

ESCOLA DE GUERRA NAVAL

CC KARIM HEBERT

A DECISÃO HUMANA ENTRE RACIOCÍNIO E INSTINTO EM FACE DA  
INTELIGÊNCIA ARTIFICIAL

O caso do emprego da força letal no mar

Rio de Janeiro

2020

CC KARIM HEBERT

A DECISÃO HUMANA ENTRE RACIOCÍNIO E INSTINTO EM FACE DA  
INTELIGÊNCIA ARTIFICIAL

O caso do emprego da força letal no mar

Dissertação apresentada à Escola de Guerra  
Naval, como requisito parcial para a conclusão  
do Curso de Estado-Maior para Oficiais  
Superiores.

Orientador: CMG (Ref<sup>o</sup>) Nilson da Silva Moreira

Rio de Janeiro  
Escola de Guerra Naval

2020

## **AGRADECIMENTOS**

A minha esposa Oriane pelo apoio, amor e compromisso para me deixar enfrentar este trabalho de longo prazo. A meus filhos Edouard e Alice que me permitem de relembrar todos os dias da alegria de ter uma família mesmo quando minhas tarefas profissionais parecem difícil.

Ao CMG (Ref<sup>o</sup>) Nilson da Silva Moreira, meu orientador, pelas sugestões que me foram decisivas par me permitir de estruturar este trabalho e pela paciência e gentileza que ele mostrou para ajudar a melhorar meu nível de português.

Ao CF (RM1) Nagashima, pelas excelentes aulas de metodologia científica que ajudaram bastante a realização deste trabalho.

## RESUMO

O propósito deste trabalho é deduzir o papel que a inteligência artificial (IA) poderia gerar entre intuição e racionalidade no campo naval e aplicado à tomada de decisões militares sobre o uso da força letal, observadas a partir da perspectiva da teoria dos dois "sistemas" de Daniel Kahneman. O estudo primeiro se concentrou em identificar as contribuições da IA para os processos de tomada de decisão humana. Lembrando a representação de Daniel Kahneman das capacidades humanas de tomada de decisão entre o "sistema 1" e o "sistema 2", pareceu durante este estudo que, devido a sua autonomia, a inteligência artificial poderia ser comparada a um "sistema 3" em interação com a mente humana. No contexto de uma decisão de usar força letal no mar, parecia que o sistema 3 poderia aumentar as capacidades dos sistemas 1 e 2. Este aumento das capacidades cognitivas humanas pela IA baseia-se em grande parte em sua capacidade de analisar enormes quantidades de dados digitais. Entretanto, este aumento apresenta riscos ligados às limitações atuais deste tipo de tecnologia: sua especialização, sua imprevisibilidade, sua dependência de grandes quantidades de dados e, finalmente, sua incapacidade a produzir um senso comum. O estudo propõe princípios que permitem conter esses riscos para cumprir o Direito Internacional dos Conflitos Armados (DICA) e a ética do soldado, preservando ao mesmo tempo o lugar do homem no processo decisório e, em particular, permitindo-lhe manter o uso de seu instinto guerreiro.

Palavras-chaves: Processo decisório. Kahneman. Inteligência artificial. Instinto. Racionalidade. Força Letal.

## LISTA DE ABREVIATURAS E SIGLAS

C2	–	Comando e Controle
DICA	–	Direito Internacional dos Conflitos Armados
GOFAI	–	<i>Good Old Fashion IA</i>
IA	–	Inteligência artificial
IHM	–	Interface Homem Maquina
SAAL	–	Sistemas de Armas Autonomous Letais
SCAF	–	Sistema de Combate Aéreo do Futuro

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO .....</b>	<b>7</b>
<b>2.</b>	<b>DECIDIR COM A IA CONSIDERANDO A TEORIA DOS DOIS SISTEMAS DE KAHNEMAN .....</b>	<b>11</b>
2.1.	DEFINIÇÕES E TÉCNICAS DE IA.....	11
2.1.1.	Definição de IA.....	11
2.1.2.	<i>Machine Learning</i> .....	12
2.1.3.	Outras técnicas .....	14
2.1.4.	Características essenciais da IA.....	15
2.2.	A TEORIA DAS DUAS VELOCIDADES DO PENSAMENTO E DA TOMADA DE DECISÃO .....	17
2.2.1.	Sistema 1 / Sistema 2 .....	17
2.2.2.	As falhas da tomada de decisão humana.....	18
2.2.3.	O senso comum e o raciocínio consciente como ferramentas de tomada de decisão .....	19
2.3.	DECIDIR COM A AJUDA DA IA .....	20
2.3.1.	Um aumento das capacidades cognitivas do ser humano .....	20
2.3.2.	Algumas possíveis funções de IA úteis para a tomada de decisões no mar .....	21
2.3.3.	Os limites intrínsecos da IA ao decidir.....	22
<b>3.</b>	<b>DECIDIR USAR A FORÇA LETAL NO MAR COM A AJUDA DA IA.....</b>	<b>26</b>
3.1.	DECIDIR USAR A FORÇA LETAL DE ACORDO COM OS DOMÍNIOS DE LUTA .....	26
3.1.1.	Antes de decidir usar a força letal, enfrente a incerteza com a IA.....	26
3.1.2.	Luta antiaérea, luta assimétrica: os domínios de luta onde predomina o "Sistema 1".....	28
3.1.3.	Luta anti superfície e luta anti submarina: os domínios de luta onde predomina o "Sistema 2".....	29
3.2.	O QUE PODE TRAZER A IA NESTE TIPO DE DECISÕES?.....	30
3.2.1.	Buscar correlações para enfrentar a incerteza para o benefício do sistema 1 ou do sistema 2.....	31
3.2.2.	Detectar / reconhecer para o benefício do sistema 1.....	32
3.2.3.	Prever e deduzir para o benefício do sistema 2.....	33
3.3.	OS LIMITES DA CONTRIBUIÇÃO DA IA PARA ESSAS DECISÕES.....	34
3.3.1.	Especialização e imprevisibilidade nas áreas de controle do sistema 1.....	34
3.3.2.	A falta de senso comum nos domínios de luta do sistema 2.....	35
3.3.3.	Os limites intrínsecos de acesso aos dados na ação.....	35
<b>4.</b>	<b>O QUE EXIGIR DAS TÉCNICAS DE IA NO CONTEXTO DO USO DA FORÇA LETAL .....</b>	<b>37</b>
4.1.	UMA EXIGÊNCIA LEGAL E ÉTICA .....	37
4.1.1.	O paradigma do “robô assassino” .....	37
4.1.2.	Os riscos legais dos vieses em um sistema 3 .....	38

4.2.	UM REQUISITO PARA O DESENHO DE ALGORITMOS E GERENCIAMENTO DE DADOS .....	39
4.3.	PRESERVANDO O LUGAR DA DECISÃO HUMANA.....	41
4.3.1.	Luta contra <i>o Nudge</i> .....	41
4.3.2.	Manter o homem conectado ao processo de decisão da IA .....	41
4.3.3.	O lugar do acaso.....	42
<b>5.</b>	<b>CONCLUSÃO .....</b>	<b>45</b>
	<b>REFERÊNCIAS .....</b>	<b>48</b>

## 1. INTRODUÇÃO

Quando ela aborda o tema da Inteligência Artificial (IA), a cultura popular contemporânea é muitas vezes inspirada pelo teste de Turing, cujo objetivo é saber como reconhecer um robô de um humano. Em 1982, no filme Blade Runner, os robôs têm uma aparência humana quase perfeita, mas não têm a capacidade de chorar. Mais recentemente, em 2014, o diretor americano Spike Jonze num filme futurístico, dirige um personagem confuso que se vê dizendo à Samantha, com voz sintética produzida por uma inteligência artificial extremamente desenvolvida, que ele a ama. A possibilidade desta confusão da natureza entre o humano pensante e a máquina pensante alimenta muitas fantasias. Estas fantasias dão origem ao medo do advento de uma humanidade que luta com uma inteligência sintética que é inquietante e fora do controle humano. Podemos definir "singularidade tecnológica"<sup>1</sup> como o momento de ruptura em que a máquina escaparia ao controle da humanidade e começaria a agir de acordo com sua própria vontade. Embora a própria possibilidade desta singularidade seja uma questão de debate, ela rega o inconsciente coletivo e ressurge regularmente em trabalhos de ficção científica.

No campo militar, a IA está alimentando o medo de que sistemas de armas autônomas letais (SAAL) ou "robôs assassinos" capazes de identificar, designar e engajar alvos por conta própria apareçam um dia no campo de batalha. Especialistas em IA, muitas vezes chefes de empresas que ganham a vida com essas tecnologias, expressaram publicamente a necessidade de proibi-las. Entre eles estão algumas celebridades, incluindo o empresário Elon Musk e o ilustre físico Stephen Hawking. A organização não governamental (ONG) *Human Rights Watch* lidera uma rede de mais de sessenta ONGs engajadas em uma campanha para proibir os

---

<sup>1</sup> Esta noção é estudada pelo filósofo Jean-gabriel GANASCIA em seu livro: *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?* Paris. Le Seuil. 2017.



"robôs assassinos"<sup>2</sup>. Estas ações ecoam a primeira das três leis da robótica imaginada pelo autor de ficção científica Isaac Azimov, proibindo-as de prejudicar os seres humanos<sup>3</sup>.

Será que a IA é hoje uma cópia imperfeita de nossas capacidades humanas? Será que ela apresenta capacidade de "raciocínio" ou "instinto"? Quais são os limites de sua aprendizagem autônoma? Estas questões são cruciais na tomada de decisões e particularmente na tomada de decisões militares. Será que essas decisões que ameaçam a vida são susceptíveis de ser delegadas, pelo menos em parte, a alguma forma de inteligência artificial? Quais seriam as consequências previsíveis?

Uma das características da tomada de decisões militares no século XXI é que ela ocorre em um mundo incerto e ambíguo. Dentro dela, a inteligência artificial oferece muitas vantagens, em particular para ajudar a dissipar a "névoa da guerra"<sup>4</sup> e acelerar o ritmo das operações para manter a iniciativa da força amiga. Ao mesmo tempo, os sistemas de armas contemporâneos são predominantemente sistemas baseados em softwares e oferecem a possibilidade de gerar grandes quantidades de dados numéricos que podem ser analisados pela IA. Entretanto, existe uma ligação muito forte entre a disponibilidade de dados e as capacidades da IA. A disponibilidade de tantos dados permitirá, a curto prazo, que a IA gere ferramentas para análise ambiental ou gerenciamento de sistemas implementados a bordo de navios de guerra de uma escala completamente nova. Mais simplesmente, ela abre a porta para o uso do aprendizado por a IA, técnicas chamadas de *Machine Learning*. Estes últimos, que não podem ser desenvolvidos sem grandes quantidades de dados, estão na origem dos avanços mais impressionantes em termos de autonomia da IA no mundo civil. Por exemplo, num futuro próximo, pode ser que

---

<sup>2</sup> Informações sobre esta campanha estão disponíveis em <https://www.stopkillerrobots.org/?lang=fr>. Acesso em: 18 de julho 2020

<sup>3</sup> lei número 1 : um robô não pode prejudicar um ser humano nem, permanecendo passivo, permitir que um ser humano seja exposto ao perigo ;  
lei número 2 : um robô deve obedecer às ordens que lhe são dadas por um ser humano, a menos que tais ordens entrem em conflito com a primeira lei ;  
lei número 3: um robô deve proteger sua existência desde que esta proteção não entre em conflito com a primeira ou segunda lei.

<sup>4</sup> Uma definição deste conceito do Karl Von Clausewitz sera proposta no capítulo 3

somente as técnicas de IA sejam capazes de cumprir a tarefa de identificar o inimigo em um prazo aceitável em um campo de batalha multidimensional. Um campo de batalha onde a quantidade de dados disponíveis para processamento, coletados por nossos sensores, é tal que escapa da capacidade de análise humana que não seria auxiliada por uma IA.

Mas tal afirmação poderia levar à difamação das capacidades humanas de tomada de decisão. A capacidade humana de decidir vai muito além da capacidade de processar uma grande quantidade de dados. Para representar isto, Daniel Kahneman define assim dois "sistemas" operando de forma complementar que oferecem tanto uma capacidade intermitente de análise como um raciocínio factual e rigoroso: o "sistema 2", mas também uma generalização rápida e muitas vezes muito eficaz baseada em poucos dados que podem ser assimilados às manifestações do instinto: o "sistema 1". Seu trabalho também se concentrou na definição de vieses cognitivos susceptíveis de levar a raciocínios automáticos enganosos, por facilidade ao usar o "sistema 1". Se tentarmos imaginar a IA como um "sistema 3", o que aumentaria as capacidades de um ou outro dos dois sistemas, podemos ver como os defeitos da IA poderiam, por si mesmos, constituir vieses para o raciocínio. A identificação desses vieses é, portanto, crítica, particularmente em termos de tomada de decisões militares.

Além disso, há uma necessidade de reflexão ética sobre as imperfeições da IA. Por sua natureza, o uso de IA pode implicar que parte da decisão seja delegada. As condições para a atribuição de responsabilidade por uma decisão tomada usando a IA merecem ser perfeitamente claras neste caso. O risco de dividir a responsabilidade do desenvolvedor do algoritmo para o tomador de decisão no mar, através dos responsáveis pela aquisição de uma técnica de IA e todos os níveis intermediários deve ser descartado. O lugar do ser humano que usa algoritmos de IA deve, portanto, ser definido. Junto com isso, deve ser definida a exigência de transparência dos algoritmos de IA e como esses algoritmos podem integrar conceitos como os do Direito Internacional Humanitário (DIH) em seu desenho.

O presente trabalho se propõe a deduzir o papel que a IA poderia gerar entre intuição e racionalidade no campo naval e aplicado à tomada de decisões militares sobre o uso da força letal, observadas a partir da perspectiva da teoria dos dois "sistemas" de Daniel Kahneman. Para atingirmos o propósito deste estudo, empregaremos um desenho de pesquisa exploratória.

O capítulo que se segue a esta introdução, segundo capítulo, se concentrará em definir e recordar a estrutura geral das capacidades da IA, bem como a teoria de Daniel Kahneman sobre as duas velocidades do pensamento. Apresentará então o que a decisão com a ajuda da IA pode apresentar como pontos fortes e fracos.

O terceiro se concentrará em identificar como a tomada de decisões militares sobre o uso da força letal no domínio naval pode evoluir através do uso da IA. Também abordará as limitações da IA em comparação com as capacidades humanas em tal contexto.

O quarto capítulo ilustrará quais requisitos básicos os sistemas de tomada de decisão de força letal baseados na IA militar devem cumprir. Também considerará o lugar que o ser humano deve reter em relação à IA neste contexto.

Finalmente, no capítulo 5 será apresentada a conclusão do trabalho.

## 2. DECIDIR COM A IA CONSIDERANDO A TEORIA DOS DOIS SISTEMAS DE KAHNEMAN

A associação dos termos "inteligência" e "artificial" é uma questão de debate. Um dos dois criadores do assistente vocal “Siri” da empresa Apple, Luc Julia, chega ao ponto de falar de erros semânticos que alimentam as fantasias (TUAL, 2019). Neste capítulo estudaremos como esta técnica pode aumentar ou restringir uma inteligência humana bem real, descrita por Daniel Kahneman. Algumas noções centrais da IA serão definidas. Em seguida, os princípios básicos da teoria das duas velocidades do pensamento (sistema 1, sistema 2) serão lembrados. Finalmente, serão estabelecidos elementos estruturantes para a tomada de decisões utilizando a IA.

### 2.1. DEFINIÇÕES E TÉCNICAS DE IA

#### 2.1.1. Definição de IA

A definição que usaremos para inteligência artificial é a de Yann Le Cun, que, juntamente com Geoffrey Hinton e Yoshua Bengio, recebeu o Prêmio Turing de 2019, o equivalente a um Prêmio Nobel em ciência da computação, graças ao seu trabalho em *Deep Learning*. O cientista, que também é diretor de pesquisa básica no *Facebook*, o define da seguinte forma:

*[...] l'intelligence artificielle est la capacité, pour une machine, d'accomplir des tâches généralement assurées par les animaux et les humains : percevoir, raisonner et agir. Elle est inséparable de la capacité à apprendre, telle qu'on l'observe chez les êtres vivants.<sup>5</sup> (LE CUN, 2019, p. 22.)*

Além da analogia com a capacidade de interagir com um ambiente percebido no qual ela pode agir de forma adequada, podemos manter a relação direta que o autor estabelece entre

---

<sup>5</sup> “[...] A inteligência artificial é a capacidade de uma máquina de executar tarefas normalmente executadas por animais e seres humanos: perceber, raciocinar e agir. É inseparável da capacidade de aprender, como observado nos seres vivos.” (Tradução Nossa)

a inteligência artificial e a capacidade de aprender. Para entender melhor a importância deste aspecto, é necessário voltar à definição de duas famílias de técnicas de IA: *IA simbólica* e *Machine Learning*, ambas lembradas no livro de Yann Le Cun.

As técnicas de IA chamadas *inteligência artificial simbólica* foram as primeiras a serem desenvolvidas. Eles não utilizam a capacidade de aprendizagem de uma máquina. As primeiras técnicas de IA são agora chamadas GOFAI, um acrônimo para *Good Old Fashioned AI*. Sua ambição era usar a linguagem de computador para reproduzir os mecanismos do raciocínio humano. Eles se baseavam principalmente em regras lógicas como árvores de decisão e manipulação de símbolos (LE CUN, 2019). Eles enfatizam o raciocínio lógico causal direto e consciente do tipo "Se A é verdade, a ação é B...". Eles são baseados na experiência humana no campo que simulam, o que torna possível estabelecer regras causais. Para serem realistas, eles exigem que o desenvolvedor do código de computador compreenda e domine suficientemente as ações humanas que ele está simulando ou que trabalhe de perto com um perito que compreenda a lógica do código.

De acordo com LE CUN, as técnicas do GOFAI e seus sucessores não refletem adequadamente a realidade da experiência humana na tomada de decisões. Considerando um exemplo do mundo marítimo, parece difícil imaginar usar estes métodos para simular um oficial da marinha manobrando um navio em um canal de navegação estreito. A aplicação de Regulamento Internacional para Evitar Abalroamentos no Mar pode ajudar, mas não é suficiente para representar todos os casos possíveis que exigem uma decisão humana de mudança de rumo. Comparativamente, as técnicas da IA usadas para desenvolver carros autônomos são baseadas nas técnicas de *Machine Learning*, possivelmente combinadas com outras técnicas das quais a IA simbólica pode fazer parte.

### 2.1.2. *Machine Learning*

Ela surgiu no final dos anos 1980. É uma mudança de paradigma. Com estas técnicas, deixamos de estabelecer regras lógicas *a priori* que se aplicam o tempo todo e, portanto, só são eficazes em um ambiente definido. Em vez disso, uma grande quantidade de dados é coletada sobre um ambiente descontrolado do mundo real e a máquina aprende a reagir analisando este tipo de dados. Ao submeter esses dados a funções matemáticas, chamadas algoritmos, é possível estabelecer uma "resposta" da máquina que analisa os dados. Estas "respostas" são comparadas a uma "resposta esperada" (LE CUN 2019). A resposta esperada é definida por um ser humano. Durante a fase de aprendizagem, os algoritmos são corrigidos até que, para um conjunto cuidadosamente construído de dados de aprendizagem, cada resposta da máquina corresponda à resposta esperada definida pelo ser humano. Chama-se aprendizado "supervisionado". O conjunto de dados de treinamento já constitui uma quantidade muito grande de dados. Uma vez verificada a capacidade da máquina de fornecer a resposta esperada no conjunto de dados de aprendizagem, podemos passar à fase de implementação, onde a máquina será capaz de responder "sozinha", analisando novos dados que nunca viu antes. Esta capacidade de "generalizar" sua resposta aos dados desconhecidos de um experimento de dados de aprendizagem é apresentada como o resultado da aprendizagem humana (LE CUN 2019). Estas técnicas são implementadas, por exemplo, para permitir o reconhecimento de imagens, de texto ou mesmo de voz por uma máquina. Para distinguir um C de um D, para reconhecer a imagem de um gato ou a voz de uma pessoa por uma assistente virtual, são amplamente utilizados (LE CUN, 2019).

Existem outros métodos de aprendizagem de máquinas. O aprendizado "reforçado", por exemplo, levará a máquina a testar uma sequência de ações que são permitidas à máquina sozinha, a fim de obter um resultado esperado. Este resultado pode ser descrito como o estado de um sistema que inclui a máquina e seu ambiente. O método requer literalmente deixar a máquina se enganar para que ela se corrija a si mesma. Após uma série de falhas, ela tende para

o desejado estado final imposto pelo homem. Se o desejo for programar um robô para trazer um objeto do ponto A ao ponto B em um ambiente controlado, estas técnicas seriam perfeitamente adaptadas. Como observado por Yann Le Cun, eles não são adequados para aplicações em ambientes não controlados, como um navio em um canal de navegação ou um carro autônomo, pois imporiam acidentes reais, pois não podem simular um ambiente suficientemente próximo à realidade para serem usados na fase de aprendizagem (LE CUN, 2019).

Um dos maiores avanços recentes no *Machine Learning* é o advento do *Deep Learning*. Seus fundadores foram inspirados pela estrutura neuronal do cérebro para desenvolver arquiteturas de rede na forma de camadas de neurônios. As máquinas são, em fim, por enquanto, imitações bastante rudimentares do cérebro humano. Cada camada tem uma capacidade de aprendizado que aumenta significativamente a complexidade das tarefas que a rede neural artificial em geral pode realizar. Assim, uma imagem C distorcida à mão ou o rastreamento automático de um móvel, apesar da má visibilidade, será mais facilmente reconhecido com este tipo de técnica do que com uma simples camada de rede neural. Assim, eles gerenciam melhor o "ruído" nos dados de entrada. É com base nestas técnicas que uma máquina foi capaz de vencer o campeão mundial do jogo de Go<sup>6</sup> em 2017. Foi também usando estes sistemas que uma máquina foi capaz de terminar partituras interrompidas de grandes compositores, respeitando todos os cânones de seus estilos. Eles estão sendo usados em sistemas de reconhecimento de voz, assistentes virtuais e uma série de outras tecnologias onde a IA está atuando cada vez mais de forma autônoma (LE CUN, 2019).

### 2.1.3. Outras técnicas

---

<sup>6</sup> Esse jogo foi inventado há cerca de 2500 anos na China antiga. Ele é considerado com um jogo extremamente estratégico. Se joga com pedras pretas e brancas sobre um grade de linhas pretas. Cada jogador deve controlar com suas pedras a área mais grande possível do tabuleiro. As combinações possíveis são ainda mais numerosos que no jogo de xadrez. Em maio de 2017, um programa baseado na IA chamado *AlphaGo*, venceu o campeão mundial deste jogo, o chinês Ke jie.

As técnicas detalhadas acima são aquelas destacadas principalmente por Yan Le Cun. O relatório sobre Inteligência Artificial a Serviço da Defesa do Ministério das Forças Armadas da França distingue duas outras grandes famílias de técnicas: pesquisa operacional e sistemas multiagentes.

A pesquisa operacional é definida da seguinte forma:

*[...] famille d'algorithmes qui consiste à rechercher une solution à un problème d'optimisation, qui peut contenir un ou plusieurs objectifs à optimiser, et une ou plusieurs contraintes que la solution doit respecter.*<sup>7</sup> (TASK FORCE IA, 2019, p. 29.)

Uma das aplicações típicas desta família seria a escolha de uma rota de acordo com muitos parâmetros que evoluem com o tempo (tráfego, clima, terreno, ameaças, ...).

Os sistemas multiagentes são definidos da seguinte forma: *"famille d'algorithmes qui servent à gérer de manière décentralisée un système dans lequel de nombreux agents interviennent."*<sup>8</sup> (TASK FORCE IA, 2019. p. 29.). Estas técnicas seriam utilizadas, por exemplo, para implementar enxames de drones ou para coordenar a ação de um número muito grande de sistemas tripulados ou não tripulados.

#### 2.1.4. Características essenciais da IA

Uma revisão dessas técnicas exige algumas observações.

A primeira é que a eficácia da IA é, para a maioria das técnicas, inseparável da quantidade e qualidade dos dados disponíveis. No caso do *Machine Learning*, a escolha do conjunto de dados de aprendizagem é crucial. Mesmo no caso de aprendizado reforçado, a quantidade de dados aumenta com o número de eventos de tentativa e erro da máquina, que devem ser muito numerosos. Uma técnica não baseada no *Machine Learning* dependeria muito

---

<sup>7</sup> “[...] família de algoritmos que consiste na busca de uma solução para um problema de otimização, que pode conter um ou mais objetivos a serem otimizados, e uma ou mais restrições que a solução deve respeitar.” (Tradução Nossa)

<sup>8</sup> “uma família de algoritmos usados para administrar de forma descentralizada um sistema no qual muitos agentes intervêm.” (Tradução Nossa)



da qualidade dos dados coletados pelos sensores, por exemplo. O mundo da IA está, portanto, intimamente ligado ao mundo dos dados digitais.

A segunda é que a intervenção humana ocorre primeiro no desenho das funções matemáticas implementadas pela máquina, mais comumente conhecidas como algoritmos. No caso do *Machine Learning*, o ser homem atua na fase de aprendizagem definindo a resposta esperada e, em última instância, as funções matemáticas que compõem os algoritmos. O ser homem tem controle sobre os algoritmos, a máquina não os produz por si só.

A terceira é que, na ação, quando a máquina for posta em contato com dados que nunca viu antes, muito provavelmente será necessário um sistema de *Comando e Controle* (C2). No caso do *Machine Learning*, seria possível controlar a capacidade de generalização da máquina. Em paralelo com a fase de implementação de uma técnica de IA baseada no *Machine Learning*, poderia impor regras lógicas para garantir esta fase.

O quarto ponto é que as técnicas de IA são numerosas e estão em constante evolução. A este respeito, o mundo da defesa acompanha a evolução do mundo civil onde a IA encontra inúmeras e potencialmente lucrativas aplicações comerciais. Trata-se, portanto, de adaptar técnicas às exigências de defesa potencialmente muito diferentes das do mundo civil, particularmente em termos de eficiência e qualificação.

Vamos analisar novamente o caso do navio no canal de navegação. Agora imagine que cada navio está equipado com todo tipo de sensores, múltiplas câmeras, posição, aceleração, gravação de áudio, radares de proximidade e radares de vigilância de longo alcance, sistemas de visão com base optoeletrônica... Vamos supor que este é o caso de centenas de navios que utilizam um determinado número de canais cuidadosamente escolhidos pela representatividade deles (força das correntes, comprimento, largura, sazonalidade, ...) e isto, por várias décadas. O projeto de uma IA baseado no aprendizado de máquinas, que permite a função de navegação de um navio em um desses canais, então parece mais realizável, particularmente se forem

adicionadas regras lógicas de segurança. Pode-se imaginar, por exemplo, reações de emergência ao cruzar uma linha de sondagem mínima, dependendo da qualidade dos dados hidrográficos disponíveis e da precisão do posicionamento. Como último recurso, talvez precise ser capaz de entregar a um ser humano.

Ainda parece difícil na fase de implementação dispensar completamente o controle humano em um ambiente altamente complexo, se apenas em caso de uma emergência comprovada (falha de máquina, incêndio a bordo, etc.).

O desafio agora é entender como tais técnicas podem interagir com o melhorar ou perturbar a inteligência humana na tomada de decisões. Para isso, começaremos pelo estudo das capacidades humanas de tomada de decisão.

## 2.2. A TEORIA DAS DUAS VELOCIDADES DO PENSAMENTO E DA TOMADA DE DECISÃO

### 2.2.1. Sistema 1 / Sistema 2

Se mantivermos como definição da inteligência humana a capacidade de perceber, raciocinar e agir para poder vinculá-la com a da IA de acordo com Le Cun, é interessante compará-la com o trabalho do Prêmio Nobel de Economia Daniel Kahneman quando ele define dois sistemas de pensamento que os humanos implementam quando confrontados com uma decisão. O Sistema 1 e o Sistema 2 são complementares e cada um oferece vantagens distintas que permitem de decidir e agir.

O sistema 1 é baseado em uma forma de intuição inconsciente que é muito rápida na ordem dos reflexos. Ele realiza operações automáticas. É o que intervém para retirar a mão de uma panela muito quente antes mesmo de se tomar consciência da dor. Ela requer pouco ou nenhum esforço para funcionar e permanece capaz de se impor mesmo quando o sistema 2 é utilizado. É usado na maioria dos gestos comuns da vida cotidiana, abrindo uma porta, escolhendo uma rota em um lugar muito familiar, dirigindo um carro. Portanto, é adequado para

emergências, mas também oferece uma solução automática e sem esforço para muitas situações encontradas pelos seres humanos (KAHNEMAN, 2016).

O Sistema 2 funciona quando você se concentra em uma tarefa complexa. Por exemplo, quando alguém se concentra conscientemente em uma demonstração, estuda um objeto cuidadosamente ou realiza um complicado cálculo mental. É mais lento e requer um esforço para ser ativado. Esta é a imagem do raciocínio humano consciente (KAHNEMAN, 2016).

Estes dois sistemas são uma representação do pensamento humano. Na realidade os dois sistemas não são independentes, são inextricáveis e interagem um com o outro. Recordemos o que Daniel Kahneman diz sobre a relação entre os dois sistemas:

*Quand nous pensons à nous-mêmes, nous nous identifions au Système 2, le soi conscient, qui raisonne, qui a des convictions, fait des choix et décide que penser et que faire. Bien que le Système 2 croie être au cœur de l'action, c'est le Système 1 automatique qui est le héros du livre. Pour moi, le Système 1 produit sans effort les impressions et les sentiments qui sont les sources principales des convictions explicites et des choix délibérés du Système 2<sup>9</sup> (KAHNEMAN, 2016, p. 34).*

### 2.2.2. As falhas da tomada de decisão humana

Na prática, o Sistema 1, mais rápido e sem esforço, muitas vezes impõe uma decisão que não foi submetida ao Sistema 2. Isto permanece verdadeiro mesmo quando certas decisões exigiriam a consciência para não ficar preso pelo que Kahneman define como "viés de raciocínio" ou "viés cognitivo". Os vieses de raciocínio funcionam como simplificações excessivas. Diante de uma situação que pode requerer o sistema 2, o cérebro recorre a automatismos por facilidade. Por exemplo, ele pode detectar ligações causais mesmo quando elas não existem. Daniel Kahneman dá muitos exemplos de interpretação de estatísticas baseadas em relações causais que a mente inventa por si só, reunindo dois fatos que poderiam

---

<sup>9</sup> “Quando pensamos em nós mesmos, nos identificamos com o Sistema 2, o eu consciente, que pensa, tem convicções, faz escolhas e decide o que pensar e o que fazer. Embora o Sistema 2 acredite estar no centro da ação, é o Sistema 1 automático que é o herói do livro. Para mim, o Sistema 1 produz sem esforço as impressões e sentimentos que são as principais fontes das convicções explícitas e escolhas deliberadas do Sistema 2.” (Tradução Nossa)

estar correlacionados, mas não estão. Além disso, muitas vezes não há provas reais de que eles são (KANHEMAN, 2011). Há muitos tipos de viés de raciocínio: viés de confirmação, viés de disponibilidade, efeito halo, ancoragem mental, ... Eles são uma das fontes da falibilidade da capacidade humana de tomar decisões.

Em ação, as consequências desses vieses cognitivos podem ser dramáticas. Muito bem estudados no mundo da aviação, eles são levados em conta no treinamento de *gestão dos recursos da tripulação*, em particular para ajudar as tripulações a evitar que fiquem presas por eles e evitar acidentes. A Marinha francesa também os apresenta durante os cursos de *gestão dos recursos da ponte de comando* que oferece a suas tripulações em navios de superfície.

### 2.2.3. O senso comum e o raciocínio consciente como ferramentas de tomada de decisão

Embora o Sistema 1 seja falível, ele também é representativo do que pode ser o senso comum ou o instinto. A partir de muita pouca informação direta, a mente generaliza e aplica uma decisão de senso comum sem a necessidade de uma ação ou análise consciente. Assim, para-se num semáforo vermelho, não se pergunta a hora do dia a uma pessoa que se considera inconscientemente perigosa, escolhe-se o caminho certo de carro em um ambiente conhecido. Como marinheiro, com experiência e graças ao Sistema 1, você sabe verificar uma área safe para atracação de embarcações à contrabordo, se o navio desvia o rumo, por quantos graus o leme tem que ser ativado para evitar um obstáculo e tudo isso sem esforço. O senso comum também descarta qualquer ideia absurda. Se nos apresentam uma menina de três anos como sendo militar, sabemos que isso não é instintivamente verdade.

Além disso, o Sistema 1 pode ser melhorado. O sistema 2 é o melhor remédio para o viés cognitivo. Daniel Kahneman escreve: "*Le Système 2 est en partie capable de modifier la*

*façon de fonctionner du Système 1, en programmant les fonctions normalement automatiques de l'attention et de la mémoire.”*<sup>10</sup>(KAHNEMAN, 2016, p. 37).

Em ação, forçar-se a usar o sistema 2 ajuda a evitar generalizações abusivas. O papel das listas de verificação nos cockpits ou dos briefings das manobras nas pontes de comando é esse mesmo. Se forem bem desenhados e compreendidos pelas tripulações, o papel deles é tanto forçá-los a verificar parâmetros quanto não basear decisões futuras em dados incompletos. Elas permitem que uma série de perguntas conscientes sejam feitas antes da ação e que uma resposta apropriada seja programada com antecedência, por exemplo, no caso de uma emergência. Como eles ocorrem antes da ação, não são limitados pela urgência e imediatismo.

### 2.3. DECIDIR COM A AJUDA DA IA

#### 2.3.1. Um aumento das capacidades cognitivas do ser humano

Além de certa complexidade, as capacidades cognitivas dos seres humanos estão ultrapassadas. Os sistemas de seleção de pilotos de aeronaves mostram isso, apesar do treinamento; o ser humano atinge a sobrecarga cognitiva dele quando uma série de tarefas deve ser realizada ao mesmo tempo. Ele pode dividir e sequenciar sua atenção, mas não infinitamente. Pode até ser doloroso (KAHNEMAN, 2016). Devido à sua capacidade de lidar com grandes quantidades de dados, é provável que as técnicas de IA aumentem essas habilidades cognitivas, substituindo a necessidade de rever constantemente os dados. Na tela de um console de sistema de combate num recente escolta de grande porte de defesa aérea, é provável que milhares de moveis estejam representadas. A habilidade do operador pode lhe permitir identificar comportamentos anormais, mais isto requer sua atenção constante. É difícil garanti-lo totalmente durante um período de vários meses para cada indivíduo que se revezar nesta função. Esses fluxos colossais de dados se prestam à análise pela IA e poderiam ser

---

<sup>10</sup> “O Sistema 2 é em parte capaz de modificar o funcionamento do Sistema 1, através da programação das funções normalmente automáticas de atenção e memória.” (Tradução Nossa)

submetidos a uma "vigilância permanente garantida" que permitiria a detecção de sinais fracos de comportamento anormal.

Da mesma forma, tarefas analíticas repetitivas de compilação de dados de situação tática poderiam ser facilmente delegadas a uma IA. Pode ser usada com grande eficiência para a análise de imagens (satélite, reconhecimento...), a coleta e apresentação de dados cinemáticos de várias fontes em um determinado móvel, a operação de uma rede de sensores remota a partir de um navio de guerra.

Nas palavras da teoria de Kahneman, ao aliviar os humanos de tarefas repetitivas desinteressantes geralmente atribuídas ao sistema 1, a IA tornaria possível limitar a ocorrência de vieses cognitivos aos quais os humanos podem cair vítimas. Da mesma forma, ao aliviar o sistema 2 de um problema excessivamente complexo que está fora de seu alcance ou requer muito investimento para ser resolvido, a IA permite que o ser humano dure e se concentre sobre dados sintetizados que lhe permitirão tomar uma decisão consciente e fundamentada.

A IA seria um "sistema 3" que aumenta as capacidades dos outros dois sistemas.

### 2.3.2. Algumas possíveis funções de IA úteis para a tomada de decisões no mar

O relatório sobre "IA ao serviço da defesa" do Ministério das Forças Armadas francês define as funções prováveis da IA na forma de um verbo infinitivo (TASK FORCE IA, 2019). Para cada um deles, apresentamos exemplos confiáveis de utilização em operações.

*Detectar / Reconhecer.* Assim, a IA permitiria explorar grandes quantidades de dados de naturezas muito diferentes: imagem, radar, espectro acústico, eletromagnético. O objetivo seria acelerar o processo de reconhecimento.

*Prever.* Dependendo da disponibilidade de bases de dados de identificação de sistemas de combate combinados com bases de dados históricos, é concebível que em uma dada situação tática, a máquina possa ser capaz de manter as previsões da evolução cinemática de uma

ameaça. Esta previsão poderia ser acompanhada por uma probabilidade de ocorrência e propostas de reação associadas.

*Otimizar.* Determinar a cinemática de um grupo de plataformas variadas com capacidades diferentes diante de uma determinada ameaça. Esta poderia ser uma aplicação útil, para um grupo de ação anti submarino composto de navios de superfície, de drones, de um helicóptero ou de uma aeronave de patrulha marítima em um estreito congestionado defendido para um submarino convencional, a fim de maximizar a eficácia dos sensores.

*Buscar correlações.* A observação por uma IA de uma área durante um longo período de tempo tornaria possível identificar comportamentos anormais no meio de uma atividade de rotina, construindo uma representação da área que integra todos os sensores disponíveis. Os "sinais fracos" de atividade suspeita poderiam ser identificados em um estágio inicial da manobra.

*Deduzir / Diagnosticar.* A capacidade de manutenção preventiva oferecida pela IA torna possível prever uma individualização mais avançada da manutenção do que a atual para cada sistema do navio. Outra aplicação pode ser o diagnóstico pela IA de um plano de engajamento de mísseis de acordo com a ameaça, o estado do sistema, o meio ambiente e propostas de melhorias.

Em todos esses casos, trata-se, de fato, de ir além das capacidades cognitivas dos seres humanos, aumentando o número de dados estudados, sintetizando as contribuições deles e apresentando-os de forma eficaz para as capacidades de raciocínio humano. Na verdade, este aumento das capacidades cognitivas será muito dependente da interface homem-máquina que será desenvolvida.

### 2.3.3. Os limites intrínsecos da IA ao decidir

O primeiro fator limitante dos equipamentos baseados em inteligência artificial é seu grau de autonomia. Quando Yann Le Cun fala sobre o futuro advento de uma máquina

"pensante", ele admite que uma grande quantidade de pesquisas fundamentais sobre o funcionamento do cérebro humano ainda será necessária para alcançar este objetivo (LE CUN, 2019). Outros refutam a possibilidade de tal hipótese. Para eles, a humanidade nunca vivenciará aquele momento de "singularidade tecnológica" em que a máquina pensará e agirá de forma completamente autônoma. Um dos principais argumentos dos defensores desta teoria é que a máquina nunca será capaz de inventar sua própria linguagem para representar o mundo. Será sempre necessário um humano para "qualificar" o que a máquina percebe quando é projetada (GANASCIA, 2017).

Na indústria automotiva, a fim de melhor compreender a gama potencial de veículos autônomos, foram criados níveis de autonomia. Há seis deles (de 0 a 5), desde um carro conduzido por seres humanos sem qualquer assistência (nível 0), até um carro sem condutor capaz de operar em qualquer ambiente: não haveria volante ou pedais neste carro. A diferença entre o nível 4 e o nível 5 reside principalmente nas condições ambientais em que o veículo opera. As experiências até o momento mostram que veículos autônomos como os da empresa Waymo, uma subsidiária da Alphabet (a empresa mãe da Google), ainda precisam operar em um ambiente controlado. As estradas devem ser largas, o tráfego deve ser leve e o clima deve ser ameno para que o carro possa operar em plena autonomia (LE CUN, 2019).

O segundo fator limitante é a especialização das máquinas baseadas em IA. Agora é possível criar uma máquina capaz de derrotar qualquer humano no jogo de Go. Também é possível criar uma máquina baseada em IA e *Deep Learning* que seja capaz de fazer música ou pinturas no estilo de um artista famoso. Uma máquina baseada em IA pode reconhecer uma voz, uma imagem. Entretanto, hoje é impossível desenvolver uma máquina que faça tudo isso ao mesmo tempo. Um dos fatores limitantes seria a capacidade computacional necessária para criar gigantescas redes neurais artificiais capazes de aprender em vários campos. Esta



capacidade teria que ser fenomenal e não pode ser criada com a tecnologia de hoje (LE CUN, 2019).

Outro fator determinante em parte relacionado à especialização das máquinas baseadas em IA é a falta de senso comum delas. Tudo tem que ser descrito a uma máquina, e os dados de aprendizagem fornecidos a ela têm que ser etiquetados para que ela possa reconhecê-la. Em um ambiente complexo e descontrolado, sua capacidade de "inferência" é zero. Se ele "vê" um objeto que não conhece, não pode, como os seres humanos, usar sua experiência e habilidades instintivas para encontrar um uso para ele (LE CUN, 2019). A máquina não adivinha fora de seu campo de conhecimento e não sabe instintivamente, observando-os, que uma panela pode estar quente, que a imagem de um edifício visto do céu em 2D corresponde na realidade a um objeto em 3D, que um objeto colocado em equilíbrio só pode cair.

O último fator determinante apresentado aqui é a imprevisibilidade de uma máquina baseada no *Deep Learning*. De fato, na fase de implementação, a máquina integra dados que nunca viu antes. A questão de como determinar o erro potencial da máquina, exceto empiricamente através da experiência, surge. Enquanto para uma máquina de reconhecimento de cheques isto pode não ser um problema, para um sistema independente que contribui para o uso da força, está fora de questão. A fim de limitar o erro é necessário, por exemplo, escolher cuidadosamente os dados de aprendizagem. Isto coloca dois problemas: o primeiro é a disponibilidade de dados em grandes quantidades. A segunda é que, além de uma certa quantidade, é ilusório acreditar que um ser humano é capaz de examiná-los individualmente. Portanto, são possíveis erros nos dados fornecidos à máquina (GANASCIA, 2017). Isto é verdade no caso da aprendizagem de máquinas, mas permanece verdadeiro no caso de um grande volume de dados usados pela IA.

Depois de descrever brevemente as capacidades da IA e sua contribuição às capacidades cognitivas humanas, conforme descrito por Daniel Kahneman, vamos aplicar este conhecimento às decisões sobre o uso da força letal no mar.

### 3. DECIDIR USAR A FORÇA LETAL NO MAR COM A AJUDA DA IA

O objetivo deste capítulo é estudar como a IA pode interagir com as habilidades cognitivas humanas no caso específico do uso da força letal no mar. Para isso, começaremos por identificar algumas características dos processos de uso da força no mar ligando-os, em parte, à teoria de Kahneman. Em seguida, vamos propor exemplos de como a IA pode fornecer contribuições úteis para este tipo de decisão. Finalmente, estudaremos os limites dessas contribuições.

#### 3.1. DECIDIR USAR A FORÇA LETAL DE ACORDO COM OS DOMÍNIOS DE LUTA<sup>11</sup>

##### 3.1.1. Antes de decidir usar a força letal, enfrente a incerteza com a IA

Em uma passagem clássica de *Von Kriege*, Clausewitz define a incerteza de todos os dados na guerra:

*Enfin, la grande incertitude de toutes les données constitue une difficulté particulière de la guerre, car toute action s'accomplit pour ainsi dire dans une sorte de crépuscule qui confère souvent aux choses comme un aspect nébuleux ou lunaire, une dimension exagérée, une allure grotesque.*<sup>12</sup>  
(CLAUSEWITZ, 1955, p. 133).

Assim, ele define uma característica do teatro das operações, comumente chamada de "névoa da guerra", marcada pela incerteza. Esta incerteza continua sendo uma característica fundamental da tomada de decisões militares nos dias de hoje. Porém Clausewitz não vivia na era do *Network Centric Warfare*<sup>13</sup> ou da <sup>14</sup>guerra da informação. O General francês Vincent Desportes, que estudou a tomada de decisões diante das incertezas mais recentemente, explica:

---

<sup>11</sup> Este estudo se concentrará em quatro domínios de luta de acordo com a tipologia da Marinha francesa: luta antiaérea, luta assimétrica, luta anti superfície, luta anti submarina

<sup>12</sup> “Finalmente, a grande incerteza de todos os dados é uma dificuldade particular da guerra, pois cada ação é realizada, por assim dizer, em uma espécie de crepúsculo que muitas vezes dá às coisas um aspecto nebuloso ou lunar, uma dimensão exagerada, um fascínio grotesco” (Tradução Nossa)

<sup>13</sup> *Network Centric Warfare* é um conceito que surgiu nos Estados Unidos durante a Guerra do Golfo, descrevendo o estabelecimento de numerosas redes de informação em todo o teatro, ligando todas as forças amigas para que elas possam compartilhar informações de forma mais eficaz.

<sup>14</sup> *A guerra da informação* descreve ações que vão desde a aquisição de informações até a degradação da capacidade do adversário de adquirir informações, até a manipulação de informações e influência. Ela assume um valor especial na era das redes sociais e da *Fake News* e está totalmente integrada na condução das operações contemporâneas.

“*On pouvait hier décider et combattre par l’information ; on se rend compte aujourd’hui que l’on est contraint de combattre en plus pour l’information avant même de pouvoir décider.*”<sup>15</sup> (DESPORTES, 2011, p. 15). De fato, o campo da informação útil para a tomada de decisões, e em particular devido aos dados digitais, está se tornando cada vez mais amplo. Além disso, esses dados têm uma capacidade até então desconhecida de influenciar o teatro das operações e as linhas de operação. Da mesma forma, parte das informações úteis para a tomada de decisões está potencialmente disponível, mas escapa da atenção do tomador de decisões. Finalmente, o estado das redes de comunicação está no centro da manobra. Por exemplo, a camada física e a camada de software do ciberespaço <sup>16</sup> condicionam o acesso à camada de informação. As fraquezas e falhas delas modificam potencialmente em tempo real o campo de informação acessível ao tomador de decisão. Além disso, todos esses elementos são objeto de uma luta por seu controle entre os oponentes das guerras contemporâneas.

A capacidade da IA de explorar quantidades muito grandes de dados é uma ferramenta perfeitamente adaptada a este contexto. Independentemente do domínio de luta no mar, ele potencialmente amplia o escopo de informações disponíveis para o tomador de decisão. Também torna possível fazer melhor uso dessas informações. Por exemplo, usando sua capacidade de buscar correlações a fim de identificar comportamentos que reflitam a vontade do adversário, pode ajudar a compreender suas intenções. Finalmente, oferece uma capacidade de otimizar os fluxos de informação para melhorar o funcionamento das redes de comunicação. Diante de uma IA ofensiva implementada pelo adversário, é provável que apenas uma IA saiba como se defender nesta luta para a informação (TASK FORCE IA, 2019).

---

<sup>15</sup> “Ontem podíamos decidir e lutar por informações; hoje percebemos que somos forçados a lutar para informações antes mesmo de podermos decidir.” (Tradução Nossa)

<sup>16</sup> O modelo mais simples de representação do ciberespaço é composto de três camadas: física (*hardware*, infraestrutura, cabos, fibra ótica, etc.) / software (o *software* o que faz a camada física funcionar e fornece acesso à informação) / informação (o conteúdo e, portanto, a informação útil).

Estas reflexões sobre a incerteza são válidas qualquer que seja o domínio de luta prevista no mar. Elas se aplicam particularmente ao conhecimento geral do teatro das operações e do ambiente geral nos níveis operacional e estratégico. No nível tático, os processos de tomada de decisão variam de acordo com os domínios de luta e influenciam as contribuições potenciais da IA para cada uma deles.

### 3.1.2. Luta antiaérea, luta assimétrica: os domínios de luta onde predomina o "Sistema 1".

A luta antiaérea e a luta assimétrica são domínios de luta estreitamente relacionadas em termos do uso de armas. Em face de uma aeronave, um míssil ou uma embarcação suicida, a tripulação de um navio de guerra tem um curto tempo de reação para tratar o alvo se for necessário. A aquisição das primeiras informações táticas úteis para o emprego de armas é descentralizada. De fato, o comandante do navio ou os oficiais encarregados do emprego de armas não são necessariamente os primeiros a perceber a ameaça. É até mesmo provável que seja um simples marinheiro.

Em vista destes elementos, o princípio da delegação de respostas à ameaça é um importante fator de sucesso para enfrentar a ameaça de forma adequada. Essas delegações existem tanto para o processamento de informações (classificação de pistas, informações à distância, comportamento, etc.) quanto para reações à ameaça (disparo de foguetes, tiros de advertência, tiros de neutralização, etc.). O problema é ainda mais complexo no caso de vários navios atuando em conjunto. Nesses casos, a interação das delegações se estende às unidades entre elas. Este será o caso, por exemplo, entre um escolta especializado<sup>17</sup> responsável pela defesa aérea e os outros navios da força amiga. A diferença no nível de emprego de todos os atores da cadeia de comando, a possível presença de várias unidades e a necessidade de ser

---

<sup>17</sup> Escolta especializada: uma escolta cujos sensores, sistemas de armas e habilidades da tripulação lhe permitem assumir um papel de liderança em uma determinada domínio de luta, em benefício de uma força de vários navios. Na França, as fragatas de defesa aérea do tipo Horizon, por exemplo, estão entre as escoltas especializadas em defesa aérea.

rápido requer um extenso treinamento a fim de evitar surpresas e enganos. Estas áreas de treinamento de combate são caracterizadas pela mecanização e pelo treino de repetição. Cada nível da cadeia do emprego das armas aprende a executar rapidamente as ações delegadas a ele, até chegar ou não a uma decisão de abrir fogo.

Lembrando a teoria dos dois sistemas, esta descrição enfoca as operações do sistema 1, conforme descrito por Kahneman. As ações devem ser rápidas, automáticas e capazes de funcionar continuamente. As operações do Sistema 1 são de fato rápidas, constantemente ativadas e automáticas. No nível tático de ação, estes domínios de luta são, portanto, em grande parte, uma questão de um Sistema 1 "treinado". A repetição de ações, a experiência de cada um e, em última instância, as situações encontradas no mundo real contribuem para a programação do Sistema 1 para reagir bem. Nesses domínios de combate, as operações do sistema 2 são utilizadas para a avaliação prévia do meio ambiente e da ameaça, bem como no projeto de treinamento. Quando as equipes realizam reuniões de debriefing, elas contribuem para a programação do Sistema 1 utilizando o Sistema 2. As operações do Sistema 2, portanto, ocorrem principalmente antes ou depois da ação.

### 3.1.3. Luta anti superfície e luta anti submarina: os domínios de luta onde predomina o "Sistema 2"

A luta anti superfície e a luta anti submarina na guerra naval são domínios de luta que são implantadas em um espaço-tempo diferente das áreas de guerra discutidas no parágrafo anterior. Em particular, uma ameaça que consiste de submarino(s) e/ou de navio(s) de superfície tem uma capacidade de ataque deslumbrante menor se sua presença for conhecida. Neste caso, eles permitem mais tempo para desenvolver uma manobra da força amiga ao longo do tempo. Estas manobras podem ser realizadas em uma grande área geográfica e são fortemente condicionadas no tempo pelo ambiente (batimetria, geometria costeira, estado do mar, etc.). Assim, a decisão de concentrar os fogos, a necessidade de usar meios e sensores complementares ou o uso de

meios aéreos tornam a construção da ideia de manobra sensível a um grande número de parâmetros que evoluem durante a manobra que leva à abertura do fogo.

Esses dois domínios de luta são caracterizados pela decisão centralizada de emprego da força. O comando tem geralmente, o tempo necessário para manter a decisão de abertura de fogo em seu nível. Muitas vezes também manterá em seu nível, a decisões das ações de gradação do uso da força (iluminação com o radar de controle de fogo, cargas de profundidade dissuasivas...). Portanto, é uma questão para as equipes que lideram a luta produzir um esforço para sintetizar e atualizar informações. Reações automáticas imediatas permanecem úteis, mas sim em caso de engajamento do inimigo, o que muitas vezes constitui o fracasso da manobra da força amiga.

Estes elementos devem ser comparados com as definições apresentadas acima no Sistema 2<sup>18</sup>. Assim, pode-se ver que as operações deste sistema são preponderantes no caso destas duas áreas de ação. De fato, a complexidade, a necessidade de concentração ao longo do tempo e o ritmo relativamente lento de ação se prestam às vantagens deste sistema. O sistema 1 não está totalmente ausente; ele é de grande ajuda na reação ao disparo de um torpedo pelo submarino inimigo ou de um míssil por um navio de superfície inimigo. Portanto, estes são domínios de luta onde o sistema 2 é predominante, mas também "treinado". Sua eficácia depende em particular da resistência da tripulação. Esta deve permanecer concentrada por um longo período de tempo, permanecendo capaz de detectar pontos de decisão durante um longo período de tempo, às vezes por vários dias. Ela deve ser condicionada a evitar de recorrer ao sistema 1, que oferecerá raciocínio automático, alguns dos quais estarão sujeitos a viés cognitivo e os fará perder o senso de ação e, em última instância, a iniciativa.

### 3.2. O QUE PODE TRAZER A IA NESTE TIPO DE DECISÕES?

---

<sup>18</sup> Ver parágrafo 2.2.1

3.2.1. Buscar correlações para enfrentar a incerteza para o benefício do sistema 1 ou do sistema 2.

Consideremos agora as aplicações das funções relacionadas à IA identificadas no parágrafo 2.3.2 no contexto da decisão sobre o uso da força no mar. A função "Buscar correlações" pode dar à IA a missão de acelerar a transmissão de informações úteis estabelecidas no nível operacional ou mesmo estratégico para o nível tático. O objetivo é construir uma representação do teatro que permita a identificação de sinais ligados a eventos como a saída de um submarino, a preparação da artilharia costeira, a preparação de um ataque aéreo ou os movimentos coordenados de uma força de superfície. Tudo isso são mudanças na situação que interessem ao teatro inteiro. Para atingir tal objetivo, é claro que estamos pensando na análise acelerada pela IA de imagens de qualquer tipo, mas a ambição de tal IA deve integrar pesquisas de fontes múltiplas e, em particular, visar a verificação cruzada das informações disponíveis no ciberespaço. A vantagem da IA sobre os processos de inteligência tradicional seria oferecer uma análise permanente e contínua do teatro, disponível em tempos de reação mais curtos. Por exemplo, para um determinado teatro, seria uma manobra de busca de correlação em quatro etapas:

- Identificar fontes de dados de todos os tipos que podem ser utilizadas pela IA;
- Construir um plano de recuperação de dados de acordo com os princípios de redundância, continuidade e permanência;
- Construir um plano de divulgação das informações obtidas a fim de alimentar o nível tático o mais rápido possível;
- Identificar os pontos fracos previsíveis em cada fonte de dados e fornecer um plano de defesa e contingência das fontes no caso da indisponibilidade de uma ou mais fontes.



Dada a complexidade dos algoritmos implementados e a capacidade computacional necessária associada a eles, este seria um apoio à decisão em terra para a tomada de decisões no mar. Ela serviria a todos os domínios de luta e contribuiria para a eficácia de ambos os tipos de sistemas de raciocínio no momento da ação.

### 3.2.2. Detectar / reconhecer para o benefício do sistema 1.

Um dos desenvolvimentos no campo da defesa aérea e da luta assimétrica é a redução do limite de detecção de ameaças, tanto em termos de velocidade quanto de capacidade de furtiva. O advento dos mísseis de hipervelocidade<sup>19</sup>, ou a miniaturização da ameaça assimétrica com a ajuda de drones <sup>20</sup>aéreos ou de superfície, demonstra isto. A retirada deste limite é expressa por uma detecção possível mas tardia (caso do míssil de hipervelocidade), mas também por uma degradação da capacidade de detecção (caso do drone comercial usado como arma). No primeiro caso, a tentação de usar uma defesa antimíssil autônoma e automática é grande, o que, na presença de tráfego civil significativo, aumenta o risco de engano<sup>21</sup>. No segundo caso, as dificuldades de detecção favorecerão o risco de surpresa. Seria uma questão de dar à IA a tarefa de manter este limite de detecção em um nível aceitável, preservando ao mesmo tempo a possibilidade de decisão humana na abertura do fogo para assegurar o reconhecimento da ameaça. Para fazer isso, a IA proposta forneceria uma capacidade colaborativa de detecção e reconhecimento multi sensor e multiplataforma que poderia fornecer designação de alvo para o benefício de sistemas de armas. Seria capaz de ter todos os sensores

---

<sup>19</sup> Por exemplo, a Rússia está desenvolvendo mísseis Zirkon hipersônicos que seriam lançados a partir de navios de superfície e depois de submarinos cuja velocidade prevista entre Mach 8 e 9 tornaria muito difícil a interceptação pelos atuais sistemas anti-mísseis. Em março de 2020, a campanha de teste deste míssil foi amplamente noticiada na imprensa. Disponível em: <https://www.naval.com.br/blog/2020/03/20/russia-anuncia-segundo-teste-do-missil-hipersonico-zircon/> Acesso em 12 julho 2020.

<sup>20</sup> Usado em terra no Afeganistão, Iraque ou Síria - como meio de transporte de explosivos, por exemplo - seu uso no mar representa um grande risco, particularmente para operações costeiras e anfíbias, ou ao cruzar pontos de estrangulamento (estreitos, canais, arquipélagos, etc.).

<sup>21</sup> Os eventos atuais de 2020 demonstraram mais uma vez quão central deve ser a consciência do risco de engano no campo da defesa aérea no estabelecimento de procedimentos e no desenvolvimento de equipamentos de defesa aérea. Disponível em: <https://oglobo.globo.com/mundo/chefe-da-forca-aerea-do-ira-assume-responsabilidade-por-derrubar-aviao-ucraniano-24184762> Acesso em 12 julho 2020.

de uma força, seja ela tripulada ou não, contribuindo. Pode-se pensar em particular na instalação de drones em um papel avançado<sup>22</sup> de detecção / reconhecimento.

### 3.2.3. Prever e deduzir para o benefício do sistema 2.

A disseminação da ameaça submarina e as estratégias de antiacesso e negação de área complicam o controle dos espaços marítimos<sup>23</sup>. O controle dessas áreas é o objetivo de qualquer contexto de luta antinavio ou antissubmarina. Uma força naval amiga por razões defensivas ou ofensivas deve manter uma mobilidade cinemática que é cada vez mais impedida perto das costas onde são aplicadas restrições "maiores", tais como o alcance máximo dos sistemas de artilharia antinavio inimigas, os alcances dos meios aéreos inimigos ou a simples presença de um submarino no mar. Entretanto, dois fatores favoráveis deveriam levar a uma revisão dessas grandes restrições, cuja relevância seria aumentada pelo uso da IA.

Em primeiro lugar, os sensores e efetores<sup>24</sup> da força amiga melhorarão com a entrada da IA. É de se esperar que sua eficácia contra certas ameaças seja aumentada, em particular devido ao aumento das capacidades de detecção/reconhecimento. Já os meios modernos antissubmarinos mais eficientes permitem, dependendo do ambiente, um rastreamento quase contínuo dos submarinos durante vários dias. Além disso, a resistência de drones ou outros sensores não tripulados controlados pela IA, provavelmente reduzirá a incerteza sobre as posições do inimigo.

Em segundo lugar, os meios do oponente também são limitados. Uma bateria em terra não se instala em qualquer lugar porque a geometria da costa nem sempre o permite. Da mesma forma, não "cobre" toda sua área de ação, por exemplo, devido à presença de tráfego civil

---

<sup>22</sup> No setor aéreo, o programa de armamento envolvendo a França, Alemanha e Espanha do Sistema de Combate Aéreo do Futuro (SCAF) está buscando desenvolver uma arquitetura de "sistemas de sistemas" combinando uma nova aeronave de combate com sensores e drones remotos.

<sup>23</sup> A instalação de mísseis anti-navio YAKHONT na Síria ou a militarização das ilhotas no Mar do Sul da China são exemplos disso. Existem agora mais de 500 submarinos operacionais no mundo e este número vem aumentando constantemente há vários anos.

<sup>24</sup> O conjunto de todas as capacidades ativas de enfrentar a ameaça: armas, capacidade de guerra eletrônica ativas...

pesado ou condições ambientais degradadas. Ao refrescar a posição das unidades no mar, sejam navios de superfície ou submarinos, cada ocorrência permite que as áreas de ação cobertas por seus sistemas de armamento evoluam, cuja eficácia varia de acordo com as condições ambientais.

O papel de uma IA neste contexto poderia ser propor ferramentas para a integração de todas estas dimensões para caracterizar no tempo e de forma mais otimizada o controle do espaço aeromarítimo pela força amiga e a ameaça. Tal IA permitiria, por exemplo, representar o espaço aeromarítimo em uma divisão de zonas estabelecida de acordo com a liberdade de manobra que eles oferecem. Isso permitiria identificar dinamicamente as fraquezas na armadura do adversário para manter a iniciativa. Com o tempo, ele identificaria pontos de decisão para ganhar vantagem sobre o adversário ou preservar a força amistosa da ameaça. Tal raciocínio combinatório, que apresentaria tantas variáveis, está além do escopo das operações do Sistema 2 somente se tiver que ser constantemente atualizado. A IA com sua capacidade de processamento de dados é capaz de atender a esta necessidade. Seria uma questão de primeiro, prever, depois deduzir uma vez contratada a força amiga, onde se posicionar e como favorecer o emprego de seus próprios efetores, minimizando ao mesmo tempo a eficácia do dispositivo inimigo.

### 3.3. OS LIMITES DA CONTRIBUIÇÃO DA IA PARA ESSAS DECISÕES.

Estas aplicações teóricas das capacidades de IA devem ser comparadas com as fraquezas intrínsecas deste tipo de tecnologia.

#### 3.3.1. Especialização e imprevisibilidade nas áreas de controle do sistema 1.

Os limites ligados à especialização da<sup>25</sup> IA, na luta antiaérea contra aeronaves com evoluções cinemáticas erráticas, na luta assimétrica contra móveis cujas formas e cinemáticas

---

<sup>25</sup> Ver o parágrafo 2.3.3.

podem ser muito diversas são preocupantes. É difícil prever a criação de uma IA que possa lidar com todo o espectro de possíveis ameaças. Tomando o exemplo de uma tecnologia baseada na aprendizagem supervisionada de máquinas<sup>26</sup>, pode-se perguntar qual conjunto de dados de aprendizagem seria suficientemente completo para permitir a generalização efetiva da IA.

As limitações associadas com a imprevisibilidade da<sup>27</sup> IA são problemáticas em relação às exigências para o reconhecimento formal de ameaças. Em particular, no contexto da aprendizagem de máquinas, agora parece difícil obter uma qualificação de um sistema que não seja empiricamente, através da extração de estatísticas de seu uso em condições reais.

### 3.3.2. A falta de senso comum nos domínios de luta do sistema 2.

Na exploração das funções de previsão ou dedução, descritas para o benefício dos campos de antinavio e/ou guerra antissubmarino, surge a questão do senso comum<sup>28</sup>. Este seria no caso, em particular, da avaliação de risco, que exige constantemente uma capacidade de inferência ou até mesmo de instinto. Enquanto uma grande quantidade de dados técnicos - se disponíveis - tornaria possível caracterizar melhor a capacidade da ameaça de uma forma dinâmica, o problema surgido seria o de ligá-la a um nível de risco. A ameaça é composta de uma capacidade de prejudicar, mas também de uma intenção de prejudicar. Esta intenção do adversário é difícil de ser parametrizada em algoritmos. Se o adversário é uma coalizão cujos interesses dos membros são semelhantes, mas não idênticos, o problema se torna ainda mais complexo. Se a IA tornasse possível prever ou inferir as capacidades do inimigo, a névoa da guerra perduraria por muito tempo sobre suas intenções.

### 3.3.3. Os limites intrínsecos de acesso aos dados na ação.

Qualquer que seja a tecnologia de IA utilizada, ela exigirá uma grande quantidade de dados para ser eficaz. O acesso a esses dados em tempo real e em preparação para a ação será

---

<sup>26</sup> Idem.

<sup>27</sup> Idem.

<sup>28</sup> Ver o parágrafo 2.3.3.

provavelmente objeto de uma luta com o adversário que tentará privar a força amiga dos mesmos. Falhas técnicas dos sensores ou sistemas de comunicação e problemas de interoperabilidade entre plataformas podem levar à redução do acesso aos dados dos sistemas implementados. Estas degradações na eficácia das ferramentas baseadas na IA constituem uma limitação que será difícil de avaliar sem uma compreensão para os membros do nível tático mais baixo de seus impactos sobre o funcionamento da IA.

Tendo vinculado as capacidades oferecidas pela IA a casos de decisões de uso de força letal no mar, parece que a IA atuaria como uma espécie de "sistema 3" que aumentaria de acordo com as missões a ela confiadas, sejam operações do sistema 1 ou do sistema 2. Agora é uma questão de definir quais requisitos e características essenciais são necessários considerar em vista das capacidades e limitações identificadas das técnicas de IA, no contexto em consideração.

## 4. O QUE EXIGIR DAS TÉCNICAS DE IA NO CONTEXTO DO USO DA FORÇA LETAL

Se considerarmos que a IA nesta estrutura atuaria como um "sistema 3" interagindo com as capacidades do espírito humano descritas a partir da perspectiva da teoria de Kahneman, uma IA estaria em posição de impor "viés de operação"<sup>29</sup> ao espírito humano. Portanto, devemos discutir que tipo de exigência deve ser definida para evitar as muitas limitações da IA, conforme descrito acima. Neste capítulo, lembramos algumas exigências óbvias relacionadas com a lei e a ética dos militares. Em seguida, nos concentraremos na definição do desenho dos algoritmos e dos recursos de gerenciamento de dados que atendam parcialmente a estes requisitos. Finalmente, discutiremos novas exigências relacionadas ao lugar dos seres humanos em relação à IA.

### 4.1. UMA EXIGÊNCIA LEGAL E ÉTICA

#### 4.1.1. O paradigma do “robô assassino”

A fim de esclarecer este tópico, é necessário primeiro considerar o que seria uma máquina baseada em IA, totalmente autônoma no uso da força letal. A definição mais comumente aceita de tal máquina chamada Sistema Autônomo de Armas Letais é um sistema de armas “that, once activated, can select and engage targets without further intervention by a human operator.”<sup>30</sup> (ANDERSON, WAXMAN, 2014, p. 2.).

Com relação ao direito internacional do conflito armado, tal sistema levanta a questão do respeito da IA pelos princípios de distinção, proporcionalidade, necessidade e humanidade. Em vista do conhecimento previamente desenvolvido sobre as capacidades do sistema 2 para modificar a forma como o sistema 1 funciona,<sup>31</sup> é fácil entender como os seres humanos seriam

---

<sup>29</sup> Distinta das descritas por Kahneman para o Sistema 1

<sup>30</sup> "uma vez ativado pode selecionar e engajar um alvo de forma autônoma sem a necessidade de intervenção humana adicional" (Tradução Nossa)

<sup>31</sup> Ver parágrafo 2.2.3

levados a aplicá-los instintivamente. De fato, estes princípios do direito são obviamente levados em conta nas escolas de treinamento militar, na concepção das operações e na elaboração das ordens antes da ação. Este é o momento em que o Sistema 2 constrói a estrutura para as futuras reações do Sistema 1. Mais tarde, ao reagir a uma ameaça súbita e difícil de prever, os militares usarão o senso comum deles para obedecer a estes princípios. Além de uma reação automática, é a apreciação visual deles da presença de um não combatente, o julgamento do grau de ameaça em relação à capacidade percebida de prejudicar, a tensão ou não das horas anteriores, mas também o resultado de uma lenta construção de estruturas morais e éticas que o levará a respeitar estes princípios. No entanto, é precisamente este senso comum que seria o aspecto mais difícil da mente humana para ensinar IA (LE CUN, 2019).

Finalmente, a questão levantada pelas SAALs é a da responsabilidade legal. Como o uso de força letal está além do controle de qualquer ser humano, a responsabilidade por um erro cometido pela máquina pode ser rejeitada por aqueles que a utilizam. Seria até possível usar a IA para evitar ter que assumir a responsabilidade pelo uso da força letal. Desta forma, o surgimento de sistemas de armas ilegais, autônomos e letais contribuiria para a eficácia da guerra híbrida<sup>32</sup>.

#### 4.1.2. Os riscos legais dos vieses em um sistema 3

As questões levantadas pelo uso do SAAL também se aplicam a sistemas de armas parcialmente autônomos. Pode-se ver pelos exemplos desenvolvidos no Capítulo 2, particularmente nas funções de detecção/reconhecimento, que a designação de um alvo ilegal é um risco a ser levado em conta. De fato, assim como o sistema 1 tem precedência sobre o sistema 2 em um contexto de emergência, é possível que o sistema 3 não possa ser questionado

---

<sup>32</sup> A guerra híbrida aqui se refere a formas de guerra que combinam ações assimétricas, guerra de informação e guerra convencional com ações econômicas, políticas e sociais com o objetivo de permanecer abaixo do limite de responsabilidade que leva a declaração de guerra.

pelo sistema 2. Pela facilidade, cansaço, falta de tempo ou falta de compreensão de como funciona, o uso do sistema 3 poderia levar à violação das leis da guerra.

#### 4.2. UM REQUISITO PARA O DESENHO DE ALGORITMOS E GERENCIAMENTO DE DADOS

A fim de proteger-se contra vieses no funcionamento da IA, parece que regras devem reger o desenho dela. Estas regras se aplicariam ao desenho dos algoritmos, mas também ao gerenciamento das enormes quantidades de dados necessários para sua operação. As regras aqui apresentadas são baseadas nos elementos de uma "IA de confiança" apresentada no relatório da Força Tarefa IA do Ministério das Forças Armadas francês (TASK FORCE IA, 2019).

O primeiro princípio que se aplicaria seria a transparência dos algoritmos. O objetivo de tal princípio é tornar o funcionamento da IA inteligível para seus usuários. Pela sua concepção, os limites das ferramentas baseadas em IA devem ser explícitos para os seres humanos. Como exemplo, seria uma questão de apresentar o grau de autonomia da IA na imagem do que é feito para os veículos autônomos. Finalmente, é uma questão de não aceitar uma técnica de IA que apresentaria uma "caixa preta" que não tenha sido explicitada por seu desenhista. Portanto, é uma questão de proceder como Kahneman faz com o sistema 1: identificar desde a fase de desenho e apresentar ao usuário os vieses do sistema 3, que é a IA. Além disso, qualquer técnica de IA precisa ter acesso a grandes quantidades de dados em tempo real na ação a fim de funcionar. A degradação do acesso a estes dados deve ser caracterizada para o usuário de tal forma que este possa reavaliar a relevância das soluções propostas pela IA.

O segundo princípio é uma exigência de qualificação. No caso de uma técnica baseada em redes neurais que envolvem um risco de imprevisibilidade, trata-se, por exemplo, de utilizar algoritmos de controle comportamental (TASK FORCE IA). O objetivo deste princípio é poder



fornecer uma demonstração estatística da operação eficiente e legal da IA, a fim de limitar o risco de imprevisibilidade a um limite aceitável.

O terceiro princípio é a necessidade de fazer evoluir a técnica ao longo do tempo de acordo com as novas fontes de dados. Um conflito armado muito provavelmente resultaria em uma tentativa do adversário de contornar a eficácia da IA. O objetivo seria apresentar-lhe dados fora de sua capacidade de generalização, como a camuflagem dinâmica. Também poderia tentar atrair um sistema de armas, mesmo que isso signifique direcioná-lo para alvos ilegais. É, portanto, uma questão de projetar técnicas de IA que evoluem com o tempo. É concebível que o feedback das operações possa levar a atualizações dos bancos de dados de aprendizagem e que os algoritmos possam ser readaptados de acordo. Novos algoritmos de controle também poderiam ser introduzidos de forma dinâmica. Isto exigiria, no caso de uma técnica a bordo de um navio de guerra, uma estrutura baseada em terra que poderia desempenhar esta função de atualização antes de redistribuir os algoritmos modificados.

O quarto princípio seria a desativação. Trata-se da possibilidade de usar os sistemas de armas relevantes sem a IA. Isto é essencial quando há a necessidade de poder fugir de uma IA mal desenhada diante de uma nova ameaça ou quando há a necessidade de limitar os efeitos de muito pouco acesso aos dados em ação. Modos degradados de sistemas de armas devem permitir o uso de força sem o uso de IA enquanto permanecem controláveis pelos operadores.

Finalmente, o princípio da cooperação e do compartilhamento em técnicas de IA deve ser aplicado. A cooperação internacional ou mesmo o estabelecimento de padrões comuns para técnicas de IA militar deve ser incentivada, particularmente dentro das organizações militares aliadas. De fato, o sucesso dessas técnicas depende, em última instância, de sua interoperabilidade com os sistemas aliados. Em comparação com o mundo civil, a eficácia da pesquisa de IA é baseada em uma cultura de compartilhamento de conjuntos de dados, mas também em algoritmos. As equipes de pesquisa se alimentam umas das outras (LE CUN, 2019).

As indústrias de defesa e as forças armadas se beneficiariam, portanto, de compartilhar, mesmo que apenas parcialmente, seus conjuntos de dados a fim de facilitar a pesquisa e a eficácia de seus sistemas de armas baseadas em IA.

### 4.3. PRESERVANDO O LUGAR DA DECISÃO HUMANA

#### 4.3.1. Luta contra o *Nudge*

Introduzida por Richard Thaler e Cass Sunskind no campo da economia comportamental, a noção de *Nudge* é amplamente utilizada em aplicações civis de IA. Esta teoria se baseia no fato de que os seres humanos são deficientes na tomada de decisões por constantes tendências comportamentais como as descritas por Kahneman. A técnica do *Nudge* consiste em modificar o ambiente do sujeito para direcioná-lo para a decisão correta sem impor. Há muitos exemplos desta técnica em aplicações civis da IA. Este é particularmente o caso nas redes sociais que coletam e exploram dados que supostamente representam os interesses de seus usuários. Eles são então capazes de selecionar elementos do ambiente publicitário que possam ser de seu interesse. Desta forma, evitam a rejeição de seus membros ou a impressão desagradável de ter propagandas que não lhes interessam impostas, conduzindo-os "naturalmente" para os produtos desejados. O campo de uso das técnicas de *Nudge* se estende com a personalização das ferramentas baseadas na IA. O risco é que parte do ambiente seja escondida, mesmo que informações valiosas sejam descartadas. Finalmente, o indivíduo abdica de parte de sua liberdade de escolher algo diferente do que a IA lhe oferece (KOENIG, 2019).

Na mesma lógica de transparência dos algoritmos, podemos pensar que as interfaces homem-máquina (IHM) das tecnologias baseadas em IA no mundo militar devem ser preservadas deste efeito de *Nudge* ou, pelo menos, propor um modo de operação que elimine este filtro do ambiente quando necessário. Neste aspecto, o desenho das IHMs é fundamental para evitar encorajar sentimentos de superconfiança nas capacidades de IA.

#### 4.3.2. Manter o homem conectado ao processo de decisão da IA

Três tipos de formas pelas quais os seres humanos podem interagir com as decisões de uma IA empregado no processo de decisão de força letal podem ser definidos como *man in the loop*, *mano on the loop* ou *man out of the loop* (SCHAUB, KRISTOFFERSEN, 2017). O homem pode, portanto, estar completamente fora do processo de decisão (*man out of the loop*). O homem pode ser integrado ao ciclo de decisão (*man in the loop*). Neste caso, sem ele, o processo não pode ser bem-sucedido por si só. Finalmente, o homem pode estar em uma situação intermediária onde, após iniciar o processo, ele o controla com um direito de veto que pode interromper a ação (*man on the loop*).

Nos dois últimos casos, o lugar da tomada de decisão humana é preservado. Eles não têm as mesmas vantagens. No caso do *man in the loop*, a intervenção humana é ad hoc e necessária em marcos selecionados no projeto. No caso do *mano on the loop*, a capacidade de interromper a ação pode ser contínua e possível até o último momento. Em contextos de tempos de reação muito curtos, isto economiza o tempo necessário para que um míssil viaje até seu alvo para mudar sua decisão de se engajar. De fato, isto envolveria a generalização de processos de IA que já se aplicam a muitos sistemas de armas em serviço<sup>33</sup>. Mas dada a complexidade dos algoritmos de IA, poderia haver arquiteturas híbridas. Algumas condições estruturais mudariam automaticamente de uma *mano on the loop* lógica para uma lógica de *man in the loop* que interrompe o processo até um ato humano. Esta seria uma das formas de respeitar o princípio de desativação em<sup>34</sup> caso de: falta de dados, ameaça desconhecida, ambiente propício enganos, ...

#### 4.3.3. O lugar do acaso

O espírito humano tem alta capacidade de reagir a um evento extraordinário que nunca encontrou antes, graças em particular à combinação sistema 2 / sistema 1. Entretanto, talvez

---

<sup>33</sup> É o caso do míssil ASTER da empresa MBDA que tem em sua operação nominal um uplink que permite que o míssil seja destruído em voo.

<sup>34</sup> Cf para 4.2

seja um dos pontos fracos da IA e, em particular, das técnicas de *Deep Learning*, o fato de que elas só são desenhadas com dados do passado. A capacidade de "generalizar" dela é inevitavelmente impactada quando ocorre um evento "fora do comum". Diante dos eventos trazidos à luz pela teoria do "Cisne Negro"<sup>35</sup> de Nassim Nicholas Taleb, a IA, assim como as organizações que dependem demais dela, teriam dificuldade em reagir. De fato, as organizações alimentadas por algoritmos calibrados para um tipo de ambiente são eficientes e otimizadas para esse ambiente. Ao contrário, são frágeis diante de qualquer transformação radical desta última (KOENIG, 2019).

Num contexto de guerra, a incerteza das decisões militares não é apenas a "névoa da guerra" mencionada acima. Os conflitos armados estão constantemente gerando novas formas de guerra que se alimentam da interação e da inventividade dos adversários. Sua forma também depende de condições sociais, políticas e econômicas sobre as quais o líder militar não tem controle. O tomador de decisão na guerra deve, portanto, deixar espaço para o acaso. Logo após definir a névoa da guerra, Clausewitz escreve como a terceira peculiaridade da mente do líder:

*Le défaut de visibilité qu'entraîne cet éclairage affaibli doit être compensé par le talent de divination, ou abandonné à la chance. En l'absence de sagesse objective, il faut donc, là encore, se fier au talent, voire à la faveur du hasard.*<sup>36</sup> (CLAUSEWITZ, 1955, p. 133).

Diante do acaso, é precisamente o instinto que nos permite continuar a decidir, mas estamos longe do advento de um instinto de síntese moldado pela IA. É aqui que o Homem retém uma vantagem definitiva sobre o sistema 3 que é a IA. O valor do instinto guerreiro é celebrado por todos os grandes líderes da guerra. Em *Le Fil de l'épée*, o General de Gaulle escreveu sobre este assunto: "Les grands hommes de guerre ont toujours eu, d'ailleurs, conscience du rôle et de la valeur de l'instinct. Ce qu'Alexandre appelle « son espérance »,

<sup>35</sup> Os "Cisnes Negros" são eventos extraordinários que mudam profundamente as suposições comumente aceitas até então. Eles geralmente constituem uma mudança de paradigma.

<sup>36</sup> "A falta de visibilidade causada por esta iluminação debilitada deve ser compensada pelo talento da adivinhação, ou deixada ao acaso. Na ausência de sabedoria objetiva, deve-se, portanto, mais uma vez, contar com o talento, ou mesmo com o acaso" (Tradução Nossa)

César « sa fortune », Napoléon « son étoile »<sup>37</sup> (DE GAULLE, 2015, p. 29). A IA não será, portanto, obrigada a mostrar instinto, mas será obrigada a não impedir que o tomador de decisões militares mostre seu instinto, pois será sempre de grande ajuda para ele em certas circunstâncias.

---

<sup>37</sup> “Os grandes homens de guerra sempre estiveram conscientes do papel e do valor do instinto. O que Alexandre chamou de sua "esperança", César de "sua fortuna", Napoleão de "sua estrela”. (Tradução Nossa)

## 5. CONCLUSÃO

Decidir é a principal função do líder militar. Diante da incerteza, ele será sempre solicitado a arbitrar a parte de suas decisões baseadas em elementos factuais, o fruto do raciocínio - possivelmente preparado por outros - e a parte de suas decisões baseadas em seu instinto, sua bússola diante do acaso.

No contexto da guerra naval, no mar, estas decisões sempre foram tomadas em um ambiente que associa uma organização humana, a tripulação, com os meios técnicos e materiais que compõem o navio de guerra. Neste sentido, o elemento humano que é o marinheiro já está "aumentado". Ele evolui em um ambiente hostil que não é natural para ele. Só o domínio tecnológico dele lhe permite sobreviver. Isto permanece verdadeiro seja qual for a embarcação, desde navios patrulha até porta-aviões nucleares. Os comandantes de navios de guerra ou forças navais tomam decisões, portanto, considerando constantemente as interações entre as pessoas e as ferramentas de trabalho deles. Os marinheiros desempenham um papel central na eficácia da ferramenta, mas a ferramenta, por sua vez, impõe suas limitações intrínsecas a eles. Os oficiais navais tomam essas decisões multifatoriais, graças às habilidades cognitivas que são o resultado de um longo processo de aprendizagem e experiência, o que torna ainda mais valioso ter uma carreira e, portanto, envelhecer dentro de uma instituição como a marinha. À luz da teoria de Kahneman, podemos ver como o instinto, emanado do Sistema 1, está relacionado às laboriosas horas passadas estudando a técnica, aprendendo, preparando a ação e depois tirando conclusões uma vez que ela acabou. Naqueles momentos, o Sistema 2 está em funcionamento na preparação do Sistema 1. Esses constantes ir e voltar entre considerações racionais, relatórios, briefings, considerações técnicas e a percepção instintiva de coisas como o moral da tripulação, a forma da ondulação ou o ruído ao redor, constituem um continuum, uma espécie de ruminação interior permanente num comandante no mar.

O advento das técnicas de IA mudou este contexto inicial. Isto se deve à autonomia da IA, à sua concepção baseada no suposto funcionamento do cérebro humano, à sua capacidade de aprendizagem, ao seu domínio potencial de um novo campo de batalha, cheio de dados e informações digitais. Estas qualidades de IA dão um novo lugar ao aspecto técnico na ruminação interna do decisor militar no mar. A IA irá interagir com suas decisões em um grau provavelmente maior do que qualquer outra técnica até agora inventada.

Esta evolução é irreversível. Isso se deve à explosão da quantidade de dados digitais disponíveis e ao fato de que a IA se desenvolve primeiro no mundo civil fora do controle do mundo da defesa. A capacidade de gerenciamento de IA dessas quantidades fenomenais de dados faz com que seja um campo irresistível de investimento tecnológica para empresas civis. O mesmo vale para as bases industriais tecnológicas de defesa nacional que encontram aí oportunidades de inovação e diferenciação.

O presente trabalho se concentrou em mostrar que a IA, considerada da perspectiva da teoria de Kahneman, interage com ambos os sistemas de pensamento humano como um sistema 3 que os aprimoraria. Este aumento aplicado a um contexto de força letal na tomada de decisões no mar oferece perspectivas interessantes para melhorar a eficácia operacional das unidades militares. No entanto, este aumento precisa ser controlado, pois traz consigo o risco de decisões humanas inconscientes serem deixadas para os vieses técnicos de operação do sistema 3.

Portanto, devem ser estabelecidos requisitos para estas técnicas de IA que incorporem os princípios de transparência do algoritmo, qualificação do sistema, evolução rápida, possível desativação e compartilhamento e cooperação. Estas parecem ser as condições prévias para uma "inteligência artificial confiável" que respeite o direito internacional do conflito armado e a ética do soldado.

Mas, mais do que uma simples exigência de desenho, o mundo da defesa e as forças armadas devem contribuir desenvolvendo sistemas de armas baseados em IA a serem

perseguidos de acordo com uma visão estratégica de "singularidade militar"<sup>38</sup>. Esta noção, explicada pelo Chefe do Estado-Maior das Forças Armadas Francesas, General François Lecointre, explica todas as facetas da relação entre as forças armadas francesas e a sociedade de onde ela provém. Longe das fantasias do "robô-assassino", trata-se da contribuição das forças armadas para o desenvolvimento das armas, integrando a IA, que fazem parte de uma continuidade histórica e cultural que faz a força e a humanidade delas.

---

<sup>38</sup> *Visão Estratégica Para uma Singularidade Positiva, General do Exército François Lecointre, Chefe do Estado-Maior das Forças Armadas.* Disponível em: <  
<https://www.defense.gouv.fr/english/content/download/542851/9295728/vision-statégique%281%29.pdf>  
>. Acesso em 19 de julho de 2020.



## REFERÊNCIAS

ANDERSON, Kenneth A. ; WAXMAN, Matthew C. *Law and Ethics for Autonomous Weapons Systems: Why a ban won't work and how the laws of war can.* Jean Perkins Task Force on National Security and Law, Hoover Institution, Stanford University, 2013. p. 33. Relatório. Disponível em : < [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2250126](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2250126) > Acesso em: 19 julho 2020.

CLAUSEWITZ, Carl Von. *De la Guerre.* Tradução de Denise Naville. Paris : Les Editions de Minuit. 1955. 760p. Título original : *Vom Kriege.* ISBN: 2-7073-0107-8

DE GAULLE, Charles. *Le fil de l'épée.* 4. ed. Paris: Perrin, 2015. *E-book.* (160 p.). EAN : 978-2-262-05134-1

DESSPORTES, Vincent, General. *Décider dans l'incertitude.* 2. ed. Paris: Economica, 2011. *E-book.* (219 p.). EAN digital: 9782717862089

FRANÇA, Júnia Lessa et al. *Manual para normalização de publicações técnico-científicas.* 8. ed. Belo Horizonte: Ed. UFMG, 2007. 255 p. (Coleção Aprender).

GANASCIA Jean-gabriel. *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?.* Paris : Le Seuil, 2017. *E-book.* (144 p.). ISBN 978-2-02-131001-6

KAHNEMAN, Daniel. *Système 1 / Système 2: les deux vitesses de la pensée.* Paris: Flammarion, Tradução : Raymond Clarinard. 2016. *E-book.* (595 p.). Título original : *Thinking, fast and slow.* ISBN: 9782081307827

KOENIG, Gaspard. *La Fin de l'Individu. Voyage d'un philosophe au pays de l'intelligence artificielle.* 1. ed. Paris : Editions de l'Observatoire/Humensis, 2019. 400 p. ISBN: 979-10-329-0720-7

LE CUN, Yan. *Quand la machine apprend.* Paris: Odile Jacob, 2019. *E-book.* (394 p.). ISBN: 978-2-7381-4932-9

MINISTERE DES ARMEES. *L'intelligence artificielle au service de la défense.* Paris. Task Force IA, 2019. 41 p. Relatório. Disponível em : <https://www.defense.gouv.fr/content/download/566813/9761182/20200108-NP-Rapport%20de%20la%20Task%20Force%20IA%20Septembre.pdf> Acesso em: 9 julho 2020.

SCHAUB Gary ; KRISTOFFERSEN Jens Wenzel. *In, On, or Out of the Loop? Denmark and Autonomous Weapon Systems.* Centre for Military Studies, University of Copenhagen, 2017. p. 45. Relatório. Disponível em : < [https://cms.polsci.ku.dk/english/publications/in-on-or-out-of-the-loop/In\\_On\\_or\\_Out\\_of\\_the\\_Loop.pdf](https://cms.polsci.ku.dk/english/publications/in-on-or-out-of-the-loop/In_On_or_Out_of_the_Loop.pdf) > Acesso em: 19 julho 2020.

TUAL, Morgane. *L'intelligence artificielle, le grand malentendu.* *Le Monde.* Paris. 13 nov. 2019. Disponível em : < [https://www.lemonde.fr/idees/article/2019/11/13/l-intelligence-artificielle-le-grand-malentendu\\_6018956\\_3232.html](https://www.lemonde.fr/idees/article/2019/11/13/l-intelligence-artificielle-le-grand-malentendu_6018956_3232.html) > Acesso em 15 junho 2020.