

# Entrevista com David J. Gunkel<sup>1</sup>

Alex Galeno<sup>2</sup>

PPGCS/UFRN: <https://orcid.org/0000-0001-5103-0339>

Fagner Torres de França<sup>3</sup>

PPGCS/UFRN: <https://orcid.org/0000-0003-2170-4288>

DOI: <https://doi.org/10.21680/1982-1662.2019v2n26ID17160>

## Apresentação

Em sua página na internet, David J. Gunkel, professor no Departamento de Comunicação da Northern Illinois University, nos Estados Unidos, aparece de óculos escuros segurando um enigmático cartaz onde se lê “Robot Rights Now”. Mundialmente reconhecido pelos seus estudos em tecnologia da comunicação e informação a partir de uma abordagem ética e filosófica, Gunkel é autor de aproximadamente 50 artigos científicos e sete livros sobre as questões colocadas pelos avanços tecnológicos na sociedade contemporânea. Ultimamente tem dedicado sua atenção ao fenômeno que chama de “invasão de robôs” em nossa vida cotidiana. “As máquinas agora estão em todos os lugares e fazendo virtualmente de tudo: conversamos com elas *online*, jogamos games digitais, colaboramos com elas no trabalho ou confiamos em suas capacidades de gerenciar vários aspectos de nossa vida complexa crescentemente orientada por dados”, afirma. Não se trata de uma projeção de ficção científica, mas algo que já está acontecendo. Em uma de suas mais conhecidas obras, *The Machine Question* (MIT Press, 2012), reflete sobre a possibilidade de atribuirmos responsabilidades morais às máquinas criadas por nós, principalmente por sua capacidade de interferir cada vez mais nos destinos da humanidade: “há máquinas aprendendo sistemas que são intencional e deliberadamente fabricados para fazer coisas que não são capazes de serem completamente controladas por seus *designers*, programadores ou usuários humanos”, constata. Na entrevista a seguir, inédita no Brasil, David Gunkel nos convida a pensar sobre os diversos aspectos e dilemas (éticos, morais, intelectuais, afetivos) envolvidos em nossa relação doravante incontornável e cada vez mais profunda com as máquinas, robôs e a inteligência artificial.

---

<sup>1</sup> Tradução: Eloisa Klein, professora da Universidade Federal do Pampa, doutora em Ciências da Comunicação pela Unisinos.

<sup>2</sup> Email: alexgalenno@gmail.com.

<sup>3</sup> Email: fagnertf@yahoo.com.br.

## Como você contextualiza o debate sobre o mundo das máquinas hoje?

**David Gunkel**

Nós estamos no meio do que parece ser uma invasão de robôs. As máquinas agora estão em todos os lugares e fazendo virtualmente de tudo: conversamos com elas *online*, jogamos games digitais, colaboramos com elas no trabalho ou confiamos em suas capacidades de gerenciar vários aspectos de nossa vida complexa crescentemente orientada por dados. Conseqüentemente, a invasão dos robôs ou da Inteligência Artificial (A. I., para *Artificial Intelligence*) não é algo que vai acontecer como nós imaginamos na ficção científica, com um exército de andróides saqueadores vindo dos céus com armas de imensurável poder. Trata-se de um evento que já está ocorrendo, com máquinas de várias configurações e capacidades vindo tomar nossas posições no mundo, através de uma incursão lenta, mas estável. Parece menos com *Battlestar Galactica* ou o *Exterminador do Futuro* e mais com *A Queda de Roma*, em termos de representação fílmica.

Conforme estes variados mecanismos venham ocupar posições crescentemente influentes em nossas culturas contemporâneas - posições nas quais eles não necessariamente sejam apenas ferramentas ou instrumentos de ação humana, mas um tipo de entidade de interação social com seus próprios direitos - nós precisaremos colocar algumas questões um tanto interessantes, mas também difíceis. Até que ponto um robô, algoritmo ou outros sistemas autônomos podem ser responsabilizados pelas decisões ou ações que iniciam? Quando, se algum dia, faria sentido dizer: “é culpa do robô”? Por outro lado, quando pode ser devido um certo nível de respeito ou de direito a um robô, ou artefato inteligente, ou outro mecanismo socialmente interativo? Quando, em outras palavras, não seria mais considerado sem sentido perguntar sobre “direitos de robôs” e fazer a pergunta: “os robôs podem ou deveriam ter direitos?”. Em minha opinião, estes são os dois vetores de investigação que definem os presentes desafios e oportunidades que encaramos na face (ou placa frontal) da máquina.

**Por que podemos falar hoje sobre uma ética para robôs? Quais os direitos e deveres dos robôs?**

**David Gunkel**

Quando tratamos da questão dos direitos e deveres a respeito da tecnologia, há uma configuração padrão. Tipicamente é somente um ser humano que possui responsabilidades e direitos. Máquinas - ainda que equipamentos simples como martelos ou sistemas mais complexos, como computadores - são considerados ferramentas neutras de decisões e ações humanas. Este jeito de pensar de senso comum e aparentemente intuitivo é persuasivo precisamente porque é estruturado e informado pela resposta que é tipicamente provida pela questão acerca da tecnologia. “Nós perguntamos a questão acerca da tecnologia”, escreve Martin Heidegger (1977, 4-5), “quando perguntamos o que ela é. Todos conhecemos duas afirmações que respondem nossa questão. Uma diz: tecnologia é um meio para um fim. A outra diz: tecnologia é uma atividade humana. As duas definições de tecnologia caminham juntas. Pois postular fins e obter e utilizar os meios para tais fins é uma atividade humana”. De acordo com a análise de Heidegger, o papel presumido e função de qualquer tipo de tecnologia - seja ela uma simples ferramenta manual, um avião a jato, ou um sofisticado robô - é o meio empregado por usuários humanos para fins específicos. Heidegger chama esta caracterização específica de tecnologia como “a definição instrumental” e indica que tal caracterização define o que é considerado o entendimento “correto” de qualquer tipo de artifício tecnológico.

O desafio (ou a oportunidade) agora mesmo é que nós temos, atualmente, mecanismos que não podem, por uma ou outra razão, ser facilmente acomodados neste quadro teórico. Por um lado, há máquinas aprendendo sistemas que são intencional e deliberadamente fabricados para fazer coisas que não são capazes de serem completamente controladas por seus *designers*, programadores ou usuários humanos. Um bom exemplo é a Tay.ai, da Microsoft. Tay era uma robô social do Twitter, designada para evoluir seu comportamento conversacional através de interações com usuários humanos na mídia social. Rapidamente após ser lançada na internet, a robô se tornou alvo de um grupo de usuários que começaram a manipular as capacidades de aprendizagem de Tay, alimentando-a com afirmações racistas e intolerantes. Como resultado, Tay começou a postar comentários racistas, de ódio, que não podiam ser controlados ou contidos, forçando a Microsoft a tirar a Tay do ar, apenas 16 horas depois de lançada. Então, a questão é: “quem ou o que poderia ser responsabilizado pelos tweets intolerantes”? Responder a essa questão, no entanto, é difícil. Logicamente, os engenheiros da Microsoft não configuraram ou criaram um robô social racista para o Twitter. E quando se olha para o que a Microsoft disse em resposta a esse evento, as coisas ficam ainda mais

confusas. Ao invés de tomar responsabilidades sobre os maus resultados, a Microsoft buscou culpar a comunidade de usuários do Twitter e a própria robô. Embora a medida com a qual se possa atribuir “agência” e “responsabilidade” a um mecanismo como o robô social do Twitter Tay prossiga um problema contestado, o que não é debatido é o fato de que as regras do jogo mudaram e há um alargamento do “vão de responsabilidade” na face destes dispositivos. Consequentemente, a teoria instrumental da tecnologia, que tinha amarrado efetivamente ação mecânica com agência e responsabilidade humana, não mais se aplica adequadamente a mecanismos que têm sido desenhados deliberadamente para operar e exibir alguma forma - não importa quão rudimentar - de ação independente ou tomada de decisão autônoma.

Por outro lado, temos (ou estamos desenvolvendo atualmente) vários tipos de companheiros robóticos que estão desenhados para ser mais que apenas ferramentas ou instrumentos. Robôs sociais como *Buddy*, um robô familiar emotivo, da *Blue Frog Robotics*, na França, são deliberadamente desenhados para ocupar uma posição social que está situada em algum lugar entre o que são considerados meros objetos e outra pessoa, que tem presença social reconhecível e um status desta presença social. Isso não é necessariamente novo ou produto de Inteligência Artificial avançada, mas já foi identificado e demonstrado em investigações empíricas com um objeto até bastante besta. Nos estudos sobre computador como ator social, conduzidos por Byron Reeves e Clifford Nass, no meio dos anos 1990, por exemplo, foi descoberto que usuários humanos dão aos computadores e outros artefatos uma posição social semelhante à de outra pessoa humana e que isso ocorre como um produto da interação social extrínseca, independentemente da composição interna real do objeto em questão. Em outras palavras, não precisa de muito esforço para um artefato socialmente interativo ser considerado outro sujeito e não apenas um objeto ou ferramenta. Em face destas entidades crescentemente sociais e interativas nós vamos ter que decidir se elas são consideradas meras coisas, como nossos outros objetos tecnológicos, ou alguém que importa e a quem nós vamos imputar responsabilidade, como outro membro da família; ou decidir se tais entidades são algo completamente diferente, que está situado entre o primeiro (uma mera coisa) e o segundo (um membro da família), como se fosse um animal de estimação. Qualquer caminho que venha a ser decidido, entretanto, esses artefatos irá, sem dúvida, desafiar nossa compreensão sobre responsabilidade e o modo como tipicamente distinguimos entre quem é que vai ser considerado outro sujeito social e o que vem a ser um mero instrumento ou objeto.

## Quais relações deveríamos estabelecer com robôs? Deveríamos amá-los?

**David Gunkel**

Em face destes desafios/oportunidades, nós podemos tentar responder como tipicamente temos respondido, tratando esses mecanismos crescentemente sociais e autônomos como meros instrumentos ou ferramentas. Como a pesquisadora de ética dos computadores Deborah Johnson (2006, 197) explicou: Os sistemas de computador são produzidos, distribuídos e usados por pessoas engajadas em práticas sociais e buscas significativas. Isso é tão verdadeiro acerca dos atuais sistemas de computadores como será sobre o futuro dos sistemas de computadores. “Não importa quão independente, automático ou interativo se comporte um sistema de computador no futuro, eles serão os produtos (diretos ou indiretos) do comportamento humano, instituições sociais humanas e decisões humanas”. Este argumento é persuasivo, precisamente porque ele desenha e é subscrito pela teoria instrumental da tecnologia. Tecnologias - não importa quão sofisticada, inteligente e social sejam, parecem ser, ou se tornarão - são e continuarão sendo ferramentas da ação humana, nada mais. Mas esta restrita reaplicação do pensamento instrumentalista, tendo em conta toda a sua utilidade e aparente simplicidade, negligencia a própria presença desses dispositivos e os efeitos dessa presença dentro das redes da cultura contemporânea. Considere-se, por exemplo, o que está acontecendo com os robôs de eliminação de explosivos, que estão sendo usados por soldados. Conforme estudos e relatórios recentes têm demonstrado, os soldados dos Estados Unidos, no Iraque e Afeganistão, formaram laços pessoais surpreendentemente próximos com suas unidades de robôs de eliminação de explosivos, dando nomes a eles, premiando-os com promoções no campo de batalhas, arriscando suas próprias vidas para proteger a vida de seus robôs, e até sofrendo luto por suas mortes. Isso acontece não porque esses dispositivos estilo tanques controlados remotamente foram desenhados para ser companhias sociais. Isso acontece como um produto do mecanismo que está situado dentro do papel que o dispositivo tem nas operações no campo de batalha. E isso acontece em direta oposição ao que de outro modo soaria como um bom senso comum: eles são apenas tecnologias - instrumentos ou ferramentas que trabalham em nosso benefício e não sentem nada.

Por esta razão, lembrar usuários que eles estão apenas trabalhando e interagindo com “coisas sem mente” pode ser uma “informação correta”, mas fazer isso é

frequentemente tão inefetivo como dizer a espectadores de cinema que a ação que veem na tela não é real. Nós sabemos disso, mas isso não necessariamente muda as coisas. O que temos, conseqüentemente, é uma situação na qual nossa teoria a respeito de tecnologia - uma teoria que carrega uma história considerável e tem sido determinada a ser aplicável tanto para dispositivos simples, como uma ferramenta manual, como um sistema tecnológico complexo - parece estar fora de sintonia com as experiências reais que temos com objetos tecnológicos em uma variedade de situações e circunstâncias. Em outras palavras, o modo de pensar instrumentalista pode ser ontologicamente correto, mas é socialmente inepto e fora de alcance para nossas experiências ordinárias. Agora, esse argumento pode não prover prova definitiva de que devemos “amar nossos robôs”, mas indica o fato de que esses dispositivos já são algo a mais que instrumentos tecnológicos ou objetos que podem ser relegados ao status de mera ferramenta.

**Se um robô olhasse ao mundo de hoje, o que ele diria a Donald Trump e Jair Bolsonaro, que se utilizaram de sistemas de análise de *big data* para suas campanhas, misturando até mesmo informações falsas?**

**David Gunkel**

Se tal coisa fosse possível, o robô poderia fazer uma afirmação direta e simples, como: “eu posso fazer melhor”. E há evidência dando suporte a tal premissa. Gerenciar o dia a dia de operações de governo de um Estado-Nação no século XXI se tornou algo crescentemente complexo e que demanda a coleta, seleção, processamento de quantidades massivas de dados sobre tudo, desde uma estratégia tributária e econômica até mudanças demográficas na população. Computadores e algoritmos não só são desenhados para suportar grandes quantidades de informação como têm, em numerosas áreas, capacidades de tomada de decisão que na verdade superam as dos humanos.

Essa capacidade fica evidente com coisas mundanas, como participar de campeonatos e jogar xadrez - como recentemente foi demonstrado com as vitórias impressionantes no campeonato com *AlphaGo* e *AlphaZero*, da *Deep Mind* -, e mais importante, em assuntos de “vida ou morte”, como diagnosticar câncer ou descobrir novas terapias e tratamentos. Já foi sugerido que a Inteligência Artificial realmente seria melhor em tomada de decisões administrativas, especialmente situações nas quais a quantidade de dados é esmagadora e as decisões poderiam ser enviadas por emoções

ou uma circunstância em que as pessoas estejam envolvidas. O roboticista Ronald Arkin, por exemplo, defendeu “robôs assassinos” no campo de batalhas. Ele argumenta que sistemas armamentistas autônomos, programados apropriadamente, podem ser melhores em seguir regras do que os militares engajados, e assim se poderia fazer mais conflitos armados com menos perdas humanas.

Mas tal capacidade pode vir a ser uma espada de dois gumes, com regimes totalitários providos de Inteligência Artificial, e líderes que querem ser autoritários com novas ferramentas para exercer controle sobre populações humanas. Um exemplo notório é o experimento do governo chinês com métricas sociais para desenvolvimento e uso de uma pontuação de crédito social para classificar cidadãos. Mas este tipo de coisa não é algo limitado ao governo da China. Nós agora reconhecemos como plataformas de mídia social, como o Facebook - que usa sofisticadas soluções de engenharia para controlar fluxos de informação e governar engajamento social - foram manipuladas em ambas as eleições de Donald Trump, nos Estados Unidos, e de Bolsonaro, no Brasil. O problema é que a aplicação de Inteligência Artificial para situações políticas e governança parece ser racional e uma solução guiada por dados, mas frequentemente acabam sendo tão limitadoras e antidemocráticas como os sistemas humanos que eles substituiriam ou aos quais se somariam.

**Considerando-se as distopias contemporâneas, tais quais aquelas retratadas em *Black Mirror* ou *Westworld*, você acha que deveríamos estar esperançosos com o mundo das máquinas?**

**David Gunkel**

O que é interessante sobre a tecnologia de Inteligência Artificial e robôs é que eles são simultaneamente o objeto de sonhos utópicos incontidos e pesadelos distópicos sombrios. Seguindo a lógica do legado do determinismo tecnológico, a inovação em tecnologia computacional mantém a grande promessa de responder e consertar problemas que parecem sem resolução. Qualquer que seja a crise - uma epidemia global, mudança climática, desigualdade social, etc. - há (ou haverá em breve) uma aplicação, um *app*, para tal. Ao mesmo tempo, no entanto, estas tecnologias têm sido objeto direto de medo e agouro. Recentemente, Elon Musk e Stephen Hawking sugeriram que a Inteligência Artificial representa uma grande ameaça existencial para as espécies humanas e para o nosso mundo.

Há dois itens importantes a serem notados neste aspecto: primeiro, a dialética entre utopia e distopia é dramática, mas também um pouco de distração. Certamente faz sentido para algumas boas histórias de ficção científica e manchetes dramáticas de jornais, mas ao limitar as coisas a um tudo ou nada de narrativas utópicas/distópicas se arrisca perder mudanças mais mundanas que a tecnologia introduz em nosso mundo. Eu estou, por exemplo, menos preocupado com a ameaça da invasão de robôs e a distopia resultante dela do que com algo como a Matrix em que estou inserido, com mudanças aparentemente desimportantes em nossas situações e circunstâncias sociais cotidianas, como algoritmos de recrutamento, desenhados para ajudar corporações a selecionar concorrentes de emprego qualificados, mas que depois acabam como foi o recente caso da Amazon, discriminando candidatas mulheres. Ao invés de focar nas mudanças grandes e dramáticas que são boas para a ficção científica, nós precisamos aprender como prestar atenção nas alterações pequenas e aparentemente insignificante que silenciosamente retrabalham e reconfiguram nossa realidade social.

Em segundo lugar, há uma boa razão pela qual nós temos visto todas essas narrativas distópicas agora - e não tem nada a ver com ficção científica e seu modo de funcionamento. A ficção científica não trata - ao menos primariamente - de previsão do futuro, mas de diagnosticar o presente. Se neste momento particular, o começo do século XXI, temos todo tipo de histórias dramáticas e distópicas sobre robôs e Inteligência Artificial, é porque nós estamos aqui e agora incertos sobre o significado social e o status da tecnologia contemporânea. Por esta razão, nós precisamos procurar retratos distópicos de robôs e Inteligência Artificial, como *Blade Runner* e séries como *Black Mirror* e *Westworld*, não para refletir especulativamente sobre um possível futuro, mas como expressões de ansiedades do presente, que precisam ser enfrentadas dentro de nossa situação social. Em outras palavras, nós precisamos seriamente tomar a ficção científica como um tipo de espelho de preocupações contemporâneas e não desfazê-la como mero entretenimento. Há um tipo de pensamento profundo e importante acontecendo na ficção e nós precisamos aprender como dar sentido a isso e potencializar suas contribuições.

**Além de uma ética, qual estética deveríamos pensar para um mundo contemporâneo, entre máquinas, robôs e seres humanos?**

Como as tecnologias de vários tipos e configurações usurpam as habilidades humanas, o



único baluarte remanescente de um excepcionalismo humano parece ser a criatividade e a arte. A Inteligência Artificial pode deslocar seres humanos das tarefas mundanas como dirigir carros e caminhões, tradução entre línguas, e básicas interações interpessoais para ambas as transações comerciais e sociais. Mas estes sistemas de A.I. são programados e controlados por designers humanos, desenvolvedores e usuários. Desde simples instruções programadas até as sofisticadas máquinas de algoritmos de aprendizagem, treinados em *big data*, os computadores apenas fazem o que os dizemos para fazer. Este é o ponto que Alan Turing chamou objeção de “Lady Lovelace”: um computador “não tem pretensão de originar qualquer coisa. Ele pode apenas fazer algo que saibamos como ordenar que performe” (TURING, 1999, p. 50). Por esta razão, parece uma aposta segura dizer que computadores, algoritmos e Inteligência Artificial vão automatizar uma ampla variedade da rotina e atividades repetitivas em muitas áreas do esforço humano. Mas - e o argumento continua - eles nunca vão poder e nunca vão ser capazes de fazer algo inventivo, inovativo e inspiracional. Eles nunca vão, em outras palavras, ser capazes de ser criativos e produzir algo completamente novo e original. Certo?

Talvez não. Atualmente, há tecnologias e sistemas capazes de produzir o que pode parecer trabalho “criativo” de arte. Existe, por exemplo, Shimon, que toca marimba, é um robô de jazz, da Georgia Tech University, que pode improvisar músicos humanos em tempo real. Há experimentos em Inteligência Musical, como um compositor digital, baseado em um computador, que pode criar novas versões de notas musicais que podem (a partir de algumas contas) ser virtualmente indistinguíveis do trabalho de mestres como Bach, Beethoven, Mozart. Tem ainda o *The Painting Fool*, um algoritmo pintor que aspira ser “tomado seriamente como um artista criativo em seus próprios direitos”; e *Narrative Science's Quill* e *Automated Insight's Wordsmith*, sistemas de geração de linguagem natural desenhados para escrever histórias/textos originais, legíveis por humanos, através do acesso, análise e montagem de dados. Consequentemente, parece que o que temos chamado de “criatividade” e “arte” pode não ser unicamente algo humano, como se poderia ter pensado inicialmente. Criatividade pode ser tão susceptível de automação como outras atividades.

## Referências

Martin Heidegger. 1977. *The Question Concerning Technology and Other Essays*. Trans. William Lovitt. New York: Harper & Row.

Deborah Johnson. 2006. Computer Systems: Moral Entities but not Moral Agents. *Ethics and Information Technology*, 8(4), 195-204. <https://doi.org/10.1007/s10676-006-9111-5>

Byron Reeves and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge: Cambridge University Press.

Alan Turing. 1999. Computing Machinery and Intelligence. In *Computer Media and Communication: A Reader*, ed. Paul A. Meyer, 37-58. Oxford: Oxford University Press.

Recebido: 10 mar 2019

Aceito: 22 mar 2019