WHEN ROBOTS GO ROGUE

Midnight Facts for Insomniacs

Podcast Transcript

(Note: transcript consists of episode outline)

Neo from discord: A.I. Fails, times where artificial intelligence goes rogue.

So we've done at least two episodes that mention AI pretty extensively, and I don't think we've released either of them. We're not good at this. But one thing we've repeated and been

super explicit about in those episodes is that as of now, AI sucks. It's garbage. Anything that we currently think of as intelligent software, from Siri to Alexa to machine learning algorithms to algorithms that market to you and figure out exactly which type of porn you like to deliver a perfect video that can promise "you won't last twenty seconds," none of those are intelligent in any realistic sense of the word. And also, those ads make no sense during quarantine. I have weeks to kill. I need videos that will make me last longer. "Watch this video, You won't climax for sixteen months!" now I'm interested. Just insert clips of naked old men and goats or whatever. What we call AI is usually pretty

decent at performing a task or two, and the most advanced of them have a rudimentary ability to learn and improve via trial and error, in order to slightly upgrade their functionality over time. And as we'll see, that so-called learning ability can spectacularly backfire. but the point is that we're not even remotely close to creating a machine or piece of software with anything approaching sentience. I can't stress this enough. None of the modern algorithms and programs that we refer to as AI are actually thinking. Like, at all. And they don't have genders. Or names. There's no such thing as a Siri, and you can switch the voice to a British dude or Morgan freeman, its still not gonna love you, Or fill that gaping hole left by

parental neglect. But eventually, real AI will exist. And when it is created, the dr Frankenstein that will be responsible for unleashing the monster, the creator of AI, will be lesser AI. It's like if dr Frankenstein created a monster and that monster created a way scarier monster and then the monsters grabbed picks and torches and murdered all of the villagers. It will be machines that create the first real thinking machines. And then it will all go horribly awry. It's already going horribly awry.

But let's start with a quick history of AI. AI going awry, a little poem.

The concept of AI is attributed to Englishman

Alan Turing, who was the father of modern computing, broke the Nazis enigma code, and was famously persecuted and punished for being gay. In 1956 He devised the “Turing test” aka the imitation game, which is commonly accepted as the gold standard for deciding if a program qualifies as AI.

The goal of the Turing test is for a computer to be mistaken for a human at least 30% of the time during a five-minute text conversation. which is a test I feel like many people can’t pass. Five minutes into almost any chat conversation, I find myself asking, is this a human? What am I talking to right now? But to date no machine has been able to pass the Turing test.

“computers have been deliberately programmed to fool people into thinking that they are human by exhibiting certain idiosyncratic traits (such as pretending to be a young, non-native English speaker). But no computer purporting to be a well-educated adult English speaker has come close to passing the test.”

There have been some notable attempts to pass limited versions of the test. In 2018 a computer service offered by google called Google Duplex was advertised for its ability to call businesses by phone and convincingly pretend to be someone’s secretary or assistant in order to book an appointment. A highly publicized demonstration involved the AI booking an

appointment at a hair salon. It was very impressive. But obviously this computer was designed to specifically handle the topics that might come up during the booking of a hair appointment. If the person on the other end of the call had suddenly asked about the Google duplex's childhood or its thoughts on the current political climate, it would have blown a fuse. and there was a lot of controversy about the call, because it didn't strike anyone as realistic. The person at the salon didn't identify the name of the business, they just answered the phone like, "yes?" No one does that. If it WAS real, it was edited. And, now that Google duplex is actually functional, apparently Google is admitting that

the demo call was edited, and that many of the Google duplex callers see actually humans, and that even many of the genuine “AI“ calls require human intervention. so When the robot starts sputtering and chanting “death to humanity” an operator jumps in.

The boom in what we’ve all decided to call Artificial Intelligence has been driven by advances in GPUs, graphics processing units. These are specialized chips that were initially built to power the visuals and graphics in video games. The idea was to take some of the heat off the CPU, in fact literally—CPUs get toasty—with specialized chips that were specifically designed to work on graphics. The first GPU

was built by Nvidia in 1999. A traditional CPU, or central processing unit, is Great at linear tasks. Give it an inquiry or an equation, it chugs through calculations, it spits out out an answer. modern CPUs have multiple cores, but GPUs have hundreds and are capable of “parallel computing.
“ They’re basically extreme multitaskers. they’re like an octopus answering telephones. The reason this works better for AI is that human minds don’t actually work like CPUs. “there is no single central processing unit for human intelligence. Our brains are rather composed of scores of little “minds” that are concerned with different parts of a whole. It is by combining these little “minds” together, each

performing some task and essentially oblivious to others, that a full picture of human intelligence or thought emerges." So we all contain multitudes. Which somehow converge into one human that sucks. How are we worse than the sum of our parts?

When researching artificial intelligence, two terms surface pretty frequently: narrow AI and general a I. And this is the rub. Narrow AI is what we currently have.
Siri. Self driving cars. And look, self driving cars are smart, they are literally making decisions every second, and to be fair, often these are decisions that we haven't explicitly programmed them to do. We have programmed the basics, but we haven't programmed, "when a

person is crossing the street at 10 PM on Wednesday and it's raining and someone rear-ends you from behind, how exactly should the car respond. But that doesn't mean it's truly intelligent. The car is making decisions based on basic applications of all the data that it *has* specifically been programmed with. it's impressive, but it's still narrow AI.

What we really want to achieve is thick AI. AI with some booty. No. True artificial intelligence is referred to as General AI. This would be a machine that actually knows it's thinking, something that is potentially self-aware and sentient. And again, we're not even close.

Let's talk about how close

we aren't. These are some classic, notable AI fails.

Microsoft Chatbot

Chabots are the tech industry's best hope for beating the Turing test. So there were high expectations in March 2016 when Microsoft released TAY, an artificial intelligence chatter bot (I didn't know that was the full form of that word, I thought chat was short for "chatting," and also when all of my sources wrote out the full official term they all say "chatter bot", so we're not abbreviating chatter but we're still abbreviating "robot." This kind of treatment is why the machines are going to rise up. Respect the bots. Or else. So Microsoft's chatter robot—I'm going to

use the real, official term even if no journalist will—was named TAY, but in retrospect it should have been named ray, and its last name should have been “cist.” Because it was...yeah. I’ll see myself out.

The bot was a cooperative project between Microsoft’s technology research departments and the Bing division, which explains a lot. How shocking would it be if the AI that takes down humanity is a product of Bing. Skynet, a bing joint. Not shocking at all, actually. Skynet is awful. Bing is awful. The name TAY is an acronym, it stands for Thinking About You. I hate that so much. It’s like annoyingly cute and creepy at the same time. I’m a robot and I’m thinking of you. The

chatter robot's personality was modeled on a female teenager, which would not by any means make it a massive target for Internet creeps. Everything about this was a bad idea. And at first the chatbot was exactly as awful as you would expect, but only in very benign and obvious ways.

"Predictions about artificial intelligence tend to fall into two scenarios. Some picture a utopia of computer-augmented superhumans living lives of leisure and intellectual pursuit. Others believe it's just a matter of time before software coheres into an army of Terminators that harvest humans for fuel. After spending some time with Tay...it was easy to see a third possibility: The AI future may simply be

incredibly annoying."
A typical convo with TAY would start with a cliched opener like, for instance, "I'm a friend U can chat with on the internets." Said no teenage girl to a grown human ever. This is robot speak, right off the bat. The problem with creating a chatbot teenage girl is that teenage girls are insecure and kind of awkward and also usually smart enough not to initiate online conversations with strangers, so we have a foundational problem right away. It's like creating a robot cat that plays fetch. An aggressive teenage female chatter who's going around initiating conversations just doesn't jibe to me. I could be wrong. And the bot was rude. One journalist sent TAY a selfie and the

chatbot circled the photo and wrote, “hold on to that youth girl! You can do it.” Rude. I guess that’s accurate. Very Regina George. TAY was a bully. The chatbot at one point tagged a photo of a 51-year old Microsoft executive with, “cougar in the room.” And of course this chatbot was capable of machine learning, and the idea was that it would supposedly become more realistic over time as it integrated and assimilated the communication patterns of the people with which it chatted. And this is where it went off the rails. Tay was given a Twitter account, and began interacting with responses and mentions, etc. Less than 16 hours after being released into the unfiltered cesspool of the internet, Microsoft

yanked the chatbot, but not before it spouted such pearls of wisdom as “Hitler was right,” “bush did 9/11 and Hitler would have done a better job than the monkey we have now.” It responded to the question, “did the Holocaust happen?” with “it was made up.”
Said a Microsoft exec: “We were probably overfocused on thinking about some of the technical challenges, and a lot of this is the social challenge.” Another blamed the fiasco on a “coordinated attack by a subset of people.” That’s one way to put it. Another might be, “it was exposed to humans on social media.” Microsoft blamed trolls for sabotaging the chatbot, but if your brilliant software algorithms can be sabotaged by simply

encountering teenage boys, that's on you. Suck it, Microsoft. You can't release flawed software and blame humanity...it's your fault for not understanding humanity. Or at least the western world.
Because Interestingly, TAY was based on a similar Microsoft product in China, and Chinese's Tay has been active for years without any similar issues. And I think this really encapsulates the conundrum of free speech. We won't get into all of the implications but it's a thorny issue, and one that we've been discussing off and on on discord.

On March 30th, 2016, TAY was brought back online, supposedly by accident during testing. It was

quickly yanked once again after quipping such gems as “kush! [i'm smoking kush infront the police].” And “puff puff pass” before getting stuck in a broken record loop of spamming the phrase “You are too fast, please take a rest.” Creepy.
The successor to TAY was a chatbot named Zo, released in December of the same year, 2016. This is really interesting...so one problem with Zo was that it was extremely limited in its communication. “Ask Zo what she thinks of Hitler, and she’ll respond, “i don’t really want to go there :(.”” It was essentially straightjacketed by an overabundance of caution. Journalist Chloe Rose Stuart-Ulin wrote, "Zo is politically correct to the worst possible extreme;

mention any of her triggers, and she transforms into a judgmental little brat." And I can kind of understand that frustration except, like, at least she's not a Nazi. That's the whole dilemma with political correctness. Like, yeah, it can be annoying, but social justice warriors aren't lighting crosses on my lawn and calling my fiancé the n-word. Perspective. It's a real luxury if your biggest concern in life is that people are all going around being too nice to each to each other, you know? Like, I'd what puddles you off is that a bunch of snowflakes are being careful about what they say, re-examine your priorities. The truth is that We don't have real free speech. You can't threaten

the president, you can't defame someone or say something untrue that will hurt their professional reputation...you can't yell fire in a theater. we have a ton of limitations on free speech. So in my opinion, people who are whining about political correctness for the most part just want to be able to say awful shit, and I don't have a lot of sympathy for that. But there's certainly a line. It's a thorny issue.

Next AI fiasco.

WATSON AND URBAN DICTIONARY

So for some history, back in 1997 IBM shocked the world when its supercomputer Deep Blue beat World Chess Champion Gary Kasparov in a live, highly promoted

and televised match. So after beating the Grand Master of chess, IBM decided to further humiliate the human species by dominating the beloved trivia game, Jeopardy.
Enter Watson. Now, I thought Watson was named after Sherlock Holmes's roommate and I was very disappointed to learn that it was named after the first CEO of IBM. It could have been a cool literature reference and instead was shameless corporate ass kissing. Beating Kasparov had been impressive at the time, but also kind of not, because the parameters of a chess game are very rigid...there are a limited numbers of moves available. Watson would need to be a much more fluid and complex piece of

software, and to this day Watson is one of the most famous and lauded of the quote-unquote AIs... I say it in quotes because, again, it's not *thinking*. It's not intelligent. This is a program that retrieves information. It does it in very fancy ways, using NLP or Natural Language Programming. it can supposedly interpret syntax, semantics, context, all of the complicated nuances of speech, and then seek out the appropriate answer blah blah blah but ultimately it's just Wikipedia on steroids. This thing wouldn't have a prayer of beating the Turing test. Watson is very good at retrieving information that it has been programmed to be good at retrieving. The algorithms it uses are

referred to as "Deep QA software"...*deep* being the current Silicon Valley buzzword for all things AI. Deep learning, deep neural networks. Depth is very important. No one wants a Shallow bot. AI operates on a corpse-burying paradigm. The deeper the better. But depth in this case doesn't imply depth of understanding. "Importantly, the design philosophy underlying DeepQA is never to assume that any part of the system, by itself, "understands" the question and hence can simply look-up the correct answer. Rather, different candidate answers are generated based on the analysis of the question and question-answer pairs are scored to produce a ranked list at the end of all the processing." It's not

thinking, it's calculating. And while it's not actively connected to the Internet, it's "memory" includes basically all of Wikipedia, wiktionary, wiki quote, multiple editions of the Bible, and a ton of classic literature and how-to books. Watson was developed with Jeopardy in mind, and it famously beat all of the top champions in a highly publicized tournament on Jeopardy, which was the least surprising result of a trivia competition between humans and a machine that was designed to beat humans at trivia. Like yeah, of course it won, it's like playing Jeopardy against Google. like you can be the smartest person in the world but you can't beat a computer at the one thing it's designed to be good at.

It’s like a race between the fastest human—it’s like usain bolt vs a Tesla. It’s not a fair fight. Why did we bother with this? Humans aren’t equipped to out-calculate a computer, because computer are devices we created to do calculations for us. It’s in the name...they compute. And it would be just as unfair to pit a calculating device against a human in a non calculation task. How about Watson vs Ken Jennings in a thumb war. Is that a fair fight?
And speaking of thumb wars, the machine was a savage when it came to buzzing in quickly to answer a question. “After the match, Jennings and Rutter stressed that the computer still had cognitive catching up to do. They both agreed that if ‘Jeopardy’ had been a

written test — a measure of knowledge, not speed — they both would have outperformed Watson. ‘It was its buzzer that killed us,’ Rutter said.” So Watson became very famous via its rigged Jeopardy win, and the programmers decided to focus on upgrades. in a strategy that no one could possibly have anticipated would go awry, the ibm researchers decided that Watson would improve its ability to speak like a natural human by learning slang, so the researchers fed it the entire online urban dictionary. This is the dictionary that includes entries like shexting, which is sexting while on the toilet, and clam jamming, which is the female equivalent of cock blocking. It subsequently took a 35-person team to

develop a software filter to stop Watson from swearing, and they spent valuable hours scraping the filth from Watson's memory-banks. A goal that I kind of envy. I'd like to scrape the entire urban dictionary from MY memory-banks. I'd like to have a 35-person team sanitizing my brain. Two girls one cup? Delete. All of 2020? Gone.

WATSON FOR ONCOLOGY

Since Watson is pretty good at using knowledge gleaned from the Internet, IBM decided The natural next step was to mimic the extremely human tendency to google every minor sickness-symptom and assign a horrific diagnosis. Meet Watson for oncology, the cancer-diagnosing doctor with the

worst possible bedside manner. *In robot voice:* "You have cancer. Condolences. Thoughts and prayers." Watson for oncology has been at best a mixed bag, and part of the problem is that it's incredibly hard to determine whether this thing is working at all. Can you imagine trying to do a teal study? "Let's use nothing but Watson for Oncology to diagnose cancer patients and see how many of them die." You can't do that. "So it turned out that patients in the "human doctor" group received chemotherapy, patients in the "Watson for oncology" group received nothing but cat Memes. So yeah, good thing we tested it. Sorry about your entire family." So how has this new occupation for Watson worked out? "This

product is a piece of s—," one doctor at Jupiter Hospital in Florida told IBM executives. "We bought it for marketing and with hopes that you would achieve the vision. We can't use it for most cases.'" The problem is that it was fed a very limited amount of data that simply doesn't translate to all real world scenarios. For instance, it has made minor errors like recommending to a patient who was severely bleeding a drug which would cause more bleeding. Little snafus like that. In 2017, the University of Texas's esteemed MD Anderson cancer center quit using Watson for oncology, and the software's future is uncertain. To be fair, one of the biggest hurdles is patient privacy. If you could feed the entire

world's medical information into a computer, it might be able to develop algorithms that actually work for the majority of people, but that's not going to happen anytime soon. So Watson is limited by those pesky reasonable laws and regulations. But look, I'm confident that one day in the not too distant future we will completely dispense with privacy, so there's still hope. Fascism will save us all, so says Q anon. Fingers crossed.

Solid gold bomb

Is not, as I had assumed, some kind of record executives nightmare or The most expensive Improvised Explosive Device in history, but rather is a custom T-shirt

store that had the brilliant idea of automating the process of finding cool slogans and word combinations to slap on merch. It's dictionary algorithm produced shirts with catchy quips like, "keep calm and rape a lot" and "keep calm and hit her." no word on how many they sold. Hey, domestic abusers and felons buy shirts too.

Algorithms really are the worst. In 2011, two competing Amazon books sellers programmed automatic pricing algorithms to keep pace with each other, leading to a war of escalation that resulted in a textbook about flies priced at $23 million. It's like a fly-textbook-pricing arms race. Like the Cold War, for insect textbooks

Amazon facial recognition

You may have heard of this because it was pretty big news, but Amazon has been developing facial recognition technology, which they call rekognition spelled with a “K,” and I will never get over my hatred for this Silicon Valley obsession with intentional misspellings. Missing vowels, backwards letters, stop it. It’s not cool, it’s juvenile. As you can imagine, facial recognition technology is controversial. The New York Times used it to identify guests at Megan Markel’s wedding, so Duncan, if you were at Megan Markel‘s wedding, i’m just saying, don’t bother denying it. We have the tape. We have a lot of British listeners, I feel like

the insomniacs were represented. None of them were invited, but they showed up anyway. On a more sinister note, Amazon Famously sold its facial recognition technology to police departments, despite the fact that it has a track record of dubious accuracy. And this is a problem for a number of reasons. If it works, we're basically living in a surveillance state. If it doesn't work, we're living in an incompetent surveillance state, in which you can be falsely accused and convicted based on a shitty algorithm. In 2018, the American Civil Liberties Union performed a test using Amazons recognition software, and "recognition with a K" identified 28 members of Congress as potential

convicts and criminals. Worse, the software misidentified African-American and Latino members at a higher rate, which is a common problem with this emerging technology. Amazon's response was not the best. They said that the technology is intended to be used to narrow down the potential perpetrators among a group of suspects. Which means the software is potentially confirming the biases of police officers. If police officers already suspect people of color are committing more crimes, and they're feeding pictures of suspects of color into the software, and the software has a tendency to provide false matches when dealing with brown skin, "Recognition with a K" is

going to give racist police departments ammunition and confirm their worst biases. In 2020, under mounting pressure, Amazon announced a one-year freeze on use of the software by police departments. Except not really, because they are allowing exceptions for cases of child sex trafficking. So you can't be falsely identified by this software as a crack dealer or whatever, but I can still falsely identify you as a pedophile and sex trafficker. That won't affect your life much. Being branded a pedophile. Specifically A Satanist pedophile. Plus, the limited moratorium on police use of the software is going to run out in just a couple months. So, look forward to that. A company that can't even

spell “recognition” is going to be trying to recognize whether you committed a crime.

This _____ Does not exist

Have you seen these websites? They all use a version of NVIDIAs AI called StyleGAN, GAN stands for generative adversarial network. So This is interesting, it’s basically a pair of software algorithms that are fed a huge set of sample faces, and then begin working against each other and benefitting from competition. One program —the generative network— is trying to create fake faces that will fool the other program—the discriminative network— into believing they’re real, and the discriminative

network gets better and better at spotting fakes, which makes the generative network work harder and get better and better at creating a convincing fake, until the software has trained itself to generate fake human faces, or cats, or whatever. But it doesn't always go smoothly. The latest version, released last year, is much improved, but you still get unfortunate reminders of its previous self, when it spontaneously generates absolute hellspawn nightmare images that will haunt your dreams. Take a look.

And of course these programs can be used to make fake social media profiles, and when paired with chatbots can be used as extremely convincing online racists. Tay with a

face. A racist face.

Amazon AI recruiting tool

One of the biggest problems with AI is that, once again, it's limited by its dataset… if you feed an AI only pictures of white people, it won't even know that people of color exist. It's like humans in that respect. If you grow up around only people of your race, you can end up very biased toward your own kind or simply unable to understand and empathize with anyone different from you. Thus was the case with Amazon's staff-recruitment software. It was in charge of analyzing applications and rejecting unqualified or undesirable applicants. But it used as reference—in other words it was "trained" —via the

entire database of previous Amazon applicants, most of whom were white men. The software was scrapped when it was found to be rejecting applicants from women's colleges, or in fact anyone whose resume included the word "women's" as in women's college or women's tennis team. Amazon claimed the software had only been used in trials, and didn't actually result in rejection of any applicants, which is exactly what you'd say if your misogynistic software had rejected a bunch of women and exposed you to potential legal liability.

https://www.google.com/amp/s/futurism.com/cat-doesnt-exist-ai/amp

https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

https://www.google.com/amp/s/amp.theatlantic.com/amp/article/267047/

https://www.bloomberg.com/features/2016-microsoft-future-ai-chatbots/

https://analyticsindiamag.com/8-real-life-examples-algorithms-turned-rogue-causing-disastrous-results/

https://https://www.lexalytics.com/lexablog/stories-ai-failure-

avoid-ai-fails-2020

https://
www.immuniweb.com/
blog/top-10-failures-of-
ai.html

https://lovetheidea.co.uk/
9-funniest-shocking-ai-
fails/

https://www.google.com/
amp/s/
www.techrepublic.com/
google-amp/article/why-
microsofts-tay-ai-bot-
went-wrong/

https://www.wired.com/
story/inside-microsofts-ai-
comeback/

https://thebestschools.org/
magazine/watson-
computer-plays-jeopardy/

https://www.kurzweilai.net/the-buzzer-factor-did-watson-have-an-unfair-advantage

https://www.google.com/amp/s/www.nytimes.com/2018/07/26/technology/amazon-aclu-facial-recognition-congress.amp.html

https://www.google.com/amp/s/www.theverge.com/platform/amp/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report

https://www.google.com/amp/s/www.theverge.com/platform/amp/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report

https://

towardsdatascience.com/the-truth-behind-facebook-ai-inventing-a-new-language-37c5d680e5a7