



A imagem retrata artisticamente a evolução de autoinvestigação e de identidade dos seres humanos conforme eles determinam como incorporar a tecnologia progressivamente mais sofisticada. O autor do artigo analisa como a responsabilidade do *jus in bello* terá que mudar radicalmente na medida em que os sistemas de armas autônomos crescem em sofisticação. (Foto cortesia de Pixabay)

O Devido Cuidado com a Robotização do Campo de Batalha

As Limitações Cognitivas dos Sistemas Autônomos de Combate e dos Seres Humanos

Maj Jules Hurst, Reserva do Exército dos EUA

Durante a Segunda Guerra Mundial, os pilotos dependiam dos cálculos analógicos e da heurística para guiar suas bombas ao alvo. Já na década de 80, os computadores de rastreamento de alvos notificaram aos pilotos exatamente quando deviam lançar suas munições. O fracasso do piloto de soltar as munições no momento certo podia resultar na falha do alvo por grandes distâncias, tudo dependente da altura, velocidade e orientação da aeronave. Atualmente, as grandes áreas aceitáveis para lançamento das munições guiadas de precisão requerem que os pilotos exerçam muito menos habilidade para colocar as bombas no alvo. Se uma munição guiada de precisão falha um alvo positivamente identificado, a culpa é, provavelmente, da ogiva, não da técnica de lançamento do piloto.

A proliferação dos sistemas semiautônomos no campo de batalha aumentará essa tendência. Na medida em que os computadores assumirem mais responsabilidades de combate, as avarias de equipamento corresponderão a uma maior porcentagem de falhas do emprego das armas. Os militares podem facilmente assimilar as taxas de erro e os prováveis erros recorrentes das armas de precisão, mas os erros que um robô militar semiautônomo poderia cometer podem ser mais difíceis de serem descobertos. Uma análise correta dessas falhas pode exigir conhecimento tanto dos algoritmos que conduziram às decisões do robô, quanto o código que as executou. Um caça F-35 *Joint Strike Fighter* já exige oito milhões de linhas de código para controlar seu voo, mas os sistemas semiautônomos futuros requererão ainda mais¹.

A complexidade do processo decisório dos robôs de combate no futuro fará, potencialmente, com que as tentativas de exercer as responsabilidades do *jus in bello* — a conduta justa na guerra — sejam progressivamente mais difíceis. Em paralelo com sua capacidade, os robôs de combate exercerão mais liberdade no campo de batalha, e os parâmetros do seu processo decisório serão estabelecidos, em grande parte, antes de saírem da fábrica, potencialmente inalteráveis pelos militares que os empregam. A sociedade impõe a responsabilidade ética proporcionalmente à capacidade de um indivíduo ou de uma organização de controlar as ações em questão. Quando os sistemas autônomos letais proliferarem no campo de batalha, os combatentes talvez não sejam a força dominante no controle das suas atividades. Em vez disso, os oficiais

responsáveis pelas aquisições do material bélico e os engenheiros que obtêm e constroem esses sistemas provavelmente terão a maioria da influência sobre o comportamento dos robôs militares no campo de batalha. Consequentemente, alguns conceitos do *jus in bello* precisam ser atualizados para levar em conta a maior responsabilidade que os não combatentes na cadeia de aquisição dos sistemas de armas terão sobre o comportamento ético no campo de batalha.

Quais Sistemas Autônomos?

Nem todos os sistemas de armas autônomos letais (*lethal autonomous weapon systems* — LAWS) irão complicar as tentativas dos militares de exercer responsabilidades éticas no combate. Se os avanços nas redes neurais e na informática fornecem aos LAWS as capacidades de tomada de decisão e a percepção sensorial semelhantes aos seres humanos, é lógico que as preocupações sobre o *jus in bello* não sejam piores do que as enfrentadas pelos combatentes humanos. Independente disso, sem dúvida haverá um período de transição entre a consecução da inteligência artificial (IA) de nível humano e o surgimento de robôs de combate no campo de batalha.

As estimativas periciais variam muito sobre quando aparecerá a inteligência artificial semelhante à dos seres humanos. Em 2012, analistas do Machine Intelligence Research Institute examinaram 257 previsões literárias, feitas por especialistas e não especialistas, sobre quando as máquinas atingiriam o desempenho cognitivo comparável aos seres humanos. As previsões variam entre 1980 e além de 2100. A maioria ficou entre 2020 e 2060². Nenhum consenso de peritos existe sobre a data da chegada da IA comparável aos seres humanos, tampouco há uma definição precisa do termo. Além disso, a criação de máquinas com inteligência semelhante à dos seres humanos não necessariamente coincidirá com as máquinas atingindo uma capacidade comparável aos humanos de perceber seus ambientes³. A inteligência e a percepção sensorial são capacidades independentes; seria delicado perguntar alguém

O Maj Jules Hurst, da Reserva do Exército dos EUA, serviu como o oficial de operações da Seção de Inovação Analítica e de Tecnologia do Comando de Operações Especiais dos EUA e como analista civil superior do 1º Batalhão, 75º Regimento de Ranger.

com deficiências de perda auditiva ou cegueira sobre isso. A IA semelhante aos seres humanos pode permanecer muito distante e, até com a sua chegada, ela pode ser alcançável somente pelo uso de um computador de tamanho de uma sala ou processadores quânticos em temperaturas abaixo de zero⁴. Além disso, provavelmente levaria muitos anos para miniaturizar, de forma rentável, a IA de nível humano para o emprego nos sistemas de armas táticas, como levou décadas para que os computadores pessoais se tornassem viáveis financeira e tecnologicamente⁵.

O Período de Transição entre a IA de Nível Humano e os Robôs Militares Capazes de Combater

Independentemente de quando os robôs atingirão a paridade intelectual e sensorial com os seres humanos, quase com certeza haverá um período de transição antes da IA ser exatamente equivalente à do ser humano, no qual as nações empregarão os LAWS no campo de batalha devido às vantagens das máquinas em desempenhar tarefas restritamente definidas⁶.

Com poucas exceções, as máquinas têm um desempenho superior aos seres humanos em qualquer papel restritamente definido. A utilização dos sistemas de armas automatizados em tarefas rotineiras e bem definidas já oferece grandes vantagens táticas aos Estados Unidos e a outras forças sofisticadas. Essas vantagens crescerão em paralelo com os avanços tecnológicos e a tendência de a cultura militar aceitar mais naturalmente seu emprego. A capacidade de empregar máquinas com níveis subhumanos de inteligência artificial em tarefas estruturadas (IA restrita) já afeta a capacidade militar de uma nação. A longo prazo, tornar-se-á ainda mais importante⁷. Será enorme a tentação de nações concederem autonomia letal aos robôs antes de atingirem os níveis humanos de cognição e percepção.

A diretriz atual dos EUA restringe o desenvolvimento de veículos robóticos àqueles que “permitem que comandantes e operadores exerçam níveis apropriados de julgamento humano sobre o uso da força”, mas as Forças Armadas dos EUA podem abandonar essa diretriz durante um conflito por razões práticas⁸. A menos que a comunidade internacional imponha, com sucesso, um regime de controle de armamento que proíba os robôs assassinos ou limite seu uso por leis internacionais ou normas estabelecidas, as nações beligerantes podem

facilmente justificar os LAWS como uma necessidade em um conflito prolongado, ou até limitado⁹. Mesmo se a comunidade internacional estabelecer um regime de controle de armamento ou uma legislação que reja o emprego dos sistemas autônomos, a História mostra que os Estados violam esses acordos quando há a necessidade ou o desejo. Nos anos 30, tanto o império japonês quanto o Terceiro Reich se retiraram, ou violaram, das provisões relacionadas às armas do Tratado Naval de Washington e do Tratado de Versalhes¹⁰. Os regimes mais liberais tendem, também, a ignorar as provisões ou os acordos internacionais quando é conveniente para eles. Há apenas 15 anos atrás, o Presidente George W. Bush retirou os Estados Unidos do Tratado de Mísseis Antibalísticos assinado com a União Soviética, em 1972, para começar a construção de um sistema nacional de defesa antimíssil¹¹.

Quando um participante significativo de um tratado de controle de armamento se retira por uma vantagem militar, os adversários potenciais dele e os signatários do tratado perdem os incentivos para cumprir com as prescrições contidas no tratado; uma ovelha negra estraga o rebanho. Se a vantagem militar de empregar sistemas letais com IA restrita parece suficientemente grande, os países podem rapidamente modificar as restrições internas ou ignorar os tratados e normas. Essas mudanças políticas podem até surgir de usos militares bem intencionados. O mundo ocidental continua a desdobrar dezenas de milhares de militares para combater às ameaças terroristas na África, na Ásia e no Oriente Médio. O emprego dos LAWS, em vez de soldados humanos, pode agradar os regimes eleitos democraticamente que precisam equilibrar as demandas por segurança nacional com a vontade pública de aceitar baixas em tempos de guerra¹².

Se a comunidade internacional não conseguir efetivamente impor as proibições dos robôs assassinos, os sistemas autônomos ainda podem ter oportunidades de tomar decisões letais no campo de batalha. Se as nações utilizarem sistemas robóticos ofensivos, mesmo com a participação humana dentro do seu ciclo de emprego, os robôs militares poderão sofrer ataques eletrônicos ou cibernéticos (inimigos ou amigos) que os isolem de seus controladores humanos¹³. As forças armadas em todo o mundo têm aceitado, em todos os escalões, a expansão das capacidades da guerra eletrônica, estimulada, em grande medida, pela dependência moderna



Um Standard Missile 3 (SM-3) é lançado do contratorpedeiro *USS Decatur*, da classe Arleigh Burke, que é equipado com o Sistema de Combate Aegis, durante um teste aéreo de míssil balístico da Missile Defense Agency (Agência de Defesa contra Mísseis), 22 Jun 07. Logo depois, o SM-3 interceptou um míssil balístico lançado do Campo de Provas de Mísseis no Pacífico, Barking Sands, Kauai, Havaí. Foi a primeira vez que tal prova foi conduzida de um contratorpedeiro da Marinha dos EUA equipado com uma defesa antimísseis balísticos. O Sistema de Combate Aegis possui um módulo automático que permite que os computadores de controle de fogos identifiquem e engajem alvos, sem seres humanos no ciclo. (Foto cortesia da Marinha dos EUA)

dos dispositivos explosivos improvisados e os veículos aéreos não tripulados (VANT). A China e a Rússia, em particular, aspiram bloquear as comunicações táticas de uma força oponente, um problema potencial para as forças dos EUA que estão acostumadas ao comando e controle desimpedido¹⁴. Os robôs militares que percorrem o campo de batalha podem ser facilmente isolados do controle humano por um ataque eletrônico ou modificados por meios cibernéticos.

Por outro lado, os danos aos robôs militares podem torná-los incapazes de receber ou processar as ações de controle dos seres humanos. Os robôs de combate ou os sistemas autônomos com avarias que estão separados dos controladores podem agir como elefantes de guerra do Século XXI, sistemas de armas capazes de infligir grandes danos aos amigos e inimigos quando sob controle tênue. Quer seja danificado ou sob ataque eletrônico, dispositivos de segurança podem ser instalados

para forçar que os sistemas autônomos cessarem as atividades letais caso o contato com os controladores humanos seja perdido, mas outras nações podem não adotar as mesmas regras de engajamento. Se os robôs de combate se tornarem essenciais para o êxito militar, a negação de uma forma de autoridade letal autônoma pode ser uma razão de perder guerras.

A colocação de seres humanos dentro do ciclo ou acima do ciclo talvez não reduza drasticamente o risco de violações do *jus in bello* pelos robôs de combate porque os humanos tendem a se submeter ao julgamento das máquinas quando as informações são limitadas ou durante situações estressantes. Os controladores humanos dos robôs de combate, assim como os operadores atuais dos VANT, podem tomar decisões letais com base em informações fornecidas a eles por meio de sensores remotos ou até de leituras não visuais em uma área limitada pela largura de banda. Os robôs militares

futuros podem talvez até possuir uma capacidade de explicar concisamente a seu operador o processo que o levou a chegar a uma determinada conclusão, mas isso não removerá necessariamente o risco de subordinação humana às máquinas se os operadores não possuírem a experiência de combate ou do sistema¹⁵. Como o Dr. John Hawley observa em seu artigo “Patriot Wars: Automation and the Patriot Air and Missile Defense System” (“As Guerras do Patriot: A Automatização e o Sistema de Defesa Antimíssil e Antiaérea Patriot”, em tradução livre), “um sistema automatizado nas mãos de uma tripulação inadequadamente treinada é, de fato, um sistema completamente automatizado”¹⁶. Outros pesquisadores têm observado o mesmo tipo de confiança nas máquinas durante as emergências. Em um estudo, de 2015, do Georgia Institute of Technology, os participantes seguiram um guia robô a uma sala de conferências e o observaram cometer muitos erros de navegação durante o trajeto. Depois de chegar à sala de conferências, os pesquisadores acionaram o alarme de incêndio. Apesar de observar o robô cometer erros de navegação apenas alguns minutos antes, todos os participantes do estudo escolheram seguir o robô para sair do potencial prédio em chamas¹⁷.

Apesar das dúvidas dos EUA sobre o emprego futuro das máquinas autônomas letais, os sistemas de defesa antiaérea já têm entrado um período de transição em que as armas “inteligentes” possuem a capacidade e a autoridade de tomar decisões letais. O Patriot Air Defense System, o Aegis Combat System e o Close-In Weapon System contêm módulos automáticos que colocam os sistemas de classificação e de engajamento de alvos nas mãos dos computadores de controle de fogos¹⁸. Todos os três já foram usados em combate, e cada sistema já classificou erroneamente e engajou um alvo amigo ou neutro, resultando em ferimentos em aliados ou não combatentes¹⁹. Em muitos desses casos, o fratricídio ou os danos colaterais resultaram da falta de entendimento dos algoritmos do sistema de defesa antiaérea pelos humanos ou dos operadores que confiaram demasiadamente na identificação de alvos do sistema de controle de fogos. Esses erros não são limitados aos sistemas dos EUA. O abatimento do Voo 17 da Malaysia Airlines por um SA-11 controlado pelos separatistas ucranianos provavelmente caracteriza o mesmo tipo de erro entre humanos e máquinas²⁰.

Os sistemas autônomos letais do futuro enfrentarão situações significativamente mais complexas dos que os sistemas automatizados de defesa antiaérea atuais porque eles terão a capacidade de mover-se autonomamente e serão distribuídos através de uma maior gama de ambientes (aéreo, terrestre, marítimo e espacial). Os sistemas de armas autônomas interagirão, também, com uma maior diversidade de ameaças e, mais importante ainda, com seres humanos em um maior número de cenários²¹. Além disso, os LAWS podem operar grandes distâncias dos seus controladores e supervisores humanos. Essa mobilidade concederá aos robôs militares a capacidade de se colocar em circunstâncias imprevisíveis, magnificando a probabilidade de que eles enfrentarão situações fora do seu entendimento lógico, além do alcance ou da capacidade de ser resolvido pelos algoritmos internos.

A Dificuldade dos Robôs Militares de Seguirem os Princípios da Guerra Justa

O reconhecimento e o emprego da força letal contra alvos potenciais indiscriminadamente no campo de batalha exigem muito menos cálculos cognitivos do que o emprego justo da força. Não importa se o combatente é humano ou robótico. Os combatentes humanos que se preocupam em seguir os princípios do *jus in bello* — distinção, proporcionalidade e necessidade militar — recorrem ao conhecimento cultural, teórico e prático para evitar infligir sofrimento desnecessário aos não combatentes. Também, desenvolvem inconscientemente heurísticas e integram rapidamente as informações que parecem não relacionadas no seu processo decisório. Em geral, as máquinas têm dificuldade em criar seus próprios modos de agir ou relacionar dados divergentes tão bem quanto os seres humanos e talvez nunca possam obter uma capacidade semelhante aos humanos de fazer isso²². As seções abaixo explicam três dos princípios do *jus in bello* e a razão que os robôs militares podem ter dificuldade em aderir a eles.

A distinção. Uma vez começada uma guerra, os militares e combatentes fisicamente aptos são sujeitos a ataques, em qualquer momento²³. Quando um indivíduo pega em armas, ele perde a imunidade civil que o impede de se tornar alvo da força letal. Conforme esses combatentes perdem seus direitos civis, eles obtêm outros, por sua vez. Agora têm a liberdade de

atacar outros combatentes. Os combatentes somente readquirem sua imunidade dos ataques quando perdem fisicamente a capacidade de fazer mal a outros (ao ficar severamente ferido) ou renunciam permanentemente ao seu papel. Se o período de alistamento termina e ele volta para casa, esse indivíduo já não é um alvo militar legítimo.

Se for incumbido com a identificação de alvos, um LAWS provavelmente terá dificuldade em diferenciar entre combatentes e não combatentes fora da guerra convencional, da mesma forma que os soldados humanos. Na grande quantidade de conflitos que vêm sendo travados por todo o mundo (e.g., Síria, Ucrânia, Iraque e Afeganistão), grande parte dos combatentes está fora das estruturas militares convencionais. Os uniformes homogêneos ou os sistemas de armas idênticos não os identificam como combatentes. Atualmente, o software de reconhecimento de objetos permite que máquinas detectem e classifiquem objetos ao compará-los a fotografias mantidas em bancos de dados internos — os robôs de combate podem classificar um fuzil AK-47 ou um carro de combate T-72 e concluir que os indivíduos associados com eles são uma ameaça.

O software, contudo, não pode chegar a essas mesmas conclusões se o combatente não está usando um sistema de armas tradicional ou um objeto de dupla utilização — um insurgente usando um telefone celular para detonar um dispositivo explosivo improvisado parece muito como uma pessoa enviando uma mensagem de texto. O poder de processamento da mente humana é necessário para fazer isso. Para fazer esse tipo de julgamento, um infante jovem tem que avaliar as intenções pelos sinais visuais do insurgente e conduzir uma análise do estilo de vida dos civis e da atividade insurgente na área, recorrendo rapidamente a informações obtidas de experiências e treinamento passados. Tudo isso ocorre em poucos segundos. Os sistemas de armas autônomos letais não desempenharão bem essa tarefa ou outras que exijam os modelos heurísticos complicados das mentes biológicas. Às vezes, o carro autoguiado do Google tem dificuldade em atravessar uma interseção de quatro ruas porque é incapaz de decifrar as intenções dos outros motoristas humanos²⁴.

Essas microanálises são importantes. Se um LAWS se deparar com um combatente ferido, como ele determinará se ele está incapacitado? Como uma máquina aceitará uma rendição de militares que portam armas

para que os prisioneiros de guerra sejam tratados justamente? Algoritmos podem ser escritos visando a fornecer às máquinas alguma capacidade para fazer isso, mas eles provavelmente ficarão aquém dos níveis humanos, em um futuro previsível. Os robôs de guerra podem identificar os combatentes em flagrante com facilidade, mas têm dificuldade em distinguir entre não combatentes e soldados irregulares²⁵.

A proporcionalidade. Além da diferenciação do emprego da força, a teoria da guerra justa exige que “a destruição necessária para atingir um objetivo militar seja proporcional à vantagem de realizá-lo”²⁶. Em resumo, usar a força apropriada para o alvo — não há necessidade de lançar uma bomba de 220 kg contra um insurgente portando um AK-47 quando um cartucho de 5,56 mm seria suficiente. A falha de acatar a esse princípio aumenta a probabilidade de danos colaterais desnecessários e de perdas excessivas de vidas dos combatentes.

Os robôs de combate terão dificuldade em tomar decisões sobre proporcionalidade, sem a contribuição humana²⁷. Os desenvolvedores de software podem projetar que um LAWS siga as regras de engajamento pré-programadas: restrições sobre o calibre das armas empregadas em zonas urbanas, zonas onde não podem disparar, etc., e comandantes podem aplicar vários tipos de limites geográficos e medidas de controle do apoio de fogo para evitar o uso excessivo da força e o fratricídio pelos robôs. No entanto, mesmo com essas restrições, inevitavelmente os LAWS ficarão aquém das expectativas se os controladores humanos não conseguirem prever quais tipos de medidas de controle serão necessários. No combate volátil, isso testará severamente a capacidade de previsão humana.

Uma estimativa dos danos colaterais, elemento essencial da proporcionalidade, exige uma infinidade de previsões e premissas. As mentes humanas apenas fazem com que pareça fácil. Se um infante recebe fogos de um prédio em um ambiente urbano, ele automaticamente considera a função dessa estrutura e a presença potencial de não combatentes antes de retaliar. Os seres humanos analisam inconscientemente o material estrutural, o design e a sinalização do prédio entre outros dados antes de compará-los a um arquivo de informações arquitetônicas desenvolvido ao longo da vida. As mentes humanas possuem uma grande capacidade de armazenamento e capacidades de lembrança incríveis,

que permitem que eles possam rapidamente recolher as informações que parecem não relacionadas e aplicá-las a novos problemas²⁸. Os sistemas de computadores atuais não fazem isso.

As determinações de proporcionalidade exigem mais do que identificação estrutural. Sem muito esforço, os seres humanos analisam os efeitos do tempo do dia, dos ambientes culturais e dos dias da semana nos padrões de vida de civis — é improvável matar muitos civis por meio de um bombardeio a meia-noite de uma terça-feira durante a maioria do ano, mas pode matar vintenas de pessoas se fosse a véspera de Natal. A capacidade dos LAWS de chegar a esses tipos de conclusões sob os cronogramas de combate permanece em dúvida. Os programadores podem escrever algoritmos para aproximar esses ciclos de decisão humanos, mas representar todas as variações possíveis será quase impossível. As diferenças entre as capacidades sensoriais robóticas e humanas apresentarão dificuldades²⁹. Além disso, os robôs militares de nível tático talvez não possuam uma memória suficiente para acomodar os bancos de dados necessários ou a capacidade computacional para fazer uma referência cruzada deles, sem ter um baixo custo-benefício ou estar conectado a uma nuvem próxima.

A necessidade militar. Os alvos da violência na guerra precisam ser legítimos, atacados para cumprir um objetivo que contribuirá para a derrota de uma força inimiga. Mesmo o assassinato de combatentes por finalidades independentes de um objetivo militar pode ser desnecessário³⁰. Essa exigência representa dificuldades para combatentes, tanto no nível estratégico quanto no tático. O curso da guerra é imprevisível. Frequentemente, é difícil avaliar se um ataque é necessário ou não.

Os sistemas de armas autônomos letais, também, terão grande dificuldades nessa determinação. Imagine que um LAWS localiza um terrorista que está planejando conduzir um carro bomba contra uma embaixada dos EUA, nos próximos dias. Seu filho de oito anos

de idade está ao lado dele. É uma necessidade militar que o LAWS o ataque agora? Ou deve esperar até que o terrorista esteja sozinho, mesmo que isso possa deixar que o ataque de carro bomba ocorra. Decisões como essas são problemáticas, mesmo se o contexto maior que exige ação seja entendido. Um tomador de decisões humano pode fazer cálculos sobre a probabilidade do terrorista executar exitosamente o ataque via carro bomba ou das forças terrestres capturarem o terrorista se o robô de combate não atacar. Um robô de nível tático provavelmente não terá acesso às informações externas ou ao conhecimento da situação mais ampla para fazer uma estimativa.



Um pesquisador de material analisa dados experimentais sobre o planejador de inteligência artificial do Autonomous Research System (ARES). O sistema desenvolvido pelo Laboratório de Pesquisa da Força Aérea usa inteligência artificial para projetar, executar e analisar experimentos em um ritmo mais rápido do que os métodos tradicionais de pesquisa científica. (Foto cortesia do Departamento de Defesa dos EUA)

Atualmente, os soldados humanos têm dificuldade em cumprir perfeitamente suas responsabilidades do *jus in bello*. Os robôs de combate, com inteligência subumana, provavelmente terão mais dificuldade. Se os controladores e supervisores humanos esperarem prevenir que os LAWS violem as regras de distinção, proporcionalidade e necessidade militar, precisarão ser completamente versados nos parâmetros que orientam a seleção e o engajamento de alvos. Será pedido aos militares a assunção de responsabilidade pelas decisões feitas pelos sistemas de armas automatizados no campo de batalha. Os sistemas de defesa antiaérea e as munições do tipo

“dispare e esqueça” têm fardos semelhantes atualmente, mas a maior autonomia dos robôs militares de manobra automática aumentará esse ônus.

Embora imperfeitos, os cães de guerra proporcionam a analogia mais próxima. Quando o condutor solta um cão de guerra no campo de batalha, ele não dispara uma arma. Ele emprega um sistema de armas que toma suas próprias decisões de engajamento. Para garantir a eficácia do cão de guerra no combate e evitar as violações éticas e o fratricídio, o animal passa por meses de treinamento intensivo ao lado de seu condutor. O condutor depende desse treinamento para orientar o animal pelas tarefas simples e aprende a entender os pontos fortes e fracos do seu cão de guerra³¹. Os controladores dos robôs militares passam por um período de treinamento semelhante aos LAWS em ambientes virtuais e físicos, mas podem ter mais dificuldades em usar essa experiência para obter discernimentos sobre os comportamentos do robô. Apesar das diferenças em escala, os sentidos compartilhados (e.g., visão, audição e olfato), os padrões de pensamento e as emoções (e.g., medo, excitação) proporcionam aos seres humanos e cães uma base evolucionária para entender um ao outro. Um ser humano pode observar um cão interagir com seu ambiente e entender, pelo menos parcialmente, sua intenção, motivações e pensamentos. As diferenças nítidas entre a percepção e a cognição humanas e mecânicas frustrarão os esforços humanos de desenvolver esse mesmo nível de entendimento com os robôs de combate.

Entendendo o Processo Decisório dos Robôs Militares

Os algoritmos — processos definidos por codificados para que computadores possam resolver problemas — são os componentes básicos de software. Os programadores podem combinar centenas ou milhares de algoritmos para criar um programa de software que toma decisões sem o conhecimento do usuário. Os usuários manipulam a interface gráfica que executa os scripts de uma linguagem de programação de um nível superior, que é, por sua vez, traduzida em entradas binárias para a unidade central de processamento. Esses níveis de abstração escondem do usuário o processo decisório real da máquina e as várias heurísticas, premissas e falhas que os codificadores incluem, intencional ou inadvertidamente, dentro dos programas. Conforme os especialistas

em IA usam progressivamente mais a aprendizagem mecânica e as técnicas de aprendizagem profunda para permitir que as máquinas elaborem seus algoritmos, essas propensões podem tornar-se até mais obscuras. As técnicas de aprendizagem mecânica permitem que algoritmos criem generalizações e analisem padrões com base na avaliação de dados de treinamento. De uma forma eficaz, os algoritmos ensinam a si próprios por meio de ensaios e erros. A dependência dos dados de treinamento por parte da aprendizagem mecânica permite que os cientistas da computação insiram, inconscientemente, propensões significativas nos algoritmos por meio de sua seleção de entradas de dados³². Por exemplo, um algoritmo de reconhecimento facial que usa o anuário do ensino médio do meio-oeste americano como dados de treinamento pode ter dificuldade na identificação de minorias étnicas quando usado devido à falta de exemplos nos dados de treinamento [está região dos EUA é, na maioria, composta de brancos — N. do T.]. No contexto comercial, esse tipo de propensão, que resulta da seleção humana dos dados de treinamento, é embaraçoso. Com respeito aos LAWS, é letal.

Imagine o design de um algoritmo para identificar um homem armado, de idade militar. Os programadores podem cumprir isso com uma variedade de maneiras. O software pode usar sensores para medir a altura de um alvo potencial para confirmar a idade adulta; buscar por pelos faciais ou medir a relação ombro-cintura para avaliar o gênero; e estimar a massa corporal do alvo potencial para comparar com as médias de adultos e crianças. Quando a precisão dos sensores permite, o programa pode até medir objetos segurados pelo alvo potencial, para determinar se esses combinam com as especificações das armas armazenadas na memória do robô militar.

Cada uma dessas determinações exigiria a execução de algoritmos que comparam e aplicam os dados obtidos pelos sensores do robô militar nos bancos de dados internos e nos algoritmos pré-desenvolvidos. Esse algoritmo pode falhar na sua tarefa de identificar os homens de idade militar, de várias maneiras: (1) Pode faltar a precisão necessária para a tomada da decisão pelo robô; seus sensores podem ser incapazes de identificar objetos ou medir alturas além de uma certa distância; e o algoritmo pode forçá-lo a tomar uma decisão de qualquer maneira. (2) Os bancos de dados do robô de combate podem conter informações



Um técnico de munições guiadas de precisão, do 23º Esquadrão de Manutenção, examina as munições de ataque direto Joint Direct Attack Munitions (JDAM), na Base da Força Aérea Moody, Geórgia, 11 Jan 18. Esse conjunto converte as bombas normais em munições "inteligentes" que são capazes de engajar autonomamente os alvos designados. (Cb Eugene Oliver, Força Aérea dos EUA)

insuficientes; o alvo, por exemplo, pode estar segurando um modelo de uma arma de fogo que não está dentro dos dados de treinamento do algoritmo ou ela pode ter alterações que fazem com que ela seja irreconhecível pela máquina. (3) As heurísticas ou as premissas do programador para a tomada de uma decisão podem ter falhas ou fazer premissas falsas. Se o algoritmo tentar identificar objetos como masculino ou feminino, com base na altura e na massa corporal, pode encontrar problemas se ele deixar de considerar a nacionalidade e/ou a afiliação étnica do alvo potencial — a mulher escandinava média é mais alta do que o homem chinês médio, e provavelmente pesa o mesmo³³. Alternativamente, um algoritmo que identifica armas pode classificar um fuzil de caça como uma ferramenta e um AK-47 como uma arma, com base na opinião de um engenheiro de software, embora os dois são letais.

Quase todos os algoritmos contêm os juízos de valor e as propensões inerentes das pessoas que os criam e dos dados de treinamento escolhidos por

eles³⁴. Os militares incumbidos com o monitoramento e o controle dos sistemas autônomos letais precisarão ser familiarizados com as propensões e os juízos de valor incutidos no software do sistema de armas, para evitar acidentes e violações das leis de conflito armado.

O entendimento desses juízos de valor e o reconhecimento de quando esses irão afetar o desempenho operacional talvez seja mais fácil do que parece. Os algoritmos que alimentam os LAWS serão incrivelmente densos. Os robôs de combate podem depender de milhões de linhas de código para operarem, e cada um deles terá propensões e premissas incutidas dos engenheiros humanos de software que o criaram. O volume de códigos fará com que seja muito difícil entender as complexidades de cada decisão algorítmica em sua plenitude e seu efeito em um robô militar, durante tempos de combate. Também, o uso das técnicas de aprendizagem mecânica aprofundará, potencialmente, essa ambiguidade.

Os militares podem experimentar até mais dificuldade em entender esses algoritmos porque eles não possuem os antecedentes técnicos dos engenheiros de software, que irão criar os LAWS operados pelos soldados. Os controladores dos robôs militares talvez nem tenham acesso aos algoritmos devido às preocupações sobre a segurança; os governos podem protegê-los como segredos privados ou nacionais³⁵. Essa dissimulação pode ser problemática. Os conceitos atuais de *jus in bello* responsabilizam os combatentes e os comandantes pelo emprego ético dos sistemas autônomos letais, embora as decisões tomadas pelas autoridades que os adquirem e pelos engenheiros de software tenham um impacto igual, ou maior, no comportamento dos robôs militares.

A Extensão das Responsabilidades do *Jus in Bello*

Considerando que a maioria dos combatentes que proporcionam orientação aos LAWS será incapaz de entender completamente como seus robôs de combate tomam as decisões letais, os engenheiros de software e as autoridades responsáveis pela aquisição e colocação desses equipamentos nas mãos dos controladores devem assumir uma parte da responsabilidade ética pelas violações do *jus in bello* dos robôs militares no campo de batalha. Os sistemas legais ocidentais não responsabilizam criminosamente os acusados se são incapazes mentalmente de entender seu crime, então a sociedade não deve responsabilizar somente os combatentes pelas decisões tomadas pelos LAWS no campo de batalha, que ficam aquém da sua capacidade de compreender³⁶. Em vez disso, a sociedade responsabiliza os indivíduos pelos eventos em proporção à sua capacidade de moldá-los. Embora os combatentes sempre tenham responsabilidade pelas violações do *jus in bello* que resultam de erros do operador, os criadores dos sistemas e as autoridades de aquisição devem assumir a culpa em proporção à sua responsabilidade pela violação³⁷.

As autoridades responsáveis pela aquisição têm o dever ético de determinar as exigências dos sistemas autônomos letais que minimizem a probabilidade que eles violem as leis do conflito armado. Especificamente, os representantes do governo precisarão estabelecer padrões rigorosos para os algoritmos de treinamento e para o teste dos sistemas robóticos em ambientes reais e simulados, para evitar erros prováveis. Os contratados

de defesa, por sua vez, têm a obrigação de fornecer os robôs militares que satisfaçam essas exigências até o nível tecnológico mais alto possível e informar os governos sobre as vulnerabilidades conhecidas do processo decisório algorítmico e da percepção dos sensores. As empresas que fabricam os carros autoguiados são responsáveis pelos acidentes causados pelas falhas de design³⁸. Por isso, as empresas que produzem as máquinas de guerra com falhas negligentes devem, devidamente, enfrentar penalidades civis e, potencialmente, criminais. Finalmente, tanto os líderes civis quanto militares têm responsabilidades para determinar políticas e regras de engajamento que minimizem as oportunidades que os combatentes têm para colocar sistemas autônomos letais em situações fora da capacidade analítica e lógica deles.

Independentemente da dificuldade, os líderes militares precisam se esforçar para preparar os soldados a entender os riscos do emprego de sistemas autônomos letais em situações de combate previsíveis³⁹. As forças armadas futuras que empregarem os LAWS poderão descobrir que muitas das vantagens econômicas da automatização dissipam à luz dos aumentados custos de treinamento de operários e dos testes necessários para aprimorar os algoritmos da IA. A automatização das tarefas no campo de batalha desempenhadas pelos seres humanos provavelmente forjará forças armadas com um menor número de combatentes que possuam maior aptidão técnica.

As forças armadas devem considerar a criação de classes especiais de militares semelhantes aos controladores aéreos táticos conjuntos que recebem treinamento avançado sobre as complexidades dos sistemas autônomos. Os comandantes podem, potencialmente, limitar o controle dos robôs militares, além de situações de emergência, até que a inteligência artificial amadureça mais⁴⁰. Uma variedade de metodologias e pessoal especializado existem para orientar comandantes no emprego de ataques aéreos e de fogos de longo alcance, como identificadores de alvos, controladores aéreos táticos conjuntos e observadores avançados. As forças armadas que empregam os LAWS devem considerar a criação de posições semelhantes. Os especialistas em robótica devem acompanhar os comandantes de campanha e ajudá-los a tomar as decisões difíceis em relação ao emprego de robôs militares no campo de batalha, da mesma forma que identificadores de alvos ajudam

comandantes a determinar o risco de danos colaterais durante ataques aéreos. Também, as forças armadas podem precisar formar células de teste e avaliação que armazenem dados de treinamento e criem simulações que permitam que os LAWS sejam calibrados para teatros de operações e regras de engajamento específicos. As metodologias para estimar o risco do emprego dos LAWS podem, também, ajudar comandantes nas suas decisões de desdobrar os robôs de combate e estabelecer normas sobre seu uso.

Durante as próximas décadas, os combatentes irão entrar em uma era em que seus sistemas de armas acumularão progressivamente maior responsabilidade devido à sua capacidade de aplicar a força, com êxito, contra alvos militares legítimos. Os conceitos éticos e as diretrizes precisam avançar em sintonia para garantir que as mudanças tecnológicas não resultem em lacunas éticas, ao distribuir proporcionalmente as responsabilidades do *jus in bello* aos atores humanos envolvidos. ■

Referências

1. "A Digital Jet for the Modern Battlespace," Lockheed Martin (website), acesso em: 31 jan. 2018, <https://www.f35.com/about/life-cycle/software>.
2. Stuart Armstrong e Kaj Sotala, "How We're Predicting AI—or Failing To," in *Beyond AI: Artificial Dreams*, eds. Jan Romportl et al. (Pilsen, Czech Republic: University of West Bohemia, 2012), p. 52–75.
3. Noel E. Sharkey, "The Evitability of Autonomous Robot Warfare," *International Review of the Red Cross* 94, no. 886 (2012): p. 788–89.
4. Um exemplo de um computador de tamanho de uma sala é o National Supercomputing Center (Centro Nacional de Supercomputação) em Wuxi, China, para mais informações, veja seu website, <http://www.nscw.cn/wxcyw/>. A IBM desenvolveu um computador quântico que pesquisadores podem usar para apurar tamanho potencial de computação. Cade Metz, "IBM Is Now Letting Anyone Play with its Quantum Computer," *Wired* (website), 4 May 2016, acesso em: 31 jan. 2018, <https://www.wired.com/2016/05/ibm-letting-anyone-play-quantum-computer/>.
5. History.com Staff, "Invention of the PC," History.com, 2011, acesso em: 31 jan. 2018, <http://www.history.com/topics/inventions/invention-of-the-pc>.
6. Patrick Lin, George Bekey e Keith Abney, "Autonomous Military Robotics: Risk, Ethics, and Design," Preparado para o Office of Naval Research, 20 Dec. 2008, acesso em: 12 fev. 2018, http://ethics.calpoly.edu/onr_report.pdf.
7. Greg Allen e Taniel Chan, "Artificial Intelligence and National Security" (Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School, July 2017), p. 17–18, acesso em: 31 jan. 2018, <https://www.belfercenter.org/publication/artificial-intelligence-and-national-security>.
8. Department of Defense Directive 3009.09, *Autonomy in Weapon Systems* (Washington, DC: U.S. Government Publishing Office [GPO], 21 Nov. 2012); Joseph Breecher, Heath Niemi e Andrew Hill, "My Droneski Just Ate Your Ethics," *War on the Rocks* (website), 10 Aug. 2016, acesso em: 31 jan. 2018, <https://warontherocks.com/2016/08/my-droneski-just-ate-your-ethics/>.
9. Os grupos como a Campaign to Stop Killer Robots (Campanha para Parar os Robôs Mortíferos) já têm começado a pressionar para uma proibição mundial dos sistemas de armas letais. Para mais informações sobre esse grupo, veja <https://www.stopkillerrobots.org/>.
10. C. Peter Chen, "Japan's Refusal of Washington Treaty, 19 Dec 1934," WWII Database (website), acesso em: 31 jan. 2018, https://ww2db.com/battle_spec.php?battle_id=45.
11. Wade Boese, "U.S. Withdraws from ABM Treaty; Global Response Muted," *Arms Control Today* (website), July/August 2002, acesso em: 31 jan. 2018, https://www.armscontrol.org/act/2002_07-08/abmjul_aug02.
12. Jules Hurst, "Intervention and the Looming Choices of Autonomous Warfighting," *War on the Rocks* (website), 25 Aug. 2016, acesso em: 31 jan. 2018, <https://warontherocks.com/2016/08/intervention-and-the-looming-choices-of-autonomous-warfighting/>.
13. "Dentro do ciclo" se refere a situações em que os seres humanos precisam dar uma ordem aos robôs de guerra para exercer a força letal. Os controladores humanos estão na cadeia de matar. "Acima do ciclo" se refere a situações em que os seres humanos supervisionam os robôs enquanto eles exercem a força letal.
14. Thomas Gibbons-Neff, "'We Don't Have the Gear': How the Pentagon is Struggling with Electronic Warfare," *Washington Post* (website), 9 Feb. 2016, acesso em: 31 jan. 2018, <https://www.washingtonpost.com/news/checkpoint/wp/2016/02/09/we-dont-have-the-gear-how-the-pentagon-is-struggling-with-electronic-warfare/>.
15. Dave Gershgorn, "We Don't Understand How AI Make Most Decisions, So Now Algorithms are Explaining Themselves," *Quartz* (website), 20 Dec. 2016, acesso em: 12 fev. 2018, <https://qz.com/865357/we-dont-understand-how-ai-make-most-decisions-so-now-algorithms-are-explaining-themselves/>.
16. John K. Hawley, *Patriot Wars: Automation and the Patriot Air and Missile Defense System*, Voices from the Field series (Washington, DC: Center for a New American Security, January 2017), acesso em: 31 jan. 2018, <https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf>.
17. Paul Robinette et al., "Overtrust of Robots in Emergency Evacuation Scenarios" (lecture, The 11th ACM/IEEE International Conference on Human Robot Interaction, Christchurch, New Zealand, 7 March 2016), p. 101, acesso em: 12 fev. 2018, <http://ieeexplore.ieee.org/document/7451740/>.

18. John Canning, "A Concept of Operations for Armed Autonomous Systems" (apresentação do PowerPoint, 3rd Annual Disruptive Technology Conference, Washington, DC, 6–7 September 2010), acesso em: 12 fev. 2018, http://www.dtic.mil/ndia/2006/disruptive_tech/canning.pdf.
19. Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]), "Report of the Defense Science Board Task Force on Patriot System Performance Report Summary" (report, Washington, DC: Department of Defense, January 2005); George C. Wilson, "Navy Missile Downs Iranian Jetliner," *Washington Post* (website), 4 Jul. 1988, acesso em: 12 fev. 2018, <http://www.washingtonpost.com/wp-srv/inatl/longterm/flight801/stories/july88crash.htm>; A. J. Plunkett, "Iwo Jima Officer Killed In Firing Exercise," *Daily Press* (website), 12 Oct. 1989, acesso em: 12 fev. 2018, http://articles.dailypress.com/1989-10-12/news/8910120238_1_iwo-jima-ship-close-in-weapons-system.
20. Simon Tomlinson, "Russian Missile Killed Pilots and Cut Jet in Half but Passengers Could Have Been Conscious for up to a Minute as Plane Plunged, Reveals Official Report into MH17 Downed over Ukraine," *Daily Mail* (website) 13 Oct. 2015, acesso em: 12 fev. 2018, <http://www.dailymail.co.uk/news/article-3270355/Doomed-flight-MH17-shot-Russian-BUK-missile-fired-rebel-held-territory-eastern-Ukraine-Dutch-investigators-set-rule.html>.
21. Andrew Ilachinski, "AI, Robots, and Swarms Issues, Questions, and Recommended Studies" (Arlington, VA: CNA, January 2017), p. vi–vii, acesso em: 12 fev. 2018, https://www.cna.org/CNA_files/PDF/DRM-2017-U-014796-Final.pdf.
22. Sharkey, "The Evitability of Autonomous Robot Warfare," p. 789.
23. Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 4th ed. (1977; repr., New York: Basic Books, 2006), p. 138.
24. Matt Richtel e Conor Dougherty, "Google's Driverless Cars Run into Problem: Cars with Drivers," *New York Times* (website), 1 Sep. 2015, acesso em: 12 fev. 2018, <https://www.nytimes.com/2015/09/02/technology/personaltech/google-says-its-not-the-driverless-cars-fault-its-other-drivers.html>.
25. Sharkey, "The Evitability of Autonomous Robot Warfare," p. 788–89.
26. Brian Orend, *The Morality of War* (Peterborough, Ontario: Broadview Press, 2006), p. 119.
27. Sharkey, "The Evitability of Autonomous Robot Warfare," p. 789.
28. Jeneen Interlandi, "New Estimate Boosts the Human Brain's Memory Capacity 10-Fold," *Scientific American* (website), 5 Feb. 2016, acesso em: 12 fev. 2018, <https://www.scientificamerican.com/article/new-estimate-boosts-the-human-brain-s-memory-capacity-10-fold/>.
29. Ruth A. David e Paul Nielsen, "Report of the Defense Science Board Summer Study on Autonomy" (Washington, DC: OUSD[AT&L], June 2016), p. 14, acesso em: 12 fev. 2018, <http://www.dtic.mil/dtic/tr/fulltext/u2/1017790.pdf>.
30. Orend, *The Morality of War*, p. 119.
31. Department of the Army Pamphlet 190-12, *Military Working Dog Program* (Washington, DC: U.S. GPO, 30 Sep. 1993), p. 24.
32. Aylin Caliskan, Joanna J. Bryson e Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science* 356, no. 6334 (2017): p. 183-86.
33. Ian Langtree, "Height Chart of Men and Women in Different Countries," *Disabled World* (website), última atualização em 19 Dec. 2017, acesso em: 12 fev. 2018, <https://www.disabled-world.com/artman/publish/height-chart.shtml>.
34. Felicitas Kramer, Kees van Overveld e Martin Peterson, "Is There an Ethics of Algorithms?," *Ethics and Information Technology* 13, no. 3 (2010): p. 251.
35. Tom Simonite, "For Superpowers, Artificial Intelligence Fuels New Global Arms Race," *Wired* (website), 8 Sep. 2017, acesso em: 12 fev. 2018, <https://www.wired.com/story/for-superpowers-artificial-intelligence-fuels-new-global-arms-race/>.
36. "The Infancy Defense: Criminal Law Basics," *The Lectric Law Library*, acesso em: 12 fev. 2018, <http://www.lectlaw.com/mj/c1032.htm>.
37. Aaron M. Johnson e Sidney Axinn, "The Morality of Autonomous Robots," *Journal of Military Ethics* 12, no. 2 (2013): p. 131.
38. Anjali Singhvi e Karl Russell, "Inside the Self-Driving Tesla Fatal Accident," *New York Times* (website), 12 Jul. 2016, acesso em: 12 fev. 2018, <https://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html>.
39. Isso é mais fácil dizer do que fazer. É provável que os comandantes somente sejam culpáveis pela falha de preparar as tropas para um cenário de combate se esse cenário já existisse como parte de um currículo de treinamento padronizado.
40. David and Nielsen, "Report of the Defense Science Board," p. 38.