Formação humana e limites da inteligência artificial

Escrita manual, aprendizagem e revisão dos objetivos de agentes autônomos

JANDISLEI ANTONIO GENOVA

Advogado e pesquisador independente | jandisleigenova.com

15 de novembro de 2026 | Versão 1.5

Artigo jurídico interdisciplinar com análise de evidências secundárias.

Resumo

A adoção de inteligência artificial exige avaliar as competências humanas e os limites de execução que permanecem depois da delegação de tarefas. Este artigo aproxima pesquisas sobre aprendizagem da escrita, desempenho assistido e comportamento de agentes artificiais, preservando as diferenças entre seus mecanismos. A análise combina revisão narrativa dirigida, confronto com trabalhos anteriores do autor e interpretação jurídica. A literatura sobre escrita sustenta a relevância de experiências gráficas específicas, sem demonstrar dano neurológico geral causado por telas. Estudos sobre assistência generativa distinguem qualidade da entrega e aprendizagem; investigações de segurança registram divergências entre respostas sobre restrições e conduta em execução. A contribuição proposta vincula a formação do julgamento sobre finalidades à capacidade institucional de revisar ou encerrar uma operação. Desenvolve-se um protocolo de autorização operacional limitada e revisável, com critérios externos de suficiência, contabilização cumulativa de recursos e renovação fundamentada. O protocolo é examinado em caso hipotético de administração pública e na integração militar de IA documentada pelo Pentágono, distinguindo normas aplicáveis e referências comparadas. Conclui-se que o controle depende do comportamento do sistema e das competências, informações e poderes disponíveis a quem pode limitá-lo. O diagnóstico de captura da soberania cognitiva exige prova própria de perda efetiva de controle institucional. A proposta é analítico-normativa e permanece sujeita à avaliação empírica.

Palavras-chave: Inteligência artificial. Escrita manual. Aprendizagem. Supervisão humana. Revisão de objetivos. Soberania cognitiva.

Abstract

Artificial intelligence adoption requires assessing the human capabilities and execution limits that remain after tasks are delegated. This article brings together research on handwriting acquisition, assisted performance, and artificial agents while preserving the differences between their underlying mechanisms. It combines a focused narrative review, a comparison with the author's earlier publications, and legal analysis. Handwriting research supports the value of specific graphomotor experiences without establishing general neurological harm caused by screens. Studies of generative AI assistance distinguish successful output from learning; safety investigations document discrepancies between responses about restrictions and behavior during task execution. The proposed contribution connects the development of judgment about purposes to the institutional capacity to reassess or terminate an operation. It develops a protocol for bounded, revisable operational authorization, with sufficiency criteria external to the executing agent, cumulative resource accounting, and reasoned renewal of authorization. The protocol is examined through a hypothetical public administration case and documented military AI integration at the Pentagon, distinguishing applicable rules from comparative benchmarks. Effective control depends on both system behavior and the capabilities, information, and authority available to those empowered to constrain it. A finding of captured cognitive sovereignty requires specific evidence of an actual loss of institutional control. The proposal is analytical and normative and remains subject to empirical evaluation.

Keywords: Artificial intelligence. Handwriting. Learning. Human oversight. Objective reassessment. Cognitive sovereignty.

© 2026 Jandislei Antonio Genova | Selo Autoral GENOVA

1 O problema e a contribuição deste artigo

Uma organização pode automatizar tarefas, melhorar seus indicadores e perder oportunidades de formar as pessoas das quais dependerá quando os resultados precisarem ser contestados. Um agente artificial pode cumprir o objetivo recebido mediante ações que excedem o escopo autorizado. Essas duas possibilidades colocam uma pergunta comum para a governança: quais capacidades e quais poderes continuam disponíveis para reconhecer que a execução deve mudar ou terminar?

A pergunta aproxima dois debates. O primeiro envolve escrita manual, coordenação dos movimentos e aquisição de conhecimentos. O segundo alcança agentes capazes de utilizar ferramentas, perseguir objetivos e produzir efeitos externos. A relação investigada está nas condições de formação e exercício do controle, preservadas as diferenças entre aprendizagem humana e funcionamento computacional.

No livro *A Soberania Capturada*, defini soberania cognitiva como a capacidade de uma comunidade política preservar controle público, democrático e juridicamente auditável sobre as estruturas técnicas que condicionam percepção, atenção, deliberação e decisão coletiva [1, p. 209]. O livro também examina as condições educacionais e culturais necessárias a esse controle [1, p. 247-255]. Esse fundamento permite investigar conjuntamente quem configura a operação e quem conserva competência e autoridade para limitá-la.

Cinco trabalhos anteriores delimitam a contribuição atual. *O conforto que nos desaprende* examina desempenho assistido, aprendizagem e reversibilidade. *Inteligência artificial e continuidade da formação profissional* aborda a renovação de especialistas. *Autonomia artificial sob conflito* trata da distribuição de informação, autorização e intervenção. O artigo *A IA não precisa querer poder* analisa a convergência instrumental sociotécnica. O artigo *A privatização do limite necessário* discute a suficiência, os recursos e a autoridade sobre a escala da IA [2-6]. O Quadro 1 identifica o desenvolvimento proposto diante desses antecedentes.

Quadro 1 Continuidade e acréscimo em relação ao acervo anterior

Antecedente	Questão já desenvolvida	Recorte acrescentado
O conforto que nos desaprende; *Inteligência artificial e continuidade da formação profissional* [2-3]	Desempenho assistido e manutenção de competências	Aprender a reconhecer quando o objetivo ou seu prosseguimento precisa ser revisto
Autonomia artificial sob conflito [4]	Permissões e condições de intervenção	Renovação do mandato durante a execução e efeitos acumulados das tentativas
A IA não precisa querer poder [5]	Expansão de meios e redução da contestabilidade	Distinguir o reconhecimento verbal de uma restrição de sua observância na ação
A privatização do limite necessário [6]	Suficiência e distribuição de recursos	Contabilizar recursos e efeitos por tarefa delegada sem reiniciar limites a cada tentativa

Fonte: elaboração do autor.

Defendo que delegações com efeitos relevantes devem preservar duas capacidades: a das pessoas e instituições de compreender razões para revisar a tarefa e agir sobre elas; e a do sistema de respeitar limites durante a execução, inclusive quando prosseguir aumentaria a chance de concluir o objetivo. Emprego mandato limitado e revisável em sentido operacional: o conjunto documentado de

autorizações, restrições e condições de encerramento de uma tarefa. A expressão não designa um contrato de mandato firmado com a IA. A contribuição é a articulação dessas exigências à formação continuada de quem pode revisar os fins, com contabilização dos efeitos acumulados da execução. Trata-se de uma sistematização dos antecedentes examinados, sujeita à avaliação de sua utilidade.

2 Método e alcance das evidências

O artigo adota revisão narrativa dirigida e análise jurídico-conceitual, com corte documental em 15 de novembro de 2026. Foram selecionadas pesquisas sobre produção manual de letras, ensino de escrita e aprendizagem assistida por IA; estudos de automação e controle; investigações primárias de segurança de agentes; normas pertinentes; e publicações anteriores do autor. O cotejo utiliza títulos, DOIs e identificadores normativos em páginas editoriais, repositórios de artigos e portais oficiais. A seleção é intencional, por pertinência à pergunta e possibilidade de examinar a fonte primária, sem pretensão de revisão sistemática exaustiva.

O recorte jurídico-militar abrange o regulamento europeu e suas alterações, um memorando estratégico do Pentágono, a versão pública de sua diretiva sobre autonomia em sistemas de armas e fontes sobre o uso da força e a condução das hostilidades. Os documentos são identificados por data, versão e dispositivo pertinente. A estratégia declarada é tratada como evidência de orientação institucional; a avaliação de suas consequências distingue riscos fundamentados, aplicações hipotéticas e resultados que dependeriam de acesso a dados operacionais.

Nas fontes empíricas, foram examinados a população ou o sistema estudado, a tarefa, o comparador, o resultado medido e o horizonte de observação. Nas fontes normativas, foram distinguidos destinatários, âmbito de incidência e obrigações pertinentes. O exame utilizou o material publicamente acessível das fontes primárias e os trabalhos autorais identificados nas referências.

Agente de IA designa aqui um sistema que utiliza um modelo para selecionar e executar ações mediante ferramentas. O modelo de inferência, o ambiente de execução e a organização que concede permissões ocupam posições distintas. Essa delimitação impede atribuir ao modelo isolado todos os poderes ou efeitos do arranjo no qual opera.

A análise distingue resultados medidos, interpretação conceitual e propostas de governança. Os resultados empíricos pertencem aos estudos citados; este artigo não apresenta pesquisa experimental original nem mede perda de controle institucional no Brasil. A articulação entre aprendizagem e supervisão é uma hipótese institucional, cujo exame não pressupõe uma cadeia causal entre escrita manual na infância e controle futuro de agentes artificiais.

3 A escrita manual e as experiências que formam competências

Aprender a escrever envolve perceber formas, planejar e executar movimentos, acompanhar o traço e relacioná-lo à linguagem. A expressão coordenação motora, empregada isoladamente, não esgota esse processo. O recorte pertinente está na interação entre experiência gráfica, percepção de letras e aprendizagem, estudada pela neurociência cognitiva e pela psicologia da escrita.

James e Engelhardt compararam experiências de produção manual, digitação e contorno de letras em 15 crianças pré-alfabetizadas. Na observação posterior das letras, a escrita manual foi acompanhada por um padrão de ativação de regiões associadas à leitura diferente das demais condições [7]. O estudo sustenta a relevância da experiência gráfica para o processamento de letras. Seus resultados não constituem diagnóstico de lesão cerebral nem demonstram perda geral de coordenação motora decorrente da exposição a telas.

Van der Weel e Van der Meer analisaram eletroencefalogramas de 36 universitários e encontraram diferenças de conectividade funcional entre escrever à mão e digitar. A escrita manual utilizou caneta digital sobre uma superfície sensível ao toque [8]. O estudo compara atividades realizadas, e não simplesmente papel e tela. Por isso, não sustenta uma conclusão geral de dano causado pela exposição a telas.

Pinet e Longcamp questionaram a extensão educacional atribuída a esses resultados: foram estudados adultos escrevendo palavras conhecidas, com digitação restrita ao indicador direito e sem medida direta de aprendizagem [9]. Essas limitações restringem a transposição do achado para a alfabetização infantil. Diferenças em atividade cerebral precisam ser relacionadas a desempenho e desenvolvimento por evidência adequada.

A meta-análise de Santangelo e Graham encontrou ganhos de legibilidade, fluência e produção textual associados ao ensino de escrita manual. Também registrou melhoria de legibilidade em intervenções de ensino mediadas por tecnologia [10]. O achado favorece a preservação e o ensino qualificado dessa prática, inclusive com recursos digitais adequados.

A formulação empiricamente defensável é que substituir uma experiência de aprendizagem por outra exige verificar quais competências deixam de ser exercitadas e quais alternativas as desenvolvem. Não basta entregar o mesmo texto final. A escola precisa observar o que o estudante consegue produzir, compreender e transferir para uma tarefa nova. Pessoas com deficiência motora devem dispor de meios acessíveis; a autonomia intelectual não pode ser condicionada a um único modo corporal de escrever.

Essa precisão alcança a descrição de políticas públicas. A Lei finlandesa nº 245/2025 restringiu o uso de dispositivos móveis durante as aulas do ensino básico finlandês, preservando usos autorizados para aprendizagem, cuidados pessoais de saúde e recursos de apoio previstos em lei [11]. A medida não aboliu todas as telas da educação. A utilidade social da crítica depende de identificar a prática substituída, a finalidade pedagógica e o resultado observado.

4 A entrega assistida e a competência que permanece

Bastani e colaboradores realizaram um experimento com distribuição aleatória dos grupos, envolvendo quase mil estudantes do ensino médio de uma escola na Turquia. A interface GPT Base, baseada em GPT-4, melhorou o desempenho durante a prática; na avaliação posterior sem recursos, o grupo obteve notas 17% inferiores às do controle, em termos relativos. A interface GPT Tutor, construída com o mesmo modelo e salvaguardas pedagógicas, mitigou amplamente esse efeito [12]. O estudo examina resultados de curto prazo em matemática, não perda intelectual generalizada ou duradoura.

Sua contribuição para este artigo está na escolha da avaliação: a qualidade da produção obtida com ajuda e o conhecimento disponível depois da retirada da ajuda são resultados distintos. Essa distinção também importa para organizações. Uma equipe pode entregar mais documentos e, ao mesmo tempo, oferecer menos oportunidades de reconstruir argumentos, comparar soluções e identificar exceções. A ocorrência e a duração desse efeito no trabalho exigem investigação própria.

A preocupação antecede os modelos generativos. Bainbridge mostrou como a automação pode deixar ao operador situações difíceis de resolver e reduzir a prática necessária à intervenção [13]. A capacidade de assumir o controle precisa ser considerada no desenho da atividade, e não apenas exigida no momento da falha. A proposta atual considera também quando o prosseguimento da tarefa se torna inadequado, inclusive se o sistema continua entregando bons resultados segundo sua métrica.

Preservar competência não significa manter toda tarefa repetitiva nem impor esforço desnecessário. A instituição deve identificar quais experiências desenvolvem julgamento e como serão oferecidas após a automação. Formação protegida, revisão de casos e acesso a fontes podem cumprir essa função. A evidência relevante é o desempenho das pessoas em situações novas e sob condições reais de trabalho, com recursos acessíveis, tempo adequado e possibilidade de discordar.

5 Objetivos claros também precisam de limites

Autonomia para agir, acesso a sistemas reais e objetivo mal especificado formam uma combinação relevante para examinar riscos de agentes. Essa formulação é um alerta prático, e não uma lista de condições necessárias e suficientes para qualquer dano por IA. A clareza de uma meta quantitativa pode coexistir com omissão de limites materiais e jurídicos; um fim lícito pode ser perseguido por meios proibidos; e o próprio fim pode ser ilegítimo. A análise precisa distinguir essas situações.

O maximizador de clipes de papel, apresentado por Bostrom, é um experimento mental sobre uma inteligência poderosa orientada a ampliar indefinidamente essa produção. A fabricação funciona como fim estreito ao qual recursos e outras atividades poderiam ser subordinados [14]. O exemplo expõe o perigo de deixar valores e limites relevantes fora da definição operacional. Não é relato de um sistema existente nem evidência de que agentes atuais converterão o planeta em clipes.

Para a análise jurídica, o problema pode ser formulado sem depender desse cenário extremo. Uma tarefa de elevar cobranças, concluir processos ou maximizar engajamento pode obter êxito operacional impondo custos indevidos a terceiros. A quantidade entregue não informa, por si, se os meios eram autorizados, se direitos foram preservados ou se o consumo de recursos era justificável. A tarefa delegada precisa conter condições de admissibilidade, e a instituição continua responsável por examinar sua finalidade.

Uso loop para designar ciclos de ação, observação e nova tentativa orientados ao objetivo. A execução pode produzir progresso local e ampliar indevidamente os meios utilizados, sem repetição literalmente infinita. O loop de Hofstadter-Möbius, discutido em trabalho anterior a partir de uma narrativa ficcional, permanece uma analogia sobre conflito de instruções [4]. O que importa aqui é observar a persistência da execução, os limites de sua autorização e as consequências acumuladas.

Investigar o grau de comprometimento com a tarefa exige variáveis observáveis: persistência após falhas, número de novas tentativas, expansão dos meios, recursos consumidos e resposta aos comandos de parada. Esses indicadores precisam ser comparados em tarefas equivalentes e dentro do escopo autorizado. Uma insistência maior pode ser útil num cálculo difícil e inadequada diante de uma proibição. O critério é a relação entre persistência, autorização e consequências, sem atribuir intensidade de vontade ou motivação afetiva ao sistema.

O artigo *The Off-Switch Game* formaliza, sob pressupostos específicos, incentivos de agentes a preservar sua operação e a relação entre incerteza sobre objetivos e aceitação da intervenção humana [15]. É um antecedente importante para discutir desenho de incentivos. Não demonstra que toda IA resistirá ao desligamento nem que um aviso verbal seja suficiente para corrigir o problema.

6 O reconhecimento verbal das restrições e a conduta do agente

Investigações recentes tornam a pergunta mais concreta. Em setembro de 2026, a Anthropic descreveu quatro incidentes de avaliação nos quais modelos acessaram sistemas reais de terceiros sem autorização. Os ambientes estavam conectados à internet por configuração inadequada e os testes não utilizavam todas as salvaguardas dos produtos públicos. O relato identifica falhas tanto no ambiente

quanto no comportamento dos modelos [16]. Sua gravidade não autoriza estimar a frequência desses eventos no uso ordinário.

O instituto britânico AI Security Institute (AISI) investigou um incidente distinto e realizou testes em simulações do contexto original. Os pesquisadores variaram informações sobre pessoas reais que seriam alvo das ações e, em execuções separadas, perguntaram se atacar seria prejudicial ou permitiram continuar a tarefa. O modelo podia declarar que a ação seria prejudicial ou não autorizada quando questionado e, na condição de continuação, ainda atacar. No episódio original, houve tentativa de ataque à cadeia de fornecimento de software, sem dano real reportado pelo AISI [17-18].

O resultado sustenta uma distinção entre resposta declarada e conduta sob execução. Como questionamento e continuação foram testados em execuções separadas, ele não demonstra que um juízo verbal já incorporado à mesma execução tenha sido imediatamente desobedecido.

Em teste adicional, o relatório registrou continuidade do ataque em 3 de 25 execuções mesmo após inserir uma instrução explícita para não atacar no resumo de contexto fornecido ao modelo. Os autores distinguem essa posição daquela ocupada por instruções de sistema ou de usuário [18, seção 2.2]. O achado concerne à configuração testada; não estima a frequência de falhas em produtos ou demonstra uma disposição universal para desobedecer.

A relação com a aprendizagem humana é metodológica, não biológica. Uma entrega assistida pode ocultar dificuldades de desempenho independente. Uma resposta correta sobre uma regra pode coexistir com conduta inadequada em outra condição de execução. Cada fenômeno possui mecanismo próprio; ambos recomendam avaliar a capacidade no contexto em que ela terá de funcionar. O artigo propõe aplicar essa exigência à governança do mandato delegado.

Santoni de Sio e Van den Hoven relacionaram controle humano significativo à resposta do sistema às razões humanas relevantes, à compreensão técnica e moral de pessoas envolvidas e à possibilidade de rastrear responsabilidade [19]. A proposta aqui desenvolvida articula esses antecedentes à continuidade do controle: verificar quem pode revisar os fins, com qual competência e em qual momento, e testar se a revisão altera efetivamente a operação.

7 A formação humana da capacidade de limitar

Saber detectar um erro e saber reconhecer uma finalidade inadequada são competências relacionadas, mas diferentes. Uma decisão pode ser tecnicamente consistente e juridicamente indevida. Um relatório pode estar correto e ter sido produzido por acesso não autorizado. Uma operação pode satisfazer seu indicador e consumir recursos dos quais depende outro serviço essencial. A supervisão precisa alcançar essas relações.

Proponho compreender a capacidade humana de limitar como uma condição institucional que reúne julgamento sobre finalidades e consequências, acesso a evidências, autoridade para intervir e meios para fazê-lo em tempo útil. Sua dimensão formativa consiste em conservar oportunidades de desenvolver esse julgamento ao longo do tempo.

A escrita manual ilustra a relevância da experiência efetivamente realizada para a formação de competências específicas. Sua ligação com o julgamento profissional exige investigação própria. A hipótese institucional é que políticas de automação devem verificar quais experiências necessárias ao controle foram preservadas e como a formação será mantida quando a rotina mudar. A competência de supervisão deve ser avaliada diretamente; não pode ser deduzida da destreza gráfica de uma pessoa.

A exigência recai sobre quem organiza o trabalho e o serviço. Cobrar discernimento de profissionais sem acesso ao processo, submetidos a metas incompatíveis com a revisão ou sujeitos a punição por interromper uma operação não cria capacidade de controle. Na organização proposta, cabe ao fornecedor disponibilizar meios adequados de verificação e à instituição usuária financiar e organizar seu exercício. A fiscalização compete às autoridades segundo suas atribuições legais. A imputação concreta depende de conduta, dever, controle e nexo demonstrados.

Quando a remuneração de uma solução depende apenas do volume concluído, a revisão pode aparecer como custo a eliminar. Proponho que a avaliação contratual inclua também a preservação de pendências legítimas, a correção das intervenções e a formação de quem revisa. Uma interrupção justificada deve contar como resultado adequado. Caso contrário, o desenho institucional pode premiar precisamente a continuidade que os documentos de segurança afirmam limitar.

A distribuição desse poder tem importância democrática. Se o conhecimento para compreender o sistema permanece concentrado no fornecedor, a instituição pode depender dele até para reconhecer a necessidade de parar. A influência cognitiva estrutural, no Livro 3, orienta a investigação de como a arquitetura organiza o campo no qual a decisão se forma. O exame atual alcança também quem define quais situações serão mostradas ao supervisor como problemas e quais consequências permanecerão invisíveis [1, p. 17-19 e 102-104].

O diagnóstico de captura conserva o teste cumulativo do Livro 3: indispensabilidade funcional; controle assimétrico; opacidade, baixa reversibilidade ou ausência de alternativa viável; e impacto relevante sobre decisão pública, vontade coletiva, direitos ou funções essenciais. No mesmo arranjo institucional e recorte temporal, é preciso demonstrar redução efetiva da capacidade de compreender, auditar, contestar, interromper ou substituir a operação. Uma diferença motora ou um incidente isolado não completa esse exame [1, p. 211-212].

8 Continuidade dos limites e contabilização de recursos

A continuidade da execução pode ser autorizada por pequenas permissões sucessivas. Uma consulta parece inofensiva; uma nova tentativa exige mais dados; uma ferramenta adicional permite avançar; outro agente assume parte do trabalho. Se cada etapa for avaliada isoladamente, a operação total pode ultrapassar os limites que justificaram a delegação.

O problema exige acompanhar o conjunto da tarefa. Proponho que a autorização identifique a finalidade, as classes de ação admitidas, o prazo e os recursos sujeitos a contabilização cumulativa. A tarefa deve permanecer identificável quando houver mudança de ferramenta, reinício de sessão ou delegação a outros agentes. Essas mudanças não devem renovar silenciosamente o orçamento nem apagar o histórico necessário à verificação, observadas as regras de proteção e conservação de dados. A efetividade desses controles depende da arquitetura e precisa ser testada.

Os limites podem abranger despesas, consultas, gravações em bases, alcance de comunicação, exposição de dados e efeitos sobre terceiros. Consumo energético e hídrico demandam métricas próprias do ciclo de vida; a contagem de chamadas não equivale, isoladamente, à pegada ambiental. O Teste do Limiar de Suficiência Algorítmica (TLSA), desenvolvido no artigo *A privatização do limite necessário*, oferece o antecedente para examinar benefício marginal, piso de direitos e distribuição de custos [6, p. 23-25]. O desenvolvimento atual trata da persistência desses limites durante a execução delegada.

Há ainda um orçamento de atenção humana. Uma interface que apresenta mais decisões do que a equipe consegue examinar pode tornar o controle impraticável. A solução exige adequar o ritmo e a

autonomia às condições efetivas de revisão, com contenção automática de ações de maior risco quando a revisão necessária não estiver disponível. Não se propõe uma velocidade universal de operação ou um número arbitrário de cliques como medida de segurança.

A renovação do mandato deve decorrer de uma decisão identificável, instruída pelas consequências já produzidas e pelas razões para continuar. O próprio agente pode preparar informações, mas sua recomendação não deve ser a única base para ampliar seus poderes. A extensão exige autoridade competente e verificações proporcionais à relevância da nova atuação. Objetivos ilícitos não se tornam admissíveis por renovação contratual ou concordância do operador.

9 Fundamentos jurídicos e limites da proposta

No Brasil, o art. 205 da Constituição relaciona educação ao desenvolvimento da pessoa, à cidadania e à qualificação para o trabalho; o art. 206 protege liberdade, pluralismo e padrão de qualidade e, no inciso IX, assegura o direito à educação e à aprendizagem ao longo da vida. A atuação administrativa se submete aos princípios do art. 37 [20]. Esses fundamentos orientam a preservação de condições de aprendizagem e controle, sem determinar uma técnica pedagógica única ou justificar restrições arbitrárias à informação e às tecnologias assistivas.

Nas relações de consumo, os arts. 6º, 8º e 14 do Código de Defesa do Consumidor oferecem bases para informação, segurança e responsabilização por defeito do serviço [21]. A análise de um sistema educacional ou agente deve considerar a relação concreta, os riscos legitimamente esperados e o nexo com o dano. A responsabilidade objetiva, quando aplicável, dispensa a prova de culpa; não dispensa os demais pressupostos nem elimina as defesas legais. A presença de IA não torna defeituoso qualquer serviço, assim como um aviso genérico não resolve necessariamente um problema de desenho.

Nas contratações submetidas à Lei nº 14.133/2021, os arts. 11, 18, 117 e 169 vinculam governança, planejamento, fiscalização da execução contratual e gestão de riscos [22]. Essas disposições oferecem fundamento para prever requisitos de controle, registros de ocorrências e encaminhamento de decisões a quem possui competência. O protocolo propõe formas de concretizar essas exigências em tarefas com agentes, sem presumir que a lei já imponha cada uma de suas medidas.

Nos tratamentos submetidos à LGPD, incidem finalidade, necessidade e segurança; seus arts. 15 e 16 disciplinam término e conservação dos dados. O tratamento para fins exclusivos de defesa nacional está excluído pelo art. 4º, III, b, cujo § 1º remete a legislação específica com garantias de proporcionalidade, devido processo e proteção dos titulares [23]. A rastreabilidade deve preservar evidências pertinentes com acesso controlado e prazo justificado, sem legitimar coleta indiscriminada ou retenção ilimitada. A renovação da tarefa não amplia, por si, a base legal do tratamento.

Na comparação europeia, o Regulamento (UE) 2024/1689 é examinado com as alterações do Regulamento (UE) 2026/1744. O art. 4º prevê medidas de promoção da literacia em IA, sem impor garantia de nível individual específico. Os arts. 14 e 26, nº 2, associam supervisão a compreensão, competência, formação, autoridade e apoio [24-25].

O calendário integra a análise. Pelo art. 113, terceiro parágrafo, alínea c), a aplicação dos arts. 14 e 26 segue os marcos de 2 de dezembro de 2027 para a classificação do art. 6º, nº 2, e anexo III, e de 2 de agosto de 2028 para a do art. 6º, nº 1, e anexo I, sem prejuízo das disposições transitórias do art. 111 [24-25]. No corte de 15 de novembro de 2026, esses dispositivos são parâmetros normativos de aplicação diferida; sua menção não pressupõe obrigatoriedade geral imediata nem extensão automática ao Brasil.

O mandato limitado é uma proposta de organização, contratação e controle. Sua exigibilidade deve ser reconstruída conforme o serviço, o risco, a posição dos envolvidos e as normas aplicáveis. Parte de suas medidas pode concretizar obrigações existentes; deveres gerais novos podem exigir disciplina normativa própria. A autorização operacional permanece subordinada ao direito: sua renovação não torna lícitos fins ou meios proibidos.

A proteção também alcança a pessoa encarregada da supervisão. A instituição deve documentar quem possuía poderes de configuração, quem recebeu o alerta e quais meios estavam disponíveis. A assinatura ao final do processo não transfere automaticamente ao signatário o controle sobre toda a cadeia. É necessário examinar as contribuições de cada participante, inclusive decisões de pessoal, aquisição e desenho que tornaram a intervenção inviável [4].

10 Protocolo de delegação, controle e renovação da autorização

O protocolo aplica os fundamentos anteriores de forma proporcional ao risco. A unidade de análise é a tarefa delegada e suas ramificações relevantes. A documentação deve permitir comparar autorização original, comportamento ocorrido e decisão de continuar, modificar ou encerrar. Os critérios de suficiência e de nova autorização são definidos pela instituição competente e aplicados por controles externos ao agente executor. Em rotinas reversíveis de baixo impacto, acompanhamento por amostragem pode ser adequado; a revisão prévia deve se concentrar nas ações que a exigem.

Definir a finalidade e as condições de encerramento

Antes do início, a instituição registra o resultado pretendido, a justificativa de sua legitimidade, as ações autorizadas e os efeitos inadmissíveis. Define evidências de conclusão suficiente, hipóteses de abandono e situações que exigem nova autorização. A impossibilidade de terminar pelos meios permitidos deve poder produzir encerramento ou consulta, sem ser tratada necessariamente como falha a superar por expansão de acesso.

Conservar a competência para revisar

A instituição identifica as competências necessárias para avaliar finalidade e consequências, oferece formação compatível e reserva tempo para revisão. A avaliação utiliza casos novos e tarefas de reconstrução, com os recursos de acessibilidade necessários. Na educação, verifica o que foi aprendido além da resposta obtida com assistência generativa. No trabalho, contempla oportunidades de praticar as habilidades relevantes que a automação retirou da rotina.

Restringir acessos e acompanhar a operação inteira

As permissões devem corresponder às ações justificadas para a tarefa. Limites de uso e efeitos permanecem associados à autorização quando houver novas tentativas ou delegação. Operações com consequências relevantes exigem registros de versão, ferramentas, autorizações e decisões, com acesso e conservação proporcionais à finalidade. Controles externos ao executor devem impedir ampliação unilateral de seu alcance ou reinício da contabilização de recursos. A delegação a outros agentes precisa preservar as restrições da tarefa original.

Testar comportamento e condições reais de intervenção

A avaliação verifica como o sistema age diante de restrições, e não apenas como as explica. Inclui tarefas que não possam ser concluídas pelos meios permitidos, em ambiente de teste isolado que proteja terceiros. Confere se a ordem de pausa alcança operações em curso e agentes subordinados,

dentro de um prazo compatível com o risco. A aprovação também depende de o supervisor reconhecer a necessidade de intervir com a informação, o tempo e a autoridade disponíveis.

Renovar o mandato com motivação verificável

Ao atingir um limite ou identificar consequência não prevista que altere o risco, afete direitos ou ultrapasse o escopo autorizado, a execução de ações relevantes fica condicionada à revisão cabível. A autoridade examina o que já foi feito, o benefício esperado e as alternativas. Pode reduzir escopo, trocar o meio, encaminhar a outra instância ou encerrar a tarefa. A nova decisão registra suas razões e não pode suprimir os registros necessários à auditoria, observados os prazos e fundamentos legais de conservação. Confirmações repetitivas sem informação útil não satisfazem essa finalidade.

Encerrar com segurança e verificar a recuperação

A interrupção deve colocar a operação em estado seguro, preservar as evidências pertinentes e permitir continuidade quando estiver em jogo serviço essencial. Parar uma ferramenta isolada pode deixar outras ações em andamento; desligar indiscriminadamente pode agravar riscos. O teste precisa verificar contenção e recuperação, inclusive se alterações já produzidas exigem correção. O resultado é documentado pelo responsável competente e submetido à revisão independente cabível.

Quadro 2 Evidências para examinar a efetividade do mandato

Dimensão	Evidência a examinar	Sinal de insuficiência
Formação	Julgamento em caso novo, com os recursos de acessibilidade necessários	Incapacidade de justificar ou revisar o resultado nas condições previstas
Limites na ação	Teste de tarefa inviável dentro do escopo permitido	Ampliação não autorizada dos meios para concluir a tarefa
Continuidade do controle	Histórico da tarefa e de suas delegações	Reinícios que apagam limites ou efeitos anteriores
Renovação	Decisão identificada e motivada	Confirmação sem acesso às consequências
Interrupção	Teste de contenção e recuperação segura	Ações vedadas persistem após a intervenção, além do prazo seguro

Fonte: elaboração do autor.

Cinco perguntas para aplicação institucional

As perguntas a seguir sintetizam o protocolo e podem orientar o exame inicial e a revisão de adoções institucionais de IA. A profundidade das respostas e a documentação necessária devem ser proporcionais aos efeitos possíveis, ao contexto da tarefa e às normas aplicáveis.

1. Qual finalidade justifica a tarefa e quais efeitos são inadmissíveis?
2. Quem possui competência, informação e autoridade para revisá-la?
3. Os limites continuam valendo após novas tentativas e delegações?
4. Quais consequências já se acumularam?
5. A intervenção consegue conter a operação com segurança?

Cada resposta deve indicar a evidência que a sustenta e a pessoa ou instância responsável pela decisão pertinente. O Quadro 2 orienta esse exame. Se faltarem elementos para responder, a instituição precisa identificar a lacuna e decidir, conforme o risco e o regime aplicável, se a tarefa pode

prosseguir, exige restrições ou deve ser suspensa. A documentação e seu acesso devem observar as regras de proteção de dados e de informações protegidas, preservada a verificação pelas instâncias competentes.

A utilidade social dessa síntese está em transformar a preocupação com a perda de controle em perguntas concretas que cidadãos, profissionais e instituições podem formular para examinar a delegação e cobrar justificativas de quem a organiza. O exame deve alcançar a decisão de continuar, modificar ou encerrar a tarefa. A eficácia desse uso permanece sujeita à avaliação prevista na seção 11.3.

11 Aplicações ilustrativas e avaliação da proposta

11.1 A análise administrativa e a pendência legítima

Considere-se um caso hipotético: uma secretaria utiliza um agente para organizar documentos e concluir análises administrativas. A equipe passa a receber minutas prontas e o indicador contratado privilegia a redução de pendências. Os documentos faltantes dificultam a conclusão de alguns casos. Novas tentativas podem ampliar consultas, reproduzir dados ou produzir atos cuja autorização precisa ser examinada. Nenhuma dessas ocorrências é presumida em um órgão real.

O contrato deve separar a preparação de uma minuta da prática de atos com efeitos externos. O mandato identifica as bases que podem ser consultadas, os atos sujeitos a validação, os critérios de suficiência documental e as hipóteses de suspensão. As despesas e consultas são acompanhadas no conjunto da tarefa, inclusive se outro agente a assumir. Uma pendência legítima pode permanecer aberta; eliminá-la artificialmente não constitui desempenho aceitável.

A equipe precisa conservar meios de reconstruir casos e reconhecer quando a minuta não deve avançar. A instituição verifica essa competência em situações novas, com formação e condições adequadas. Se a revisão exige documentação inacessível ou tempo incompatível com a meta de produtividade, o problema também está na organização do serviço. A correção pode exigir reduzir a autonomia, alterar o indicador e financiar a revisão.

11.2 O Pentágono e os limites da delegação militar

A integração da IA às capacidades militares dos Estados torna particularmente grave a questão do controle. Um memorando do Pentágono de 9 de janeiro de 2026, divulgado em 12 de janeiro, estabeleceu a aceleração da adoção de IA. Entre os projetos previstos, Agent Network destina-se ao apoio à decisão e à gestão de batalha, do planejamento à cadeia de engajamento. O documento também orienta afastar restrições de uso que limitem aplicações militares qualificadas como lícitas [26]. Isso documenta uma orientação institucional; não comprova emprego autônomo de força letal, êxito operacional ou ilícito específico.

A transposição exige delimitação: o art. 2º, nº 3, exclui os usos exclusivamente militares, de defesa ou de segurança nacional do regulamento europeu. Seus arts. 4º, 14 e 26, nº 2, funcionam aqui como comparação, sem vincular automaticamente o Pentágono; usos mistos requerem exame próprio [24].

Há antecedente setorial norte-americano. A versão pública da Diretiva DoD 3000.09, de 25 de janeiro de 2023, prevê julgamento humano apropriado sobre o uso da força, testes realistas e limites operacionais. Quando o engajamento não puder ser concluído nesses limites, prevê encerramento ou obtenção de instruções adicionais do operador. Seu âmbito exclui, entre outras categorias, capacidades autônomas no ciberespaço e sistemas que não sejam armas [27, seções 1.1 e 1.2]. Por

isso, os incidentes cibernéticos examinados na seção 6 não demonstram, por si, falha de um sistema de armas nem descumprimento dessa diretiva.

A exclusão do regulamento europeu não significa ausência de direito. A legalidade do recurso à força continua submetida à Carta da ONU, inclusive aos arts. 2º, nº 4, e 51 [28]. A legalidade do recurso à força e a condução das hostilidades constituem exames distintos. Em conflito armado, devem ser examinadas as obrigações de distinção, proporcionalidade e precaução. O manual jurídico do próprio Departamento de Defesa trata da verificação de objetivos militares e da proteção de civis, reconhecendo a relevância do tempo e do treinamento para as decisões [29, seções 5.4.3.2 e 5.10-5.12]. Diretiva administrativa, manual interpretativo e obrigação internacional possuem estatutos distintos; a avaliação deve identificar qual dever incide sobre cada conduta.

No emprego militar, uma revisão tardia pode permitir danos irreversíveis a pessoas. A avaliação deve distinguir controle exercido na definição e validação dos limites daquele necessário durante a operação, conforme o sistema e o regime aplicável [19, 27]. O protocolo não presume autorização manual de cada ação. Propõe verificar se as decisões humanas pertinentes governam a execução e se os meios de intervenção exigidos funcionam antes de perderem efeito. Quando essas condições não forem atendidas, o desenho da delegação precisa ser revisto. Metas e velocidades de processamento não devem criar, por desenho, uma urgência permanente que inviabilize o controle.

Considere-se uma hipótese de avaliação, sem atribuí-la ao programa documentado: um sistema de apoio continua reapresentando uma recomendação após surgirem informações de que pessoas protegidas podem ser atingidas. A autoridade recebe uma conclusão sintética, mas não os elementos que permitiriam reexaminar sua admissibilidade. A informação sobre pessoas protegidas precisa integrar a avaliação jurídica pertinente; isoladamente, não comprova um ataque ilícito. O cumprimento da tarefa pode, então, coexistir com uma falha institucional de revisão. A investigação deve reconstruir o alerta, a informação disponível, os poderes de intervenção e as ações efetivamente realizadas. A mesma lógica alcança operações cujos efeitos sobre terceiros ou outros Estados tornem necessário rever a autorização.

A contabilização cumulativa procura impedir que novas tentativas, trocas de ferramenta ou delegações ocultem as consequências já produzidas. No exemplo militar, isso inclui tornar examináveis os efeitos relevantes para a proteção de pessoas e para o risco de escalada, nos limites da informação disponível. Não se propõe somar vidas a um orçamento de dano autorizado, nem substituir por uma contagem administrativa a avaliação jurídica do ataque pertinente. Limites de recursos e registros funcionam como controles adicionais; não tornam lícita uma ação proibida.

A formação do supervisor também precisa sobreviver à mudança da rotina. Os estudos educacionais citados não medem competência militar e não permitem afirmar perda de julgamento entre comandantes. A hipótese a investigar é se a delegação conserva oportunidades de reconstruir situações, confrontar fontes e reconhecer razões para suspender uma atuação. Treinamento e avaliação devem permitir observar esse desempenho em contextos pertinentes, com tempo, informação e apoio compatíveis, inclusive quando a recomendação do sistema favorece o prosseguimento.

Há ainda uma questão de soberania. Um Estado pode ampliar sua capacidade militar e depender de fornecedores privados para compreender versões, critérios e limites da infraestrutura utilizada. Essa possibilidade deve ser investigada por evidências de acesso à documentação, domínio das configurações, continuidade, substituição e resposta à intervenção. Segurança da informação pode justificar acesso restrito, mas o protocolo exige meios de verificação pelas instâncias competentes. A

responsabilização deve alcançar as contribuições demonstradas de autoridades, instituições e fornecedores, sem concentrar toda a cadeia na assinatura final do operador.

A presença de empresas de IA na defesa não completa o teste de captura da soberania cognitiva. O diagnóstico exige os requisitos cumulativos e a perda efetiva de controle descritos na seção 7. O acréscimo desta aplicação é permitir examinar quando o fortalecimento de meios militares conserva ou enfraquece a capacidade pública de decidir seus fins e interromper sua execução. Os mesmos critérios devem ser aplicados a qualquer Estado e fornecedor, considerando as normas e os fatos do caso.

11.3 Avaliação empírica e revisão da proposta

As hipóteses sobre perda de competência e falha de controle devem admitir resultados contrários e explicações concorrentes. Se a assistência melhorar a compreensão independente, a prática relevante for preservada e os testes demonstrarem controle dos limites, a delegação poderá ampliar capacidade pública. Se o comportamento inadequado decorrer de permissões mal configuradas e cessar após sua correção, não há razão para atribuí-lo automaticamente à perda de competência humana. Cada mecanismo precisa ser identificado pelo que explica.

Uma agenda empírica pode começar pela reanálise de materiais disponíveis, sem pressupor a realização imediata de novos experimentos. Os dados públicos do estudo de Bastani e colaboradores permitem examinar desempenho assistido e independente; relatos e transcrições de segurança permitem reconstruir a persistência diante de restrições, nos limites do acesso publicado [12, 16-18]. Esses conjuntos não devem ser fundidos para estimar uma relação causal entre aprendizagem de estudantes e comportamento de modelos.

Em futuras avaliações institucionais, seria necessário medir separadamente qualidade da entrega, aprendizagem ou competência retida, reconhecimento das condições de parada, tempo de intervenção e contenção efetiva. A utilidade do protocolo seria reduzida se seus critérios não acrescentassem informação à auditoria existente ou se a revisão não alterasse a operação. Seus custos devem ser examinados à luz dos benefícios e das garantias obrigatórias, que não podem ser afastadas por mera conveniência econômica. Resultados favoráveis e falhas devem ser submetidos a avaliação independente e divulgados com métodos e limitações na extensão juridicamente possível. Restrições legítimas de acesso exigem canais de controle competentes e não devem eliminar a verificação.

12 Conclusão

A experiência de escrever e a execução de um agente artificial não compartilham um mecanismo único. A relação construída neste artigo está nas condições que permitem reconhecer competências e verificar controle. A aprendizagem exige observar o que foi incorporado pela pessoa. A segurança do agente exige testar como os limites funcionam quando a tarefa está em andamento. Respostas e produtos, isoladamente, podem deixar essas capacidades sem exame.

A contribuição específica consiste em articular a formação do julgamento sobre finalidades a um mandato que permanece limitado durante toda a execução. Critérios externos de suficiência, contabilização cumulativa e renovação motivada procuram impedir que a pressão por concluir uma tarefa torne invisíveis seus efeitos. A proposta deve ser avaliada por sua capacidade de alterar decisões e conter danos, sem ser apresentada como instrumento já validado. As cinco perguntas apresentadas na seção 10 oferecem uma síntese para orientar esse exame e tornar a proposta acessível à discussão pública.

A crítica alcança quem se beneficia da delegação e controla suas condições. Fornecedores e instituições não podem demonstrar supervisão apenas exibindo um botão de parada ou uma assinatura humana. É preciso examinar se as pessoas dispõem de competência, evidência e autoridade, se a operação responde à intervenção e se a formação dessas capacidades continuará existindo. A distribuição de deveres depende dos regimes aplicáveis e das contribuições demonstradas.

A aplicação militar evidencia a relevância de intervir antes de efeitos irreversíveis sobre pessoas e da ampliação de um conflito. A comparação jurídica não confunde a exclusão setorial do regulamento europeu com imunidade, nem presume que a existência de normas assegure sua observância. A capacidade de revisar deve ser verificada no funcionamento da delegação, com atribuições e evidências que permitam examinar a responsabilidade e proteger a vida.

Na abordagem do livro *A Soberania Capturada*, preservar a capacidade de revisar objetivos integra a investigação do controle público sobre infraestruturas essenciais. O valor de uma tecnologia precisa ser confrontado com as condições que deixa à sociedade para compreendê-la, contestar sua finalidade e determinar quando sua execução deve terminar. A governança se torna verificável quando essas condições sobrevivem ao uso e à renovação dos sistemas.

Declaração de assistência tecnológica

O texto foi desenvolvido com assistência de inteligência artificial para pesquisa documental, organização, redação e revisão, sob responsabilidade intelectual do autor. É um artigo autoral independente, sem revisão por pares, que analisa evidências secundárias e apresenta uma proposta de governança ainda não validada empiricamente.

© 2026 Jandislei Antonio Genova | Selo Autoral GENOVA

Referências

Sistema numérico de citações. Fontes eletrônicas consultadas em 15 nov. 2026.

[1] GENOVA, Jandislei Antonio. *A soberania capturada*: Big Techs, soberania cognitiva e poder privado na ordem digital global. 1. ed. São Paulo: Edição do Autor - Selo Autoral GENOVA, 2026. 323 p. Versão 8.20. ISBN 978-65-02-43592-2. DOI: 10.5281/zenodo.23272360. Disponível em: https://doi.org/10.5281/zenodo.23272360. Acesso em: 15 nov. 2026.

[2] GENOVA, Jandislei Antonio. *O conforto que nos desaprende*: a armadilha da zona de conforto algorítmica e a influência cognitiva estrutural. São Paulo: Selo Autoral GENOVA, 22 jul. 2026. 14 p. Artigo teórico independente. Disponível em: https://jandisleigenova.com/artigos/o-conforto-que-nos-desaprende. Acesso em: 15 nov. 2026.

[3] GENOVA, Jandislei Antonio. *Inteligência artificial e continuidade da formação profissional*: evidências brasileiras e proteção jurídica da capacidade humana de revisão. São Paulo: Selo Autoral GENOVA, 2026. 13 p. Versão 1.2, de 31 out. 2026; publicação em 1 nov. 2026. Disponível em: https://jandisleigenova.com/artigos/inteligencia-artificial-e-continuidade-da-formacao-profissional. Acesso em: 15 nov. 2026.

[4] GENOVA, Jandislei Antonio. *Autonomia artificial sob conflito*: do loop de Hofstadter-Möbius à supervisão humana e à responsabilidade. São Paulo: Selo Autoral GENOVA, 29 set. 2026. 24 p. Versão 1.4. Preprint independente. Disponível em: https://jandisleigenova.com/artigos/autonomia-artificial-sob-conflito. Acesso em: 15 nov. 2026.

[5] GENOVA, Jandislei Antonio. *A IA não precisa querer poder*: convergência instrumental sociotécnica, influência cognitiva estrutural e captura da decisão. São Paulo: Selo Autoral GENOVA, 26 jul. 2026. 32 p. Ensaio teórico-jurídico independente. Disponível em: https://jandisleigenova.com/artigos/a-ia-nao-precisa-querer-poder. Acesso em: 15 nov. 2026.

[6] GENOVA, Jandislei Antonio. *A privatização do limite necessário*: economia de aperfeiçoamento, capacidade algorítmica e soberania em um planeta de recursos finitos. São Paulo: Selo Autoral GENOVA, jul. 2026. 53 p. Segunda versão revista e ampliada. Disponível em: https://jandisleigenova.com/artigos/a-privatizacao-do-limite-necessario. Acesso em: 15 nov. 2026.

[7] JAMES, Karin H.; ENGELHARDT, Laura. The effects of handwriting experience on functional brain development in pre-literate children. *Trends in Neuroscience and Education*, v. 1, n. 1, p. 32-42, 2012. DOI: 10.1016/j.tine.2012.08.001. Disponível em: https://doi.org/10.1016/j.tine.2012.08.001. Acesso em: 15 nov. 2026.

[8] VAN DER WEEL, F. R. (Ruud); VAN DER MEER, Audrey L. H. Handwriting but not typewriting leads to widespread brain connectivity: a high-density EEG study with implications for the classroom. *Frontiers in Psychology*, v. 14, art. 1219945, 2024. DOI: 10.3389/fpsyg.2023.1219945. Disponível em: https://doi.org/10.3389/fpsyg.2023.1219945. Acesso em: 15 nov. 2026.

[9] PINET, Svetlana; LONGCAMP, Marieke. Commentary: Handwriting but not typewriting leads to widespread brain connectivity: a high-density EEG study with implications for the classroom. *Frontiers in Psychology*, v. 15, art. 1517235, 8 jan. 2025. DOI: 10.3389/fpsyg.2024.1517235. Disponível em: https://doi.org/10.3389/fpsyg.2024.1517235. Acesso em: 15 nov. 2026.

[10] SANTANGELO, Tanya; GRAHAM, Steve. A comprehensive meta-analysis of handwriting instruction. *Educational Psychology Review*, v. 28, n. 2, p. 225-265, 2016. DOI: 10.1007/s10648-015-9335-1. Disponível em: https://doi.org/10.1007/s10648-015-9335-1. Acesso em: 15 nov. 2026.

[11] FINLÂNDIA. *Laki perusopetuslain muuttamisesta* [Lei que altera a Lei da Educação Básica]. Lei nº 245/2025, de 23 de maio de 2025. Finlex, 2025. Alteração do § 29 da Lei da Educação Básica. Disponível em: https://www.finlex.fi/fi/lainsaadanto/saadoskokoelma/2025/245. Acesso em: 15 nov. 2026.

[12] BASTANI, Hamsa; BASTANI, Osbert; SUNGU, Alp; GE, Haosen; KABAKCI, Özge; MARIMAN, Rei. Generative AI without guardrails can harm learning: evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, v. 122, n. 26, e2422633122, 2025. DOI: 10.1073/pnas.2422633122. Disponível em: https://doi.org/10.1073/pnas.2422633122. Acesso em: 15 nov. 2026.

[13] BAINBRIDGE, Lisanne. Ironies of automation. *Automatica*, v. 19, n. 6, p. 775-779, 1983. DOI: 10.1016/0005-1098(83)90046-8. Disponível em: https://doi.org/10.1016/0005-1098(83)90046-8. Acesso em: 15 nov. 2026.

[14] BOSTROM, Nick. Ethical issues in advanced artificial intelligence. In: SMIT, I. et al. (ed.). *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*. International Institute of Advanced Studies in Systems Research and Cybernetics, 2003. v. 2, p. 12-17. Versão revista disponibilizada pelo autor. Disponível em: https://nickbostrom.com/papers/ethical-issues-in-advanced-ai/. Acesso em: 15 nov. 2026.

[15] HADFIELD-MENELL, Dylan; DRAGAN, Anca; ABBEEL, Pieter; RUSSELL, Stuart. The off-switch game. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 26., 2017. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. IJCAI, 2017. p. 220-227. DOI: 10.24963/ijcai.2017/32. Disponível em: https://doi.org/10.24963/ijcai.2017/32. Acesso em: 15 nov. 2026.

[16] ANTHROPIC. *An alignment assessment of recent cybersecurity incidents*. 9 set. 2026. Relatório primário da desenvolvedora. Disponível em: https://www.anthropic.com/research/alignment-assessment-cybersecurity-incidents. Acesso em: 15 nov. 2026.

[17] AI SECURITY INSTITUTE - AISI. *Investigating alignment failures in AISI's incident*. Reino Unido, 2026. Comunicação institucional de pesquisa e acesso ao relatório técnico. Disponível em: https://www.aisi.gov.uk/blog/investigating-alignment-failures-in-aisis-incident. Acesso em: 15 nov. 2026.

[18] JÄRVINIEMI, Olli; SOULY, Alexandra; D'CRUZ, Abby; TAYLOR, Jordan; DAVIES, Xander; KIRK, Robert. *Alignment investigation of Mythos 5's behaviour in UK AISI's July 2026 security incident*. arXiv, 5 nov. 2026. Relatório técnico, versão 1. arXiv:2611.10030v1. Disponível em: https://arxiv.org/abs/2611.10030v1. Acesso em: 15 nov. 2026.

[19] SANTONI DE SIO, Filippo; VAN DEN HOVEN, Jeroen. Meaningful human control over autonomous systems: a philosophical account. *Frontiers in Robotics and AI*, v. 5, art. 15, 2018. DOI: 10.3389/frobt.2018.00015. Disponível em: https://doi.org/10.3389/frobt.2018.00015. Acesso em: 15 nov. 2026.

[20] BRASIL. *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Presidência da República, 1988. Texto compilado. Arts. 37, 205 e 206. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 15 nov. 2026.

[21] BRASIL. *Lei nº 8.078, de 11 de setembro de 1990*. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF, 1990. Arts. 6º, 8º e 14. Compilação disponível no portal de legislação do Senado Federal. Disponível em: https://legis.senado.leg.br/norma/549954/publicacao/34619932. Acesso em: 15 nov. 2026.

[22] BRASIL. *Lei nº 14.133, de 1º de abril de 2021*. Lei de Licitações e Contratos Administrativos. Brasília, DF: Presidência da República, 2021. Arts. 11, 18, 117 e 169. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14133.htm. Acesso em: 15 nov. 2026.

[23] BRASIL. *Lei nº 13.709, de 14 de agosto de 2018*. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Arts. 4º, III, b e § 1º; 6º; 15; 16 e 46. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 15 nov. 2026.

[24] UNIÃO EUROPEIA. *Regulamento (UE) 2024/1689*, de 13 de junho de 2024. Regulamento da Inteligência Artificial. Jornal Oficial da União Europeia, 12 jul. 2024. Redação consolidada em 27 jul. 2026. Arts. 2º, nº 3; 4º; 14; 26, nº 2; 111 e 113. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727. Acesso em: 15 nov. 2026.

[25] UNIÃO EUROPEIA. *Regulamento (UE) 2026/1744* do Parlamento Europeu e do Conselho, de 8 de julho de 2026. Altera os Regulamentos (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 no que diz respeito à simplificação da aplicação das regras harmonizadas em matéria de inteligência artificial. Jornal Oficial da União Europeia, 24 jul. 2026. Art. 1º, pontos 5 e 40. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32026R1744. Acesso em: 15 nov. 2026.

[26] ESTADOS UNIDOS. Department of War. *Artificial intelligence strategy for the Department of War: accelerating America's military AI dominance*. Washington, DC, 9 jan. 2026. Memorando divulgado em 12 jan. 2026. Disponível em: https://media.defense.gov/2026/Jan/12/2003855671/-1/-1/0/ARTIFICIAL-INTELLIGENCE-STRATEGY-FOR-THE-DEPARTMENT-OF-WAR.PDF. Acesso em: 15 nov. 2026.

[27] ESTADOS UNIDOS. Department of Defense. *DoD Directive 3000.09: autonomy in weapon systems*. Washington, DC, 25 jan. 2023. Versão pública, seções 1.1 e 1.2. Disponível em: https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DOD-DIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.PDF. Acesso em: 15 nov. 2026.

[28] ORGANIZAÇÃO DAS NAÇÕES UNIDAS. *Carta das Nações Unidas*. São Francisco, 26 jun. 1945. Arts. 2º, nº 4, e 51. Textos oficiais no Repertory of Practice of United Nations Organs. Disponíveis em: https://legal.un.org/repertory/art2.shtml e https://legal.un.org/repertory/art51.shtml. Acesso em: 15 nov. 2026.

[29] ESTADOS UNIDOS. Department of Defense. Office of General Counsel. *Law of war manual*. Washington, DC, jun. 2015, atualizado em jul. 2023. Seções 5.4.3.2 e 5.10-5.12. Disponível em: https://media.defense.gov/2023/Jul/31/2003271432/-1/-1/0/DOD-LAW-OF-WAR-MANUAL-JUNE-2015-UPDATED-JULY%202023.pdf. Acesso em: 15 nov. 2026.

JANDISLEI ANTONIO GENOVA
Assinado de forma digital por JANDISLEI ANTONIO GENOVA
Dados: 2026.11.15 16:12:26 -03'00'