TechXplore

# Microsoft's racist chatbot Tay highlights how far AI is from being truly intelligent

March 28 2016, by David Glance, University Of Western Australia

Tay. Credit: Microsoft

t has been a nightmare of a PR week for Microsoft. It started with the head of Microsoft's Xbox division, Phil Spencer, having to apologise for having scantily clad female dancers dressed as school girls at a party thrown by Microsoft at the Game Developers Conference (GDC). He said that having the dancers at this event "was absolutely not consistent or aligned to our values. That was unequivocally wrong and will not be tolerated"

The matter was being dealt with internally and so we don't know who would have been responsible and why they might have thought this was going to be a good idea.

But things were going to get much worse for Microsoft when a chatbot called Tay started tweeting offensive comments seemingly supporting Nazi, anti-feminist and racist views. The idea was that the artificial intelligence behind Tay would learn from others on Twitter and other social media networks and appear as an average 19 year old female person. What happened however was that the experiment was hijacked by a group of people from the notorious "pol" (politically incorrect) bulletin board on 4chan and 8chan who set about training Tay to say highly inappropriate things.

This time it was down to Peter Lee, the Corporate Vice President of Microsoft Research who had to say "We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for".

Tay was taken down and the tweets deleted but not before some of the most offensive of them were captured and spread even further on the

Internet.

Apparently, the researchers at Microsoft thought that because they had successfully developed a similar AI chatbot called XiaoIce that has been running successfully in China on the social network Weibo, that Tay's experience on Twitter with a western audience would follow the same path.

Caroline Sinders, an AI interaction designer working on IBM's Watson computer has written a good explanation of how the developers of Tay should have anticipated this outcome and protected against it. There hadn't been enough testing of the bot and certainly the developers of the technology did not have the sociological skills to understand the range of communities on the Internet and what they would do once the technology was released into the wild.

The disturbing outcome of Tay was that Microsoft's Peter Lee saw the problem with the Tay "experiment" as being a technological one that could be solved with a simple technology fix. He missed entirely that the problem was a sociological and philosophical one which unless addressed from that perspective, will always result in technology that sounds superficially human but will always stop well short of displaying any real intelligence.

Chatbots are designed to learn about how language is constructed and to use that knowledge to create words that are contextually correct and relevant. They are not taught to understand what those words actually mean, nor to understand the social, moral and ethical dimensions of those words. Tay did not know what a feminist is when it suggested that "they should all die and burn in hell", it was simply repeating a construct of words that it had input as parts of sentences that it could reformat with a high probability of sounding like it made sense.

It is a testament to human nature's ability to anthropomorphise technology that we make the leap from something that sounds intelligent to an entity that actually is intelligent. This was recently the case with Google's AI software AlphaGo which beat a world-class human player at the complex game of Go. Commentary on this suggested that AlphaGo had exhibited many human intelligence characteristics instead of what it did do, which was efficiently searching and computing winning strategies out of the many millions of games that it had access to.

Even the term "learning" that is applied to AI leads many, including the developers of AI itself to assume wrongly that it is equivalent to the learning processes that humans go through. This in turn leads to the risks of what AI experts like Stuart Russell and Peter Norvig have warned about for many years that an "AI system's learning function may cause it to evolve into a system with unintended behavior".

The experiment with Tay highlighted the poor judgement of developers at Microsoft as much as the limitations of chatbots. It seems that in this case both humans and software might not learn the real lessons from this unfortunate incident.

*This article was originally published on* The Conversation. *Read the* original article.

Source: The Conversation

Citation: Microsoft's racist chatbot Tay highlights how far AI is from being truly intelligent (2016, March 28) retrieved 8 May 2024 from https://techxplore.com/news/2016-03-microsoft-racist-chatbot-tay-highlights.html