**)
They're really good at doing a lot of things, but the things they're good at are different than the things humans are good at. And so that potential, that combination of ability and diversity portends really good things for society, in my opinion.

There's almost too much optimism at first, but then there's too little optimism about the long-term impact. And the reason why is if you look at how these things play out initially, people use the new technology to do what we're currently doing maybe a little bit better. Then what happens is people realize, wait, I can do these other things that I couldn't do before. So people start seeing new functionalities that they didn't expect. And then eventually, what will happen is society will reorganize itself in structural ways around this new technology.

Not the physical world, but the world of information in the world of work, the types of jobs people have are going to get restructured. (prompt engineering)

Short and Long-term Economic Impacts of ChatGPT

So everybody's human capital has suddenly just been kind of lifted up and in particular dimensions where you're just really bad at it, you're suddenly at a pretty decent baseline level

The challenge is how do we use universities, how to employers, how to community colleges, how does their educational system. How does the government help people recognize that this is a tool they can use to as sort of sort of augment their own human capital to enable them to live more productive lives.

Short and Long-term Societal Impacts of ChatGPT

In terms of what it means socially. I think that's so much harder to predict because there's functionalities that this has in terms of just asking it.
What's the impact on the economy, what's the impact on society? I think you want to view this from a very kind of evolutionary perspective and that is people are experimenting like crazy. There's going to be these kind of like survival of the fittest dynamics, like things that are really cool, things that are really productive. They're going to get copied and they're going to just grow in terms of whether it's within business, whether it's in society that you're just going to get this kind of mimicry. This is a really good use for this.

Two big questions:

- whether it stays passive or becomes active in the sense that right now my phone isn't interrupting me.
- How is it going to affect just the structure of our minds in the sense that there's a thing called the Flynn effect basically shows that people are getting smarter than they used to be.

Crystallized intelligence which is just kind of a knowledge base. And the second is **fluid intelligence**, how good am I like solving logic troubles, finding connections.
Crystallized intelligence has gone up over the last 50 years, basically because of education and nutrition. But over the last 25 years, it's flattened. It's just going to baseline.
But fluid intelligence just keeps going up and up because we're just exposed to LEGO sets, puzzles, games,, all these like things where we think and interact, this is just going to ramp that up. So our brains will probably get even better at kind of the because we're being asked to be really fluid, really creative, ask questions, do things, we're not being asked to give facts. And so it will literally be the case that our brains will be restructured as a result of this with a swing towards more fluid intelligence, which should be.

Michael Wellman (Econimic Responses to AI)
Technologies like ChatGPT are potentially quite disruptive to a lot of economic activity. This is really true for a lot of other AI technology today.
If it suddenly becomes much cheaper and easier to do by automated means through AI, it might first enable all kinds of new applications that were not cost-effective to try before.
It can also enable a lot of AI decision-making that was possible before but limited in its applicability because of the need to communicate results by language. If all of a sudden AI is able to communicate fairly credibly, fairly fluidly with people, it's going to enable many applications that we didn't see before.
How will that affect current economic actors and the economy?

- One is substitution. If there's some other way that something is being done that way maybe pushed out by the new, cheaper, more effective way perhaps.
- Other possibility is compliments. When a technology can amplify the effectiveness or reduce the costs of a person doing something, they can do it better, more efficiently by the addition of AI that may lead to more opportunities for those people to be more productive and to produce better products.

I think typically both of those factors are in play.

Effect on Human Capital

The effect that human capital will be very much affected by the balance of the substitute and complement effects. To the extent that the AI can do exactly what a human creative knowledge worker could do before, that may have some displacement, or it may shift the kinds of tasks that the human does versus what the AI does. To the extent that they're complementary, it can really amplify the effects of the content creator.

Impact on Education

A lot of writers have trouble just getting started. I think that's where a lot of the value is with ChatGPT is that it gets you over the hurdle of the white blank page that we we all fear when you have to write your essay for the first time, it just starts writing for you.