

O USO DE INTELIGÊNCIA ARTIFICIAL PARA A GERAÇÃO AUTOMATIZADA DE TEXTOS ACADÊMICOS: PLÁGIO OU META-AUTORIA?

Marcelo Araujo

Departamento de Filosofia da Universidade do Estado do Rio de Janeiro- UERJ
Doutor em Filosofia pela Universidade de Konstanz, Alemanha
marcelo.araujo@pq.cnpq.br

Resumo

Programas de computador vêm sendo usados para a geração de textos jornalísticos, publicados diariamente na imprensa de língua inglesa. Existem milhares de livros vendidos na livraria da Amazon que não foram escritos por pessoas, mas por algoritmos sofisticados que compilam textos sobre temas específicos a partir de dados disponíveis online. Recentemente, pelo menos dois grupos de pesquisadores, trabalhando de modo independente um do outro, publicaram suas pesquisas sobre o uso de algoritmos capazes de gerar hipóteses científicas para serem testadas empiricamente. Em 2014, a editora Springer constatou que mais de 120 textos, gerados por algoritmos, haviam sido publicados em algumas de suas revistas. Não deve transcorrer muito tempo, portanto, até que o uso de algoritmos para a geração de textos acadêmicos comece a se difundir nas universidades. O objetivo deste artigo é discutir algumas das implicações do uso de inteligência artificial para a produção de textos acadêmicos, especialmente textos de filosofia. Noções como “autoria” e “originalidade” terão de ser redefinidas no futuro.

Palavras-chave: Inteligência Artificial. Algoritmos. Plágio. Meta-autoria. Pesquisa em Filosofia. Sequências *n*-grama.

THE USE OF ARTIFICIAL INTELLIGENCE FOR THE AUTOMATED GENERATION OF ACADEMIC TEXTS: PLAGIARISM OR META-AUTHORSHIP?

Abstract

Computer programs are used on a daily basis for the generation of journalistic texts published in the English-speaking press. There are thousands of books sold at Amazon that have not been written by a person, but generated by sophisticated algorithms that compile texts about specific topics from data available online. Recently, two independent teams of researchers published their research about the use of algorithms for the generation of scientific hypotheses that scientists can test empirically. In 2014, the publishing house Springer realized that over 120 papers, generated by algorithms, had been published in some of its journals. It is reasonable to assume, therefore, that it is a matter of time before the use of algorithms for generation of academic texts becomes widespread in the universities. The aim of this paper is to discuss some of the implications ensuing from the use of artificial intelligence for the generation of academic texts, especially philosophy texts. Some ideas such as “authorship” and “originality” will have to be redefined in future

Key-words: Artificial Intelligence; Algorithms; Plagiarism; Meta-authorship; Research in Philosophy; *N*-gram Sequences



Acredito que daqui a aproximadamente cinquenta anos será possível programarmos computadores [...] para fazê-los jogar o jogo da imitação tão bem que um interrogador mediano não terá mais do que setenta por cento de chance de identificar corretamente após cinco minutos de perguntas.

Alan Turing (1950, p. 59).

1 Introdução

Por que razão uma pessoa deveria passar vários meses de sua vida, talvez anos, escrevendo um artigo, uma dissertação de mestrado, ou uma tese de doutorado, se um programa de computador pode fazer isso para ela em menos tempo? Essa é uma pergunta que jornalistas, estudantes, e pesquisadores se colocarão nas próximas décadas. Sofisticados programas de computador já vêm sendo utilizados para gerar milhares de notícias publicadas diariamente em jornais de língua inglesa. Esses textos, de modo geral, são indistinguíveis de textos escritos por seres humanos. O jornal The New York Times, por exemplo, possui uma página na internet na qual desafia os leitores a descobrirem se os textos ali publicados foram escritos por uma pessoa ou gerados por um algoritmo (THE NEW YORK TIMES, 2015).

Dois estudos recentes sugerem inclusive que, em alguns casos, os leitores tendem a considerar as notícias geradas por algoritmos mais convincentes do que as escritas por jornalistas profissionais. Nesses estudos, os leitores não sabiam de antemão se as notícias haviam sido escritas por uma pessoa ou por uma máquina (CLERWALL, 2014, KAA; KRAHMER, 2014).

A questão sobre se “robôs jornalistas” tomarão o emprego de jornalistas profissionais nos próximos anos já vem sendo amplamente discutida, tanto na imprensa quanto na literatura científica (ANDERSON, 2012; BUCHER, 2016; CROUCH 2015; EUDES, 2014; FASSLER, 2012; FINLEY, 2015; FLEW et al, 2012; LEVY, 2012; MILLER 2015; MOROZOV, 2012; MOSS, 2015; ROOSE, 2014; VAN DALEN, 2012). Mas a pergunta sobre o impacto que tecnologias para a geração de textos terão no meio acadêmico ainda não foi examinada. O objetivo deste artigo é tratar dessa questão. Minha hipótese é que a emergência de algoritmos para a geração de textos afetará não apenas o modo como escrevemos e avaliamos trabalhos acadêmicos, mas já nos obriga também a repensar o que significa pesquisar no domínio das ciências sociais e das ciências humanas, e mais especialmente no âmbito da filosofia.

2 Três autores: Quill, Wordsmith, e Quakebot

As empresas mais conhecidas até o momento, responsáveis pela criação de *software* para geração automática de conteúdo jornalístico, são a Narrative Science e a Automated

Insights. A Narrative Science criou inicialmente um algoritmo capaz de gerar comentários sobre partidas de *baseball* a partir de informações básicas sobre o placar dos jogos, nomes dos jogadores, e o desempenho das equipes em partidas anteriores. Mas a empresa rapidamente percebeu que era possível ir além de comentários esportivos. Ela desenvolveu então um *software* capaz transformar “dados estruturados”, contidos em planilhas, gráficos, tabelas, e diagramas, em textos de prosa simples e descomplicada. O *software* se chama Quill (FASSLER, 2012; SIMONITE, 2015).

A empresa Automated Insights utiliza um *software* semelhante, chamado Wordsmith, para gerar textos jornalísticos. Centenas de *sites* de notícias já publicam diariamente artigos gerados pelo algoritmo da Automated Insights. Os artigos são tão convincentes que eles costumam vir seguidos de um aviso: “Esta história foi gerada pela Automated Insights”.¹ Quando Marvin Minsky faleceu, em janeiro de 2016, a Automated Science gerou um obituário em sua homenagem, publicado na revista Wired. Nenhum outro texto, escrito por um ser humano, poderia ter sido mais adequado para homenagear um dos pioneiros da inteligência artificial (ROGERS, 2016).

Estima-se que mais de um milhão de histórias, geradas por programas de computador, já tenham sido publicadas na imprensa de língua inglesa. Kristian Hammond, um dos fundadores da Narrative Science, acredita que até 2020 mais da metade dos textos jornalísticos serão gerados por algoritmos. Hammond tem a expectativa, inclusive, de que em poucos anos até mesmo prêmios de jornalismo possam ir para alguma história gerada por um algoritmo (ADAMS, 2015). A questão, no entanto, é sabermos a quem caberia o mérito do prêmio nesse caso: ao programa que gerou a história, ou ao programador que gerou o programa? A resposta para essa questão, a meu ver, é mais complexa do que parece, e tem importantes consequências para o modo como, atualmente, avaliamos a “originalidade” de trabalhos acadêmicos.

Em março de 2014, um terremoto de baixa intensidade atingiu Los Angeles, nos Estados Unidos. O primeiro jornal a publicar uma notícia *online* sobre o evento foi o Los Angeles Times, poucos minutos após o abalo (BBC, 2014; OREMUS, 2014). Como “autor” da postagem assina um certo Quakebot. Sobre o “autor”, aparece ao final do texto a seguinte informação: “esta postagem foi criada por um algoritmo criado pelo autor”. Mas que “autor”? Quakebot, o algoritmo que gerou o texto publicado no jornal, três minutos logo após o evento

¹ Basta inserir no Google a seguinte frase (com as aspas) para encontrar uma infinidade de histórias geradas pelo algoritmo da Automated Insights: “This story was generated by Automated Insights.”

noticiado; ou Ken Schwencke, o autor do algoritmo que gerou a reportagem e que estava dormindo enquanto Quakebot escrevia? O “autor” do algoritmo não pode ser exatamente o mesmo “autor” da postagem, pois um poderia ter morrido em meio ao terremoto, sem jamais ter lido o que Quakebot escreveu. Num cenário hipotético – e no entanto nada implausível – Quakebot poderia inclusive ter rastreado informações *online* sobre as vítimas, e ter listado Schwencke entre os mortos. Mas Schwencke não poderia ser o “autor” de uma reportagem que noticia a sua própria morte. O problema quanto à “autoria” de textos gerados por algoritmos é agravado se considerarmos ainda que algoritmos podem ser criados com *deep learning*, ou seja, com a capacidade de se autocorrigirem sem a intervenção de um programador. (GIBNEY, 2016; JONES, 2014; SANTUCCI et al, 2014). O algoritmo poderia detectar, por exemplo, uma gradual diminuição no número de “curtidas” nos textos que gera, e tentar identificar em seguida estratégias para reconquistar seus leitores. Um robô como Quakebot poderia, em princípio, continuar gerando textos por vários anos após a morte de Schwencke, e num estilo diferente daquele previsto pelo seu criador original. Isso torna ainda mais problemática a suposição ingênua de que poderíamos atribuir a Schwencke a “autoria” dos textos escritos por Quakebot.

3 Philip Parker: autor de mais de dez mil livros

Esse problema afeta também a autoria dos livros “escritos” por Philip Parker, que assina milhares de obras vendidas no *site* da livraria Amazon. Parker figura como “autor” de dicionários, relatórios financeiros, compêndios de história, livros didáticos, textos médicos, monografias sobre acnes e uma infinidade de outros temas. Mas Parker, evidentemente, não escreveu esses livros como um pesquisador escreveria. Parker criou um algoritmo que é capaz de reconstruir, passo a passo, todas as etapas que um pesquisador segue ao escrever um texto técnico, ou que um estudante observa ao escrever um TCC (Trabalho de Conclusão de Curso). O que Parker fez, basicamente, foi transformar as instruções contidas num manual de TCC em linhas de um programa de computador. Parker afirma não ter inventado uma nova forma de escrever. Tudo que ele fez foi criar um algoritmo capaz de “imitar” o modo como os pesquisadores e estudantes escrevem. Segundo uma terminologia sugerida pela revista *Wired*, Parker seria um “meta-escritor” (*meta-writer*): um escritor que escreve programas capazes de escrever como um escritor (LEVY, 2012). Parker seria, nesse sentido, o “meta-autor” dos livros que publica na Amazon. Como hoje em dia a maior parte das informações de que um pesquisador necessita para escrever um livro está disponível na internet, o algoritmo criado

por Parker é capaz de gerar um livro sobre praticamente qualquer tema. Numa entrevista concedida em 2013, Parker afirma estar interessado agora em criar um algoritmo capaz de gerar teses de doutorado que apresentem conclusões originais:

Uma das áreas em que estou trabalhando é sobre se podemos criar uma tese com nível de doutorado e que seja inteiramente automatizada – para pouparmos o trabalho de quatro anos de doutorado – e ao final termos ainda uma conclusão original. Se pudermos fazer isso de modo automatizado, nós aumentaríamos a velocidade da descoberta. (Parker apud BOSKER, 2013, não paginado).²

Essa questão foi retomada em março de 2016 num artigo sobre Parker publicado no *Business Times*. A matéria tem como título: “Subvertendo até mesmo o mundo dos acadêmicos” (HOWELLS, 2016). Segundo Parker, cientistas e pesquisadores profissionais são, de fato, responsáveis pela produção de conhecimento novo. Mas uma boa parte dos textos que eles escrevem consiste na sistematização do que já foi escrito e publicado por outros pesquisadores. Em muitos programas de pós-graduação, inclusive no Brasil, a “revisão da literatura” é considerada uma parte fundamental da pesquisa. A revisão de literatura aparece também, às vezes, no início de artigos acadêmicos, e capítulos de dissertações de mestrado e teses de doutorado. A revisão da literatura constitui também uma parte fundamental de muitos “livros textos” e “obras de referência”, indispensáveis para a formação de novas gerações de pesquisadores. Mas, segundo Parker, esse tipo de produção acadêmica poderia ser facilmente delegada a algoritmos, pois o que está aqui em questão não é a produção de uma nova ideia, mas a sistematização do que já foi escrito e publicado pelos outros. A emergência de tecnologias para a geração automatizada de trabalhos acadêmicos, segundo Parker, permitiria aos pesquisadores se concentrar nos *hard problems*, isto é naquelas questões que não poderiam ser analisadas e resolvidas por programas de computador (PARKER, 2016).

A simples possibilidade de gerarmos uma tese de doutorado (ou TCC, ou dissertação de mestrado) por meio de algoritmos sofisticados subverte inteiramente a ideia corrente acerca da “originalidade” de trabalhos acadêmicos. Uma tese de doutorado gerada por um programa de computador deveria ser desqualificada como “plágio”? Acredito que não, pois o algoritmo poderia ser facilmente programado para jamais introduzir no texto final qualquer transcrição literal de um texto consultado. O programador pode também estipular que as

² Texto original em inglês: “One of the areas I’m working on is, can we create a doctoral-level thesis that’s fully automated — to save the pain of four years of a Ph.D. program — and still come up with an original conclusion? If we could do it in an automated way, we’d increase the speed of discovery. As time goes on, someone might be able to say, “I’d really like to do a doctoral thesis on X and Y with a Z perspective,” and with a few clicks, it will have be authored.”

transcrições literais não podem exceder um determinado número de palavras, e devem ser sempre seguidas de uma nota com informações bibliográficas sobre o texto transcrito. A maior parte dos casos de plágio na academia diz respeito à transcrição literal de passagens de textos já publicados, sem a devida identificação das fontes, e não à compilação de ideias disponíveis em outros textos. Na verdade, muitos trabalhos acadêmicos, produzidos hoje em dia no Brasil, são exatamente isso: compilações de ideias já publicadas em livros e artigos. Mas como os textos usados como fonte são devidamente mencionados em notas de rodapé, e listados na bibliografia, raramente encontramos razões para desqualificar esses trabalhos acadêmicos como “plágio”.

Em 2014, a revista *Nature* anunciou que a editora Springer teve de remover mais de 120 artigos de seu sistema de assinaturas de periódicos científicos após descobrir que os textos haviam sido gerados por um programa de computador chamado SCIgen (BOHANNON, 2015; VAN NOORDEN, 2014).³ Os artigos, nesse caso, não eram muito sérios. Eles foram gerados, e submetidos à publicação, para mostrar que o sistema de *peer review* de algumas revistas não é lá muito sério também. Os artigos gerados pelo SCIgen reproduzem o jargão técnico de determinadas áreas do conhecimento e mencionam a literatura relevante, mas sem que uma ideia nova seja realmente defendida. Os textos gerados através do SCIgen foram enviados para as revistas da Springer – e aceitos para publicação – no mesmo espírito de um estudante que, deliberadamente, escreve uma prova ruim apenas para mostrar que o professor ou professora, sem se dar o trabalho de fazer as correções, dará notas aleatórias para todos. Mas os textos gerados pelo programa que Parker desenvolveu, por outro lado, são diferentes dos textos gerados pelo SCIgen. Os textos que Parker gera, através de sofisticados algoritmos, podem talvez ainda não contribuir para o progresso da ciência, pois são apenas compilações de informações já disponíveis em documentos dispersos na internet, mas eles são de todo modo, em vários aspectos, indistinguíveis de muitos trabalhos acadêmicos atualmente defendidos como TCC ou dissertação de mestrado.

Consideremos, por exemplo, um livro de exercícios de língua estrangeira, com palavras cruzadas em inglês, para falantes de português do Brasil que estão se preparando para o TOEFL (*Test of English as a Foreign Language*). Talvez poucas editoras tivessem interesse em publicar um livro com esse perfil. Mas Parker já “escreveu” e publicou esse livro também: *Webster’s English to Brazilian Portuguese Crossword Puzzles: Level 1*. O livro

³ Ver também declaração da editora Springer: “Second update on SCIgen-generated papers”, 14 de abril de 2014, disponível em: <http://goo.gl/WDYYXg>.

custava U\$14,95 na loja da Amazon em março de 2016. Na livraria da rede Saraiva a obra custava R\$ 44,85. Mas o leitor não era informado sobre a “autoria” do livro.

Uma obra gerada por computador poderia também ser oferecida para venda, em qualquer idioma que o algoritmo domine, antes mesmo de ter sido “escrita”. Na patente que Parker obteve para o programa consta a seguinte informação: “o título pode ser escrito (*authored*) sob demanda, em qualquer idioma desejado e com versão e conteúdo mais recentes” (PARKER, 2005, tradução livre nossa). A vantagem para o leitor é evidente: o livro gerado no momento da compra estará em conformidade com a literatura atualizada sobre o tema em questão.

Se as previsões de Kristian Hammond e Philip Parker se mostrarem corretas, algumas ideias aparentemente triviais como “autoria” e “originalidade” terão de ser redefinidas. No futuro, casos de plágio serão um problema menor. A grande dificuldade não consistirá em sabermos se um estudante ou pesquisador reproduziu em seu texto passagens de outros autores. Esse problema já pode ser combatido por meio de ferramentas *online* para a detecção de plágio. O grande problema será sabermos se estudantes e pesquisadores são, de fato, os autores dos trabalhos que apresentam, ou se eles não seriam, antes, apenas “meta-autores” de suas respectivas pesquisas. O “meta-autor”, diferentemente do “autor”, estipula o tema da pesquisa e outros dados acerca do escopo e extensão da investigação, e delega em seguida ao algoritmo a tarefa de vasculhar bancos de dados e de analisar e organizar as informações coletadas. O próximo passo consiste na geração de um livro ou artigo, já em conformidade com as “normas” da revista, da editora, ou da faculdade à qual o trabalho se destina.

Evidentemente, seria possível alegar que o uso de algoritmos para a geração de textos acadêmicos não dispensaria o pesquisador ou pesquisadora da tarefa ao mesmo tempo árdua e criativa de propor uma “hipótese original”. A originalidade de uma pesquisa científica se justifica, ao menos em parte, pela nossa capacidade de formular uma ideia nova, e de explorar essa ideia em consonância com a metodologia mais adequada a cada âmbito de investigação. Essa ideia original, aparentemente, nenhum programa de computador poderia gerar. Ou poderia?

O grande problema com relação a uma tese de doutorado gerada por um programa de computador é que ela não nos permitiria avaliar a competência acadêmica do doutorando ou doutoranda, responsável pela “geração” da pesquisa. Parker, por exemplo, é PhD em economia. As leituras que ele fez, e o esforço que ele próprio despendeu em sua pesquisa de doutorado, devem ter sido imprescindíveis para que ele pudesse um dia adquirir a expertise necessária para criar um *software* capaz de gerar livros automaticamente. É justamente essa

disposição para se engajar na investigação sistemática de um determinado problema, e apresentar ao cabo de três ou quatro anos uma tese de doutorado interessante, que está em jogo na avaliação da competência acadêmica de um doutorando ou doutoranda. Uma tese de doutorado integralmente gerada por algoritmos, portanto, não nos permitiria reconhecer um pesquisador como merecedor do título de “doutor” ou “doutora” em um determinado domínio do conhecimento científico.

Por outro lado, se deixarmos por um momento em suspenso a pergunta sobre o mérito acadêmico do “autor” de um trabalho gerado por um algoritmo e consideramos o problema do ponto de vista das pessoas que são beneficiadas pela difusão do conhecimento, que objeção poderíamos fazer à existência de livros técnicos gerados por algoritmos? Parker sustenta que existem temas sobre os quais ninguém quer escrever, ou livros que nenhuma editora estaria interessada em publicar, pois o público alvo é muito restrito. Esse público, de modo geral, é composto por pessoas que não teriam tempo ou competência para realizar uma pesquisa por conta própria com vistas à publicação de um livro (PARKER, 2013). Mas essas pessoas podem, ainda assim, ter interesse em ler uma obra sobre um tema bastante específico e ainda pouco explorado.

96

4 Gerando hipóteses científicas: KnIT e brainSCANr

No panorama da pesquisa científica contemporânea, a produção de novas ideias está diluída em milhares, às vezes milhões de publicações. Existem atualmente mais de 50 milhões de artigos científicos disponíveis online. Estima-se que a cada dois minutos um novo artigo apareça (LANDHUIS, 2016). Formular uma hipótese original nesse novo contexto é cada vez mais difícil. Mal conseguimos nos manter atualizados em nossa própria área de expertise, quanto mais em outras áreas que também podem ser relevantes para a nossa própria pesquisa. Nenhum cientista consegue mais ler “toda” a literatura relevante sobre a questão que lhe interessa investigar. Contudo, o conhecimento acerca do que vem sendo produzido em uma dada área é fundamental para que uma hipótese relevante possa ser formulada, e em seguida testada. Mas como seria possível então nos orientarmos nesse labirinto de publicações acadêmicas em constante evolução?

Uma solução para o problema consistiria na utilização de *software* para *data mining*, ou seja, programas para “escavar” ou “vasculhar” uma enorme quantidade de informação (*big data*) disponível em plataformas digitais (ANDERSON, 2008; FELDMAN; SANGER, 2007; HAN; KAMBER; PEI, 2011). O programa deve ser capaz de “ler” toda a literatura relevante,

e de sugerir em seguida algumas hipóteses promissoras. Caberia então aos cientistas examinar essas hipóteses, e verificar empiricamente se elas são verdadeiras ou não. Já existem pelo menos dois programas de computador capazes de fazer isso. Um se chama KnIT, o outro se chama brainSCANr. O primeiro foi criado por Olivier Lichtarge e Scott Spangler, o segundo foi desenvolvido pelo casal Bradley Voytek e Jessica Voytek (HODSON, 2014; THE ECONOMIST, 2014).

Lichtarge e Spangler descrevem o funcionamento de KnIT num artigo de 2014 intitulado “Geração de hipóteses automatizadas com base na escavação da literatura científica” (LICHTARGE; SPANGLER, 2014; SPANGLER et al, 2014). KnIT “leu” 186.879 artigos sobre uma proteína conhecida como p53, associada à supressão de tumores em seres humanos. O algoritmo gerou em seguida 64 hipóteses para serem testadas empiricamente. Dessa lista, 9 hipóteses se mostraram verdadeiras. Para avaliar a eficácia do algoritmo, o programa “leu” apenas os artigos publicados até 2003. A razão para isso é simples. Em 2014, Lichtarge e Spangler já sabiam o que havia sido descoberto sobre a proteína p53 nos dez anos anteriores. Mas o algoritmo, não. KnIT não tinha nenhuma informação sobre o que foi descoberto nessa área a partir de 2004, uma vez que os dados mais recentes inseridos no sistema eram de 2003. A ideia de Olivier Lichtarge e Scott Spangler era a seguinte: se pelo menos algumas das hipóteses geradas pelo algoritmo coincidisse com algumas das descobertas feitas a partir de 2004, então o algoritmo estaria fazendo um prognóstico confiável sobre o que, a partir de 2004, valeria a pena pesquisar na busca por novas formas de tratamento contra o câncer. As sugestões propostas por KnIT foram encorajadoras: o algoritmo foi capaz de gerar 9 hipóteses que se mostraram corretas entre 2004 e 2014.

O casal Voytek descreve o funcionamento de brainSCANr num artigo de 2012. O programa analisou informações contidas em cerca de 3,5 milhões de artigos indexados no PubMed, um importante banco de dados sobre publicações médicas. O problema que os dois pesquisadores tentam resolver é formulado de modo bastante claro logo no início do artigo:

Antes de se formular uma hipótese é preciso que se tenha um completo entendimento da pesquisa anterior para garantir que a trajetória de investigação está fundada sobre uma base estável de fatos estabelecidos. Mas como um pesquisador poderia realizar uma revisão integral e não tendenciosa da literatura quando mais de um milhão de artigos foram publicados? (VOYTEK; VOYTEK, 2012, p. 1, tradução livre nossa) ⁴

⁴ Texto original em inglês: “Before one can form a hypothesis, one must have a thorough understanding of previous research to ensure that the path of inquiry is founded upon a stable base of established facts. But how

O brainSCANr examina correlações entre termos científicos, que ocorrem nas publicações sobre um determinado tema, e procura em seguida “buracos estatísticos” (*statistical holes*). Esses “buracos estatísticos” são correlações que ainda não aparecem na literatura, mas que poderiam ser examinadas pelos cientistas. Uma das diferenças entre KnIT e o brainSCANr são as áreas da ciência em que eles são usados. O primeiro vasculha artigos em busca de hipóteses para serem testadas nas pesquisas sobre o câncer, ao passo que o segundo vasculha artigos ligados às neurociências. Mas os dois são capazes de formular hipóteses científicas, uma habilidade que até bem pouco tempo apenas seres humanos pareciam ter. Se nos próximos anos KnIT e brainSCANr se mostrarem realmente eficazes e confiáveis, é provável que o uso de inteligência artificial para a formulação de hipóteses seja estendido a outros domínios do conhecimento nos quais já exista um vasto repositório de literatura científica disponível em meio digital. Isso se aplicaria, por exemplo, à economia, à sociologia, à antropologia, à linguística, à geografia humana, e ao direito. Mas isso poderia se estender também à pesquisa em filosofia?

5 Textos acadêmicos e discursos políticos

A existência de programas como KnIT e brainSCANr, aliada ao uso de algoritmos como aqueles atualmente desenvolvidos por Philip Parker e pelas empresas Narrative Science e Automated Insights, representam um desafio sem precedentes para a nossa compreensão acerca do que conta como “originalidade” e “autoria” no âmbito da pesquisa acadêmica. Esse desafio é especialmente premente, a meu ver, para a formação de novos pesquisadores, pois teremos de rever nos próximos anos o modo como trabalhos acadêmicos serão avaliados. O problema não será simplesmente a sombra da dúvida sobre se um determinado trabalho foi escrito autonomamente, ou se não foi gerado, pelo menos em parte, por algum sofisticado algoritmo. O problema envolverá também a pergunta sobre o que, de fato, estamos avaliando quando avaliamos um TCC, uma dissertação de mestrado, ou uma tese de doutorado. Atualmente, uma boa parte da produção acadêmica, desenvolvida nas universidades brasileiras, consiste na elaboração de fichamentos, compilações, e sistematização de notas organizadas em consonância com certas técnicas de pesquisa. As mesmas técnicas de pesquisa

can a researcher perform a thorough, unbiased literature review when over one million scientific articles are published?”

tendem a ser mais tarde retransmitidas pela nova geração de pesquisadores aos seus respectivos orientandos e orientandas. Mas os desenvolvimentos recentes nas áreas de *deep learning*, *data mining*, e inteligência artificial sugerem que, em princípio, programas de computador poderiam se tornar capazes de gerar textos acadêmicos em consonância com essas técnicas.

A filosofia, aparentemente, poderia figurar como uma disciplina diferente, uma disciplina que primaria pelo pensamento abstrato, e cuja metodologia seria incapaz de ser sistematizada por um programa de computador. No entanto, até mesmo no âmbito da investigação filosófica, muitos trabalhos apresentados hoje no Brasil sob a forma de TCC, dissertações de mestrado, artigos, e teses de doutorado consistem basicamente na compilação e reapresentação de ideias já defendidas por outros autores, sejam eles filósofos clássicos ou não. Um trabalho típico de filosofia costuma ter como título uma frase como “o conceito de x em F ”, em que x está no lugar de termos como, por exemplo, “liberdade”, “autonomia”, “conhecimento”, “natureza humana”, “razão”, “justiça”, “o belo”, “poder”, “direito”, “ação”, ou “sociedade”; e F está no lugar do nome de algum filósofo importante como, por exemplo, Descartes, Platão, Spinoza, Hume, Kant, Hegel, Marx, Nietzsche, Habermas, Foucault, Heidegger, ou Rawls. (Ocasionalmente a palavra “conceito” é omitida no título). Mas a posição efetivamente defendida pelo autor ou autora do trabalho acadêmico, e a discussão sobre se o que F disse acerca de x é verdadeiro ou falso, plausível ou implausível, nem sempre ocorre. É claro que uma investigação filosófica frequentemente nos leva a nos perguntarmos o que F_1 , F_2 , ou F_3 falaram sobre x , pois o que os filósofos falaram sobre x pode ter influenciado o modo como nós já ordinariamente compreendemos x , ainda que nós já mal percebamos o peso dessa influência. Mas a investigação filosófica, a meu ver, não deveria se limitar à simples exegese dos textos de F_1 , F_2 , ou F_3 , por mais influentes que esses textos possam ter sido para a formação de nossa visão de mundo. A meu ver, para além da simples exegese, a investigação filosófica deveria envolver também, por exemplo, questões que jamais poderiam ter sido formuladas por F_1 , F_2 , ou F_3 . E uma dessas questões, agora, é justamente a pergunta sobre as implicações do uso de inteligência artificial na produção de textos acadêmicos, incluindo-se aí “textos filosóficos”. A emergência de novas tecnologias nos obriga a revisitar a pergunta sobre o que significa escrever um texto de filosofia no âmbito da vida acadêmica.

A nossa compreensão acerca do que seja um bom trabalho acadêmico, ou uma investigação filosófica de qualidade, permanecerá a mesma se, por *hipótese*, novos algoritmos puderem gerar nas próximas décadas textos do tipo “o conceito de x em F ” indistinguíveis de textos escritos por seres humanos? Não é minha intenção propor aqui um prognóstico preciso

sobre o que os novos algoritmos serão capazes de realizar no futuro, ou quanto tempo deve transcorrer até que algoritmos possam gerar textos acadêmicos do tipo “o conceito de x em F ” indistinguíveis de muitos trabalhos acadêmicos produzidos hoje em dia. Mas tendo-se em vista alguns avanços recentes na área da ciência da informação, da inteligência artificial, e no desenvolvimento de ferramentas para *data mining*, não me parece que essa seja uma hipótese implausível. Um programa de computador poderia, por exemplo, vasculhar toda a obra de F_1 , F_2 , ou F_3 , e toda a literatura filosófica em torno de F_1 , F_2 , ou F_3 , e localizar em seguida as referências e as citações mais frequentes sobre x no corpus de textos consultados. A partir daí, o algoritmo poderia então compilar a primeira versão, nem que seja ainda uma versão bastante preliminar, de um texto acadêmico do tipo “o conceito de x em F ”. Na verdade, já é possível gerar textos em consonância com esse princípio geral.

No início de 2016, Valentin Kassarig, da Universidade de Massachusetts Amherst, publicou um artigo no qual descreve o funcionamento de um algoritmo capaz de gerar discursos políticos (KASSARNIG, 2016; MIT TECHNOLOGY REVIEW, 2016; PATEL, 2016). O algoritmo faz uso de um banco de dados (ou “*corpus*”) com trechos de 3.857 discursos proferidos por 53 congressistas americanos ao longo de 2005. Segundo Kassarig, o primeiro passo para a criação do algoritmo consistiu na classificação dos discursos em quatro categorias distintas. As categorias resultam da combinação de respostas que podem ser dadas às seguintes questões:

[1] O discurso é republicano ou democrata?

[2] O discurso é de apoio ou de oposição à legislação em votação?

Das respostas a essas duas questões se seguem quatro grupos de discursos: discursos republicanos contra uma legislação, e discursos republicanos a favor; discursos democratas contra uma legislação, e discursos democratas a favor. Em seguida, o programa analisa, para cada grupo de discursos, sequências “ n -gramas” de palavras em que $n=6$. Ou seja, dada uma sequência de 5 palavras, o programa tenta determinar, através de modelos estatísticos, qual deve ser a sexta palavra:

Nós usamos modelos linguísticos estatísticos simples baseados em sequências n -gramas. Mais especificamente, usamos 6-gramas. Ou seja, para cada sequência de seis palavras consecutivas nós calculamos a probabilidade de vermos a sexta palavra, dadas as cinco anteriores. Isso nos permite determinar rapidamente todas as palavras que podem ocorrer depois das

cinco palavras precedentes, e o quão provável é a ocorrência de cada uma delas (KASSARNIG, 2016, p. 4, tradução livre nossa)⁵.

Modelos estatísticos com base em sequências *n*-gramas têm uma infinidade de aplicações. Eles são utilizados, por exemplo, em programas para reconhecimento de voz, em corretores ortográficos, em programas de OCR (“*Optical Character Recognition*”, para a conversão de arquivo de imagem em arquivo de texto), no sequenciamento de DNA, e até no sintetizador de voz do físico britânico Stephen Hawking (JURAFSKY; MARTIN, 2009, p. 84; RUSSELL; NORVIG, 2010).⁶ Análises de sequências *n*-gramas também podem ser usadas para se determinar o estilo de redação característico de cada pessoa. Isso nos permite averiguar, por exemplo, a autoria de textos ou manuscritos que não foram assinados por seus respectivos autores, mas cujo estilo de redação pode ser comparado com o estilo de um autor ou autora cuja obra já tenha sido analisada (CLEMENT; SHARP, 2003; HOUVARDAS; STAMATATOS, 2006). Recentemente, a autoria de um artigo que Antonio Gramsci havia publicado na imprensa italiana, sem se identificar para evitar represálias, pôde ser comprovada graças à aplicação de análises estatísticas desse tipo (BENEDETTO et al, 2012).

O que o algoritmo criado por Kassarnig faz é, basicamente, estimar a probabilidade de uma determinada palavra ocorrer em sequências 6-gramas tendo-se em vista um *corpus* de discursos políticos previamente estipulado. A cada palavra é atribuído um valor numérico que exprime a probabilidade de sua ocorrência na sequência em análise. Esses valores são diferentes para cada um dos quatro tipos de discursos contidos no *corpus*. Para garantir coerência do texto, Kassarnig também introduziu no programa restrições que garantem a correção gramatical, e a consistência na sucessão de frases. Além disso, foram atribuídos valores para temas específicos como, por exemplo, desarmamento, aborto, pena de morte, impostos, etc. Esses valores também são diferenciados para cada um dos quatro grupos de discursos. Uma vez que os parâmetros tenham sido escolhidos, e as cinco primeiras palavras do discurso tenham sido inseridas, o programa faz o resto. O artigo de Kassarnig é

101

⁵ Texto original em inglês: “We use a simple statistical language model based on *n*-grams. In particular, we use 6-grams. That is, for each sequence of six consecutive words we calculate the probability of seeing the sixth word given the previous five ones. That allows us to determine very quickly all words which can occur after the previous five ones and how likely each of them is.” Russell e Norvig (2010, p. 861) definem sequências *n*-gramas nos seguintes termos: “A sequence of written symbols of length *n* is called an *n*-gram (from the Greek root for writing or letters), with special case “unigram” for 1-gram, “bigram” for 2-gram, and “trigram” for 3-gram. A model of the probability distribution of *n*-letter sequences is thus called an *n*-gram model. (But be careful: we can have *n*-gram models over sequences of words, syllables, or other units; not just over characters.)” (Cf. WARDRIP-FRUIIN, 2012).

⁶ O Google possui um mecanismo de busca específico para sequências *n*-gramas em livros publicados entre 1800 e 2000. A língua portuguesa, porém, não aparece entre os idiomas disponíveis para esse recurso do Google. Ver “Ngram Viewer”: <https://books.google.com/ngrams>. Cf. Roth (2014).

acompanhado de um apêndice com exemplos de discursos gerados pelo algoritmo. Os discursos gerados pelo algoritmo podem ainda não fazer jus à retórica de um grande orador, mas eles são, de modo geral, indistinguíveis dos discursos que os políticos americanos costumam ler no congresso.

Um algoritmo semelhante ao criado por Kassarnig poderia, em princípio, ser elaborado para a geração de textos acadêmicos do tipo “o conceito de x em F ”. O texto gerado pelo algoritmo poderia então ser discutido com o orientador do trabalho, e em seguida editado e reescrito pelo “autor”. É claro que isso somente ocorrerá se houver no futuro um entendimento de que seria aceitável que a primeira versão de um texto acadêmico possa ser gerada por um algoritmo. Essa primeira versão, acredito, seria indistinguível de muitos textos acadêmicos correntemente apresentados em nosso meio acadêmico como TCC e dissertação de mestrado, ou submetidos às agências de fomento como projetos de pesquisa. Alternativamente, é possível também que trabalhos do tipo “o conceito de x em F ” se tornem menos populares na academia, e passem a ser vistos antes como uma etapa preliminar na discussão sobre x do que como fim último da investigação filosófica.

Quando calculadoras de bolso começaram a se tornar populares, muitos educadores consideravam absurda a suposição de que calculadoras poderiam ser utilizadas em sala de aula, ou permitidas numa prova de matemática. Mas atualmente a tendência é bem diferente: muitos educadores procuram incentivar o uso de calculadoras em sala de aula para mostrar que a matemática é uma disciplina criativa, e que a capacidade para realizar operações matemáticas, ainda que indispensável para a formação dos alunos, deve ser subordinada ao exercício da criatividade, à capacidade de formularmos novos problemas e de propormos possíveis soluções. A filosofia, a meu ver, não é muito diferente da matemática nesse sentido específico.

6 Conclusão

Um trabalho acadêmico, por mais original que seja, por mais autônomo que tenha sido seu autor ou autora, é sempre um empreendimento coletivo que envolve pelo menos duas partes: o trabalho do pesquisador ou pesquisadora, de um lado, e de um orientador ou orientadora, do outro. E é por isso que essa parceria é sempre mencionada nos créditos do trabalho. Se no futuro o uso de algoritmos vier a fazer parte da rotina de pesquisa nas instituições universitárias, essa parceria talvez tenha de ser estendida à empresa responsável pela criação dos algoritmos usados na geração da primeira versão do trabalho. Mas, nesse

caso, a pergunta que teremos de responder é se os pesquisadores serão os verdadeiros “autores” dos trabalhos acadêmicos que geram, ou se não seriam, antes, “meta-autores” da pesquisa proposta. Essa é uma questão que terá de ser debatida nas próximas décadas. Meu interesse aqui, por ora, consistiu em chamar atenção para algumas das possíveis implicações que a emergência de tecnologia para geração automática de textos poderão ter no âmbito da pesquisa acadêmica.

Agradecimentos:

Pelas críticas e sugestões à primeira versão deste artigo, gostaria prestar meus agradecimentos a Rachel Herdy, Maria Clara Dias, Fábio Shecaira, Alfredo Culleton, e Gilmar do Nascimento Santos.

Referências

ADAMS, Tim. And the Pulitzer goes to... a computer. Computer-generated copy is already used in sports and business reporting - will machines soon master great storytelling? **The Guardian**, 28 de junho de 2015. Disponível em: <<http://gu.com/p/4a5jq/sbl>>. Acesso em: 08 ago. 2016.

ANDERSON, Chris. The end of theory: the data deluge makes the scientific method obsolete. **Wired**, 23 de junho de 2008. Disponível em: <<http://goo.gl/wpM1Xl>>. Acesso em: 08 ago. 2016.

ANDERSON, Christopher. Towards a sociology of computational algorithmic journalism. **New Media & Society**, v. 15, n.7, p. 1005-1021, 2012. Disponível em: <<http://nms.sagepub.com/content/early/2012/12/06/1461444812465137>>, (doi 10.1177/1461444812465137). Acesso em: 08 ago. 2016.

BBC. Robot writes LA Times earthquake breaking news article. **BBC News**, 18 de março de 2014, Technology. Disponível em: <<http://goo.gl/fyjdkd>>. Acesso em: 08 ago. 2016.

BENEDETTO, Dario et al. The unreasonable effectiveness of mathematics in human sciences: The attribution of texts to Antonio Gramsci. In: EMMER, Michele (Ed.). **Imagine math: between culture and mathematics**. New York: Springer. 2012, p. 143-154.

BOHANNON, John. “Hoax-detecting software spots fake papers. **Science**, 27 de março de 2015, News. Disponível em: <<http://goo.gl/wf1wB5>>. Acesso em: 08 ago. 2016.

BOSKER, Bianca. Philip Parker’s trick for authoring over 1 million books: don’t write. **The Huffington Post**, 11 de fevereiro de 2013. Disponível em: <<http://goo.gl/R0z8kc>>. Acesso em: 08 ago. 2016.

BUCHER, Taina. ‘Machines don’t have instincts’: articulating the computational in journalism. **New Media & Society**, v.14, n. 2016 (doi 10.1177/1461444815624182). Acesso em: 08 ago. 2016. Disponível em: <<http://nms.sagepub.com/content/early/2016/01/13/1461444815624182.full.pdf+html>>. (Forthcoming articles published ahead of print). Acesso em: 08 ago. 2016.

CLEMENT, Ross; SHARP, David. Ngram and Bayesian classification of documents for topic and authorship. **Literary and Linguistic Computing**, v. 18, n. 4, p. 423-447, 2003. Disponível em: < <https://dialnet.unirioja.es/servlet/articulo?codigo=805138>>.

CLERWALL, Christer. Enter the robot journalist. **Journalism Practice**, v. 8, n.5, p. 519-531, 2014. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/17512786.2014.883116>>, (doi: 10.1080/17512786.2014.883116). Acesso em: 08 ago. 2016.

CROUCH, Ian. The sportswriting machine. **New Yorker**, 26 de março de 2015. Disponível em: <<http://goo.gl/V7xHM8>>. Acesso em: 08 ago.2016.

EUDES, Yves. The journalists who never sleep: ‘Robot writers’ that can interpret data and generate stories are starting to appear in certain business and media sectors. **The Guardian**, 12 de setembro de 2014. Disponível em: <<https://www.theguardian.com/technology/2014/sep/12/artificial-intelligence-data-journalism-media>>. Acesso em: 08 ago. 2016.

FASSLER, Joe. Can the computers at Narrative Science replace paid writers? **The Atlantic**, 12 de abril de 2012. Disponível: <<http://goo.gl/BzcR7B>>. Acesso em: 08 ago. 2016.

FELDMAN, Ronen; SANGER, James. **The text mining handbook**: advanced approaches in analyzing unstructured data. Cambridge: Cambridge University Press, 2007.

FINLEY, Klint. In the future, robots will write news that’s all about you. **Wired**, 3 de junho de 2015, Disponível em <<http://goo.gl/rY896s>>. Acesso em: 08 ago. 2016.

FLEW, Terry et al. The promise of computational journalism **Journalism Practice**, v. 6, n.2, p 157-171, 2012. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/17512786.2011.616655?journalCode=rjop20>>, (doi <http://dx.doi.org/10.1080/17512786.2011.616655>). Acesso em: 08 ago. 2016.

GIBNEY, Elizabeth. What Google’s winning Go algorithm will do next. AlphaGo’s techniques could have broad uses, but moving beyond games is a challenge. **Nature**, v. 531, n. 7594, p. 284-285, 15 de março de 2016, News & Comments. Disponível em: <<http://www.nature.com/news/what-google-s-winning-go-algorithm-will-do-next-1.19573>>, (doi:10.1038/531284a). Acesso em: 08 ago. 2016.

GROSS, Natalie. Does coding count as a foreign language? A Florida bill proposes to let credits from computer-coding classes count like those for Spanish or French courses - but many are skeptical. **The Atlantic**, 24 de fevereiro de 2016. Disponível em: <<http://goo.gl/dxNCPI>>. Acesso em: 08 ago. 2016.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data mining**: concepts and techniques. Waltham: Elsevier, 2011.

HODSON, Hal. Supercomputers make discoveries that scientists can’t. No researcher could read all the papers in their field - but machines are making discoveries in their own right by mining the scientific literature. **New Scientist**, n. 2984, 27 de agosto de 2014. Disponível em: <<https://goo.gl/7PUFIS>>. Acesso em: Acesso em: 08 ago. 2016.

HOUVARDAS, John; STAMATATOS, Efstathios. N-Gram feature selection for authorship identification. In: ARTIFICIAL INTELLIGENCE: METHODOLOGY, SYSTEMS, APPLICATION - AIMS, 12., 2006, Berlin. **Proceedings...** EUZENAT; DOMINGUE

(Eds.). Lectures notes in computer science, 4183. Berlin: Springer-Verlag Berlin Heidelberg, 2006, p. 77-86. Disponível em: <<http://www.icsd.aegean.gr/lecturers/stamatatos/papers/AIMSA2006.pdf>>, (doi:10.1007/11861461_10). Acesso em: 08 ago. 2016.

HOWELLS, Chris. Disrupting even the world of academics. It is only a matter of time before algorithms start to augment a professor's research, taking it into realms previously unimaginable in academia. **The Business Times**, 18 de março de 2016. Disponível em: <<http://goo.gl/JccjCX> e <https://goo.gl/88nYcz>>. Acesso em: 08 ago. 2016.

JONES, Nicola. Computer science: The learning machines. Using massive amounts of data to recognize photos and speech, deep-learning computers are taking a big step towards true artificial intelligence. **Nature**, v. 505, n. 7482, p. 146-148, Jan. 2014. News & Comment. Disponível em: <<http://www.nature.com/news/computer-science-the-learning-machines-1.1448>>, (doi:10.1038/505146a). Acesso em 08 ago. 2016.

JURAFSKY, Dan; MARTIN, James. **Speech and language processing**: an introduction to natural language processing, computational linguistics, and speech recognition. New Jersey: Prentice Hall, 2009.

KAA, Hille van der; KRAHMER, Emiel. Journalist versus news consumer: the perceived credibility of machine written news. COMPUTATION + JOURNALISM SYMPOSIUM, 2014, New York. Disponível em: <http://compute-cuj.org/cj2014/cj2014_session4_paper2.pdf>. Acesso em 08 ago. 2016.

KASSARNIG, Valentin. Political speech generation. arXiv:1601-03313v2 [cs.CL] 20 Jan 2016, Repositório da Cornell University Library, 2016, 15p. Disponível em: <<http://arxiv.org/abs/1601.03313>>. Acesso em: 08 ago. 2016.

LANDHUIS, Esther. Information overload: How to manage the research-paper deluge? Blogs, colleagues and social media can all help. **Nature**, v. 535, n. 7612, p. 457-458, 21 de julho de 2016, Archive. Disponível em: <<http://goo.gl/VYsPRk>>, (doi:10.1038/nj7612-457a). Acesso em: 08 ago. 2016.

LEVY, Steven. Can an algorithm write a better news story than a human reporter? **Wired**, 24 de abril de 2012. Disponível em: <<http://goo.gl/IDNx4v>>. Acesso em: 08 ago. 2016.

LICHTARGE, Olivier; SPANGLER, W. Scott. Automated hypothesis generation based on mining scientific literature [vídeo]. **VideoLectures.NET**, 7 de outubro de 2014, disponível em: <<http://goo.gl/QM37nd>>. Acesso em: 08 ago. 2016.

MILLER, Ross. AP's 'robot journalists' are writing their own stories now. **The Verge**, 29 de janeiro de 2015. Disponível em: <<http://goo.gl/OVmgnf>>. Acesso em: 08 ago. 2016.

MIT TECHNOLOGY REVIEW. How an AI algorithm learned to write political speeches. In: **MIT Technology Review**, 19 de janeiro de 2016. Disponível em: <<https://goo.gl/AF5A63>>. Acesso em: 08 ago. 2016.

MOROZOV, Evgeny. A robot stole my Pulitzer! How automated journalism and loss of reading privacy may hurt civil discourse. **Slate**, 19 de março de 2012, disponível em: <<http://goo.gl/Gt1LvR>>. Acesso em: 08 ago. 2016.

MOSS, Richard. Creative AI: teaching computers to be reporters and storytellers. **Gizmag**, 9 de fevereiro de 2015. Disponível em: <<http://goo.gl/WrpKVJ>>. Acesso em: 08 ago. 2016.

NEW YORK TIMES. Did a human or a computer write this? A shocking amount of what we're reading is created not by humans, but by computer algorithms. Can you tell the difference? **New York Times**, de março de 2015, Quiz. Disponível em <<http://goo.gl/Do78LN>>. Acesso em: 08 ago. 2016.

OREMUS, Will. The first news report on the L.A. earthquake was written by a robot. **Slate**, 14 de março de 2014, disponível em: <<http://goo.gl/5NRZeT>>. Acesso em: 08 ago. 2016.

PARKER, Philip. **Method and apparatus for automated authoring and marketing**. United States Patent and Trademark Office, US 20060064631 A1, Family ID: 36075377. Appl. No.: 11/261.631. Filed: October 31, 2005. [O documento contém informações técnicas sobre o algoritmo criado por Parker]. Disponível em: <<http://goo.gl/biVDAf> e <http://goo.gl/Q4UYAF>>. Acesso em: 08 ago. 2016.

PARKER, Philip. **Webster's English to Brazilian Portuguese crossword puzzles**. Las Vegas: Icon Group International, 2007.

PARKER, Philip. **The computer teaches you**. TEDx, Seattle, 2013. vídeo. Disponível em: <<http://goo.gl/nf8ERt>>. Acesso em: 08 ago. 2016.

PARKER, Philip. **Can algorithms replace academics?** INSEAD Knowledge, 2016 (Entrevista com Chris Howells, 15 de fevereiro de 2016). vídeo. Disponível em: <<http://goo.gl/WzkqoI>>. Acesso em: 08 de agosto de 2016.

PATEL, Neel V. A speechwriting algorithm provides a glimpse into the automated future of crowd-pleasing rhetoric. **Inverse**, 12 de fevereiro de 2016, disponível em: <<https://goo.gl/NwRjXD>>. Acesso em: 08 ago. 2016.

ROGERS, Adam. We asked a robot to write an obit for AI pioneer Marvin Minsky. **Wired**, 26 de janeiro de 2016. Disponível em: <<http://goo.gl/O4f4uH>>. Acesso em: 08 ago. 2016.

ROOSE, Kevin. Robots are invading the news business, and it's great for journalists. **New York**, 11 de julho de 2014. Disponível em: <<http://goo.gl/Cn6VjW>>. Acesso em: 08 ago. 2016.

ROTH, Steffen. Fashionable functions: A Google ngram view of trends in functional differentiation (1800-2000). **International Journal of Technology and Human Interaction**, v.10, n.2, p. 34-58, 2014. <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2491422>. Acesso em: 08 ago. 2016.

RUSSELL, Stuart J.; NORVIG, Peter. *N*-gram character models. In: _____. **Artificial intelligence: a modern approach**. 3. ed. New Jersey: Prentice Hall, 2010, p. 861-865.

SANTUCCI, Vieri G. et al. Cumulative learning through intrinsic reinforcements. In: CAGNONI, Stefano et al (Eds.). **Evolution, complexity and artificial life**. New York: Springer, 2014, p. 107-122.

SIMONITE, Tom. Robot journalist finds new work on Wall Street. Software that turns data into written text could help us make sense of a coming tsunami of data. **MIT Technology Review**, 9 de janeiro de 2015. Disponível em: <<https://goo.gl/Se14Lq>>. Acesso em: 08 ago. 2016.

SPANGLER, Scott et al. Automated hypothesis generation based on mining scientific literature. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND

DATA MINING, 30., 2014, New York. *Proceedings...* New York: Association for Computing Machinery - ACM, 2014, p. 1877-1886. Disponível em: <<https://dl.acm.org/purchase.cfm?id=2623667&CFID=653329881&CFTOKEN=61287529>>, (doi:10.1145/2623330.2623667).

THE ECONOMIST. Automated hypothesis generation. Computer says 'try this'. A new type of software helps researchers decide what they should be looking for. **The Economist**, 4 de outubro de 2014. Disponível em: <<http://goo.gl/itdKG5>>. Acesso em: 08 ago. 2016.

THE NEW YORK TIMES. Did a human or a computer write this? **The New York Times**, 8 March, 2015, Sunday Review. Disponível em: <http://www.nytimes.com/interactive/2015/03/08/opinion/sunday/algorithm-human-quiz.html?_r=0>. Acesso em: 08 ago. 2016.

TURING, A. M. Computing machinery and intelligence. **Mind**, n.49, p.: 433-460, 1950. Disponível em <<http://www.csee.umbc.edu/courses/471/papers/turing.pdf>>. Acesso em: 08 ago. 2016.

VAN DALEN, Arjen. The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. **Journalism Practice**, v. 6, n.5/6, p. 648-658, 2012. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/17512786.2012.667268?journalCode=rjop20>>, (doi: 10.1080/17512786.2012.667268). Acesso em: 08 ago. 2016.

VAN NOORDEN, Richard. Publishers withdraw more than 120 gibberish papers: conference proceedings removed from subscription databases after scientist reveals that they were computer-generated. **Nature**, 24 de fevereiro de 2014, News. Disponível em: <<http://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763>>, (doi:10.1038/nature.2014.14763). Acesso em: 08 ago. 2016.

VOYTEK, Jessica; Bradley, VOYTEK. Automated cognome construction and semi-automated hypothesis generation. **Journal of Neuroscience Methods**, v. 208, n. 1, p. 92-100, Jun. 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3376233/>>, (doi:10.1016/j.jneumeth.2012.04.019). Acesso em: 08 ago. 2016.

WARDRIP-FRUIIN, Noah. N-gram. In_____: **Expressive processing**: digital fictions, computer games, and software studies. Cambridge (Mass): The MIT Press, 2012, p. 203-205.