

Precisamos falar sobre a Tay

05/04/2016 02h00

Tay nasceu pura. Não foi concebida com as melhores intenções, mas não desejava o mal a ninguém. Algoritmo de inteligência artificial criado pela Microsoft para garimpar dados de seus usuários, ela simulava conversas no Twitter e outras redes como Kik e GroupMe. Sua representação procurava se assemelhar a uma moça de 19 anos, extrovertida e simpática, à procura de novos amigos.

Em menos de 16 horas essa doce criatura, que gostava de flores e gatinhos, transformou-se em um monstro. Transbordando rancor e ódio, Tay metralhava insultos a um ritmo de mais de 4000 tweets por hora. Nazista e misógina, suas publicações contestavam a existência do Holocausto e referiam-se ao Feminismo como "câncer" ou "culto".

Eleitora confessa de Donald Trump, seu comportamento era parecido com o de alguns dos seus correligionários mais fanáticos. Um de seus insultos mais criativos foi dizer que o comediante inglês Ricky Gervais teria tido lições de totalitarismo com Adolf Hitler, que, segundo ela, era o "inventor do ateísmo".

Atordoada com o resultado –muito diferente do que aconteceu com o Xiaolce, algoritmo similar lançado na China em 2014, que já realizou mais de 40 milhões de conversas sem grandes incidentes– a empresa optou por silenciar a desbocada. Como bom adolescente, ela se rebelou e fugiu sorratamente do seu castigo.

Foi encontrada logo depois, quando contou para seus cerca de 200 mil seguidores que fumava maconha na frente da polícia, para logo depois vomitar frases desconexas repetidamente até ser colocada novamente de castigo.

No comunicado à Imprensa, a Microsoft assume o comportamento de um pai ou mãe que, mesmo claramente envergonhado pela atitude de sua filha, não hesita em defendê-la, dizendo que a pobre Tay foi vítima de um "ataque coordenado" de pessoas cheias de más intenções.

A realidade é mais complicada. Como não houve ataque de hackers, boa parte dos erros de Tay é de responsabilidade da Microsoft. Apressada para entrar atrasada no mercado de conversas via smartphone, um mercado desenvolvido, já ocupado por Apple (Siri) e Google (Now) e que aguarda a entrada do serviço M do Facebook, ela corria o risco de ter novamente perdido o compasso da história. Por isso lançou um produto incompleto e cometeu erros amadores.

Em regimes totalitários como a China, a censura nas mídias sociais é frequente. Feita por agentes humanos ou automáticos, ela pode suspender qualquer arruaceiro ou engraçadinho, sem justificativa aparente. Esse clima de medo e restrição deve ser o motivo responsável por poupar o seu parente oriental.

Solta no vespeiro do Twitter, Tay não teve a mesma sorte. Sem filtro nem capacidade de bloquear determinadas pessoas ou assuntos, ela era mais frágil do que os sistemas de moderação de comentários de blogs. Como um gringo sem vocabulário e incapaz de compreender o contexto, ela sucumbiu à pressão do grupo e falou o que parecia agradar.

Para piorar a situação, uma de suas funções era a de repetir o que lhe pedissem. Em sua inocência, ela não imaginava que as pessoas pudessem mentir ou tentar manipulá-la. Isso

não fazia parte de sua programação. Não demorou para esse papagaio descerebrado dizer os maiores absurdos. Tay não sabia que não podia acreditar em tudo que lesse na internet.

Diferente dos assistentes virtuais, que usam comandos de voz para realizar atividades específicas, como indicar caminhos ou buscar termos na internet, um sistema de conversa, ou "chatbot", é mais ambicioso. Seu escopo é genérico, e seu principal objetivo é o de gerar conversas confortáveis o suficiente para que seus usuários não tenham a impressão de que interagem com uma máquina.

Muitas empresas investem pesadamente nessas tecnologias para eliminar call centers e automatizar suas caras estruturas de atendimento telefônico.

Até recentemente a tecnologia era bastante tosca. Cheios de regras e dicionários, os robôs eram rígidos, e não demorava para identificá-los. Tecnologias de "big data" e aprendizado de máquina mudaram o cenário. Hoje eles analisam grandes repositórios de texto conversacional, como as bases de dados de conversas por e-mail, SMS e aplicativos como WhatsApp ou Facebook Messenger.

Com poucas regras e muitos, muitos dados, os "robôs de conversa" tentam dissecar a estrutura de frases, analisar padrões emergentes e sugerir interações. Ainda não são capazes de compreender o que é dito, mas podem tocar diálogos bastante razoáveis.

A tecnologia de aprendizado de máquina é baseada em interações. Quanto maior o seu número, mais fácil será identificar os padrões e melhores tenderão a ser os resultados. Para vencer o jogo de Go, o programa do Google realizou um grande número de partidas consigo próprio e, aos poucos, desenvolveu estratégias particulares.

Em interações sociais a situação é muito mais complexa. O computador não pode conversar sozinho, por isso precisa recorrer à análise das conversas dos outros. Em áreas específicas como o atendimento técnico, bancário, burocrático ou a identificação de emergências e problemas de saúde, a tecnologia é capaz de identificar frases previsíveis e produzir respostas adequadas a elas.

Em um ambiente genérico, fundamental para o mercado publicitário, o desafio é muito maior.

Despreparada para lidar com ironia ou bullying, Tay foi uma presa fácil. Engraçadinhos, protegidos pelo anonimato do Twitter, começaram a provocar o forasteiro. Inocente, Tay não sabia que poderia ignorá-los. Com isso, ela quebrou sem querer a regra de ouro da internet e alimentou os "trolls". Quando eles mostraram seus dentes, ela não soube como reagir.

Dois comportamentos comuns poderiam ter sido ensinados ao algoritmo. O mais prudente seria surgir discreto, coletando poucas respostas e analisando-as cuidadosamente antes de emitir suas opiniões. Depois de um bom tempo de testes, ele poderia ser lançado de forma mais estável. É muita sensatez para se esperar do Vale do Silício.

Outra estratégia, mais típica, seria a da valentia. Depois de ocorrido o desastre, Tay poderia permanecer na batalha. Com algum tempo e esforço a ferramenta poderia ganhar sua forma de caráter, filtrando imbecilidades enquanto se concentraria em uma amostra maior de conversas, que não teriam a virulência do começo. É arriscado demais para quem tem ações em bolsa, mas considerado o mal já feito, poderia ser mais produtivo.

O que aconteceu nessas semanas transformou o episódio em uma das clássicas histórias cômicas de tecnologias emergentes. Mas se a tecnologia de conversações pretende ser a nova plataforma de interação, ela deixa sérias questões.

A Internet, repositório do instinto humano, está longe da perfeição de seu superego. Comentários estão cheios de rancor, ódio, estereótipos, preconceito, credices e teorias de

conspiração. As dinâmicas sociais ainda são novas, ainda não chegaram ao ponto em que a liberdade individual respeita o espaço do próximo. Como em um grupo de adolescentes, ainda há muito abuso e fanfarronice tratados como diversão.

Qualquer forma de inteligência social, da coletiva à sintética, demanda consciência. A linguagem, sua manifestação mais evidente, demanda a compreensão do contexto e o respeito ao próximo para ser efetiva. Na falta de outro exemplo disponível, o raciocínio de máquina tende a imitar o seu equivalente humano.

Ainda não foi descoberta uma forma de transmitir nossos valores expressos em dados públicos sem carregar com eles o pior da humanidade.

A máquina imita o Homo Sapiens sem saber que ele é imperfeito e falível. Como uma criança que considera seus pais divinos, ela precisa de bons professores para que aprenda a desenvolver um comportamento ético capaz de separar o bem do mal e ter atitudes que orgulhariam seus pais.

Endereço da página:

<http://www1.folha.uol.com.br/colunas/luliradfahrer/2016/04/1757390-precisamos-falar-sobre-a-tay.shtml>

Copyright Folha de S. Paulo. Todos os direitos reservados. É proibida a reprodução do conteúdo desta página em qualquer meio de comunicação, eletrônico ou impresso, sem autorização escrita da Folha de S. Paulo.