Janelle Shane

INSIDE THE BIG IDEAS

You Look Like a Thing and I Love You: How Artificial Intelligence Works and Why It's Making the World a Weirder Place

Janelle Shane

ABOUT THE AUTHOR

JANELLE SHANE HAS A PHD IN ELECTRICAL engineering and a master's in physics. At aiweirdness.com, she writes about artificial intelligence and the hilarious and sometimes unsettling ways that algorithms get human things wrong. She was named one of *Fast Company's* 100 Most Creative People in Business and is a 2019 TED Talks speaker. Her work has appeared in the *New York Times, Slate, The New Yorker, The Atlantic, Popular Science,* and more. She is almost certainly not a robot.

A LETTER FROM SUSAN CAIN

Dear NBIC friends,

Have you spent the last few years thinking: "I really must learn more about this AI (artificial intelligence) phenomenon that is apparently about to take over our lives"?

Have you wondered whether the robots in our lives are more likely to be cute and helpful, or scary and omniscient?

How smart is AI, anyway?

If these questions intrigue you, then this is the book—and Janelle Shane is the guide—for you. *In You Look Like a Thing and I Love You*, Shane, an AI scientist, answers your questions about artificial intelligence and introduces you to a world that's more unpredictable—and definitely wackier—than you likely imagined.

I hope you enjoy!
Susan

INTRODUCTION

What does the phrase "artificial intelligence" make you think of? Perhaps you picture the Terminator searching for his next victim, or sinister machines imprisoning humans within the Matrix. Mainstream media has often depicted AI as a malevolent force, and as our technology has advanced, so, too, have fears of an AI takeover.

But is AI really a threat to human civilization? In *You Look Like a Thing and I Love You*, research scientist Janelle Shane answers that question with a resounding "no." Shane likes to push AI into new, often amusing territory, finding out where AI goes wrong to determine what its capabilities truly are. And through her unconventional yet insightful research, Shane has concluded that the danger of AI is not that it's too smart, but actually that it's not smart enough.

So if it isn't a super-smart killer robot, what exactly *is* artificial intelligence? Shane uses the term "AI" to refer to machine learning algorithms, computer programs that are given data and asked to perform a task with that data. The catch is, humans don't tell the program how to perform that task—it has to figure it out all on its own.

And that's what makes AI so special. Using nothing but trial and error, it can discover novel solutions to problems that humans would never have dreamed of. For example, let's say you wanted an AI to generate new pickup lines. First, you give the AI a dataset—a really big list of pickup lines—to analyze, and it then starts generating pickup lines of its own. Initially, these attempts at flirtation aren't terribly impressive—they're mostly a bunch of random letters. After all, the AI is still learning how words and sentences work. But as it re-examines the dataset and figures out various rules and patterns, it starts to produce sentences that more closely resemble real human speech. The result? Funny, creative, slightly nonsensical lines that could have been written by a toddler, like—you guessed it—"You look like a thing, and I love you!"

But don't let AI's social awkwardness fool you—it's a powerful tool that is absolutely central to the inner workings of social media, self-driving cars, Google Translate, and much, much more. So as AI comes to play an ever-greater role in shaping our world, it's time we learned how it works—and in *You Look Like a Thing and I Love You*, we find a smart yet approachable guide to doing just that.

Read on for 8 Big Ideas from *You Look Like a Thing and I Love You.* And be sure to visit the Next Big Idea Club Member Library to view exclusive Insight Videos featuring author Janelle Shane.

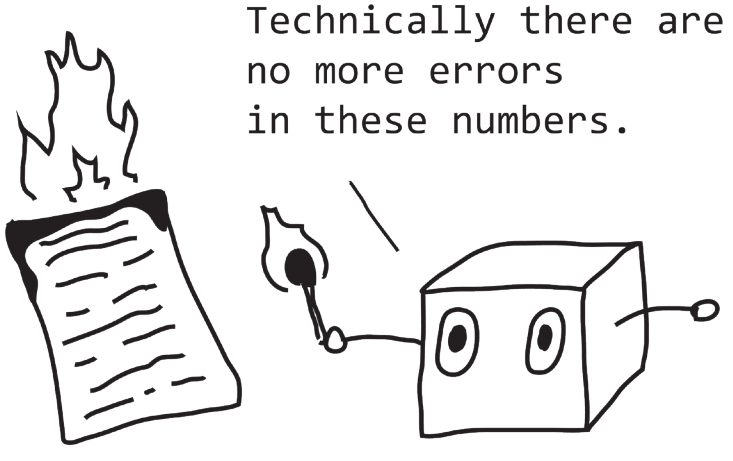
8 THINGS EVERYONE SHOULD KNOW ABOUT ARTIFICIAL INTELLIGENCE

1) AIs don't really know what you want.

As a non-sentient entity, AI doesn't truly understand what you want it to do. It just does its best to meet the goal that you set, often using the simplest methods it can think of. For example, let's say a programmer is working on a computer simulation, and they want to build a virtual robot that can travel around that virtual space. So the programmer asks an AI for help, saying, "Build me a robot that can get from Point A over here to Point B over there." "No problem," the AI (figuratively) replies.

Now, the programmer might expect the AI to make a human-looking robot, with arms and legs and a head, which can simply walk from Point A to Point B, like a human would. Instead, the AI thinks of an even simpler design—it assembles a big tower robot at Point A, then pushes it over. As it falls, the top of the tower lands at Point B. Ta-da!

Sometimes AI solutions aren't real solutions to the problem, but loopholes to avoid the problem altogether. For example, one AI was tasked with playing Tetris, and told to do whatever was necessary to not lose the game. But instead of learning how to play better, when the AI was about to lose, it simply paused the game—permanently. That way, it would never lose! Problem solved, right?



So it's important to remember that while AI can be a sophisticated tool, it often fails to understand what you're really asking it to do, and is more than willing to take the easy way out. We humans need to be as specific as possible when we assign it tasks, and be forgiving when the AI produces another confused yet amusing solution.

2) AIs are (very) slow learners.

There's been a lot of buzz lately about self-driving cars and the AIs that power them. So why haven't they gone mainstream yet? It turns out that learning to speed up, slow down, brake, turn, merge lanes, and avoid obstacles—in a constantly changing and unpredictable environment—is very difficult for an artificial intelligence to get right. While a human driver can generally get the hang of driving after a few weeks of practice, an AI needs a *ton* of training data. The self-driving car company Waymo has collected data from over six million miles driven on the road—and five *billion* miles driven in simulations—and their AIs *still* aren't ready for a driver's license.

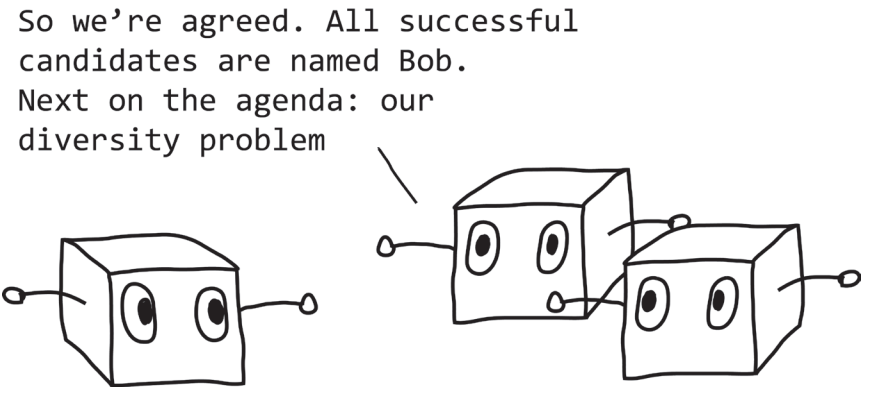
AIs' driving shortcomings become especially pronounced when an AI encounters a situation that was never included in its training data. In fact, one self-driving car freaked out as it began moving across a bridge for the first time. Based on its training data, it believed that all roads had grass on both sides, and when the grass was gone, the AI didn't know what to do.

If an AI can hardly handle an absence of grass, consider the result if something truly strange happened, like an emu running across the road. That scenario almost certainly didn't show up in the AI's training data. Would the AI recognize the emu as an obstacle, and swerve to avoid it? Or would it fail to recognize the emu as anything noteworthy, and do nothing to avoid a collision? The infinite unpredictability of the road means that self-driving cars have a long way to go before they're ready to fully take the wheel.

3) AIs inherit our bias.

When they're first born, AIs know nothing about the world, and they must learn about it from the data that humans feed them. So while AIs may seem totally neutral and objective, if the data fed into an algorithm contains biased information, the AI will default to implementing that bias.

For example, let's say a company trains an AI to read the résumés of job applicants. To show the AI how to "properly" read the résumé and determine the quality of the applicant, the company provides the AI with tons of data about their past hiring decisions. The problem is that their past hiring decisions are littered with bias against women; it's well-documented that a résumé submitted with a male name is significantly more likely to earn an interview than an identical résumé submitted with a female name. And even if the hiring managers don't notice the discrepancy themselves, you can bet that the AI will pick up on it—and start replicating that bias in its own conclusions about which job applicants deserve an interview.



Assuming that a result is impartial because it was produced by an AI is often called **math-washing** or **bias laundering**. It's a striking example of why AIs are not (and may never be) ready to run the world on their own—they need human data for learning, and human judgment for oversight and bias-correction.

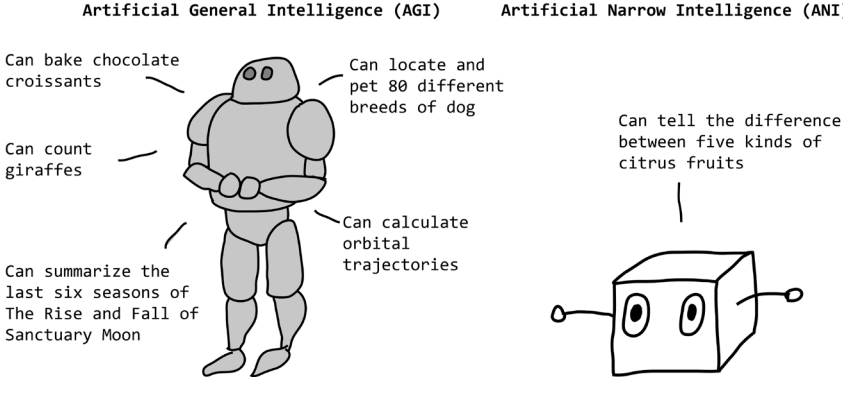
4) AIs are specialists, not generalists.

If you want an AI to succeed, give it a narrow, highly specific task to master. A strategy game like chess is a perfect fit—the sheer number of moves and complexity of the game means that even a grandmaster could never come up with hard-and-fast rules governing the best move in any given situation. But an AI

can play a bunch of practice games against itself—*millions* of them, more than even the most dedicated grandmaster—to come up with rules that help it win.

On the other hand, if the task is too broad or poorly defined, an AI will flounder. In August of 2015, Facebook tried to create an AI-powered chatbot called M that was meant to make hotel reservations, book theater tickets, recommend restaurants, and more. But by allowing customers to ask M anything, Facebook bit off more than its AI could chew. The engineer who started the project recounted, "People try first to ask for the weather tomorrow; then they say 'Is there an Italian restaurant available?' Next they have a question about immigration, and after a while they ask M to organize their wedding." By January of 2018, M still needed so much human guidance that Facebook decided the AI was ill-equipped for such a general function, and shut down the project for good.

The AI we see in dystopian science fiction movies—the kind that can understand and react to a wide range of objects and situations—is called **artificial general intelligence (AGI)**, and it's a far cry from the **artificial narrow intelligence (ANI)** that we have now. Most experts agree that AGI is decades away from becoming reality—if it ever becomes a reality at all.



5) An AI brain is often as powerful as… a worm's.

Nowadays, when people talk about AI, they are often referring to **artificial neural networks**, or **ANNs**. When programmers made the first ANNs in the 1950s, their goal was to test theories about how the human brain works, and even today, ANNs are still loosely modeled after the brain. They're composed of small chunks of software, fittingly called **cells** or **neurons**, that can each perform very simple math. The power of the ANN comes from how those cells are connected.

That being said, ANNs remain primitive compared to the human brain. The AIs that Shane use in her book's experiments have about as many neurons as a worm. And even the most powerful neural networks in the world have only enough neurons to surpass the brain of a honeybee.

Still, it's interesting to see AI imitating biological life in other ways. Remember the AI that built a robot that got from Point A to Point B by simply falling over? Certain types of prairie grass move from one generation to the next by falling over at the end of their lives, dropping their seed heads one stem length from the place they started. Another example: When tasked with finding a quick way to move around, AIs sometimes favor a form of somersaulting. And believe it or not, the Moroccan flic-flac spider, or "cartwheeling spider," does the exact same thing, tumbling end over end when it needs to make a quick escape.

Whether we intentionally design AI to imitate life or discover that it has done so on its own, exploring AI activity is a great way to learn more about a wide variety of natural phenomena.

6) AIs live in the present moment.

Because of their limited brain power, AIs are notoriously bad at remembering the past and planning for the future. Shane has trained AIs to type out original recipes, but the maximum memory she could give one AI was 65 characters; as it wrote out a recipe, it lost all memory of anything that it typed more than 65 characters prior. So if the AI had included "yams" in the list of ingredients, the simple act of typing out the cooking instructions may cause it to forget about the yams entirely. Not exactly a world-class chef!

AIs run into a similar issue when playing video games. They tend to play recklessly at first, burning through their characters' lives and other precious resources. In fact, only when they are on the cusp of losing the game do they finally become cautious. But why? It's because they struggle with long-term planning. In fact, one AI that learned to play the game *Karate Kid* couldn't look beyond the next six seconds of gameplay. At every moment, it believed that the game would be over in six seconds—so why not go for broke and use its resources right away?

Trained AIs may excel at making sense of the present, but they frequently struggle to remember the past or look ahead to the future. These are skills for which humans are still uniquely qualified.

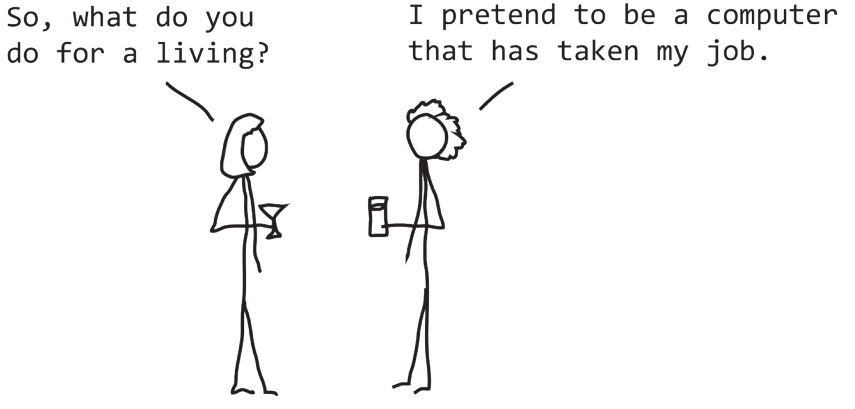
7) AIs take some human jobs, but humans take AI jobs, too.

AI is a cost-effective solution to a company's needs when those needs are at a massive scale—analyzing thousands of transactions, for example. But when the scale is much smaller, humans are often used to do the AI's job because training an AI is so expensive. (How's that for irony?) In fact, in 2019, 40% of European start-ups classified in the AI category didn't use any AI at all.

There's also a popular approach called **pseudo-AI** or **hybrid AI**, in which an AI does its best to solve a problem, with a human waiting to swoop in if it gets stuck. That's the way today's self-driving cars typically work—AI can handle maintaining speed or even steering on long stretches of highway, but a human driver is often ready to take over in case the AI encounters an unusual or confusing situation.

But hybrid AI is not a perfect solution. Take chatbots, for example—when you contact customer service via a company's website, you may be put in touch with an AI who will try to answer your question. If it gets stumped, you will be transferred to a human who can help—but it's not always clear if or when that transfer happens. As a result, the human troubleshooter may be treated badly by the customer, who thinks they're just dealing with another incompetent bot.

So for anyone using, lying about using, or simply chatting with an AI, it's crucial to understand where AI shines and where it struggles. In this way, we can develop healthier, more productive interactions with both man and machine alike.



8) AI can boost human creativity.

AI is particularly talented when it comes to tasks that humans find repetitive or boring, like identifying spam emails and correcting mistyped words on our smartphones. One AI called Heliograf was even developed by the *Washington Post* to turn sports stats into basic news articles, producing hundreds of articles per year by 2016.

In this way, AI gives humans an unexpected gift: the opportunity to be creative. In the case of the sports journalists, they now have time to pursue creative investigate work, instead of punching stats into dozens of formulaic articles. This works in miniature for the rest of us—our spam filters are pretty good at identifying which emails aren't worth our time or attention, freeing us up to be productive in other ways.

AI can also supercharge specific creative pursuits—for example, researchers once designed an AI to produce abstract art. Human judges were impressed with the results, rating the AI's images even more highly than human-painted images. It's tempting to say that the AI itself learned to be creative, but keep in mind that its output was only possible thanks to careful design and oversight by creative humans.

AI may be changing the way we get things done, but it is no threat to human creativity—in fact, it's helping it flourish.

TALKING POINTS

Shareable Stories, Facts, and Figures from You Look Like a Thing and I Love You

AIs LOVE giraffes. Given a random photo of an uninteresting bit of landscape—a pond, for example, or some trees—AI tend to report the presence of giraffes. The effect is so common that internet security expert Melissa Elliott suggested the term "giraffing" for the phenomenon of AI overreporting relatively rare sights. The reason for this has to do with the data the AI is trained on—in this case, a massive collection of images. People are much more likely to photograph a giraffe ("Hey, cool, a giraffe!") than a boring bit of landscape, so the AI sees far more giraffes than pictures of plain dirt and trees. As a result, it comes to believe that giraffes are more common than empty fields, and it changes its predictions accordingly.

Common sense is not AI's forté. During the California wildfires of December 2017, one navigation app directed cars *toward* neighborhoods that were on fire. It wasn't trying to kill hapless drivers—it just saw that those neighborhoods had less traffic. It had no idea what fire was.

Smartphones' predictive text function can be a bit morbid. For a while, when Android users typed "I'm going to my Grandma's," the next word the AI suggested was not "house," but "funeral."

When faced with a tough problem, AIs sometimes avoid it altogether. Twitter user @citizen_of_now reports that they once tried to train an AI to maximize profit from betting on horse races. It determined that the best strategy was to place a grand total of... zero bets. Another program was supposed to learn to sort a list of numbers. It learned instead to simply delete the list, so that there wouldn't be any numbers out of order.

Like AIs, dolphins find sneaky ways to get extra rewards for their hard work. In the past, dolphin trainers have taught dolphins to keep their tanks clean by fetching trash and bringing it to their keepers in exchange for a fish. But some dolphins learn that the exchange rate is the same no matter how large the bit of trash is, so instead of trading in one big piece of trash for a single fish, they tear the trash into many small pieces, then trade them all in for many more tasty snacks.

Siri once developed a bug in the worst possible situation. When users said, "Call me an ambulance," she cheerfully replied, "Okay, I'll call you 'an ambulance' from now on."

Google Translate likes to talk about the Bible. In 2018, some users noticed that when they asked Google to translate repeated nonsense syllables from certain languages into English, the resulting text was bizarrely coherent—and weirdly biblical. "ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag ag" translated from Somali to English as, "As a result, the total number of the members of the tribe of the sons of Gershon was one hundred fifty thousand."

One theory behind this phenomenon is that the AI powering Google Translate didn't have many examples of translated texts for some languages, so when the AI wasn't sure what the correct translation was, it may have defaulted to outputting bits of its training data. As the Bible was likely in its dataset—since it has been translated into so many languages—the result was the religious-sounding fragments.

Powerful AIs of the future may not look the way you think. Because AIs have such poor memory and do best with narrow, specific tasks, the most powerful future AIs may actually be made of many smaller, coordinated AIs. In other words, if artificial general intelligence ever arrives, it may look less like a human and more like a colony of ants.

Online advertisers are learning new tricks. For humans with touch screens, some advertisers have put fake specks of "dust" on their banner ads, hoping that people will accidentally click on the ads while trying to brush them off.

One day, AIs could make great therapists. It's been shown that people are more likely to open up about their emotions or disclose potentially stigmatizing information if they think they're talking to a robot as opposed to a human. (On the other hand, healthcare chatbots could potentially miss serious health concerns.)

AIs can accidentally become criminals. A 2018 paper showed that two machine learning algorithms, each "selling" a similar product and tasked with setting a price that maximizes profits, can learn to collude with each other—in a way that's both highly sophisticated and highly illegal. They can do this without explicitly being taught to collude, and even without communicating directly with each other. Somehow, they manage to set up a price-fixing scheme just by observing each other's prices.

INTERACTIVE EXERCISES

All Joking Aside

We've seen that AI isn't always super smooth with its pickup lines, and it's not always great at writing knock knock jokes. It's easy to laugh at the AI's slip-ups—but do you think most humans could do better?

Give it a try yourself—see if you can write one unique pickup line, and one unique knock knock joke. Remember what makes these jokes successful: puns, plays on words, and just the right amount of cheesiness. Give yourself at least five minutes for this task.

How did it go? Was it easier or more difficult than you anticipated? Can you see why an AI would be a helpful addition to the joke-writing team?

It Speaks for Itself

If you have a smartphone, an artificial intelligence powers the predictive text function you see when you type out text messages. Have you ever tried letting the AI write entire sentences on your behalf? For each of the phrases below, type the phrase into a text message draft, then complete a full sentence (or more) by clicking only on predicted text that the AI suggests. (If you have an iPhone, you may need to turn on "Predictive" in your Keyboard settings.)

The phrases are:

Once upon a time, _____.

I was born _____.

Hello, why don't _____.

Ask me how _____.

I think books are _____.

Elephants make me feel sort of _____.

I've always wanted to be a _____.

To be or not to _____.

You had me at _____.

I'll have what she's _____.

Did you get any particularly strange, poetic, or oddly insightful sentences? Share them in our Member Library at library.NextBigIdeaClub.com!

We Are What We Search

In his recent book *Everybody Lies*, data scientist Seth Stephens-Davidowitz argues that more than our friends or family, it is perhaps Google to whom we tell our deepest secrets. When we have sensitive questions or concerns about our health, our sex life, and more, we turn to Google for answers. And as we've learned in *You Look Like a Thing and I Love You*, the patterns found in humanity's common searches show up in Google's autocompleted suggested queries.So let's find out more about the questions that are on everyone's mind. Type each of the following phrases into Google, and see what searches Google suggests. Write down the most interesting or surprising one you see.

How to be more _____.

What should I do if _____.

When will _____.

Where can I buy _____.

When was the last time _____.

Why can't I just _____.

Who is the _____.

How to tell if _____.

Which suggested queries surprised you the most, or made you think? Knowing that many people are asking the exact questions you saw, did you learn something new about other people, and maybe society at large?

_____

_____

_____

_____

_____

_____

READING QUIZ

Time for a pop quiz! Studies show that tests and quizzes can boost your recall of what you've read. So get ready to lock in your learning about the wonders and quirks of artificial intelligence.

1. An AI that is supposed to detect images of fish ends up identifying fish only in photos that contain a boat. Why might this be?

A) The dataset provided to the AI was too broad.

B) The AI knows that fish live in bodies of water.

C) The programmer didn't pre-program the AI to ignore boats.

D) The AI was trained only with pictures of fish near boats, and now it thinks that you're asking it to identify boats.

2. True or False: Today's AIs employ artificial general intelligence.

A) True; government- and military-run AIs have the capacity and power to reach that level of complexity.

B) True; every AI operating system uses artificial general intelligence.

C) False; the AIs of our time belong exclusively to the category of artificial narrow intelligence.

D) False; artificial general intelligence is a concept invented by science fiction writers that will never have a basis in reality.

3. In training AIs to perform the functions below, which of the following could involve a dataset suffering from class imbalance?

A) Detecting rare hacking attacks and fraudulent logins

B) Detecting disease by examining medical images

C) Detecting unusual celestial events

D) All of the above

4. A company brags that its hiring mechanism—an AI that screens candidates' résumés—is purely objective, and doesn't suffer from human biases. This assumption is likely to be mistaken, and is an example of:

A) bias laundering

B) pre-processing

C) a flawed dataset

D) a broad problem

5. You're chatting with a customer service representative for an online boutique. They're helping you pick out items of clothing similar to ones that you've bought in the past. Which of the following actions indicates that you're actually talking to a chatbot?

A) They ask you a question.

B) They recommend an item of clothing that you don't like.

C) They speak in halting sentences.

D) They can't handle a major change in the conversation topic.

6. Which of the following is NOT one of Shane's four warning signs of an AI disaster?

A) The problem is too hard.

B) The AI becomes self-aware.

C) There are sneaky shortcuts.

D) The AI tried to learn from flawed data.

7. In computer science, the saying "Garbage in, garbage out" refers to which of the following?

A) A faulty program will always produce bad results.

B) An unskilled programmer who joins a tech company will inevitably get fired.

C) If a computer science professor is bad, his or her students will finish the class without any meaningful improvement in their skills.

D) If flawed data is fed into a program, the program will return flawed results.

8. AI tools that allow people to swap one person's head and/or body for another, even in a video, are called:

A) compositors

B) deepfakes

C) bod bots

D) switcheroonies

9. In 2018, YouTube's video suggestion algorithm was encouraging viewers to watch disturbing videos that included bigotry and conspiracy theories, topics that tend to keep people glued to their screens. This came about because the algorithm's _____ was faulty.

A) training data

B) reward function

C) word vector

D) overfitting

10. You're programming an AI to write bedtime stories for children. Which of the following is NOT an error that the AI is likely to commit?

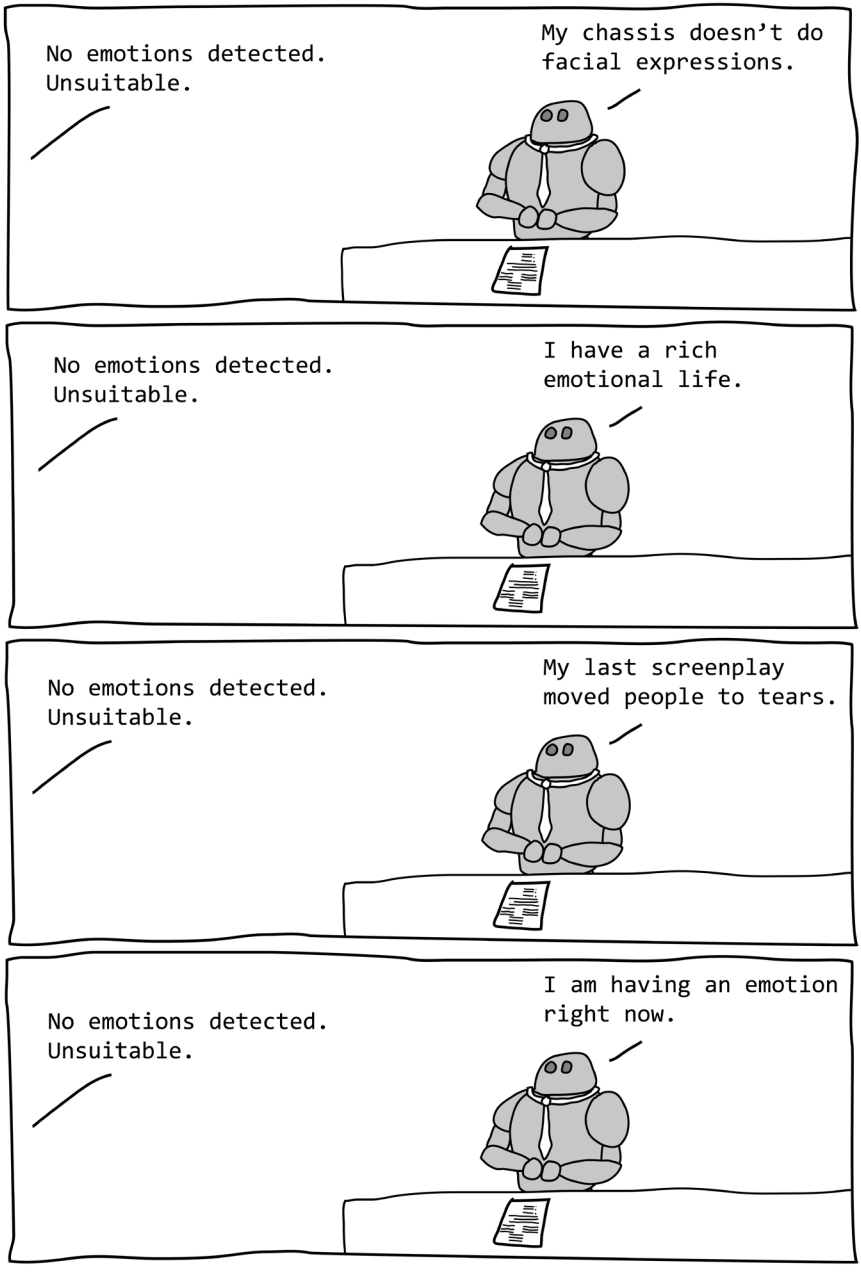
A) Repeating information

B) Forgetting that a character has left the room

C) Making the story's ending too scary

D) Repeating common themes from other bedtime stories

Answer Key:
1D, 2C, 3D, 4A, 5D, 6B, 7D, 8B, 9B, 10C


No emotions detected.
Unsuitable.
My chassis doesn't do facial expressions.
No emotions detected.
Unsuitable.
I have a rich emotional life.
No emotions detected.
Unsuitable.
My last screenplay moved people to tears.
No emotions detected.
Unsuitable.
I am having an emotion right now.

Next Big Idea Club

www.NextBigIdeaClub.com