Science AMA Series: I'm the MIT computer scientist who created a Twitterbot that uses AI to sound like Donald Trump. During the day, I work on human-robot collaboration. AMA!

BRADLEY_HAYES [R/SCIENCE](#)

Hi reddit! My name is [Brad Hayes](#) and I'm a postdoctoral associate at MIT's [Computer Science and Artificial Intelligence Lab (CSAIL)](#) interested in building autonomous robots that can learn from, communicate with, and collaborate with humans.

My research at MIT CSAIL involves developing and evaluating algorithms that enable robots to become capable teammates, empowering human co-workers to be safer, more proficient, and more efficient at their jobs.

Back in March I also created [@DeepDrumpf](#), a Twitter account that sounds like Donald Trump using an algorithm I trained with dozens of hours of speech transcripts. (The handle has since picked up nearly 28,000 followers)

Some Tweet highlights:

https://twitter.com/DeepDrumpf/status/705480367239659520
https://twitter.com/DeepDrumpf/status/705480113018707969
https://twitter.com/DeepDrumpf/status/705465462721744896

I'm excited to report that this past month DeepDrumpf [formally announced its "candidacy" for presidency](#), with [a crowdfunding campaign](#) whose funds go directly to the awesome charity ["Girls Who Code"](#).

DeepDrumpf's algorithm is based around what's called "deep learning," which describes a family of techniques within artificial intelligence and machine learning that allows computers to to learn patterns from data on their own.

It creates Tweets one letter at a time, based on what letters are most likely to follow each other. For example, if it randomly began its Tweet with the letter "D," it is somewhat likely to be followed by an "R," and then a "A," and so on until the bot types out Trump's latest catchphrase, "Drain the Swamp." It then starts over for the next sentence and repeats that process until it reaches 140 characters.

The basis of my approach is similar to [existing work](#) that [can simulate Shakespeare](#).

My inspiration for it was a report that analyzed the presidential candidates' linguistic patterns [to find that Trump speaks at a fourth-grade level](#).

Here's [a news story that explains more about Deep Drumpf](#), and [a news story written about some of my PhD thesis research](#) For more background on my work feel free to also [check out my research page](#). I'll be online from about 4 to 6 pm EST. Ask me anything!

Feel free to ask me anything about

DeepDrumpf

Robotics

Artificial intelligence

Human-robot collaboration

How I got into computer science

What it's like to be at MIT CSAIL

Or anything else!

EDIT (2:30pm ET): I'm here to answer some of your questions a bit early!

EDIT (3:05pm ET): I have to run out and do some errands, I'll be back at 4pm ET and will stay as long as I can to answer

your questions!

EDIT (8:30pm ET): Taking a break for a little while! I'll be back later tonight/tomorrow to finish answering questions

NOTE FROM THE MODS Guests of /r/science have volunteered to answer questions; please treat them with due respect. Comment rules will be strictly enforced, and uncivil or rude behavior will result in a loss of privileges in /r/science.

Many comments are being removed for being jokes, rude, or abusive. Please keep your questions focused on the science.

Hi, Brad. As a first year college student who is planning on a degree in computer science, what are some ways I'm able to get into AI out of college? Thanks for taking time to do this AMA.

WubWubWubzy

Don't wait until you're out of college! Start learning from the tremendous amount of resources online now. Regardless of your focus, as a Computer Science major I would say one of the most important things you can do is to build lots of things and write lots of code. Your CS education will hopefully give you perspectives and theoretical tools to succeed, but they will be of limited use to you if you don't practice applying them! If you're interested in research, there are lots of university research labs out there that are willing to take undergraduate researchers -- if there are any at your school, the sooner you can get involved the better.

If there are AI research groups at your school:

In my experience, undergraduates that have dedicated the time to doing research with the same lab throughout their college years have always ended up getting published, with some having first-author papers (which can greatly boost your grad school prospects). I recommend finding some lab websites, asking professors if it's alright to show up to their lab meetings, and talking to some of the people working there to see if they're working on anything interesting to you and if there's any way you can contribute.

If there aren't, or none are a great fit:

Start now! There has never been a better time to get started in Computer Science or AI in general than today. If you have the discipline, working through some online coursework during your free time will help you a lot -- but more than anything else I recommend that you actually pick a small project and try to make something. Even if you have no idea how to do it yet, it will keep you focused and give you a nail to build a hammer for. I've always found hands-on experiences to be more motivating and informative than reading blog posts/papers/lectures by themselves. Finding bits of sample code and playing with them is a great way to learn, as well as working through tutorials that others have posted, but I would say above all to start small. I recommend looking for beginner tutorials and playing with them.

If you have a little bit of background in Computer Science already, I recommend learning some Python and working through the fantastic TensorFlow tutorial series. I had success with two exceptionally bright high school interns who were able to learn some Python and make their way through a good bit of Stanford's CS231n Convolution Neural Networks for Visual Recognition course over the course of a few months (with a bit of guidance) without much of an advanced coursework background.

TL;DR -- Go build lots of cool stuff!

I have trouble understanding the "deep learning" concept. After writing all the most likely letters and the 140 characters tweet is formed, does it have to check grammar and syntaxe or is it complex/good enough to create real sentences 100% of the time ?

Dalyos

It may be easiest to not focus so much on the 'deep learning' aspect of the language model and just view it more generally as something trying to capture statistical structure. Deep Learning is just one (powerful) tool we have to do such things -- a more approachable place to start might be looking at [Markov Chains](). Recurrent Neural Networks are more expressive than these, but the intuition is still valuable.

As some of the other commenters have pointed out, there are [many]() [great]() [resources]() out there for learning about recurrent neural networks! In the earliest days of the bot, there was no postprocessing done, meaning the raw output of the neural net was being posted onto Twitter. As I've had a bit of time to devote to improving the bot's output, there is now a fair bit of postprocessing to do things like correct minor spelling errors, use named entity recognition to identify and occasionally replace certain people/places/things with other named entities (e.g., identifying and replacing state/city names), and prune out sentences that don't have the right grammatical components.

I've tried to say transparent with respect to both what the model is being primed with and how the text that gets tweeted is selected -- I'm actually sampling far more than 140 characters (typically ~3000) and choosing a subset from there. At this point, the overwhelming majority of the output is sensical but it's not necessarily relevant or funny. I act as a layer between the model and the outside world for two important reasons: 1) The training data (especially early on) made it occasionally produce threats, and 2) Humor is difficult to pick out and automate. As far as I'm aware, we don't really have great humor classification models, which is actually an incredibly tricky problem (and relies on having a lot of knowledge about the world). Of course, letting the model loose to just output whatever it wants on a regular schedule is an option, but I wouldn't expect anyone to want to spend the time sifting through it all for the occasional bit of humor.

Hi Brad (or [/r/science]() guests)! I think your work is very fascinating and cool, even though I don't really understand much (or any!) of the science behind it.

I'm a high school senior and I'm planning to go into computer science. **What type of math do you usually use in AI?** When I did HTML, CSS, and the beginning of Java on Codeacademy, there wasn't of "math math" (such as functions, derivatives, trig, etc.), it was just writing in commands in a specific syntax. Does the more complex math come in later?

The reason I'm asking is because even though I am a good math student, I'm not 100% confident I can handle the math in a computer science course. I'm doing fine in my Calculus class but I have to wrestle with the material a bit and I also didn't do ~great on the ACT/SAT Math. I like computer science a lot but I'm afraid I'm not smart enough for it.

Another question, is MIT as hard as everyone (including MIT students) says it is? A lot of bloggers on MIT's undergrad admissions blog say that MIT is the hardest thing they've ever done, and their super smart! I was just wondering what your experience is like!

Thanks!

[Yusapip]()

Hi! You're definitely doing the right things -- Codecademy is a great resource to get started with.

I'm not 100% confident I can handle the math in a computer science course.

I promise you that **you and anyone reading this are capable of handling the math/the material in any kind of university course**. I cannot stress that enough.

Different people sometimes need different presentations of the material for it to click, but universally speaking you **are** capable of it if you're prepared to devote the time and effort to it. Asking for help from

your instructors or peers (or strangers on the Internet) will definitely make your path a bit easier, but the biggest mistake to make would be giving up because you feel you're struggling too much (or more than those around you). Math and Computer Science are both challenging subjects, even moreso if you don't intrinsically enjoy what you're learning, and learning them is often painful at some point for everyone. Ultimately, I can only tell you that I found it is worth the struggle and that I'm sure that everyone who sticks with it long enough goes through such difficulty.

I think that the most important things you can take away from any university course (or experience in general) are new strategies to grasp/learn concepts and new perspectives for approaching problems.

What type of math do you usually use in AI?

To get a good understanding of the popular techniques in machine learning, I would strongly encourage you to develop a background in statistics, linear algebra, and calculus, but practically speaking you'll be far better served with a good intuition for how things work than being able to regurgitate formulas on paper.

I'd say there's a pretty big difference between doing research, where you're trying to expand the frontier of knowledge about a topic, versus the majority of the time where you'll be using existing components in interesting ways to build something new. If you download the Python library Scikit-Learn, you could start building machine learning systems without having any practical understanding of the underlying math! Of course, you have a higher likelihood of picking the right tools/methods and being successful if you understand how they work, but strictly speaking it isn't completely necessary to be able to code them from scratch. When you're building things, you'll likely be doing a lot of using other people's software libraries (instead of implementing your own) -- this is a great habit to get into since new code is far more likely to be buggy than something that's been widely used.

Another question, is MIT as hard as everyone (including MIT students) says it is? A lot of bloggers on MIT's undergrad admissions blog say that MIT is the hardest thing they've ever done, and their super smart! I was just wondering what your experience is like!

MIT, like most univerisities, can be as difficult as you let them be. Some places may push you harder by default than others, but ultimately the goal is to force you to become a better learner. If you have an inefficient learning process, places like MIT try to identify that (i.e., you won't be able to keep up with your workload) and force you to adapt new strategies. I did my undergraduate education at Boston College and my PhD at Yale, and found plenty of challenges at both places that I also see shared by the students at MIT.

Most importantly, try not to let talk like that intimidate you -- I don't believe that the people at MIT are intrinsically smarter than anyone else, but they are very effectively trained how to learn, how to problem solve, and are given amazing opportunities to test and explore those abilities as far as they're willing to push. **Even if you don't end up at a top school known for its difficulty, it would be a tremendous mistake to assume that your experiences and challenges are any less important or meaningful for it.**

If you have any follow-up questions, feel free to send me an e-mail. My address is on my personal site

Do you ever think computers will ever have 'intentionality', or only 'secondary intentionality' imparted to them by programmers? I.e., will computers ever have a conscience?

sdamaandler

If I'm understanding the question properly, in that you're asking whether computers will have desires/goals of their own versus only those dictated by their programmers, I would say that it may become easy to confuse the two and that the distinction can become fuzzy as the originally

programmed goal is increasingly far away.

Let's say a robot's is programmed to bring you a cup of coffee. If it takes the garbage out at some point during the process, it may be easy to overlook that the robot is only doing that because it thinks it's full and won't be able to throw away the coffee filter otherwise. As a human watching this process, we may not see that connection (especially early in the process or without the same information the robot has) and mis-attribute it as intentional.

The question of a robot/computer system having a conscience is more open-ended -- what is the minimum set of requirements for something to be considered as exhibiting a conscience? If we give some kind of accident/hazard avoidance capabilities to a manufacturing robot, I don't think anyone would say that it has a conscience merely because it doesn't do actions that would otherwise harm humans around it. All the same, these are complicated questions and it's important that people are thinking about these issues / keeping them in mind.

Xheotris also makes a good point about needing to be careful with respect to injecting our own biases.

Is it easier for an algorithm to learn to speak at a fourth grade level, or as if it were Shakespeare?

Overthinks_Questions

I would actually say it may be more difficult to learn to speak at a fourth grade level than to mimic Shakespeare, if only because (from my naive perspective) the constraints of "speaking like a fourth grader" are less well defined than "mimicing Shakespeare". As another commenter points out, the availability of labeled data also heavily contributes to my intuition for this question.

What would you say is the "tone" of an AI? with Trump, there's only one tone to imitate/parody. But if you were to, say, imitate Shakespeare, doesn't this "follow the letter with the letter most commonly used after it" approach fall through? Shakespeare's works have irony, comedy, melancholy in their tones. It seems to me that for an AI to imitate Shakespeare, it would have to "choose" a tone to imitate (because a sentence cobbled from two very different situations will probably have no resemblance to Shakespeare's writing), and "write tragedy like Shakespeare", or "write comedy like Shakespeare". How does it successfully, tonally imitate Shakespeare with the kind of approach you describe?

5_9_0_8

From my perspective, it comes down to the statistics underlying the output. If you were indeed trying to mimic Shakespeare and wanted to separate the stylistic elements of his comedy writing from his tragedy writing, you might need two different models. With a single model you'll probably get some cross-talk between the two higher-level distributions (tragic / comic writing) that you're encapsulting in a single model.

Style is a tricky question in the domain of writing. A fantastic visual analogue is the work in Gatys et al.'s Neural Style paper (see page 5 for the pretty pictures). They're able to use machine learning to capture and isolate the basis of an image's style, then use those same elements to reconstruct new images as if they were also done in the same style. Applying this same technique to writing would require quite a bit of work to ground the reconstruction within the space of grammatically correct / plausible language, as images tend to be far more forgiving of noise than writing.

DeepDrumpf is hilarious, but its tweets seem a bit more on-the-nose than I would expect from an LSTM-RNN. That is, taken at face value, it's like the network has learned the concept of irony. How

much manual filtering are you performing on your output to get just the right tweet for the day?

stochastic_forests

Thanks! I can guarantee you that the network has no understanding of irony, but it is certainly producing output that would seem like it sometimes. As I mentioned in a different response, I'm manually picking out a subset of a larger block of text that's generated from the model. In general, I usually end up with text for a tweet before I figure out who to reply to (rather than the other way around), but that's primarily because I'm trying not to direct the model's output any more than I have to.

Many people think that AI and machine learning would replace humans in many aspects of our life and economy , hence we should stop our population growth and invert the trend with global policies instead ; as a person working in the field :

- What should we do to prepare for a future where humans don't need each other anymore (at least as far as "normal" jobs are concerned) ?

- Aren't human valuable and thus worth keeping around (the more the better) up until the very second before we switch on a recursively improving artificial general intelligence?

- Don't you think that we should first understand consciousness before switching on an AGI ? Doing so we could assign to such AGI the only clear goal of protecting our consciousness/flow of consciousness (whatever that is) and let it figure out how

- In the future there will certainly be many people who would try to rush things up with AI/AGI because they'd fear that they might miss out on the benefits of such enormous advancement , how can we address such scenarios and make sure that we proceed with extreme caution?

- Even if a global effort were to be made to build an AGI (no competition and/or secrecy between nations/companies) an individual or a group of people would get there before all the others , how can people be sure that those who get there first would share for free the benefits of such tech considering how throughout the history of our specie that has never been the case ? Should we accept and embrace this arms race as the final act of natural selection ? Are we looking at an "every man for himself " kind of situation?

AjaxFC1900

I'm not sure I see the connection between population growth and AI displacing jobs -- if anything, the more popular concerns that I encounter about post-scarcity economies would suggest that the benefits of such systems would free us from concern about things like population growth. This is pretty far outside my scope of expertise, as I would say most of this falls into philosophy, but I'll give them a shot! The short version is that I don't view AGI as a likely outcome and I don't think this is a pressing enough concern to actually worry about right now.

What should we do to prepare for a future where humans need don't need each other anymore (at least as far as "normal" jobs are concerned) ?

I'm not sure it's reasonable to expect a future where humans don't need to cooperate to succeed (for some complicated definition of what it means to succeed), but if the question is more meant to get at what to do in the face of mass unemployment: Plenty of smart people are looking at solutions like 'basic income', though there's a fair bit of skepticism about its practicality or effectiveness.

Aren't human valuable and thus worth keeping around (the more the better) up until the very second before we switch on a recursively improving artificial general intelligence?

I'd say humans are generally valuable and worth keeping around even past the scenario of an infinitely

improving intelligence. From my perspective as a roboticist, humans are experts at manipulation/navigating our world and robots generally have a pretty hard time with it. So even in the worst case scenario where all human cognitive capability is made unnecessary, the system that did so would still have to solve some pretty difficult problems.

Don't you think that we should first understand consciousness before switching on an AGI ? Doing so we could assign to such AGI the only clear goal of protecting our consciousness/flow of consciousness (whatever that is) and let it figure out how

Personally I don't think we have much to fear here given that I think an AGI in the science fiction sense is very unlikely. I think it's a lot more important to focus on immediate-term dangers of runaway optimization for systems that we actually have today or will have in the near future... even if they're not quite on par with the paperclip maximizer scenario. Rather, we should make sure that we include appropriate penalty terms such that systems always prioritize human safety in task/motion plans over efficiency, for example to avoid harming someone for the sake of trimming a few seconds off of a delivery robot's transit time.

In the future there will certainly be many people who would try to rush things up with AI/AGI because they'd fear that they might miss out on the benefits of such enormous advancement , how can we address such scenarios and make sure that we proceed with extreme caution?

I've heard arguments characterizing the value proposition for solving intelligence as effectively infinite, so it makes sense that people are chasing it. Personally I don't view this as a reasonable concern for a lot of reasons, high among them the many steps required before such a system could even have control over something that may cause harm (but there are many very intelligent people who don't agree with my stance). Unfortunately, if this is a big concern for you, I don't think there's much to do to make people proceed with caution apart from detailing the danger scenarios and hoping they listen.

Even if a global effort were to be made to build an AGI (no competition and/or secrecy between nations/companies) an individual or a group of people would get there before all the others , how can people be sure that those who get there first would share for free the benefits of such tech considering how throughout the history of our specie that has never been the case ? Should we accept and embrace this arms race as the final act of natural selection ? Are we looking at an "every man for himself " kind of situation?

This is pretty philosophical so I'd say my opinion here isn't really worth more than anyone else's, but I would say that you have no guarantees that anyone would even reveal that they have such a technology (I've read arguments about the benefits of trying to keep it a secret, and thought experiments about how to discover if someone even had one). I'd also say that even if someone did manage to create something like what you're describing, they're not under any obligation to share. That said, I strongly, strongly urge you not to characterize AI research and advancements as part of an "arms race".

Hi Brad!

We're a group of college/grad students who made a Trump chatbot on Slack using Hidden Markov Models, tf-idf, and Latent Semantic Indexing as part of PennApps this year. We even had Berniebot debate Trumpbot!

Afterwards we were discussing how to improve Trumpbot and wanted to ask your advice:

- Were characters or words the basis of your features for DeepDrumpf?
- What kind of text pre-processing is required for RNN/LTSM? From our reading it seems like you don't need to clean stopwords as RNN will magically learn the structure.

- What comprises your dataset? We used speeches, debates, and tweets but still feel like we're not approaching the realism of DeepDrumpf.
- Does DeepDrumpf generate a bunch of quotes that you hand-select to tweet out, or is it "fully automated"?
- Lastly, would a move to RNN alone be enough to see a huge improvement in realism?

Thanks!

#makeslackgreatagain

strobe8

That's an awesome project! To answer your questions:

DeepDrumpf uses characters as a basis. I've also trained models using words, but it has less room for creativity with respect to creating new words (e.g., Scamily or Russiamerica).

No text pre-processing except for making sure you're consistently using the same type of quotes and apostrophes throughout. Even then it's not required, but doing so will make your model better.

It sounds like we're probably using the same general dataset, but I'm doing a bit of post-processing on it to make the output more sensical. Also, since I'm only tweeting out sporadically, I get to hand-pick the best subset of the model's output.

I typically generate a big paragraph and hand-select what I think will be funny to post, otherwise there'd be a lot of tweets that are plausible but repetitive or boring.

Probably not, as it seems that you're getting quite a bit of quality out of your existing model. I think if you could provide specifics on what you'd like to improve it'd be a bit easier to answer, but I would suggest forcing your model to learn distributions on a per-topic basis to constrain responses to be "relevant" to the prompt/input.

Again - great job!

So... far aside from theoretical considerations : on a very practical level, what language(s) and tool(s) did you use to achieve this? Would you mind posting your code (on github)?

Do you believe that science should be open-source / freely available?

What books would you recommend to total beginners who are eager to code Machine Learning / AI apps? Preferably non-abstract/theoretical. Thx

hintre

DeepDrumpf is written in Python, and uses TensorFlow, NLTK, and PyEnchant. I have every intention of posting my code and dataset on GitHub -- I'm in the process of writing a research paper about it and the world's (widely varying) responses to it. For a while I was training the model on an older NVidia GeForce GTX 680, but was thankfully able to find a GTX 1080 on Craigslist that I could afford, which let me sample models over a much larger parameter space.

I personally believe that all publicly funded work should be freely available and that scientific knowledge should be shared (and made easily reproducible where possible), but I don't think I have enough information about opposing viewpoints to make a fair argument against them. I suspect it's largely a practicality issue, with someone having to pay for the hosting, curation, formatting, etc.

If you're looking to get started, I wouldn't even necessarily pick up a textbook. I strongly suggest finding a brief/simple tutorial and following it through, then trying to pick a very small scale project to guide

your exploration. I gave a bit more verbose of an answer here.

What role do you predict AI will play in renewable energy / global warming?

nikolabs

I imagine that AI and machine learning will play a large role in the energy sector in general. Seeing articles like this one about successfully using machine learning to improve energy efficiency are very exciting, since it shows that we can do quite a lot with the infrastructure we already have in place. I'm particularly excited to see the effects that distributed energy storage networks powered by devices like the Tesla Powerwall have on our national power infrastructure.

We'll certainly be able to take advantage of machine learning and AI techniques to aid in the development and testing of new materials and technologies. AI is all about solving the problem of solving problems -- we have powerful general purpose tools that often require considerable effort to tailor to specific applications, but it's a safe bet that it will play a large role in this industry.

what was your career path? and how did you end up at MIT? I'm interested in HCI (human computer interaction) and applying for my masters programs right now but having a hard time deciding whether this is the career path for me (or if HCI may become obscure in the future)

keysandpencils

My career path has been a lot of fun, but planned with a relatively short horizon. I knew I wanted to do a Computer Science degree since before I went to college, but the shift to AI/Robotics didn't surface until later. I had been encouraged to seek out summer internships during my undergraduate years, and was lucky enough to have the opportunity to do internships at IBM, IBM Extreme Blue, and Microsoft. By the time I was a senior undergraduate, my interests shifted somewhat from launching a startup immediately following college to wanting to get experience with some real problems at the intersection of computer vision and HCI. I was interested in some of the work coming out of the MIT Media Lab at the time, but given my disinterest in research and interest in building things, was convinced to pursue these goals instead at BAE Systems -- a defense contractor with an office near Boston that was working on some really interesting problems.

I absolutely learned a lot when I was there, and was encouraged to go back to school for a PhD based on my interests in AI/Machine Learning. Joining a robotics lab was somewhat happenstance, as I was primarily interested in AI/ML and initially saw robots merely as an interesting application domain for it. I'm really glad I ended up in a robotics lab though, as I found (with help from my advisor) that I particularly enjoyed building systems and solving the problems intrinsic to human-robot collaboration, a subset of human-computer interaction.

If anything, I'd say that the lesson I learned is to not be too afraid of trying small diversions from what you think is your best path forward, since otherwise I wouldn't have ended up where I am now. I sincerely doubt HCI will become less important as time goes on. If anything, my intuition is that as we build increasingly complex systems, HCI and Human Factors work will become even more important.

what happened when you applied the algorithm to hillary's tweets?

Herxheim

I haven't tried applying it to Hillary yet, but I did make one for Bernie Sanders called @DeepLearnBern. Unfortunately, it was a lot harder to get short, funny quotes from that model, despite having more

training data. I eventually decided to focus on making one parody the best I could rather than splitting my limited time across many. I chose Bernie over Hillary because my intuition was that his style/platform lended itself a bit easier to be parodied/taken to an extreme, and I didn't have time to try them both (collecting training data takes a fair bit of time).

That said, in case there was a tremendous popular demand for it, I said I would make a Hillary bot if enough people donated at the $10 tier to the charity fundraiser.

Hi Dr. Hayes. I'm a current undergrad computer science student. I'm very interested in AI and robotics, but my school doesn't have any resources for either. Where would you recommend someone start to learn those?

Also, I looked at your web page and noticed you have had internships at IBM and Microsoft. What kind of projects and experience did you have to get those? Thanks in advance.

Ipuncholdpeople

I commented a bit about this here and elsewhere in the thread -- the best thing to do is start looking for tutorials online and pick a small project to do!

My internships each successively helped me get to the next steps of my career. I had originally misread an IBM internship post meant for rising seniors, applied anyway, and was given the chance to interview -- I passed the interview and took the internship. My team there was working on some internal tools for IBM, I recall doing a lot of XML parsing in Java, but it's been over a decade so I don't completely remember the specifics.

My experience at IBM Cambridge gave me the contacts and experience on my CV to make me a competitive applicant for IBM Extreme Blue, which I applied for in December 2005 and was summarily rejected from within 48 hours. Months later I was called to see if I was still interested and went through multiple interview rounds in a few days, eventually getting accepted. IBM Extreme Blue was an incredible experience that I recommend to anyone, since they give you a lot of support for learning to give effective presentations and to develop ideas into products in small teams. My team worked on tools for helping to automate regulatory compliance checks (think HIPAA or Sarbanes-Oxley) for enterprise customers.

As a rising senior, I applied for an internship at Microsoft where I worked with the Anti-Malware Lower Engine Team. My project there involved creating a scripting language and interpreter that could be used to allow security experts to quickly design and test malware detectors that responded to behavior patterns. Much like all my other internships, I had no background experience with the problem I was supposed to help solve, but with some guidance and a lot of work on the side, I was able to finish my project.

You don't need to have internships to get internships/exciting jobs/etc., though having a portfolio of projects that you've completed is definitely helpful for showing off some of your experience (and mitigating the risk of hiring you).

Have you ever considered to study how people with Asperger's syndrome process social information, facial cues, and small talk? I have wanted to suggest this to someone who works in the field because their success in learning and accurately processing this information is achievable and their struggle could be studied and beneficial to AI research, what do you think?

7billionpeepsalready

This isn't my field, though my old lab collaborated on some work studying Autism Spectrum Disorder. The best resource I can point you toward is the work of Dr. Fred Shic, who has studied social information such as facial cues, prosody, etc. in early ASD diagnosis.

Have you seen the Japanese robot called Robi? Any chance for recreational companion robots to hit the US market without breaking the bank? Also do you know anything about the Google AI that created its own encryption? Is it something to be worried about with the further development of AI? Could this event fuel malicious activity using these new encryption techniques? Could malicious AI become a thing?

zackingels

I had not seen Robi! Thanks for letting me know about it.

Any chance for recreational companion robots to hit the US market without breaking the bank?

Absolutely -- just give it a bit of time for the market to develop. Once the individual components become cheaper, I'm confident that you'll see a lot more social robots out there for purchase.

Also do you know anything about the Google AI that created its own encryption? Is it something to be worried about with the further development of AI? Could this event fuel malicious activity using these new encryption techniques?

I don't know much other than the handful of articles I've read -- definitely a very cool application of generative adversarial networks! The encryption result they have here isn't anything to worry about, and there's no connection to maliciousness. We already have encryption techniques with proven hardness, so if someone wanted to do something malicious and hide it they would be better off choosing a method that is guaranteed to be mathematically sound (guaranteed difficult to crack).

Could malicious AI become a thing?

Sure, but this is a broad term. Creating a program that learns to wake your friend up 20 minutes before their alarm goes off in the morning can be seen as malicious. Should we be worried that someone can make that program? I'd say probably not.

Have you ever read or listened to moral philosopher Sam Harris's views and opinions on AI? He has an interesting point specifically on the idea that if we eventually build a super AI that is more intelligent and capable than the human brain, would we be able to guarantee it would behave exactly as intended in the sense of it being obedient to the humans that built it? Is there a control problem to factor in if we build an intelligence far superior to our own we could become obsolete as a species to the super AI?

P.S I promise this isn't me trying to disguise a question about The Terminator coming true and Skynet taking over.

hesabigladinhe

Sorry, I have not! Though I will say that the problem of guaranteeing that systems behave as we expect is not something that is limited to future concern -- modern robot systems absolutely have the same issues, with possibly disastrous consequences.

As we develop increasingly powerful statistical methods to give us better control systems, we can quickly lose the ability to look at the model and be able to guarantee its output for any given input. For example, some of the work on learning autonomous helicopter tricks is able to produce super impressive results, but guaranteeing that the underlying controller behaves as we want it to (or even

that we can specify what it "should be doing") are incredibly difficult problems.

Brad - Thrilled to see the success you have had since Boston College. Are you still doing any work with Computer Vision?

Question 2 - What is the most interesting project you have worked on?

BCGrad09

Thanks! Did we know each other at BC? Drop me an e-mail and we can catch up!

I am still working with computer vision, though not as exclusively as I was for my senior thesis project. I recently submitted a paper about activity recognition and prediction (with tie-in to human-robot collaboration), but haven't heard back from my reviewers yet.

I'd say the most interesting project I've worked on so far has been the final part of my PhD thesis, where I developed a method allowing a robot to frame the problem of assisting others as an optimization process, using models of both its task and collaborators to figure out creative ways of how to be helpful.

I'm also incredibly excited about my ongoing work developing interpretable machine learning methods for collaboration (currently under blind review, otherwise I'd be happy to get into details).

In terms of your robotics research, do you assume that the behaviour of the cooperating humans is an invariant thing (i.e. robots have to be capable to deal with any kind of cooperative or non-cooperative behaviour), or would you enforce basic rules of conduct and cooperation when interacting with robots? I work in autonomous service robotics myself and am always torn between the fully non-intrusive "the robot will adapt to everything" and the "some care needs to be taken and some rules apply" approach to deploying robots in human domains. It is a complex machine after all and no one would dream of operating industrial machinery without proper training and controlled environments, yet with robots this typically isn't how people approach it. Anyway, just interested in your opinion. Cheers!

chris_jump

do you assume that the behaviour of the cooperating humans is an invariant thing (i.e. robots have to be capable to deal with any kind of cooperative or non-cooperative behaviour), or would you enforce basic rules of conduct and cooperation when interacting with robots?

The assumptions I made really depend on the theme of the research paper. In general, I think robots need to be capable of handling non-compliant interaction partners, even if it means just disengaging from the task until they stop being uncooperative. For manufacturing tasks, as an example, I think you can assume that the team is goal-aligned and that you can make those "everyone will be cooperative"-style assumptions. At the same time, safety is non-negotiable and should always be the top priority for the controller. Robots can quickly become dangerous if this is neglected.

For home robotics, or robots that interact with the general public, I don't think it's fair to assume any kind of compliance or even basic decency will occur.

Regarding human-robot collaboration, when might deep AI be able to help those with disabilities?

themusicdan

We have techniques for that right now, and deep learning will definitely be a useful tool to incorporate.

Improving people's autonomy is definitely one of the most exciting areas in robotics -- check out the [Rehabilitation Institute of Chicago](#) for some really awesome examples.