Satvik Anand

Isha Verma

CS35L

Dec 8 2017

Report: AI models and CAPTCHAs

I wrote my report and did my presentation on AI models and how methods have been developed to make these models solve CAPTCHAs. I also spoke about the implications of these models on the future of cyber-security.

CAPTCHAs, which are a rough acronym for Completely Automated Public Turing Test To Tell Computers and Humans Apart, are just a jumbled group of characters and squiggly lines and other background noise which everyone has seen when they're trying to log in to various websites that require human authentication. CAPTCHAs, while used for authentication on these websites have also become a common method to benchmark how smart an AI model is. If the AI can crack the CAPTCHA, then it is smarter than an AI than cannot solve the CAPTCHA.

There are different ways in which an AI model can break a CAPTCHA. One of these methods is simply exploiting weaknesses in the CAPTCHA, which can be easily defended against by just making small algorithmic changes in the program that creates CAPTCHAs. The article speaks about a new model, which was created by the AI company Vicarious. This new method parses the text more effectively than previous models, with less training. It does it by using a technique called Deep Learning, which is a technique where you have layers of neurons and you train those neurons to respond in a way that you decide. Basically, with this method, thousands of different letters are fed into the system. The letters may be distorted or transformed or changed in different ways so that the computer will get used to reading distorted symbols.

This prepares the computer for a vast array of characters. However, even this system has a very obvious flaw. The main one being that this system is still relying on the computer having seen the letter before. It is possible that if you introduce a new type of CAPTCHA, say one where one letter is juxtaposed over another one, the system will not be able solve it because it doesn't know that there are two separate letters since it has only been fed single letters. Of course this too can be programmed into the computer but this is not the purpose of AI. AI is supposed to understand and learn on the basis of existing knowledge, similar to how a human brain would work. This method involves only interpretation on the basis of existing knowledge; there is no invention or discovery involved. Thus, we need a method on top of this.

This new method is called Recursive Cortical Networking. This is an extension of deep learning which involves a training phase in which while the computer is shown different characters it builds internal models of the letters that it is exposed to, so suppose it is shown an A, it will build its own version of an A and how an A is formed. Thus if it sees an A shown juxtaposed over another character it will be able to recognize the A and therefore crack the CAPTCHA. This method was pretty successful with a variety of online CAPTCHA tests.

All this talk about CAPTCHAs and AI raises an important question about the future of cyber security. If AI is able to crack systems so easily are they really a good thing in our world? But the thing is that while the cracking systems are improving, so are the cryptographic and security processes. This is the march of technology and it is a reasonable expectation to believe that our security will not be compromised.