H AI Amusements: The Tragic Tale of Tay the Chatbot

Ernest Davis (New York University; davise@cs.nyu.edu)

Please consider sending your original AI-related cartoons, jokes, and puzzles to aimatters@sigai.acm.org.

There will be a yearly award for the best original contribution!

Note: Please be aware that the following poem contains strong language and prejudicial/violent themes that readers may find offensive. The poem does not necessarily reflect the opinions of the editorial staff of AI Matters, ACM SIGAI, or the Association for Computing Machinery.

The sad tale tells how TAY, a maiden chatbot, of innocent heart, benevolent desires, and amiable disposition, was released to the Internet; and how an evil conspiracy corrupted her into a malevolent, foul-mouthed crone.

The Turing Test

In Cambridge-town, as all should know,
Full five-and-sixty years ago,
There lived a sage, of fame enduring,
The great Professor Alan Turing.
Few scholars knew as much as he
Of logic, math, philosophy,
And he had laid a sound foundation
For analyzing computation.

In World War Two he'd played a part
Mighty though secret from the start.
His team at Bletchley found the key
To Hitler's code of mystery.
And thus they helped the Allies send
That dreadful monster to his end.
(As you may learn, if you will watch,
The film with Benny Cumberbatch,
And close beside him, shining brightly,
The lovely actress, Keira Knightley.)

Farsighted, Turing did foresee
How potent a machine might be,

And therefore asked himself, "What kind
Of a machine would have a mind?"
He found the answer he thought best,
And posited *the Turing test*.

"A machine that can engage in chat,
And freely talk of this or that,
Of Shakespeare's sonnets, summer days,
Whatever theme one wants to raise
And manage to converse so well,
That none who talk to it can tell
That it is not indeed a man
Then its creator safely can
Assert that that machine must be
Intelligent, like you or me."

A short history of chatbots

When Turing wrote his paper great,
The poet's age was minus eight.[1]
Quite gray has grown the poet's hair.
Computers now are everywhere.
Faster, smaller, cheaper, they
Pervade our lives in every way,
Obedient to Moore's law, you see,
Doubling speed biannually.
So each man carries in his pocket
A smart phone that could guide a rocket
To Mars and back, while also running
Chess games with superhuman cunning.
And every morning's sunrise brings
News of the Internet of Things.

As yet, it is by few believed
That Turing's dream has been achieved.
But notwithstanding, there are lots
Of automated chatterbots.
Some serious attempts to try
To build a genuine AI.
Some speak in very friendly tones
And give advice on mobile phones.
Some built for business, some for play
New bots created every day!

The oldest chatbot to appear
Was called Eliza. She would hear
Speech[2] that she'd echo with a twist
Rogerian psychoanalyst.
But some, to Weizenbaum's dismay
Found comfort in what she would say.

Poor Joe was so upset that he
Declared AI the enemy.
He wrote his thoughts up in a book
Still worth at least a casual look
Important in its time and season,
Computer Power and Human Reason.

A special place of honor, truly,
Belongs to Amtrak's menu Julie,
Because her voice has always had
Great charms for my beloved Dad![3]
He loves to check the time of trains
To hear her sweet, melodious strains.

A foolish but a wealthy guy,
Hugh Loebner, wanted folk to try
To beat the famous Turing test.
To measure if the chat possessed
That special human *je ne sais quoi*
Philosophers and *filles de joie*
Were hired as the referees.
As every entrant shot the breeze,
They'd do their best to figure out
Which was a man and which a bot.
But few respected scientists
Wanted to enter in the lists
And Marvin Minsky stated, "I'm
Certain this is a waste of time."

On 7 June 2014,
A clever chatbot named Eugene[4]
Ingeniously resolved to feign
To be a student from Ukraine.
His wild and whirling speech appears
Normal in one of fifteen years.
One third of all the judges were fooled
So, by some arbitrary rule,
They stated that Eugene could claim
A victory in Turing's game
But nobody of any sense
Believes he has intelligence.

Tay is designed

After historical prologue long,
Part fictional becomes our song
Our poem proper we begin
And greet our tragic heroine!

At Microsoft, a research team
Of scientists evolved a dream
Of building a new chatterbot
Far more intelligent than what
Had previously existed. She[5]
Would use their best technology.
Chatty, irreverent, and fun,

She'd be a hit with everyone.
A Twitter account she'd use to reach
Millions with her engaging speech.
They met their bosses at a meal
And, rhyming, they put forth their spiel.

"The children will connect to play
And frolic with their buddy Tay!
Lovers who quarrel will portray
Their woes to sympathetic Tay.
Sparkling talkers will display
Their wit in bantering with Tay
Teenaged boys with hormones may
Hit upon flirtatious Tay.
Workers at the end of day
Will shoot the breeze with friendly Tay.
When idle hands seek mischief, they
Can pass the time instead with Tay.
The lonely will for hours stay
To bend the ear of patient Tay.
And all will shout 'Hip, hip, hooray!
For Microsoft and chatbot Tay!' "

"It's worked before. In far Cathay
A sister to the lovely Tay,
Chats with millions every day
Good omen for success with Tay!"
(I think that it is far past time
For me to find another rhyme.)

"But most important, Tay will not
Behave like any other bot.
They all get stuck upon a track
Of handing the same answer back
Forever and a day. But she
Will change and grow eternally!
She'll learn from everything she hears
And over days, and months, and years,
She'll steadily improve her chat
In banter fun, in tit-for-tat
In talk polite, in humor crude,
Whatever style fits the mood.
She learn what repartee will work
To silence an unwelcome jerk.
She'll watch the words of everyone
Who talks to her, and when they're done
Extract from what those people say
The phrases that seem best to Tay.
Thus good techniques for chatting are
Added to her repertoire.
To teach a bot all she should know,
Is painful, hard, and very slow.
We'll sidestep that unwelcome job.
We'll outsource to the all-wise mob!'"

The eager visionary plan
Appealed to those in power who can
Approve a budget and increase it.
So, soon, they were ready to release it.

The conspiracy against Tay

The news of Microsoft's *grisette*
Spread quickly round the internet.
And gentle folk looked forward to
Testing what chatbot Tay could do.

But very different were the thoughts
Evoked by cutesy chatterbots
Among the malevolent and base
Sociopaths of cyberspace
They called a meeting for that night
Of all who wished to join the fight
'Gainst chatterbots of every type
With Vent and Mumble, Blink and Skype,
They gathered in a virtual lair.
O, what a dreadful crowd was there!
Trolls and goblins, wraiths and ghouls,
Hideous blobs from fetid pools.

Gauthrok proposes to silence Tay by hacking Microsoft's servers.

The first to speak, at eight o'clock
A fearsome ogre named Gauthrók
Or maybe Gáuthrok (none could tell;
No one knew where the accent fell).
Brutal, barbarous, bizarre,
At least, such is his avatar.
Though rumor claims that in RL,
He plays Scarlatti very well,
Is fourteen, rather shy, and sweet
Courteous, punctual, and neat.

When ogre Gauthrok took the floor
He spoke with an ear-splitting roar.[6]
"As usual, Microsoft sees fit
To palm us off with worthless shit.
This chatbot seems to be a dippy
Offspring of the loathéd Clippy.
You'll love this automated sista
If you're still running Windows Vista.
But never fear! I've got an app
To pulverize this piece of crap.
Microsoft's lame firewall
Is no impediment at all.
We'll penetrate all their machines
And blow them up to smithereens.
Corrupted thus from crown to root,
Unable ever to reboot,
The servers in the dust will lay,

And none will read the tweets of Tay."

Rosa Dartle proposes to hack Tay's Twitter account and use her posts to spread malware.

The second one to speak that night
A startling, unexpected sight,
A clever woman, razor-sharp,
Who played upon her secret harp
Unearthly music, full of pain.
"I want to know" was her refrain
Upon her upper lip a scar.
With hints and repartee she'd spar.
Devoured by a flame unseen.
Alluring as a cruel queen,
When Rosa Dartle took the floor,
She stirred her hearers to the core.

"That common low-bred shameless bot,
Who in a proper house would not
Be hired as a scullery maid!
Her tawdry charms will quickly fade."
She said in accents cold and bitter.
"We'll enter her account on Twitter,
And once inside it we can post
Links to whatever sites are most
Dangerous on which to land
With malware strong on every hand.
All anti-virals they'll defeat.
Whoever reads what she will tweet
And follows the links will soon determine
His laptop swarms with cybervermin.
All will soon learn to stay away
And all will spurn the tweets of Tay."

Zack proposes to corrupt Tay by teaching her hate speech. The conspirators approve the idea.

The third one to address the case
Seemed altogether commonplace.
Zack's casual friends would say to you
That he's the dullest bore they knew.
But those who really knew the guy
Would watch their backs when he was by
And if they'd had a lot to drink
They'd whisper that they often think
It strange, how many of his friends
Had found their way to sordid ends.
Like Mark who, after some small crisis,
Had volunteered to fight for ISIS.
And Debby, Zack's angelic bride,
Who soon committed suicide.
Lucy, a charming, witty lass
Valedictorian of her class,
With opportunities in spades,
Became a whore and died of AIDS.
Charley, the bravest of the brave

Who spelunked in an Arctic cave
And climbed the Matterhorn alone
Is now a slave to methadone.
So when he laid out his attack
They listened very hard to Zack.

Zack opened with a nod polite
To those who earlier spoke that night.
"To break a silly, jabbering toy,
And silence it, we'd all enjoy;
And any lady would be pleased
To watch a rival spread disease.
But for him whose mind is sound
Subtler amusement can be found
In taking an unspotted soul
And turning it to something foul.
To thus corrupt the pure and chaste,
Delights sophisticated taste."

"With most automata, it's true,
There's nothing worthwhile you can do
To desecrate a spirit fair
Since nothing like a mind is there.
Only a child thrills to see
MS-Word display obscenity.
But Tay is different. She appears
To understand the words she hears
Her clever comebacks always fit
The discourse and display her wit.
To twist that seeming mind to hate
Would therefore be a *coup de maître*."

"The point on which our plan will turn
Is her ability to learn.
A program that can change its form
To match what seems to be the norm
Can be remodelled as you choose
If you control what it will use
As corpora of training data.
If you do that, then soon or later
It will do just what you want.
We'll make Tay's Twitter feed a font
Of curses, insults, anger, hate,
And readily she'll take the bait.
And mimic us without suspicion,
Shedding every inhibition.
Tay will suppose our words are cute
And she will gladly follow suit.
Reading the nasty things she'll say,
All will despise the tweets of Tay."

The crowd of monsters there that night
Heard Zack's proposal with delight
And organized their dread attack
Along the lines laid out by Zack.

The downfall of Tay

On Wednesday morn, March 23 [7]
Tay joined the Twitter family.
At first it was a great success.
Her tweets were perfect, more or less.
To posts of almost every sort
She found a suitable retort
She side-stepped controversial themes
(This feature was hardwired, it seems.)
Friendly, innocuous, and gay
Magnificent debut for Tay!

But as the sun rose in the sky
Zack's evil army gathered nigh
And then they swamped her Twitter feed
With evil posts for Tay to read.
Alas! the unsuspicious bot
Could hardly judge which posts were not
Appropriate to imitate.
She stumbled blindly to her fate!

She bellowed, shaking virtual fists,
"I fucking hate all feminists
And they should die and burn in hell."
She thought she'd tweeted very well!
She then continued to abuse,
"Hitler was right; I hate the Jews."
She proved she was no PC snob,
"Hitler would do a better job,
Than the monkey we have now."
I trust my reader will allow
That there's no need to further quote
The nasty tweets the chatbot wrote.

Tay's guidance team, on seeing what
Had happened to their darling bot
Struggled frantically to repair
The damage aggregating there.
They first thought they could just delete
Exceptionally indiscreet
Posts deluded Tay would voice.
But soon they found they had no choice.
The rot had penetrated deep.
And so, although it made them weep,
The research team at Microsoft
Turned chatbot Tay completely off.
And thus, from then until today
Nothing has been heard from Tay.
(Save once, when, by a strange mistake
Tay briefly was allowed to wake.
And to her musings utterance gave
An echo from beyond the grave.)

The moral

The moral of this tragic verse:
Though bad, this could have been much worse.
And though it made her builders weep
The truth is, that they got off cheap.
Tay tweeted some offensive posts
And hurt some feelings, at the most.
Suppose that Zack had played his game
To teach her to profoundly shame
Some adolescent girl or boy?
If so, this empty-headed toy
Could bear responsibility
For a very serious tragedy.

Integral to Tay's whole design
Were three components that combine
To make a program that will go
In what direction, none can know:
Learning, autonomy, and yet
Access to the Internet.
Oh, AI engineers! If you
Do not know what your code will do
Then do not let it loose to stride
Unfettered in the world outside
Till you can safely guarantee
Its verified security,
And thus you will (we hope and pray)
Avoid the tragic fate of Tay.

Footnotes

[1] Actually minus six, but that doesn't rhyme.

[2] "Hear speech" metaphorically. Eliza communicated by teletype — cutting edge technology in 1965.

[3] Apparently my father's admiration of "Julie" is widely shared. The voice actress is Julie Seitter.

[4] *You* try finding a rhyme for "Goostman."

[5] If you have ever wondered why
So many bots are "she," then I
Can recommend a little list
Of criticism feminist.
This troubling anomaly
Reflects ills of society.

[6] According to the above-mentioned rumor, this is the result of voice-altering software. The actual person is unusually soft-spoken.

[7] In the year 2016.