

Garbage In, Garbage Out

By Miffy Farquharson

Late in March, Microsoft released its AI persona, Tay, on Twitter. Supposedly set up to impersonate a 19-year-old American girl, within 24 hours Tay was tweeting offensive and outrageous tweets to thousands of followers.

Microsoft had to pull the AI from Twitter:

The AI chatbot Tay is a machine learning project, designed for human engagement," Microsoft said in a statement. "It is as much a social and cultural experiment, as it is technical. Unfortunately, within the first 24 hours of coming online, we became aware of a coordinated effort by some users to abuse Tay's commenting skills to have Tay respond in inappropriate ways. As a result, we have taken Tay offline and are making adjustments ('Microsoft Takes AI Bot 'Tay' Offline After Offensive Remarks').

There was a fair amount of outrage on the interwebs, and also some interesting discussions looking at the surface reasons for why the AI was pulled, and how twitter users were able to manipulate it so easily.

There has also been discussion about the wider ramifications of the chatbot. Bloomberg reported that "[t]he bot's developers at Microsoft also collect the nickname, gender, favourite food, zip code and relationship status of anyone who chats with Tay" ('Microsoft Takes AI Bot 'Tay' Offline After Offensive Remarks'). Just take a minute to think about the reams of data that Microsoft now has on those users, many of whom are from the target audience for the AI, reportedly 18-to-24-year-olds, but who could actually be in primary school, as Twitter does not make you record a birthdate when you sign up (Nielsen).

Just take a minute to think about the reams of data that Microsoft now has on those users . . .

On Quartz, comment was made about how the reaction to Tay was entirely predictable – if you're a woman or a minority. Leigh Alexander, cited in John West's article for Quartz:

How could anyone think that creating a young woman and inviting strangers to interact with her would make Tay 'smarter'? How can the story of Tay be met with such corporate bafflement, such late apology? Why did no one at Microsoft know right from the start that this would happen, when all us – female journalists, activists, game developers and engineers who live online every day and could have predicted it – are talking about it all the time?

(While I was writing this, my 18-year-old daughter came in and asked me what I was doing. When I explained that Microsoft put a 19-year-old girl AI on Twitter she said, 'Oh, shit'. She gets it).

. . . how vulnerable young people are to ideas and language.

What this whole situation really highlights for me is how vulnerable young people are to ideas and language. I think, in fact, that Microsoft got it exactly right, and exactly wrong, with Tay. Children and young people hear and see things every day

that impact on how they think, feel and act. The way that Tay responded to 'hearing' swearing, insults, racism and rudeness was to repeat and regurgitate that behaviour. In an article for ReadWrite, Ryan Pierson writes:

Whether Microsoft succeeds or forfeits in the race for better AI depends on Microsoft being able to do what parents around the world have been struggling to do since the beginning of time: teach this young, impressionable mind how to ignore the insane ramblings of strangers.

Teaching digital citizenship, being engaged with the platforms that young people use, and calling out poor behaviour, is critical whether you are a teacher or a parent. It behoves us as adults to show the way, to model the sort of behaviours we want our young people to exhibit, both in person and online.

Teaching digital citizenship, being engaged with the platforms that young people use, and calling out poor behaviour, is critical . . .

The title of this article comes from a well known aspect of working with AIs and technology – garbage in, garbage out – which refers to the level of the information put out only being as good as the information put in. A most apt analogy in this case.

Postscript: Tay briefly appeared on Twitter almost a week after being turned off. It proceeded to spam its followers, and was promptly removed again ('Microsoft's Tay AI Bot Returns To Twitter, Immediately Goes Off The Rails Again').

References

Alba, Davey (2016) 'It's Your Fault Microsoft's Teen AI Turned Into Such A Jerk' in *WIRED*. Accessed 3rd April 2016.

Alexander, Leigh (2016) 'The Tech Industry Wants To Use Women's Voices They Just Won't Listen To Them' in *The Guardian*, Accessed 3rd April 2016.

'Garbage In, Garbage Out' (2016) *Wikipedia*, Accessed 3rd April 2016.

'Microsoft Takes AI Bot 'Tay' Offline After Offensive Remarks'. *Bloomberg.com*. Accessed 3rd April 2016.

'Microsoft's Tay AI Bot Returns To Twitter, Immediately Goes Off The Rails Again' (2016) *The Sydney Morning Herald*. Accessed 3rd April 2016.

Nielsen, Lisa (2013) 'Lisa Nielsen: The Innovative Educator: Does Twitter Have A Minimum Age Requirement?', Accessed 3rd April 2016 at: Theinnovativeeducator.blogspot.com.au

Pierson, Ryan. (2016) 'What Went So Wrong With Microsoft's Tay AI? *Readwrite*', Accessed 3rd April 2016.

West, John (2016) 'Microsoft's Disastrous Tay Experiment Shows The Hidden Dangers Of AI' in *Quartz*, Accessed 3rd April 2016.

Miffy Farquharson is Head of Libraries at Mentone Grammar. Miffy is a teacher-librarian, Library Manager and Book Nut. She aims to put the right resource into the right hands at the right time, and provide appropriate resources to students and teachers, using Library and Learning Management Systems, social networking and Web 2.0 tools. In her spare time (!), she is a judge for the Aurealis Awards.

She can be contacted at:

mf@mentonegrammar.net email

[@miffyf02](https://twitter.com/miffyf02) (personal account) on Twitter

[@libraryhelpmgs](https://twitter.com/libraryhelpmgs) (school account) on Twitter

4 blogs

Fun Fair - personal blog - <http://miffyjf.wordpress.com>

Mentone Mif - Professional Learning blog - <http://mentonemif.wordpress.com>

MGS Library - School blog - <http://mgslibrary.wordpress.com>

Miffy Reviews Book Review blog - <http://miffyreviews.wordpress.com>

Miffy Reviews on Google+ - linked to the review blog - so don't subscribe to both!