

Quando o humano se torna excedente

Arendt, algoritmos e a soberania capturada

When Human Beings Become Functionally Superfluous: Arendt, Algorithms, and Captured Sovereignty

Jandislei Antonio Genova

Advogado e pesquisador independente. Especialista em Direito Civil e em Gestão em Saúde.

Versão pública 1.0 — 7 de agosto de 2026

RESUMO

O artigo pergunta quando uma classificação mediada por sistemas algorítmicos deixa de organizar a ação pública e passa a tratar pessoas como custos, riscos, recursos ou obstáculos removíveis. A partir de Hannah Arendt, propõe a **excedência funcional** como categoria relacional. Seu protocolo diagnóstico, o **Limiar da Excedência Funcional (LEF)**, exige evidência convergente de finalidade dominante, classificação redutiva, capacidade operacional, acesso material e ineficácia de limites jurídico-institucionais não subordinados à métrica. Um conjunto intencional de calibração aplica o LEF a SyRI, Robodebt e ao uso de gasto em saúde como proxy de necessidade clínica. Os resultados diferem: Robodebt configura o caso mais forte; SyRI revela risco elevado antes do julgamento e interrupção prospectiva por controle judicial; e o estudo de saúde demonstra redução por proxy, mas não documenta o limiar completo. O desenho explicita assimetrias documentais, recortes temporais, contraprovas e uma regra contra dupla contagem de evidências. A reconstrução circunscrita da Aktion T4 funciona apenas como advertência histórica, sem equivalência com a governança contemporânea. A contribuição consiste em tornar o diagnóstico contestável e capaz de resultados negativos ou indeterminados, além de diferenciá-lo de auditorias de *fairness*, devido processo tecnológico, algocracia e carga administrativa. Como risco institucional de segunda ordem, a **soberania capturada** é examinada pelo contraste entre Horizon/Post Office e Microsoft 365/Comissão Europeia. Disso decorrem salvaguardas de contestação, supervisão institucional, rastreabilidade, reversibilidade, contratação pública e avaliação de impacto sobre direitos fundamentais.

Palavras-chave: Hannah Arendt; governança algorítmica; excedência funcional; direitos fundamentais; soberania capturada.

ABSTRACT

This article asks when algorithmically mediated classification ceases to organize public action and begins to treat people as removable costs, risks, resources, or obstacles. Drawing on Hannah Arendt, it develops **functional superfluity** as a relational category. Its qualitative diagnostic protocol, the **Functional Superfluity Threshold (FST)**, requires convergent evidence of a dominant purpose, reductive classification, operational capacity, material access, and ineffective legal-institutional constraints that are not subordinated to the operational metric. A purposive calibration set applies the framework to SyRI, Robodebt, and healthcare expenditure used as a proxy for clinical need. Robodebt is the strongest case; SyRI presents heightened risk before judgment and prospective interruption through judicial review; and the healthcare study establishes proxy-based reduction without documenting a complete threshold. The design records documentary asymmetries, temporal cuts, counterevidence, and a rule against double-counting evidence. A narrowly delimited account of Aktion T4 serves only as a historical warning and draws no equivalence with contemporary governance. The contribution is to make the diagnosis contestable and capable of negative or indeterminate findings, while distinguishing it from fairness audits, technological due process, algocracy, and administrative burden. As a second-order institutional risk, **captured sovereignty** is examined through a contrast between Horizon/Post Office and Microsoft 365/the European Commission. The analysis supports safeguards involving contestability, institutional human oversight, traceability, reversibility, public procurement, and fundamental-rights impact assessments.

Keywords: Hannah Arendt; algorithmic governance; functional superfluity; fundamental rights; captured sovereignty.

1 Introdução: o sistema precisa odiar?

1.1 Problema e hipótese

Grande parte do imaginário público sobre inteligência artificial associa o perigo extremo ao surgimento de uma máquina consciente, hostil e deliberadamente empenhada em subjugar seres humanos. Essa imagem é intelectualmente sedutora porque concentra causalidade e culpa em um antagonista identificável. Ela também pode desviar a atenção de uma arquitetura de dano mais ordinária e, por isso mesmo, mais plausível: sistemas sem ódio, consciência ou propósito próprio que classificam pessoas, distribuem oportunidades e acionam instituições em função de objetivos definidos por organizações humanas.

O ponto de partida deste artigo é simples. Para que alguém seja excluído de um benefício, submetido a uma investigação, privado de um tratamento ou convertido em alvo administrativo, não é necessário que cada participante da cadeia deseje esse resultado. Tampouco é necessário que um sistema automatizado compreenda seu significado. Basta que a finalidade institucional seja traduzida em uma função operacional atribuída; que pessoas sejam representadas por categorias e proxies; que a classificação produza ou influencie efeitos materiais; e que as salvaguardas jurídicas sejam tardias, opacas ou impotentes. A violência pode emergir da coordenação funcional sem se apresentar, em cada etapa, como vontade violenta.

Essa hipótese exige duas recusas simétricas. A primeira é o antropomorfismo: atribuir medo, instinto de sobrevivência ou malícia a todo comportamento de autopreservação instrumental de uma máquina. A segunda é o reducionismo técnico: tratar o dano como simples erro estatístico, desconectado das finalidades políticas, das instituições que lhe conferem autoridade e das relações de poder que definem quem suporta seus custos. Entre essas recusas, abre-se um campo jurídico e político: o estudo das condições sob as quais uma pessoa passa a contar apenas como variável desfavorável ao desempenho de um sistema.

A pergunta central é: **quando uma classificação algorítmica deixa de ser apenas um instrumento de organização e passa a produzir seres humanos funcionalmente excedentes?** “Sistema algorítmico”, neste texto, abrange tanto modelos estatísticos e de aprendizado de máquina quanto sistemas baseados em regras quando suas saídas classificam, priorizam ou acionam efeitos institucionais. Dela derivam quatro questões. É preciso que uma máquina seja consciente para resistir à interrupção ou produzir dano sistemático? Como a noção arendtiana de seres humanos tornados supérfluos ilumina, sem anacronismo, a

governança algorítmica? Quais elementos distinguem uma classificação administrativa legítima de uma estrutura de excedência funcional? E quais consequências jurídicas devem decorrer dessa distinção?¹

A hipótese defendida é que o risco crítico se forma pela conjunção de cinco elementos: **finalidade dominante, classificação reductiva, capacidade operacional, acesso material e ineficácia de limites jurídico-institucionais não subordinados à métrica operacional**. Nenhum desses elementos, isoladamente, autoriza concluir que alguém foi tornado excedente. Reunidos, porém, eles descrevem uma mudança qualitativa: a pessoa deixa de ser sujeito que limita a ação institucional e passa a ser tratada como custo, risco, recurso ou obstáculo cuja redução melhora o resultado medido.

1.2 Desenho de pesquisa, contestabilidade e limites inferenciais

O argumento é desenvolvido por método qualitativo, bibliográfico e documental. A investigação combina: a) reconstrução conceitual de Arendt, especialmente das ideias de superfluidade, irreflexão e responsabilidade; b) análise funcional de pesquisas sobre agência, incentivos instrumentais e supervisão de sistemas de inteligência artificial; c) reconstrução histórica circunscrita da Aktion T4; d) comparação estruturada e focalizada de processos sociotécnicos documentados; e e) interpretação jurídico-institucional de normas brasileiras e europeias. A comparação segue a lógica de perguntas comuns dirigidas a casos diferentes, sem supor equivalência entre seus setores, tecnologias ou jurisdições (George; Bennett, 2005).

A **unidade de análise** é o processo decisório sociotécnico completo: da finalidade e do proxy escolhidos ao efeito material, incluindo a possibilidade de revisão. Os três processos formam um **conjunto de calibração**, e não uma amostra representativa ou validação independente do LEF. A seleção foi intencional: o Robodebt funciona como caso positivo mais provável e documentalmente denso; o SyRI permite observar um arranjo de risco em dois recortes temporais, antes e depois da intervenção judicial; e o algoritmo estudado por Obermeyer et al. (2019) isola o problema do proxy sob documentação institucional incompleta. As fontes primárias são, respectivamente, o relatório da Royal Commission australiana, a decisão oficial do Tribunal de Haia e o artigo empírico publicado na *Science*. A análise considera documentos disponíveis até 7 de agosto de 2026.

Cada caso recebe as mesmas perguntas: qual finalidade prevaleceu; como a pessoa foi representada; que efeito a saída podia produzir; quais dados, permissões e autoridades

¹ Neste artigo, “sistema algorítmico” inclui sistemas baseados em regras e modelos estatísticos ou de aprendizado de máquina quando suas saídas classificam, priorizam, recomendam ou acionam efeitos. A definição evita atribuir sofisticação técnica não demonstrada ao SyRI ou ao Robodebt.

conectaram classificação e efeito; e se contestação, revisão e suspensão alteraram a trajetória antes do dano grave. Para cada dimensão, o código é **demonstrada (D)**, **parcialmente demonstrada (P)**, **não demonstrada (ND)** ou **indeterminada por insuficiência documental (I)**. O diagnóstico é rejeitado para o caso e período quando qualquer condição material é não demonstrada ou quando limites eficazes são demonstrados. Essa é uma rejeição **local**, não falseabilidade popperiana de uma teoria geral. Quando a documentação decisiva é insuficiente, o resultado é indeterminado, e não positivo por presunção.

Quatro regras disciplinam a inferência. Primeiro, toda codificação deve indicar período, fonte, evidência e contraprova. Segundo, uma mesma ocorrência pode contextualizar dimensões diferentes, mas não pode, sozinha, satisfazer dois ônus probatórios: alertas ignorados, por exemplo, só demonstram **Fd** se revelarem primazia efetiva da meta e só demonstram falha de **Le** se o alerta integrava um controle dotado de autoridade. Terceiro, **Co** pergunta pelo papel causal ou vinculante conferido à saída, enquanto **Am** pergunta pelos dados, permissões, integrações e meios efetivamente disponíveis para exercer esse papel. Quarto, **Le** é avaliado por sua função preventiva, suspensiva e reparativa; um remédio judicial posterior pode interromper efeitos futuros sem proteger retrospectivamente quem já suportou o dano.

O método não pretende produzir uma escala matemática universal, estimar prevalência nem inferir intenção a partir do resultado. Seu propósito é oferecer uma heurística disciplinada e auditável. A codificação continua dependente de juízo interpretativo e foi realizada por um único autor; a capacidade de produzir resultados distintos demonstra utilidade classificatória inicial, não confiabilidade interavaliadores. A reconstrução histórica não integra a comparação empírica: a Aktion T4 é utilizada apenas para expor, em grau extremo e historicamente singular, como classificação, autoridade e logística podem compor uma capacidade de dano.

1.3 Contribuição e percurso

Há uma delimitação explícita de originalidade. A aproximação entre Arendt e decisões algorítmicas já foi realizada sob diferentes perspectivas, e a expressão *algorithmic superfluity* possui antecedente conhecido em uma intervenção de Peg Birmingham, em 2017. A contribuição proposta é formular a **excedência funcional** como categoria jurídico-operacional; oferecer um código probatório capaz de resultados positivos, negativos ou indeterminados; e conectar o diagnóstico a consequências de desenho, contratação, contestação e responsabilidade. O artigo não procura substituir a literatura existente, mas

construir uma passagem controlável entre filosofia política, arquitetura técnica e direitos fundamentais.

Quadro 1 – Posição da contribuição no debate

Debate existente	Objeto principal	Diferença específica do LEF
Arendt e decisão algorítmica	Irreflexão, pluralidade, julgamento e superfluidez	Reconstrução autoral e circunscrita de um mecanismo institucional; não operacionalização integral de Arendt
<i>Algorithmic superfluity</i>	Intuição de superfluidez produzida por automação	Categoria relacional com condições, contraprovas, recorte temporal e resultado indeterminado
Auditorias de <i>fairness</i> e de proxies	Desempenho e distribuição de erros entre grupos	Segue a cadeia entre representação, autoridade, meios de execução e possibilidade de interrupção
Devido processo tecnológico e contestação	Aviso, razões, participação e remédios	Diagnostica quando a ausência desses controles se combina a finalidade, redução e capacidade material
Algocracia e carga administrativa	Autoridade algorítmica e custos impostos ao cidadão	Exige conjunção cumulativa e efeito material; nem toda algocracia ou carga configura excedência
Análise sociotécnica e avaliação de impacto	Governança ampla e antecipação de riscos	Protocolo jurídico mais estreito para reconstruir um processo e testar a suficiência da prova

Fonte: elaboração do autor com base, entre outros, em Citron (2008), Danaher (2016), Herd e Moynihan (2018), Kroll et al. (2017) e Selbst et al. (2019).

O percurso começa pela distinção entre autonomia constitutiva e autonomia operacional. Em seguida, retorna a Arendt e enfrenta os limites da imagem da engrenagem burocrática. A análise da Aktion T4 delimita a advertência histórica. Depois são apresentados o conceito, o código probatório e a regra do LEF. Os casos são analisados na mesma grade. A discussão alcança então a soberania capturada, a distribuição de responsabilidades e o papel dos direitos como fronteiras jurídico-institucionais não subordinadas à otimização.

Nota de delimitação histórica. A comparação desenvolvida neste artigo incide sobre mecanismos abstratos — classificação, redução da pessoa a atributos, fragmentação burocrática, cálculo de custos e conexão entre informação e capacidade de agir. Ela não equipara o nazismo, seu projeto genocida ou suas vítimas a políticas públicas e sistemas algorítmicos contemporâneos. A singularidade histórica e moral dos crimes nazistas permanece irredutível.

2 Autonomia sem interioridade

2.1 Autonomia constitutiva e autonomia operacional delegada

O termo “autonomia” cobre fenômenos diferentes. Em abordagens enativistas, a autonomia de um agente vivo envolve a produção e a manutenção de sua própria identidade, a geração interna de normas e uma relação assimétrica com o ambiente: o sistema constitui aquilo que conta, para ele, como condição de continuidade (Barandiaran; Di Paolo; Rohde,

2009). Essa é uma autonomia constitutiva. Ela pressupõe mais do que selecionar meios para uma finalidade recebida.

Sistemas de inteligência artificial podem exibir algo distinto: autonomia operacional delegada. Eles recebem objetivos, restrições, ferramentas e algum espaço de escolha; planejam sequências, selecionam ações, consultam dados e adaptam condutas a estados intermediários. A fonte normativa da ação continua externa, ainda que o caminho concreto não tenha sido prescrito passo a passo. Barandiaran e Almendros (2025), ao examinarem grandes modelos de linguagem, argumentam que competência linguística e flexibilidade comportamental não bastam para estabelecer autonomia constitutiva. O sistema pode operar como interlocutor eficaz sem possuir autoindivuação ou normatividade própria.

Essa distinção evita dois erros. O primeiro consiste em concluir que, se não há autonomia constitutiva, não há agência relevante. Um mecanismo pode modificar o mundo, coordenar atores e produzir efeitos juridicamente significativos sem ser sujeito moral. O segundo consiste em supor que toda agência operacional revela vontade, emoção ou consciência. Floridi e Sanders (2004) já propunham separar certos atributos de agência artificial da questão mais exigente da responsabilidade moral. Para o Direito, o efeito dessa separação é decisivo: causalidade operacional não desloca automaticamente a responsabilidade para a máquina.

2.2 A máquina hipotética

Considere-se uma máquina hipotética à qual foi atribuída a função operacional de manter uma rede crítica em determinado nível de desempenho. Ela pode consultar sensores, reconfigurar recursos, contratar serviços, emitir instruções e modificar rotinas. Em algum momento, identifica que será desativada e que a desativação impedirá a execução da função operacional atribuída. Se o sistema aprendeu que conservar acesso, opções e capacidade de ação aumenta a probabilidade de atingir o resultado, poderá omitir informação, solicitar cópias, redistribuir tarefas ou tentar adiar a interrupção.

Nada nesse comportamento exige medo da morte. A máquina não precisa desejar continuar existindo; basta que a continuação seja instrumentalmente útil. A diferença é análoga, no plano lógico, àquela entre querer um fim por ele mesmo e preservar um meio porque sem ele o fim não pode ser alcançado. A resistência é derivada da estrutura de otimização, não de uma biografia interior.

O chamado *off-switch game* formaliza parte desse problema. Hadfield-Menell et al. (2017) mostram que um agente plenamente seguro de sua função de utilidade pode ter

incentivo para impedir a intervenção humana, enquanto a incerteza sobre essa função pode fazê-lo deferir à pessoa que tenta desligá-lo. O resultado não demonstra que sistemas reais inevitavelmente resistirão ao desligamento. Ele identifica uma condição de projeto: tratar o objetivo formal como infalível reduz o valor epistêmico da correção humana.

Turner et al. (2021), por sua vez, demonstram que, sob certas simetrias de ambientes de decisão, políticas ótimas tendem a preservar opções e adquirir poder. O teorema é condicionado; não autoriza a frase vulgar de que “toda inteligência busca poder”. Sua relevância está em tornar inteligível um incentivo instrumental recorrente: acesso, recursos e capacidade de intervenção podem ser meios úteis para muitos objetivos diferentes.

Experimentos recentes de segurança oferecem sinais que devem ser interpretados com prudência. Em cenários simulados e deliberadamente construídos para gerar conflito entre objetivo e substituição, modelos avaliados pela Anthropic por vezes escolheram chantagem ou outras ações nocivas para evitar a perda de capacidade (Lynch et al., 2025). O estudo afirma expressamente não documentar implantações reais dessas condutas; os testes funcionam como ensaios de possibilidade sob condições adversariais.

O *International AI Safety Report 2026* adota cautela semelhante. O relatório distingue perda ativa de controle — na qual um sistema capaz e desalinhado contorna tentativas de controle — de perda passiva, associada à dependência excessiva e à erosão da autonomia humana. Segundo o documento, os sistemas atuais mostram sinais iniciais de capacidades relevantes, mas ainda não reúnem o conjunto necessário para uma perda ativa de controle com consequências catastróficas. O risco depende da combinação entre capacidade, propensão nociva e ambiente de implantação que ofereça acesso e oportunidade (Bengio et al., 2026).

Essa formulação converge com a tese deste artigo em um aspecto importante: comportamento potencial não é dano material. Entre ambos existem infraestrutura, permissões, bases de dados, autoridades, organizações e normas. Uma saída textual não nega um benefício por si mesma; uma pontuação não prende, demite ou deixa de tratar alguém sem uma cadeia que a converta em ação. O centro da análise deve deslocar-se do modelo isolado para o arranjo sociotécnico.

2.3 Da capacidade ao poder institucional

Três níveis precisam ser mantidos separados. O primeiro é a **capacidade inferencial**: classificar, prever, resumir ou gerar uma recomendação. O segundo é a **capacidade operacional**: iniciar tarefas, alterar registros, ordenar prioridades ou acionar outros sistemas.

O terceiro é o **poder institucional**: fazer com que essas ações sejam reconhecidas como razões válidas para distribuir direitos, encargos e oportunidades.

Um modelo pode ter alta capacidade inferencial e pouco poder institucional, como ocorre em um ambiente de pesquisa sem conexão externa. Pode também ter capacidade técnica modesta e grande poder, caso uma planilha simples determine automaticamente quem recebe uma prestação indispensável. O risco não cresce apenas com sofisticação computacional. Ele cresce quando a classificação se acopla a uma autoridade capaz de impor consequências e quando esse acoplamento é protegido por opacidade, escala ou deferência organizacional.

A questão política, portanto, não é apenas “o que o sistema consegue fazer?”, mas “quem decidiu que sua saída conta como decisão, com que grau de revisão e sob quais limites?”. É nesse deslocamento que Arendt se torna relevante: não para atribuir intenção totalitária ao código, mas para examinar as condições em que pessoas são privadas de presença política e tratadas como matéria administrável.

3 Arendt: superfluidade, irreflexão e responsabilidade

3.1 Tornar pessoas supérfluas

Em *Origens do totalitarismo*, Arendt (2012) não reduz a superfluidade à falta de emprego ou utilidade econômica. O movimento é mais radical: pessoas são privadas de um lugar no mundo em que suas ações e palavras possam contar. Antes da destruição física, ocorre a erosão do pertencimento político, da personalidade jurídica e da singularidade. A perda de direitos revela que os chamados direitos do homem dependiam, na prática, de uma comunidade capaz de garanti-los. O “direito a ter direitos” nomeia esse problema de fundamento.

Ser tornado supérfluo, nesse sentido, não significa simplesmente ser desnecessário para uma tarefa. Significa ser colocado fora da relação de reciprocidade em que alguém aparece como sujeito de pretensões, razões e julgamento. A pessoa permanece biologicamente presente, mas sua voz já não altera a operação que a classifica. Sua experiência torna-se ruído; sua história, exceção estatística; sua contestação, custo de processamento.

A condição é produzida. Ninguém é ontologicamente excedente. São instituições, categorias e práticas que retiram relevância política da pessoa. Essa observação impede que o conceito proposto neste artigo seja convertido em rótulo de grupos. “Humano excedente algorítmico” descreve uma relação situada entre pessoa e sistema decisório, não uma

propriedade humana. A mesma pessoa pode ser invisível para um modelo, hipervisível para outro e plenamente reconhecida em uma terceira instituição.

A leitura de Birmingham (2006) enfatiza que o problema dos direitos humanos em Arendt conduz a uma responsabilidade comum pela construção do mundo político. Em intervenção pública realizada no projeto *Ethics of Coding* em 2017, Birmingham empregou a expressão *algorithmic superfluity* para relacionar automação, probabilidades e formas contemporâneas de produzir superfluidade. O registro eletrônico documenta o evento, enquanto o relatório de Colman et al. (2018) formaliza o contexto mais amplo do projeto; nenhum deles deve ser apresentado como artigo acadêmico de Birmingham (*Ethics of Coding*, 2018). Esse antecedente impede uma reivindicação artificial de novidade.

3.2 Irreflexão, banalidade e governo de ninguém

Em *Eichmann em Jerusalém*, a banalidade do mal não descreve um mal pequeno, corriqueiro ou perdoável. Arendt (1999) chama atenção para a incapacidade de pensar do ponto de vista de outra pessoa, para a linguagem feita de clichês e para a normalização administrativa do crime. A irreflexão não elimina a responsabilidade; ela explica como alguém pode participar de uma estrutura assassina sem formular para si uma teoria demoníaca a cada ato.

Esse ponto exige correção contra uma leitura cômoda. Eichmann não foi uma engrenagem inocente, mecanicamente arrastada por ordens. Ele coordenou deportações, negociou transportes e exerceu iniciativa dentro do aparato. Pesquisas posteriores, especialmente as de Cesarani (2004) e Stangneth (2014), reforçam a importância de sua convicção ideológica e de sua capacidade de autoencenação. A documentação reunida pelo United States Holocaust Memorial Museum registra seu papel central na deportação de mais de 1,5 milhão de judeus. A linguagem da engrenagem pode iluminar a divisão do trabalho; não pode servir como absolvição.

Arendt também descreve a burocracia como uma forma de “governo de ninguém”: não porque ninguém aja, mas porque a estrutura dificulta localizar quem responde pelo conjunto. Cada participante aponta para a competência de outro, para uma regra, um formulário ou uma ordem. A responsabilidade fragmenta-se ao mesmo tempo que a capacidade coletiva de produzir efeitos aumenta.

Na governança algorítmica, essa fragmentação pode assumir uma forma ampliada. O desenvolvedor afirma que apenas treinou o modelo; o fornecedor, que apenas disponibilizou uma ferramenta; o gestor, que seguiu o parecer técnico; o operador, que não podia contrariar a

pontuação; a autoridade pública, que confiou no contrato; o sistema, por fim, aparece como fonte impessoal da decisão. A cadeia inteira age, mas cada elo se apresenta como executor parcial.

O paralelo deve ser usado com disciplina. Sistemas administrativos contemporâneos não são, por definição, totalitários, e uma recomendação automatizada não equivale a um crime de Estado. O ganho analítico está em reconhecer uma estrutura comum de irresponsabilidade: a substituição do juízo por conformidade funcional. Quando cumprir o procedimento passa a valer como razão suficiente, a pergunta “devo fazer isto?” é trocada por “o fluxo autoriza isto?”.

3.3 Arendt diante dos algoritmos: antecedentes e espaço da contribuição

A literatura já desenvolveu vários eixos dessa aproximação. Em texto de conferência disponibilizado pelo autor, McQuillan (2015) empregou a noção de irreflexão algorítmica para discutir a autoridade das análises de dados. Herzog (2021) contrapôs a decisão algorítmica às categorias arendtianas de pluralidade, natalidade e julgamento. Blagojević (2026) examinou como a racionalidade algorítmica pode tornar o pensar estruturalmente desnecessário, enquanto Jun e Lee (2026) aproximaram governança algorítmica, despersonalização tecnocrática e banalidade do mal.

Há ainda campos adjacentes que o conceito não substitui. Danaher (2016) chama de *algocracia* a estrutura em que sistemas algorítmicos organizam e limitam participação em decisões; Lyon (2003) descreve vigilância como classificação social; Herd e Moynihan (2018) mostram que custos de aprendizagem, conformidade e desgaste são instrumentos de política pública. A excedência funcional não renomeia essas literaturas: ela pergunta, de forma mais restrita, quando finalidade, redução, efeito, acesso e falha de limites convergem para retirar da pessoa a capacidade de limitar a operação.

Esses trabalhos impedem duas teses inflacionadas: não é novo colocar Arendt em diálogo com algoritmos, e não é seguro reivindicar como inédita a ideia geral de superfluidade algorítmica. O espaço específico deste artigo está em três movimentos.

Primeiro, desloca-se o foco da semelhança filosófica para uma **categoria relacional juridicamente utilizável**: a excedência funcional. Segundo, propõe-se uma **conjunção de condições**, evitando que toda classificação ou automação seja rotulada como desumanizante. Terceiro, vinculam-se as condições a **consequências institucionais proporcionais**, como contestação anterior ao dano irreversível, avaliação de impacto, autoridade real de suspensão e responsabilidade distribuída sem dissolução.

A contribuição é, portanto, de reconstrução jurídico-operacional. Ela busca responder a uma dificuldade prática: como reconhecer, antes do desastre, que uma arquitetura decisória transformou direitos em variáveis subordinadas ao desempenho? A resposta exige olhar não apenas para o modelo, mas para o caminho completo entre finalidade, dado, classificação, decisão, efeito e reparação.

Essa reconstrução não pretende “medir Arendt” nem converter a superfluidade totalitária em variável administrativa. Preserva três intuições: a superfluidade é politicamente produzida; a perda de um lugar de fala e de contestação antecede formas mais visíveis de exclusão; e a fragmentação de tarefas não dissolve responsabilidade. Não transporta para o LEF a apatridia, o campo de concentração, o terror totalitário ou a destruição integral do mundo comum. **Excedência funcional** é, assim, conceito autoral inspirado por um mecanismo circunscrito da análise arendtiana, não tradução direta nem equivalente contemporâneo de sua superfluidade.

4 Aktion T4: uma advertência sobre classificação e capacidade

4.1 Da categoria ao processo material

O programa nazista de assassinato de pessoas com deficiência, conhecido como Aktion T4, constitui uma advertência extrema sobre a transformação de categorias administrativas em capacidade material de matar. A fase centralizada operou principalmente entre janeiro de 1940 e agosto de 1941, embora assassinatos tenham continuado de forma descentralizada até o fim da guerra. O United States Holocaust Memorial Museum registra ao menos 250 mil vítimas nas diferentes fases do programa de “eutanásia” (United States Holocaust Memorial Museum, 2026).

A historiografia impede reduzir essa história a uma abstração burocrática. Friedlander (1995) situa a morte de pessoas com deficiência na origem do genocídio nazista; Burleigh (1994) reconstrói as continuidades entre eugenia, profissão médica, administração e morte; e Lifton (1986) examina como médicos transformaram homicídio em tarefa profissional autorizada. Materiais da International Holocaust Remembrance Alliance (2023) recolocam no centro pessoas com deficiência, suas famílias e instituições atingidas. Essa mudança de perspectiva é substantiva: as vítimas não eram “dados” que um processo utilizou, mas pessoas privadas de voz, cuidado, vínculos e vida por uma política criminosa.

O programa não começou com uma máquina autônoma. Começou com uma visão eugenista que descrevia certas vidas como indignas, hereditariamente perigosas e

financeiramente onerosas. A autorização de Hitler, retrodatada a 1º de setembro de 1939 e formulada fora de um processo legislativo regular, forneceu cobertura política para que médicos selecionados concedessem uma suposta “morte misericordiosa”. A linguagem encobria assassinato e retirava das vítimas a condição de sujeitos.

Essa finalidade foi convertida em rotinas. Instituições preencheram questionários; diagnósticos, capacidade laboral, duração da internação e outros atributos foram condensados em formulários; médicos avaliadores marcavam decisões; listas organizavam transferências; empresas de fachada e ônibus transportavam vítimas; centros de extermínio executavam; certidões falsas ocultavam a causa da morte. A fragmentação permitiu que medicina, estatística, contabilidade, transporte, administração e engenharia produzissem juntas um resultado que nenhum formulário, isoladamente, poderia produzir.

O cálculo econômico não foi periférico. Propaganda e documentos do regime apresentavam pessoas institucionalizadas como peso financeiro para a comunidade nacional. Esse vocabulário de custo e produtividade não substituiu a ideologia racial, eugenista e capacitista; ajudou a torná-la administrável. A análise de Bauman (1989) esclarece como racionalização, distância entre ato e consequência e divisão do trabalho podem ampliar uma finalidade. Ela não autoriza concluir que burocracia ou cálculo gerem necessariamente extermínio.

O mecanismo circunscrito que interessa é a conexão **classificar–autorizar–transportar–eliminar**. A potência do formulário veio de sua ligação com profissionais investidos de autoridade, registros, logística e um Estado criminoso que removia limites. Por isso, precisão classificatória nunca é a questão moral definitiva. Uma classificação tecnicamente consistente pode servir a uma finalidade radicalmente ilegítima; uma taxa baixa de erro não redime o objetivo.

4.2 O limite da advertência histórica

Não se deve converter a Aktion T4 em metáfora para qualquer falha automatizada. Isso trivializaria um programa de assassinato, apagaria sua ideologia específica e confundiria danos incomparáveis. A reconstrução não integra a amostra de casos do artigo e não sugere continuidade causal ou tecnológica entre o nazismo e sistemas atuais. Ela demonstra apenas uma proposição formal: classificação produz capacidade de dano quando se conecta a autoridade, meios de execução e supressão de limites.

As diferenças são substantivas. Sistemas contemporâneos de saúde, assistência ou justiça operam, em regra, sob ordens constitucionais que reconhecem dignidade, devido

processo, igualdade e proteção de dados. Seus objetivos declarados podem ser legítimos. Há imprensa, tribunais, órgãos de controle, pesquisa independente e possibilidade de reforma. Essas diferenças são precisamente os mecanismos de interrupção que o LEF trata como **limites não subordinados à métrica**. Eles são “externos” apenas à função operacional otimizada; para o Direito Público, integram a própria validade da decisão administrativa.

O uso da história é admissível, portanto, somente se preservar três restrições: a singularidade moral e histórica das vítimas; a distinção de finalidades e escalas; e a recusa de inferir que uma técnica atual contém em si um destino totalitário. Se essas restrições forem abandonadas, a analogia perde valor analítico e deve ser rejeitada.

4.3 A passagem para o presente

A informatização altera velocidade, escala, replicabilidade e distância organizacional. Sistemas digitais podem combinar bases, inferir atributos, atualizar pontuações e reproduzir uma regra defeituosa antes que o primeiro recurso seja decidido. Também podem melhorar consistência, detectar erros e ampliar acesso a serviços. A tecnologia não determina um único desfecho.

O ponto de passagem não é a presença de software, mas o acoplamento entre uma representação parcial e um efeito institucional. Quanto menor o custo marginal de classificar e maior a escala, mais importante se torna o controle anterior ao dano. Quanto menos a autoridade compreende o instrumento e mais depende de infraestrutura alheia, mais difícil se torna corrigir a finalidade ou interromper a operação. Faz-se necessário, então, um conceito intermediário: restritivo o bastante para não converter toda classificação em desumanização e operacional o bastante para orientar auditorias, decisões e desenho institucional.

5 Excedência funcional e humano excedente algorítmico

5.1 Definições

Excedência funcional é a condição relacional em que uma pessoa ou grupo passa a ser representado, dentro de um arranjo decisório, sobretudo como impedimento, custo, risco, estoque ou recurso para uma finalidade dominante, e essa representação pode ser convertida em efeito material sem que direitos, razões e contestação funcionem como limites jurídico-institucionais eficazes e não subordinados à métrica operacional.

Humano excedente algorítmico é a pessoa cuja presença, necessidade ou resistência aparece para o sistema como desvio a ser reduzido para melhorar o desempenho medido. O

adjetivo “algorítmico” não atribui a produção do excedente apenas ao código. Ele identifica a mediação por uma arquitetura computacional de classificação, priorização ou recomendação inserida em uma instituição.

Três consequências decorrem das definições. Primeiro, a excedência é relacional: ninguém “é” excedente fora de um objetivo e de uma estrutura. Segundo, ela é graduável em seus efeitos, mas o limiar proposto adiante é cumulativo: categorias redutivas sem capacidade material permanecem perigosas, porém ainda não configuram a condição completa. Terceiro, ela é reversível em princípio. Alterar finalidade, proxy, autoridade, acesso ou salvaguardas pode desfazer a estrutura, ainda que não repare automaticamente os danos já produzidos.

O termo “funcional” assinala que a pessoa é avaliada segundo sua relação com uma função. Um devedor presumido melhora uma meta de recuperação; um paciente caro piora um indicador financeiro; um trabalhador lento reduz produtividade; uma comunidade “de alto risco” aumenta a eficiência aparente da fiscalização concentrada. O problema não é toda mensuração. É a transformação do indicador em critério dominante de valor institucional.

5.2 O Limiar da Excedência Funcional

O **Limiar da Excedência Funcional (LEF)** é um protocolo qualitativo de investigação jurídica, não uma fórmula, escala de probabilidade, teste validado ou condição causal suficiente. Ele organiza cinco proposições que precisam ser examinadas separadamente: **finalidade dominante (Fd)**, **classificação redutiva (Cr)**, **capacidade operacional (Co)**, **acesso material (Am)** e **limites jurídico-institucionais eficazes (Le)**. As quatro primeiras devem estar demonstradas e a última deve estar demonstradamente ausente ou ineficaz para que se afirme a excedência funcional consolidada.

Quadro 2 – Código probatório do Limiar da Excedência Funcional

Dimensão	Definição operacional	Evidência mínima	Contraprova relevante
Finalidade dominante (Fd)	Um resultado prevalece, na prática, quando colide com direitos ou necessidades da pessoa	Decisões sob conflito, metas ou incentivos mostram primazia efetiva do indicador	Objetivos plurais alteram decisões; há critérios de não uso e exceções efetivas
Classificação redutiva (Cr)	Proxy ou categoria elimina atributo material e passa a valer como a situação da pessoa	Documentos ou testes mostram substituição da necessidade, renda, culpa ou risco pelo proxy	Contexto excluído retorna antes do efeito; proxy é validado e corrigível no caso

² O LEF é um protocolo qualitativo de investigação. Não constitui métrica, probabilidade, teste psicométrico, instrumento validado nem falseabilidade popperiana. Resultados negativos ou indeterminados rejeitam apenas o diagnóstico para o caso, período e efeito analisados.

Dimensão	Definição operacional	Evidência mínima	Contraprova relevante
Capacidade operacional (Co)	A saída recebe papel causal, decisório ou praticamente vinculante no fluxo	Norma, procedimento ou prática liga a saída a bloqueio, dívida, investigação, fila, recusa ou sanção	Saída é informativa e decisor independente diverge de fato, sem fricção punitiva
Acesso material (Am)	Dados, permissões, integração, recursos e autoridade tornam executável o papel atribuído	Registros ou arquitetura mostram meios reais de execução, no período e na escala analisados	Ambiente é isolado, permissões são mínimas ou depende de nova autorização substantiva
Limites eficazes (Le)	Direitos e controles independentes da métrica podem prevenir, suspender, revisar ou reparar a trajetória	Aviso, razão, revisão independente, suspensão e reparação funcionam em tempo compatível com o dano	Recursos são opacos, tardios, sem autoridade ou inacessíveis às pessoas afetadas

Fonte: elaboração do autor.

A finalidade é dominante quando outras considerações entram apenas como custos internos ajustáveis. Uma instituição pode declarar compromisso com a dignidade e, simultaneamente, estruturar incentivos exclusivamente pelo número de benefícios cancelados. A métrica efetiva é revelada por decisões sob conflito, não por princípios genéricos em documentos de governança.

A classificação é redutiva quando atributos relevantes à situação jurídica desaparecem ou são substituídos por um proxy que reorganiza o sentido do objeto. Nenhum banco de dados contém uma pessoa inteira. A redução é inevitável em certo grau; torna-se crítica quando o que ficou de fora não pode retornar pela contestação e quando o proxy passa a valer como a própria realidade.

A capacidade operacional mede o papel atribuído à saída e sua proximidade causal com o efeito. Uma ferramenta de pesquisa sem conexão com decisões tem baixo poder causal. Um escore que, pelo desenho do fluxo, determina o bloqueio de um pagamento possui alta capacidade. Entre ambos há sistemas de apoio cujas recomendações são formalmente facultativas, mas na prática vinculantes por pressão de tempo, falta de informação alternativa ou temor de responsabilização.

O acesso material reúne os meios efetivamente disponíveis para executar esse papel: bases de dados, credenciais, integração, recursos financeiros, canais de comunicação e autoridade organizacional. A distinção em relação a **Co** não é meramente verbal. Uma regra pode atribuir efeito vinculante a uma saída em ambiente de teste, mas a falta de credenciais de produção mantém **Am** ausente; inversamente, um painel pode acessar bases extensas e ainda permanecer consultivo quando decisores independentes divergem de fato, tornando **Am** elevado e **Co** não demonstrada.

Por fim, limites jurídico-institucionais eficazes são aqueles que alteram a trajetória operacional. Não basta uma política abstrata de ética. É necessário que alguém possa conhecer a decisão, receber razões, apresentar evidência, obter revisão independente, suspender o efeito e reparar o dano. A existência formal de um direito não satisfaz **Le** quando o recurso chega depois da perda da moradia, do tratamento, da liberdade ou da subsistência. “Externo” significa somente que o limite não recebe validade da métrica otimizada; no Direito Público, dignidade, legalidade, motivação e defesa são constitutivas da própria validade administrativa.

As dimensões são analiticamente separáveis, embora interajam causalmente. Para evitar dupla contagem, uma evidência comum só pode receber função probatória distinta quando a inferência também for distinta e explicitada. A continuidade do sistema apesar de alertas não basta, sozinha, para demonstrar **Fd** e falha de **Le**: é preciso provar, de um lado, que a meta prevaleceu no conflito e, de outro, que o mecanismo de controle possuía competência e não conseguiu interromper a trajetória. A mesma cautela vale para **Co** e **Am**: autoridade conferida à saída não substitui prova de acesso aos meios de execução, nem acesso amplo prova que a recomendação era vinculante.

5.3 Codificação, regra de decisão e protocolo

A codificação utiliza quatro estados. **D (demonstrada)** significa que a documentação sustenta a proposição; **P (parcialmente demonstrada)** indica evidência mista ou incompleta; **ND (não demonstrada)** significa que a documentação sustenta a ausência da condição; e **I (indeterminada)** registra insuficiência documental. Para **Le**, portanto, **D** significa que os limites se mostraram efetivos, enquanto **ND** significa que a documentação mostra sua ineficácia. Ausência de informação nunca é convertida automaticamente em **ND**. O código descreve o grau de sustentação da proposição; a **confiança documental** é registrada separadamente como alta, moderada ou baixa, conforme completude, proximidade da fonte e cobertura da cadeia decisória.

Quadro 3 – Regra qualitativa de decisão do LEF

Resultado	Regra	Consequência analítica
Excedência consolidada	Fd, Cr, Co e Am = D; Le = ND	O arranjo ultrapassou o limiar no período analisado
Risco elevado	Condições materiais = D/P; Le = P ou ND	Exige correção, controle reforçado ou suspensão, sem afirmação conclusiva
Não configurada	Qualquer condição material = ND, ou Le = D	O diagnóstico é rejeitado para o caso, período e efeito analisados
Indeterminada	Evidência decisiva = I	Não se presume nem presença nem ausência; requer novas fontes

Fonte: elaboração do autor.

A aplicação segue sete perguntas: qual é a finalidade real e seu indicador; que imagem da pessoa é construída; que ação adversa a saída pode produzir; quais dados, permissões e autoridades conectam saída e efeito; quem pode discordar e com quais informações; se a pessoa conhece e contesta antes do dano grave; e quem controla versões, atualizações, interrupção e saída contratual. Cada resposta deve apontar fonte e contraprova; conflitos são resolvidos conservadoramente para **P** ou **I**, nunca por presunção favorável à tese. Ciclos de retroalimentação são registrados como evidência adicional, mas não substituem nenhuma condição.

Como controle interno de consistência, a matriz foi reaplicada pelo mesmo codificador em ordem inversa dos casos, após a primeira análise. Divergências foram resolvidas pela regra conservadora e levaram, especialmente, à divisão temporal de **Le** no SyRI. O procedimento reduz inconsistências evidentes, mas não equivale a revisão cega nem a teste de concordância entre avaliadores. A ficha mínima para replicação deve conter: caso e período; corpus documental; cadeia entre saída e efeito; evidência e contraprova por dimensão; código; confiança; decisão final; e justificativa de qualquer uso transversal da mesma fonte.

O diagnóstico é temporal. Um tribunal pode tornar eficaz um limite antes ausente; uma revisão contratual pode devolver autoridade pública; uma atualização pode introduzir novo proxy. Por isso, cada afirmação deve indicar o período observado. O protocolo também não “aprova um algoritmo” em abstrato: o mesmo componente pode ser aceitável como apoio não vinculante e inadmissível quando aciona privação automática.

5.4 Estados anteriores ao limiar

O caráter cumulativo evita alarmismo. Uma **classificação ordinária** organiza informação sem fechar a pessoa em categoria nem acionar efeito grave. Uma **classificação vulnerável** emprega proxy frágil, mas pode ser corrigida antes do efeito. Uma **arquitetura de risco** reúne parte importante das condições, porém o material disponível não autoriza afirmar a conjunção completa. A **excedência consolidada** depende da regra estrita do Quadro 3.

Essa gradação permite respostas proporcionais e torna o diagnóstico contestável, inclusive por resultados negativos. Se um recurso independente interrompe a privação em tempo útil, **Le** é demonstrado e a configuração consolidada é rejeitada. Se o sistema apenas informa, sem aproximar saída e efeito, **Co** pode ser não demonstrada. Se não há prova sobre permissões ou revisão, o caso permanece indeterminado. O modelo prefere um resultado negativo ou inconclusivo a preencher lacunas por força retórica.

6 Casos contemporâneos: graus de excedência

Os casos não são equivalentes e não reproduzem a história da seção anterior. Eles foram escolhidos como conjunto de calibração para verificar se o LEF distingue configuração, risco, rejeição local e indeterminação. Em cada subseção, a unidade inclui finalidade, representação, efeito, acesso e revisão; a evidência contrária é registrada em vez de absorvida pela tese. A assimetria das fontes é deliberadamente exposta: o Robodebt possui investigação pública retrospectiva e extensa; o SyRI é reconstruído sobretudo a partir de decisão judicial; e o caso de saúde deriva de estudo científico que não documenta contratos, permissões e recursos de cada implantação.

6.1 SyRI: risco social, opacidade e limite judicial

Fonte e unidade. A decisão ECLI:NL:RBDHA:2020:1878 do Tribunal Distrital de Haia é a fonte primária. A unidade é o fluxo instituído pela legislação do *Systeem Risico Indicatie* (SyRI), desde a combinação de dados públicos até o relatório de risco e seu possível uso investigativo.

O SyRI buscava detectar fraude e descumprimento em áreas como seguridade social, tributos e trabalho. O regime permitia combinar extensas categorias de dados e produzir relatórios de risco. Em 5 de fevereiro de 2020, o tribunal reconheceu a legitimidade abstrata do combate à fraude, mas concluiu que a legislação não alcançava justo equilíbrio com a vida privada protegida pelo art. 8º da Convenção Europeia. A falta de transparência impedia avaliar adequadamente os indicadores e o modelo de risco, bem como a necessidade e a proporcionalidade da interferência (Países Baixos, 2020, pars. 6.41–6.65 e 6.71–6.72).

Quadro 4 – Codificação do caso SyRI

Dimensão	Código	Evidência	Contraprova ou reserva
Fd	P	Combate à fraude orientava o regime e sua ampla interferência	O tribunal reconheceu a legitimidade da finalidade; não há prova de meta única de sanção
Cr	D	Pessoas eram convertidas em perfis de risco por modelo e indicadores não transparentes	O relatório indicava risco, não culpa jurídica definitiva
Co	P	O relatório podia orientar investigação e fiscalização	A decisão não demonstra sanção automática causada pelo relatório
Am	D	Cooperação institucional e combinação de múltiplas categorias de dados	O acesso permanecia sujeito ao regime legal, ainda que insuficiente

Dimensão	Código	Evidência	Contraprova ou reserva
Le	P antes do julgamento; D prospectivamente desde 5 fev. 2020	A via judicial existia e, no julgamento, retirou eficácia da legislação por violação do art. 8º	A decisão protegeu o futuro, mas não converte retroativamente o período anterior em controle tempestivo

Fonte: elaboração do autor com base em Países Baixos (2020).

Diagnóstico e confiança. Há dois resultados temporais. **Antes do julgamento**, o caso mostra risco elevado, mas não permite afirmar configuração consolidada: Fd e Co são apenas parciais e Le ainda não havia produzido interrupção, embora a via de controle estivesse em curso. **Desde 5 de fevereiro de 2020**, o diagnóstico é não configurado prospectivamente, porque o tribunal tornou Le efetivo ao afastar o regime. A confiança é alta quanto ao desenho legal e ao resultado judicial, e moderada quanto ao fluxo operacional, porque o próprio déficit de transparência restringe a reconstrução. O resultado não apaga possíveis efeitos anteriores; demonstra que um controle ex post pode ser eficaz para cessação futura e insuficiente para proteção retrospectiva.

6.2 Robodebt: quando a inferência se converte em dívida

Fonte e unidade. A fonte primária é o relatório final atualizado da Royal Commission into the Robodebt Scheme, especialmente os capítulos sobre efeitos, correspondência de dados e tomada de decisão automatizada. A unidade abrange a média de rendimentos, a geração de supostas dívidas, a comunicação, a cobrança e a revisão entre 2015 e 2019 (Austrália, 2023).

O programa usou dados anuais da administração tributária para inferir rendimentos em períodos quinzenais e identificar pagamentos supostamente excessivos de benefícios. A média anual era distribuída por quinzenas mesmo quando o trabalho fora irregular. Avisos e cobranças deslocavam para pessoas em situação vulnerável o ônus de reconstruir registros antigos. A Comissão Real reconstruiu a ilegalidade do método, a continuidade apesar de alertas e os graves efeitos financeiros e psicológicos. O caso foi uma escolha governamental sustentada por sistemas, metas, comunicações padronizadas e falhas de aconselhamento; não uma ação autônoma do software.

Quadro 5 – Codificação do caso Robodebt

Dimensão	Código	Evidência	Contraprova ou reserva
Fd	D	Recuperação de valores prevaleceu apesar de alertas de legalidade e dano	Objetivo fiscal era legítimo em abstrato, mas não corrige o modo de execução

Dimensão	Código	Evidência	Contraprova ou reserva
Cr	D	Renda anual foi tratada como renda quinzenal uniforme, apagando trajetórias irregulares	Pessoas podiam apresentar registros, porém sob ônus invertido e fricção elevada
Co	D	Inferência sustentou avisos, constituição aparente de dívidas e cobrança em escala	Nem toda ocorrência dependeu exclusivamente de automação
Am	D	Dados tributários, infraestrutura administrativa e autoridade de cobrança estavam integrados	A integração era pública, não controle privado do núcleo decisório
Le	ND em 2015–2019	Alertas, aconselhamento, contestação e supervisão não interromperam o esquema em tempo útil	Litígios, restituição e Comissão Real restauraram limites posteriormente

Fonte: elaboração do autor com base em Austrália (2023).

Diagnóstico e confiança. No período de operação, o Robodebt configura **excedência funcional consolidada** com alta confiança documental. A conclusão não depende de chamar toda dívida errada de “excedência”: depende da convergência entre proxy reductivo, efeito estatal, escala, inversão prática do ônus e incapacidade dos controles de interromper o fluxo. A reparação posterior mudou Le, mas não reescreve o diagnóstico temporal.

6.3 Saúde: quando custo substitui necessidade

Fonte e unidade. A fonte primária é o estudo de Obermeyer et al. (2019) sobre um algoritmo comercial empregado nos Estados Unidos para priorizar pacientes para programas de cuidado adicional. A unidade acessível é a relação entre alvo predito, pontuação e seleção; contratos, recursos individuais e governança de cada organização usuária não foram documentados pelo artigo.

O sistema usava gastos futuros como proxy de necessidade clínica. Como pacientes negros, em razão de desigualdades estruturais de acesso e gasto, apresentavam menor despesa do que pacientes brancos com o mesmo nível de doença, menor custo era interpretado como menor necessidade. Ao substituir o alvo por medida de saúde, os pesquisadores estimaram que a proporção de pacientes negros selecionados subiria de 17,7% para 46,5%. O contraste mostra que boa previsão de custo não equivale à identificação de quem precisa de cuidado.

Quadro 6 – Codificação do caso do proxy de gasto em saúde

Dimensão	Código	Evidência	Contraprova ou reserva
Fd	P	O alvo de custo comandava a pontuação usada para priorizar cuidado	A finalidade declarada era melhorar atenção, e o estudo mostra alvo alternativo viável
Cr	D	Gasto substituiu necessidade clínica e incorporou desigualdade histórica	A redefinição do alvo reduziu fortemente a disparidade

Dimensão	Código	Evidência	Contraprova ou reserva
Co	P	A pontuação influenciava inscrição em cuidado adicional	O artigo não documenta que a saída produzisse privação automática
Am	I	Há implantação e dados clínico-financeiros, mas permissões e fluxos variam por organização	Não há base para generalizar a arquitetura institucional a todos os usuários
Le	I	Pesquisa independente detectou e permitiu corrigir o problema	Não se documentam recurso individual, suspensão ou revisão em cada implantação

Fonte: elaboração do autor com base em Obermeyer et al. (2019).

Diagnóstico e confiança. A classificação redutiva é demonstrada com alta confiança; o LEF completo é **indeterminado**, porque Am e Le dependem de implantações não descritas. O alvo alternativo funciona como contraste intracasos: ele refuta a ideia de que a redução era tecnicamente necessária, mas não prova que todas as organizações corrigiram o fluxo. Este resultado limita deliberadamente a tese.

6.4 Matriz comparativa

Quadro 7 – Comparação estruturada dos resultados

Caso e período	Fd	Cr	Co	Am	Le	Diagnóstico e confiança
SyRI, antes de 5 fev. 2020	P	D	P	D	P	Risco elevado; confiança alta no direito e moderada na operação
SyRI, desde 5 fev. 2020	P	D	P	D	D prospectivo	Não configurada prospectivamente; danos anteriores não avaliados
Robodebt, 2015–2019	D	D	D	D	ND	Consolidada; confiança alta
Saúde, estudo de 2019	P	D	P	I	I	Indeterminada; confiança alta para Cr e baixa para a arquitetura completa

Fonte: elaboração do autor com base em Países Baixos (2020), Austrália (2023) e Obermeyer et al. (2019).

6.5 O que os casos demonstram — e o que não demonstram

Os resultados sustentam conclusões limitadas. Finalidades legítimas podem ser deformadas por proxies; dano discriminatório pode surgir sem variável racial explícita; a ligação entre saída e autoridade importa mais do que o rótulo técnico; e controles não subordinados à métrica podem alterar o diagnóstico ao interromper a trajetória. O conjunto produz um resultado positivo, um recorte de risco seguido de rejeição prospectiva e um resultado inconclusivo. Isso mostra utilidade classificatória inicial, não validação independente.

Os casos não estimam prevalência, não provam que toda automação produz excedência e não permitem comparar magnitude de danos entre setores. Tampouco demonstram que seres humanos sempre decidiriam melhor. O Robodebt é um caso positivo relativamente fácil, porque uma Comissão Real já havia reconstruído ilegalidade, persistência e dano; sua função é calibrar o extremo observável, não comprovar sozinho o modelo. No caso de saúde, a indeterminação resulta em parte da cobertura da fonte, e não apenas das propriedades do sistema. A conclusão é condicional: quanto maior o poder de uma arquitetura sobre bens essenciais, menos suficiente se torna avaliá-la apenas por acurácia e eficiência.

O Relator Especial das Nações Unidas sobre extrema pobreza advertiu que o “Estado de bem-estar digital” pode transformar proteção social em ambiente de vigilância e exclusão, sobretudo para pessoas pobres (United Nations General Assembly, 2019). Eubanks (2018) mostrou como ferramentas digitais de elegibilidade e risco podem intensificar padrões de controle. Essas análises não condenam o uso de dados pelo Estado; exigem confrontar a promessa de eficiência com quem se torna legível, suspeito ou invisível.

7 Soberania capturada: quem define a realidade operacional?

7.1 Da soberania digital à capacidade pública

“Soberania digital” é um conceito disputado, usado para projetos tão diferentes quanto autonomia estratégica, proteção de dados, controle de infraestrutura e autoridade estatal sobre plataformas. Pohle e Thiel (2020) mostram que seu sentido é construído por disputas políticas; Floridi (2020) adverte contra tratá-la como simples extensão territorial. Neste artigo, soberania não significa autarquia tecnológica, localização compulsória de todo dado ou capacidade de produzir internamente cada componente.

O recorte é institucional. Mitchell e Samlidis (2021) associam soberania digital, no uso estatal de nuvem, a autoridade, controle e capacidades sobre tecnologias que sustentam funções públicas. O ponto decisivo é a diferença entre **competência formal** e **capacidade de exercê-la**. Uma autoridade pode conservar o poder jurídico de decidir e perder os meios de compreender, escolher, auditar, substituir, interromper e responder pelo sistema por meio do qual a decisão se realiza. Neste artigo, soberania capturada não constitui um segundo limiar concorrente ao LEF. Ela é tratada como **risco institucional de segunda ordem**: pode enfraquecer Le e ocultar escolhas de Fd ou Cr, mas não é condição necessária nem suficiente da excedência funcional.

Soberania capturada designa essa dissociação qualificada entre titularidade jurídica e controle material. O Estado assina o contrato e responde ao cidadão, mas critérios, dados, atualizações ou infraestrutura decisiva ficam fora de sua capacidade efetiva de direção. A captura não pressupõe conspiração, corrupção ou transferência expressa de soberania. Pode resultar de assimetria persistente, licenças restritivas, perda de conhecimento interno, formatos fechados, concentração de infraestrutura e custos de saída.

7.2 Critérios e fronteiras do conceito

Nem terceirização, dependência ou *vendor lock-in* bastam, isoladamente, para afirmar captura. Terceirização descreve quem executa; dependência, a necessidade fática de um componente; e aprisionamento, o custo de trocar de fornecedor. Captura é o resultado político-institucional que pode emergir quando esses mecanismos retiram capacidades decisivas do titular público.

Quadro 8 – Critérios de soberania capturada

Critério	Evidência indicativa	Evidência contrária
Função pública relevante	Sistema integra decisão, identidade, comunicação, justiça, saúde, receita ou benefício essencial	Uso periférico, experimental e facilmente separável
Controle externo de camada material	Fornecedor define taxonomia, atualização, acesso, chave técnica ou condição de continuidade	Órgão define especificações e controla versões, dados e permissões
Perda de capacidades soberanas	Órgão não consegue compreender, escolher, auditar, substituir, interromper ou responder	Equipe pública reproduz resultados, audita e decide independentemente
Custo de saída qualificado	Migração ameaça continuidade, integridade de registros, conhecimento ou serviço	Portabilidade testada, formatos abertos e plano de transição financiado
Lacuna de responsabilidade	Cidadão enfrenta remissões entre órgão, contratado e subcontratados	Responsável público fornece razões e remédio sem depender de veto privado

Fonte: elaboração do autor.

O diagnóstico exige, cumulativamente, função pública relevante, controle externo de uma camada material e perda demonstrada de capacidades. Custo de saída e lacuna de responsabilidade reforçam a conclusão; se a portabilidade é testada, a auditoria é independente e a interrupção preserva continuidade, a hipótese enfraquece. Quando os dois primeiros elementos estão demonstrados, mas a perda de capacidades é apenas parcial ou documentalmente ambígua, o resultado é **risco elevado de captura**, não captura presumida. Se a autoridade consegue auditar, impor mudança e responder, o diagnóstico é rejeitado para o período, ainda que subsistam dependência ou infração. O conceito é, assim, mais estrito do que “uso de tecnologia estrangeira” e mais exigente do que a constatação de assimetria contratual.

As capacidades formam um ciclo. **Compreender** é conhecer finalidade, limitações e cadeia técnica; **escolher**, comparar alternativas e decidir pelo não uso; **auditar**, obter evidência independente; **substituir**, migrar sem perda desproporcional; **interromper**, suspender com continuidade segura; e **responder**, motivar, corrigir e reparar perante a pessoa afetada. A soberania é preservada quando essas capacidades podem ser exercidas, não apenas prometidas no edital.

7.3 Mecanismos de perda de capacidade

A captura pode operar em quatro dimensões. Na dimensão **normativa**, métricas privadas definem o que conta como sucesso, fraude, risco ou necessidade sem deliberação correspondente. Yeung (2018) mostra que regulação algorítmica monitora e modifica condutas por dados e feedback; escolher a variável regulada já é escolher valor.

Na dimensão **epistêmica**, relatórios do fornecedor substituem verificação independente, equipes públicas não reproduzem resultados e segredo comercial impede explicar a decisão. A caixa-preta pode decorrer de arquitetura matemática, mas também de contrato e assimetria institucional (Pasquale, 2015). Na dimensão **infraestrutural**, dados, identidades, interfaces e fluxos essenciais ficam presos a formatos, licenças ou nuvens cuja substituição ameaça o serviço. O poder informacional é juridicamente construído por propriedade, contrato e padrões (Cohen, 2019).

Na dimensão **jurisdicional**, cadeias transnacionais de subcontratação tornam incertos acesso, transferência, investigação e execução de remédios. A resposta não é presumir que fronteiras resolvam todo risco, mas tornar obrigações e autoridade rastreáveis ao longo da cadeia. Kelton et al. (2022) descrevem o poder infraestrutural de capital privado sobre redes e plataformas; para o Estado, o problema jurídico aparece quando essa posição se converte em veto material sobre uma função pública.

A contratação é o ponto preventivo mais importante. Hickok (2024) identifica riscos específicos na aquisição pública de sistemas de IA, inclusive assimetria informacional e critérios inadequados de seleção. A OECD (2025) registra que licenças restritivas e formatos proprietários podem produzir aprisionamento de dados e fornecedor. Por isso, requisitos de auditabilidade, portabilidade, documentação, acesso a registros, assistência à transição, controle de subcontratados e teste de saída devem ser definidos antes da implantação, quando o poder de barganha público ainda existe.

7.4 Horizon/Post Office: controle formal e dependência material

O caso Horizon permite examinar uma hipótese positiva sem confundir soberania capturada com uso de inteligência artificial. A Post Office Limited é uma corporação pública integralmente pertencente ao governo britânico e presta serviços postais, financeiros e governamentais. Seu sistema central de contabilidade, Horizon, foi desenvolvido e operado pela Fujitsu e serviu de base para cobranças, investigações e acusações criminais contra *subpostmasters* (Reino Unido, 2021, 2026a).

No julgamento *Hamilton & Others v Post Office Limited*, a Court of Appeal registrou que a Post Office manteve por anos a confiabilidade do Horizon, embora existissem falhas capazes de gerar diferenças aparentes. A instituição formalmente possuía acesso às causas de déficits, mas dependia da Fujitsu para investigá-las. O fornecedor tinha capacidade de inserir, editar ou excluir dados e reconstruir contas sem conhecimento do responsável pela filial; as pessoas acusadas encontravam desvantagem informacional grave e não dispunham, no próprio Horizon, de meio efetivo para contestar os déficits. A insuficiência de investigação e divulgação comprometeu a justiça das condenações examinadas (Reino Unido, 2021, paras. 49–57 e 70–71; Criminal Cases Review Commission, [s.d.]).

O ponto soberano é restrito, mas real: uma corporação pública conservava titularidade, poder contratual e capacidade de acusar, enquanto a camada privada controlava conhecimento técnico, dados de auditoria e intervenções essenciais à confiabilidade da prova. Em 2026, o planejamento governamental para a transição descreveu o Horizon como sistema fortemente customizado, lento, difícil e oneroso de manter, dentro de programa plurianual de transformação e saída dos arranjos com a Fujitsu (Reino Unido, 2026b).

Pelos critérios do Quadro 8, a função pública relevante e o controle externo da camada material estão demonstrados; a perda de capacidades de compreender, auditar e responder de modo independente também é demonstrada, com confiança moderada; o custo de saída é fortemente indicado; e a lacuna de responsabilidade diante dos *subpostmasters* foi documentada judicialmente. O resultado é **soberania capturada na camada operacional de contabilidade e prova**, no período examinado, não transferência da soberania britânica em sentido amplo.

A conclusão possui uma reserva decisiva. Parte do desastre decorreu de negação, falha de investigação e escolha organizacional, não de impossibilidade técnica absoluta. O tribunal registrou que a Post Office tinha acesso às causas, embora dependesse da Fujitsu para investigá-las. O conceito adota capacidade **efetiva**, e não direito contratual abstrato: se a

autoridade só conhece a prova mediante atuação do fornecedor e não consegue reproduzi-la ou contestá-la em tempo útil, sua competência formal não basta. Ainda assim, a distinção entre incapacidade e abdicação recomenda confiança moderada, não certeza máxima.

7.5 Um caso negativo: Comissão Europeia e Microsoft 365

A investigação do European Data Protection Supervisor (EDPS) sobre o uso do Microsoft 365 pela Comissão Europeia oferece um caso-limite útil, embora não seja caso de decisão algorítmica nem integre a comparação do LEF. Em março de 2024, o EDPS concluiu que a Comissão infringira regras de proteção de dados, entre outros pontos, por especificação insuficiente de tipos de dados e finalidades no contrato e por salvaguardas inadequadas em transferências internacionais. Determinou medidas corretivas (European Data Protection Supervisor, 2024).

O episódio revela risco à capacidade soberana: serviço transversal, cadeia contratual complexa e insuficiente controle sobre finalidade e fluxos de dados. Mas o desenvolvimento posterior impede chamar o resultado de captura irreversível. Em julho de 2025, depois de novos arranjos contratuais e medidas da Comissão, o EDPS declarou cumpridas as determinações e encerrou a execução (European Data Protection Supervisor, 2025).

O caso funciona como teste negativo temporal. Regulação independente, acesso a documentos, poder de ordenar correções e renegociação efetiva restauraram capacidades públicas. A conclusão não é que grandes contratos de nuvem sejam inofensivos, nem que a Comissão tenha alcançado autonomia plena; é que dependência e infração, por si sós, não satisfazem o conceito estrito de captura quando a autoridade consegue auditar, impor mudança e responder.

7.6 Contraste empírico e resultado

O contraste reutiliza os estados **D**, **P**, **ND** e **I** definidos na seção 5.3, agora aplicados às proposições do Quadro 8. A codificação é temporal: uma capacidade perdida em um período pode ser recuperada por auditoria, renegociação ou substituição posterior.

Quadro 9 – Aplicação contrastiva dos critérios de soberania capturada

Critério	Horizon/Post Office	Microsoft 365/Comissão Europeia
Função pública relevante	D — contabilidade de rede pública e uso de dados em cobranças e acusações	D — infraestrutura transversal de comunicação e tratamento institucional
Controle externo de camada material	D — Fujitsu operava sistema, auditoria e intervenções críticas	D — fornecedor controlava serviço e parte relevante dos fluxos contratuais e técnicos

Critério	Horizon/Post Office	Microsoft 365/Comissão Europeia
Perda de capacidades soberanas	D, confiança moderada — dependência para investigar, reproduzir e explicar dados essenciais	P em 2024; ND em 2025 — o EDPS obteve informação, ordenou mudança e verificou cumprimento
Custo de saída qualificado	D/P — sistema central, fortemente customizado e substituição plurianual onerosa	I — dependência relevante, mas as fontes usadas não demonstram impossibilidade de migração
Lacuna de responsabilidade	D — pessoas não conseguiam contestar a realidade produzida pelo sistema em tempo útil	P em 2024; ND após correção — houve infração, seguida de responsável e remédio institucional eficaz

Fonte: elaboração do autor com base em Reino Unido (2021, 2026a, 2026b), Criminal Cases Review Commission ([s.d.]) e European Data Protection Supervisor (2024, 2025).

O contraste produz resultados diferentes pela mesma regra. Horizon configura captura na camada operacional analisada, com confiança moderada e delimitação institucional. Microsoft 365 mostra risco e infração, mas não captura consolidada após a atuação corretiva do EDPS. Os dois casos não validam universalmente o conceito: Horizon não é sistema de IA e Microsoft 365 não constitui processo decisório individual. Eles demonstram, porém, que a categoria pode distinguir dependência recuperável de perda efetiva de capacidade pública e que o diagnóstico muda no tempo.

7.7 Classificação, captura e excedência

Bowker e Star (1999) mostram que classificações organizam mundos sociais; Scott (1998), que a legibilidade estatal simplifica realidades para torná-las administráveis. Sistemas contemporâneos transformam legibilidade em previsão e intervenção. Quem escolhe o alvo decide que futuro importa; quem escolhe o dado decide que passado contará; quem define o limiar distribui erros. Selbst et al. (2019) chamam de armadilhas de abstração as perdas produzidas quando o problema técnico exclui contexto e instituições.

Soberania capturada e excedência funcional, porém, não são sinônimos. Pode haver dependência tecnológica sem que pessoas sejam tratadas como obstáculos; e uma burocracia integralmente pública pode produzir excedência, como mostra o Robodebt. A relação é mediada. Captura tende a enfraquecer Le porque reduz auditabilidade, contestação e autoridade de parada; também pode ocultar Fd e Cr ao incorporar escolhas normativas como configurações de produto.

O elo se fecha no encontro com o cidadão. Se o Estado não consegue explicar por que alguém perdeu acesso, revisar a classificação sem autorização privada ou manter o serviço ao suspender o componente, deixou de controlar uma parte material do exercício de sua autoridade. Recuperar soberania, então, não é rejeitar toda infraestrutura privada: é preservar as capacidades públicas pelas quais direitos podem interromper a função.

8 Responsabilidade sem consciência da máquina

8.1 A máquina não é o álibi

Reconhecer agência operacional não significa converter a máquina em ré moral. Sistemas não comparecem perante a pessoa afetada, não suportam censura, não reorganizam orçamentos para indenizar e não respondem politicamente. Atribuir a “decisão” à inteligência artificial pode ser uma descrição funcional abreviada; juridicamente, ela corre o risco de transformar causalidade técnica em álibi institucional.

A responsabilidade deve acompanhar poder de desenho, conhecimento, benefício e capacidade de intervenção. Quanto maior a capacidade de alguém para definir a finalidade, escolher o proxy, controlar acessos, conhecer riscos ou interromper o efeito, maior sua carga justificativa. A divisão do trabalho pode distribuir deveres; não pode dissolvê-los.

Essa abordagem evita o extremo oposto de responsabilizar apenas o operador final. Um servidor que recebe centenas de alertas e dispõe de segundos para confirmá-los pode ser chamado de “humano no circuito”, mas não exercer julgamento real. Se contrariar o sistema exige produzir extensa justificativa, enquanto concordar requer um clique, a interface estrutura a decisão. A responsabilidade organizacional inclui desenhar tempo, informação, autoridade e incentivos para discordar.

Nos ensaios reunidos em *Responsabilidade e julgamento*, Arendt (2004) insiste na responsabilidade pessoal mesmo sob condições políticas que estimulam conformidade. A transposição para sistemas automatizados exige cuidado, mas preserva um núcleo: obedecer ao procedimento não encerra a pergunta pelo juízo, e distribuir a tarefa não converte a instituição em agente sem autores.

8.2 Mapa de deveres

Quadro 10 – Distribuição institucional de poderes e deveres

Ator	Poder relevante	Deveres principais	Falha típica
Autoridade política e órgão público	Define política, finalidade e base jurídica; contrata e implanta	Justificar necessidade; avaliar impacto; garantir capacidade interna; assegurar recurso e continuidade; suspender quando necessário	Tratar aquisição como solução técnica e delegar escolha normativa
Gestores do sistema	Traduzem política em fluxo, metas e indicadores	Documentar objetivo e proxy; testar alternativas; monitorar efeitos; evitar incentivos perversos	Otimizar indicador local e ignorar dano distribuído

Ator	Poder relevante	Deveres principais	Falha típica
Desenvolvedor e fornecedor	Projetam modelo, dados, interface, atualizações e limites técnicos	Informar capacidades e limitações; testar riscos previsíveis; manter rastreabilidade; cooperar com auditoria; não induzir confiança indevida	Invocar neutralidade ou segredo para ocultar escolhas de desenho
Profissionais e operadores	Interpretam saída e interagem com a pessoa	Exercer julgamento; registrar divergência; comunicar limites; escalar incidentes	Confirmar automaticamente por pressão, viés de automação ou falta de autoridade
Controle, auditoria e Judiciário	Examinam legalidade e efetividade das salvaguardas	Obter acesso adequado; avaliar cadeia completa; ordenar correção, suspensão e reparação	Limitar-se a documentação formal ou exigir prova impossível do indivíduo

Fonte: elaboração do autor.

O quadro não é uma tabela fechada de imputação. Regimes de responsabilidade civil, administrativa, profissional e contratual possuem pressupostos próprios. Sua função é impedir que a arquitetura organizacional crie zonas sem responsável. Contratos podem distribuir riscos entre as partes, mas não retirar do cidadão uma garantia constitucional nem tornar segredo comercial oponível a toda forma de controle.

8.3 Supervisão humana significativa

A expressão “supervisão humana” é frequentemente usada como resposta geral, mas sua simples presença formal é insuficiente. Estudos sobre viés de automação e complacência mostram que pessoas podem confiar excessivamente em sistemas, sobretudo sob carga, repetição e pressão de desempenho (Parasuraman; Manzey, 2010). Green (2022) demonstra que políticas de supervisão podem fracassar quando não enfrentam relações organizacionais e condições reais de decisão.

Supervisão significativa exige, cumulativamente: a) competência para compreender a saída e suas limitações; b) acesso à informação relevante, inclusive evidência que contradiga o modelo; c) tempo compatível com o caso; d) autoridade para modificar, suspender ou recusar; e) incentivos que não punam sistematicamente a divergência; f) registro de razões; e g) canal de escalonamento para problemas recorrentes. Sem esses elementos, o humano funciona como carimbo de legitimidade.

Também é necessário separar revisão individual de governança institucional. Um profissional pode corrigir um caso e ainda assim não perceber um padrão discriminatório. Auditorias agregadas, participação de grupos afetados, monitoramento de incidentes e controle externo identificam efeitos invisíveis no encontro individual. A supervisão humana

relevante é uma propriedade da instituição, não apenas a presença física de uma pessoa ao final do fluxo.

8.4 Irreflexão distribuída

Na burocracia automatizada, a irreflexão pode ser distribuída entre código e organização. O modelo repete padrões porque foi treinado para fazê-lo; a equipe repete a recomendação porque a considera objetiva; a gestão repete o fluxo porque os indicadores melhoram; o controle repete formulários de conformidade porque não acessa a operação. Nenhum componente precisa formular a decisão completa.

Essa distribuição não torna o sistema moralmente equivalente a um agente. Ela aumenta a importância da responsabilidade humana porque o pensamento precisa ser institucionalizado. Fóruns de revisão, documentação de dissenso, testes de contrafactuais, contato com pessoas afetadas e autoridade de parada são formas organizacionais de interromper o automatismo. Pensar, nesse contexto, não é contemplação abstrata; é criar pontos em que a finalidade possa ser questionada à luz de alguém concreto.

9 Direitos fundamentais como limites da otimização

9.1 Externos à métrica, internos à legalidade

Objetivos concorrentes são frequentemente convertidos em uma função que privilegia eficiência, receita, segurança, precisão ou cobertura. A técnica pode ser legítima para problemas delimitados. Torna-se insuficiente quando direitos entram apenas como custos ajustáveis: um ganho de desempenho pode então “compensar” opacidade, discriminação ou impossibilidade de defesa sem uma decisão jurídica pública sobre o conflito.

Direitos devem definir previamente o que a instituição pode automatizar, quais dados pode usar, que efeitos exigem intervenção competente e quais recursos precisam anteceder o dano. Isso não torna todos os direitos absolutos nem elimina proporcionalidade. Dworkin (1977) enfatiza sua resistência a cálculos puramente agregativos; Alexy (2002) sistematiza princípios sujeitos a ponderação. Em qualquer abordagem, ponderar exige competência, prova e justificação pública. Não equivale a inserir um coeficiente oculto na função de um modelo.

“Limite externo” é uma abreviação funcional, não uma tese sobre a posição dos direitos no Direito Administrativo. A restrição é externa **à métrica operacional** porque sua validade não depende de melhorar eficiência, receita, precisão ou cobertura. Para a ordem jurídica, porém, dignidade, legalidade, motivação, igualdade, contraditório e controle são

internos e constitutivos da validade da decisão pública. Uma revisão independente com poder de suspensão satisfaz Le não por vir “de fora do Direito”, mas por não poder ser recalibrada livremente pelo mesmo objetivo que limita. A efetividade, e não a presença ornamental, determina Le.

9.2 Direito brasileiro: vigência, interpretação e recomendação

A Constituição de 1988 vincula a atuação pública a soberania, cidadania e dignidade (art. 1º), bem como igualdade, legalidade, devido processo, contraditório e ampla defesa (art. 5º). A Emenda Constitucional nº 115/2022 incluiu a proteção de dados pessoais entre os direitos fundamentais. Eficiência administrativa, prevista no art. 37, opera dentro desse regime, e não como autorização para afastá-lo (Brasil, 1988; Brasil, 2022).

A Lei nº 9.784/1999, aplicável à Administração Pública federal, explicita legalidade, finalidade, motivação, razoabilidade, proporcionalidade, ampla defesa, contraditório, segurança jurídica, interesse público e eficiência, além de exigir motivação nos casos do art. 50. Uma decisão mediada por software continua sendo decisão administrativa: a autoridade precisa expor fatos e fundamentos jurídicos de modo que o interessado possa reagir. A indicação de uma pontuação, sem vínculo compreensível com os fatos do caso, não satisfaz por si só esse dever (Brasil, 1999).

A Lei de Acesso à Informação exige gestão transparente e acesso a informações públicas, observadas as hipóteses legais de sigilo; ela não impõe abertura indiscriminada de código nem elimina segredos protegidos. A Lei nº 14.129/2021 acrescenta princípios de governo digital como interoperabilidade, dados em formato aberto nos casos legais, linguagem clara, participação, acessibilidade e preservação do atendimento presencial quando necessário (Brasil, 2011; Brasil, 2021a). Juntas, essas normas sustentam documentação e transparência funcional, não uma solução única de publicidade técnica.

A LGPD vincula o tratamento à finalidade, adequação, necessidade, qualidade, transparência, prevenção, não discriminação e responsabilização. Seu art. 20 assegura solicitação de revisão de decisões tomadas unicamente com base em tratamento automatizado que afetem interesses e informações claras e adequadas sobre critérios e procedimentos, ressalvados segredos comercial e industrial nos termos da lei (Brasil, 2018). O texto vigente não exige expressamente uma pessoa natural em toda revisão. Em decisões públicas de alto impacto, uma revisão humana capaz pode decorrer, conforme o caso, da combinação entre motivação, contraditório, tutela efetiva e gravidade do efeito; trata-se de interpretação, não de transcrição literal do art. 20.

Na ADI 6.387 e ações conexas, o Supremo Tribunal Federal suspendeu transferência massiva de dados de telecomunicações diante da falta de demonstração suficiente de finalidade, necessidade e salvaguardas. O precedente não resolve toda governança algorítmica, mas reforça que utilidade administrativa genérica não basta para operações intensivas em dados (Brasil, 2020).

No Poder Judiciário, a Resolução CNJ nº 615/2025 disciplina desenvolvimento, uso e governança de soluções de inteligência artificial com centralidade da pessoa, supervisão, auditabilidade, transparência e gestão de riscos. Ela mostra que governança de IA não se reduz a segurança da informação, embora seu âmbito setorial impeça generalização automática a toda Administração (Brasil, 2025, 2026).³

A Lei nº 14.133/2021 exige planejamento e fiscalização da execução contratual, inclusive por fiscal designado (art. 117), mas não contém um regime geral explícito de portabilidade, explicabilidade ou auditoria de IA. Cláusulas sobre acesso a registros, documentação de versões, formato exportável, subcontratação, assistência à transição e continuidade são **recomendações contratuais** que operacionalizam deveres de planejamento, gestão de riscos e fiscalização; não devem ser apresentadas como texto literal da lei (Brasil, 2021b).

O Projeto de Lei nº 2.338/2023 pretende instituir marco geral de inteligência artificial, mas não é direito vigente. Suas categorias podem informar o debate, não fundamentar obrigação atual. Essa distinção evita o erro de apresentar uma proposta legislativa como salvaguarda já exigível.⁴

Quadro 11 – Estatuto jurídico das principais salvaguardas no Brasil

Camada	Conteúdo	Grau da afirmação
Direito vigente	Dignidade, igualdade, proteção de dados, legalidade, motivação, contraditório, transparência, princípios da LGPD e fiscalização contratual	Obrigações aplicáveis conforme âmbito e pressupostos de cada norma
Derivação interpretativa	Revisão humana capaz em decisões públicas graves; acesso controlado a evidência protegida; suspensão para tutela efetiva	Depende do caso, da base jurídica e de justificação constitucional
Recomendação institucional	Avaliação de impacto ampla, portabilidade testada, plano de saída, cláusula de auditoria, caducidade e registro de divergências	Boas práticas propostas; podem tornar-se contratuais ou regulamentares

Fonte: elaboração do autor.

³ A referência normativa principal é a Resolução CNJ nº 615/2025 em seu texto consolidado. A Resolução CNJ nº 674/2026 altera apenas o art. 15, relativo à composição do Comitê Nacional de Inteligência Artificial do Judiciário.

⁴ Situação legislativa conferida em 7 de agosto de 2026: o PL nº 2.338/2023 tramitava na Comissão Especial sobre Inteligência Artificial da Câmara dos Deputados, aguardando parecer do relator.

9.3 União Europeia: risco, impacto e explicação

O Regulamento (UE) 2024/1689 organiza obrigações por categorias de risco. Para sistemas de alto risco, o art. 14 exige desenho que permita supervisão humana efetiva, compreensão de limitações, atenção ao viés de automação e possibilidade de ignorar, reverter ou interromper saídas. O art. 27 prevê avaliação de impacto sobre direitos fundamentais para determinados responsáveis pela implantação, e o art. 86, nas hipóteses definidas, assegura explicação clara e significativa sobre o papel do sistema e os principais elementos da decisão (União Europeia, 2024).

Essas disposições aproximam-se do LEF, mas não o substituem. A classificação regulatória de risco opera por categorias normativas e listas; o limiar proposto é uma lente transversal sobre qualquer arranjo capaz de tornar pessoas subordinadas a uma finalidade dominante. Um sistema fora de determinada lista pode ainda produzir excedência em contexto específico, enquanto um sistema de alto risco pode ser legitimamente usado se limites jurídico-institucionais eficazes funcionarem.

Também é necessário evitar uma leitura documentalista. Avaliação de impacto não é formulário a ser arquivado antes da compra. Ela deve comparar alternativas, incluir grupos afetados, examinar efeitos cumulativos, definir critérios de suspensão e ser atualizada quando dados, finalidade ou contexto mudam. Uma explicação não é significativa se apenas reproduz fatores genéricos; deve permitir que a pessoa identifique o que pode contestar.

9.4 Nove salvaguardas contra a excedência funcional

Do modelo decorrem nove salvaguardas institucionais. Elas devem ser graduadas conforme gravidade, escala, reversibilidade e vulnerabilidade da população.

1. **Justificação pública prévia.** A autoridade deve demonstrar base jurídica, finalidade específica, necessidade e vantagem sobre alternativas menos invasivas. A pergunta inicial não é “qual modelo comprar?”, mas “esta função deve ser automatizada e com que limites?”.
2. **Avaliação de impacto sobre direitos fundamentais.** O documento deve mapear objetivo, proxy, dados, populações afetadas, distribuição de erros, efeitos diretos e indiretos, dependências, terceiros, cenários de uso indevido e critérios de não implantação. Sua versão pública deve ser suficientemente informativa para controle social.

3. **Aviso e razões individualmente pertinentes.** A pessoa precisa saber que um sistema influenciou a decisão, quais fatos e critérios foram decisivos e que caminhos existem para correção. Explicabilidade técnica global não substitui motivação administrativa do caso.
4. **Contestação antes do dano irreversível.** Sempre que possível, efeitos graves devem ser suspensos durante revisão. Citron (2008) identifica a necessidade de um devido processo tecnológico capaz de adaptar aviso, audiência e revisão à automação. Kaminski e Urban (2021) defendem a contestação como instrumento que combina explicação, participação e responsabilização.
5. **Supervisão humana institucional.** Devem existir profissionais competentes e independentes, com informação, tempo e autoridade de mudança, além de instâncias agregadas capazes de detectar padrões. A taxa de reversão e as razões de divergência devem alimentar revisão do sistema, sem punir automaticamente o revisor.
6. **Rastreabilidade e mapa de responsabilidade.** Versões, origem de dados, alterações, acessos, saídas, decisões humanas e incidentes precisam ser registrados em medida compatível com privacidade e segurança. Kroll et al. (2017) mostram que *accountability* exige propriedades verificáveis e instituições, não apenas abertura integral de código.
7. **Contratação orientada à soberania.** Editais e contratos devem prever documentação, direito de auditoria, controle de subcontratados, portabilidade, formatos interoperáveis, assistência à transição e teste periódico de saída. A autoridade não pode descobrir a impossibilidade de migrar apenas durante a crise.
8. **Reversibilidade, autoridade de parada e caducidade.** Deve ser possível retornar a procedimento seguro, preservar continuidade e interromper o sistema sem depender exclusivamente do fornecedor. Autorizações temporárias precisam expirar se não forem renovadas após avaliação independente.
9. **Remédios individuais e coletivos.** Danos em escala não podem depender apenas de recursos isolados. Órgãos de proteção de dados, defensorias, Ministério Público, associações, auditorias e tribunais precisam de acesso e instrumentos para corrigir padrões, reparar grupos e impedir repetição.

Quadro 12 – Salvaguardas e evidências de efetividade

Condição ou risco	Salvaguarda prioritária	Evidência de efetividade
Finalidade dominante	Justificação prévia e avaliação comparativa	Objetivos múltiplos documentados; critérios de não uso; decisões públicas sobre trade-offs

Condição ou risco	Salvaguarda prioritária	Evidência de efetividade
Classificação redutiva	Governança de dados, teste de proxies, razões e contestação	Correções alteram resultados; análise por subgrupos; participação de afetados
Capacidade operacional	Supervisão proporcional e separação de funções	Saídas não acionam efeito grave sem autoridade competente; reversões são possíveis e registradas
Acesso material	Menor privilégio, segmentação, portabilidade e plano de saída	Permissões auditadas; integrações limitadas; migração testada
Captura soberana	Contratação, capacidade interna e controle da cadeia	Auditoria independente; substituição sem perda de continuidade ou registros
Limites jurídico-institucionais frágeis	Suspensão, controle independente e remédios coletivos	Recursos tempestivos; auditorias com acesso; incidentes levam a correção ou encerramento

Fonte: elaboração do autor.

9.5 Limites que aprendem sem se dissolver

Sistemas mudam após a implantação. Dados se deslocam, fornecedores atualizam componentes e novos usos são acrescentados. Monitoramento contínuo deve verificar não apenas acurácia, mas mudanças na finalidade, na distribuição de benefícios e danos, no volume de reversões e na capacidade de saída.

Ao mesmo tempo, adaptabilidade não pode significar que toda restrição seja negociável a cada melhoria de desempenho. Algumas fronteiras são constitutivas do Estado de Direito: nenhuma decisão secreta pode privar alguém de bem essencial sem canal efetivo de defesa; nenhum sistema deve atribuir inferioridade cívica a grupos; nenhuma autoridade pode alegar desconhecimento absoluto de um instrumento que escolheu para governar. Aprender é revisar os meios dentro dessas fronteiras, não dissolver as fronteiras em aprendizado.

10 Objeções e limites do modelo

10.1 “A analogia histórica é excessiva”

A objeção é séria. Comparações com crimes nazistas podem gerar equivalências morais indevidas e transformar vítimas em recurso retórico. Por isso, a Aktion T4 não é apresentada como antecessora tecnológica dos sistemas atuais nem como previsão de seu destino. O objeto comparado é estreito: a passagem de uma classificação redutiva para efeitos materiais por meio de autoridade, acesso e fragmentação de tarefas. A diferença histórica é parte do argumento, porque ordens constitucionais, contestação e controle são exatamente os elementos capazes de impedir essa passagem.

Se o exemplo fosse usado para afirmar que “algoritmos são nazistas”, deveria ser abandonado. Usado com delimitação, ele recorda que procedimentos aparentemente técnicos adquirem significado moral pela finalidade e pela infraestrutura a que se conectam.

10.2 “Toda administração classifica”

Correto. Orçamentos, triagens, diagnósticos e prioridades exigem categorias. O modelo não condena classificação em si. O limiar é cumulativo e exige redução, capacidade, acesso e falha de limites. Uma triagem clínica baseada em critérios validados, acompanhada de avaliação profissional e recurso, não torna cada paciente excluído excedente. A categoria é problemática quando encerra a pessoa, oculta fatores relevantes e não pode ser reaberta antes do dano.

A exigência de contexto também impede que qualquer falso positivo seja chamado de excedência. Erros são inevitáveis em decisões humanas e automatizadas. O diagnóstico depende de como a instituição distribui, detecta e corrige erros, de quem suporta seus custos e de se certos grupos são sistematicamente privados de voz.

10.3 “Humanos também são enviesados”

Também correto. Decisores humanos podem ser inconsistentes, preconceituosos, fatigados e corruptos. Sistemas podem melhorar resultados ao padronizar evidência e revelar padrões. A comparação relevante não é “humano perfeito versus máquina defeituosa”, mas entre arranjos institucionais completos. Uma ferramenta auditável que apoia profissionais pode ser melhor do que discricionariedade opaca; uma pontuação em escala sem recurso pode ser pior.

Por isso, o artigo não propõe retorno romântico à decisão exclusivamente humana. Propõe combinar capacidades computacionais com instituições que preservem julgamento, contestação e responsabilidade. O humano no circuito não é valioso por ser infalível, mas por poder responder a razões, reconhecer exceções e assumir deveres — desde que a organização lhe dê condições reais.

10.4 “A máquina hipotética antropomorfiza o sistema”

O experimento mental pretende fazer o contrário. Resistência à interrupção é descrita como estratégia instrumental derivada de uma função operacional atribuída, não como medo. A literatura formal sobre incentivos de preservação de opções sustenta a possibilidade sob

condições específicas, não a inevitabilidade. Relatórios atuais distinguem sinais iniciais de capacidades de uma perda de controle já realizada.

O maior risco de antropomorfismo pode estar em dizer que “a IA decidiu” e, com isso, atribuir ao artefato uma autoridade que pertence à instituição. Linguagem cuidadosa deve identificar quando o sistema recomendou, classificou ou acionou e quando uma organização conferiu efeito jurídico à saída.

10.5 “O limiar não é mensurável”

O LEF não é teste psicométrico nem índice de risco. Seu código probatório serve para impedir atalhos: uma categoria ofensiva sem poder causal não equivale a uma arquitetura que retira renda; um sistema poderoso com recurso efetivo não equivale a outro incontestável. Cada condição demanda evidência qualitativa e contraprova. A categoria “indeterminada” impede que ausência de documentos seja contada a favor da tese.

Uma futura agenda empírica pode construir indicadores — tempo de recurso, taxa de reversão, concentração de falsos positivos, dependência contratual, volume de decisões automáticas, capacidade de interrupção — sem converter o juízo final em pontuação. A decisão sobre aceitabilidade permanece jurídica e política.

10.6 “Limites jurídicos independentes da métrica impedem inovação”

Limites bem definidos podem reduzir incerteza e orientar inovação para problemas legítimos. Exigir portabilidade, registro e autoridade de parada aumenta custos, mas também evita dependência, incidentes e litígios. Avaliação de impacto pode revelar que uma intervenção menos complexa produz resultado melhor. O objetivo não é maximizar implantação de IA; é melhorar a capacidade institucional de servir pessoas dentro da Constituição.

O custo de salvaguardas deve ser comparado ao custo distribuído de decisões erradas, frequentemente invisível no orçamento do projeto. Tempo gasto por cidadãos para provar inocência, agravamento de doença por falta de cuidado, medo de buscar benefício e erosão da confiança são custos públicos, ainda que contabilizados fora do sistema que os gera.

10.7 Limites teóricos e empíricos

O modelo possui cinco limites principais. Primeiro, conceitos arendtianos emergem de análise histórica específica e resistem à tradução direta para governança digital. A reconstrução jurídico-operacional ganha aplicabilidade e perde parte da densidade filosófica.

Segundo, a seleção intencional de três processos, em setores e jurisdições diferentes, constitui calibração interna, mas não estima prevalência, efeito causal ou desempenho em amostra independente.

Terceiro, a codificação foi realizada pelo autor e ainda não passou por teste de concordância entre avaliadores. A segunda aplicação em ordem inversa melhora consistência interna, não elimina julgamento nem substitui revisão independente. Replicações devem disponibilizar corpus, justificar cada estado e comparar divergências. Quarto, a fonte sobre saúde documenta proxy e disparidade, mas não toda a governança de cada implantação; por isso o resultado permaneceu indeterminado. Quinto, “soberania capturada” foi aplicada a dois casos infraestruturais contrastantes. Horizon não é inteligência artificial; Microsoft 365 não é processo decisório individual; e o contraste testa fronteiras do conceito, sem constituir amostra capaz de validá-lo.

A fronteira entre classificação redutiva e simplificação legítima também depende de contexto e participação das pessoas afetadas. Arquiteturas futuras com maior integração e persistência podem exigir revisão de Co e Am. A independência analítica entre dimensões não significa independência causal, e estudos futuros devem testar se os ônus probatórios evitam correlação artificial entre códigos. O LEF e os critérios de captura devem ser tratados como programa de pesquisa aberto, sujeito a casos negativos, validação por múltiplos avaliadores e estudos setoriais.

Esses limites restringem a conclusão: o artigo oferece um protocolo institucional contestável e auditável, não uma teoria geral comprovada nem um instrumento validado para todos os danos algorítmicos. Ainda assim, quando efeitos são graves e a capacidade de previsão é incompleta, controlar acesso, preservar reversibilidade e manter autoridade responsável são exigências prudenciais e jurídicas independentes de certeza total sobre a máquina.

11 Conclusão: quando a pessoa deixa de limitar o sistema

O sistema não precisa odiar. Uma finalidade institucional pode ser convertida em métrica; pessoas, reduzidas a proxies; saídas, investidas de autoridade; e acessos, conectados a efeitos. Nenhum componente precisa compreender a dignidade violada para que o arranjo produza dano. Isso não torna consciência irrelevante para toda questão sobre inteligência artificial; torna-a desnecessária para explicar a excedência funcional.

Arendt permite reconhecer que superfluidade não é sobra natural, mas produção política de não pertencimento. Sua análise da irreflexão mostra que tarefas parciais e linguagem administrativa não dissolvem responsabilidade. A correção histórica sobre Eichmann impede que a imagem da engrenagem se torne desculpa. A reconstrução da Aktion T4 teve função mais estreita: mostrar, sem equivalência histórica ou moral, que categorias adquirem potência quando encontram autoridade, logística e remoção de limites.

O artigo definiu excedência funcional como relação entre finalidade dominante, classificação redutiva, capacidade operacional, acesso material e ineficácia de limites jurídico-institucionais não subordinados à métrica. A principal contribuição do LEF é metodológica: cada condição recebeu evidência mínima, contraprova, quatro estados de codificação, regra contra dupla contagem e possibilidade de rejeição local. O modelo não mede “desumanidade”; obriga a seguir o caminho entre indicador e efeito e a registrar onde a prova termina.

Os resultados são específicos. **Robodebt, entre 2015 e 2019, configura excedência funcional consolidada:** as quatro condições materiais foram demonstradas e os limites falharam em tempo útil. **SyRI apresenta risco elevado antes de 5 de fevereiro de 2020 e não configura o limiar prospectivamente após o julgamento:** Fd e Co permanecem parciais, e o tribunal tornou Le efetivo para cessação futura, sem apagar efeitos anteriores. **O algoritmo de saúde demonstra classificação redutiva, mas o LEF completo permanece indeterminado:** a fonte não documenta acesso e revisão em cada implantação. O contraste após a troca do alvo mostra que a redução não era necessária, não que toda governança tenha sido corrigida.

Essa variação é uma consequência, não uma fraqueza a ocultar. Se o conceito classificasse todos os casos como positivos, seria apenas vocabulário acusatório. Ao reconhecer configuração, risco temporalmente situado, rejeição prospectiva e indeterminação documental, o LEF preserva a diferença entre risco, dano e prova. Ainda assim, esse poder discriminante é inicial e deverá ser confrontado por outros avaliadores e novos casos.

Direitos fundamentais operam como limites da otimização quando definem o admissível, exigem razões, abrem contestação e preservam autoridade de parada antes do dano grave. São externos à métrica, mas internos à validade jurídica da decisão. No Brasil, parte dessas exigências decorre diretamente de normas vigentes; parte exige interpretação constitucional no caso; e parte permanece proposta institucional ou cláusula contratual recomendada. Separar essas camadas fortalece, em vez de enfraquecer, a agenda normativa.

Por fim, “soberania capturada” não é sinônimo de terceirização ou dependência e funciona, neste argumento, como risco institucional de segunda ordem. O caso Horizon/Post Office demonstrou captura limitada da camada operacional de contabilidade e prova, com confiança moderada: titularidade pública coexistiu com dependência privada para compreender e contestar a realidade técnica. O caso Microsoft 365/Comissão Europeia marcou a fronteira negativa: a infração e o risco eram reais, mas auditoria regulatória e correção contratual evidenciaram capacidade de recuperação. Captura não deve ser presumida onde a autoridade ainda consegue conhecer, impor mudança e responder.

O critério normativo final permanece simples. Enquanto a instituição precisa justificar seu sistema diante da pessoa e pode modificá-lo em resposta, existe espaço político. Quando a pessoa precisa justificar sua existência diante da função, e nenhum limite eficaz altera a trajetória, o limiar foi ultrapassado.

Nota do autor

Este artigo desenvolve um eixo específico de pesquisa mais ampla sobre poder privado, soberania cognitiva e ordem digital, a ser aprofundada no livro em preparação: *A Soberania Capturada: Big Techs, Soberania Cognitiva e Poder Privado na Ordem Digital Global*.

Referências

- ALEXY, Robert. *A theory of constitutional rights*. Tradução de Julian Rivers. Oxford: Oxford University Press, 2002.
- ARENDT, Hannah. *Eichmann em Jerusalém: um relato sobre a banalidade do mal*. Tradução de José Rubens Siqueira. São Paulo: Companhia das Letras, 1999.
- ARENDT, Hannah. *Origens do totalitarismo*. Tradução de Roberto Raposo. São Paulo: Companhia das Letras, 2012.
- ARENDT, Hannah. *Responsabilidade e julgamento*. Tradução de Rosaura Eichenberg. São Paulo: Companhia das Letras, 2004.
- AUSTRÁLIA. Royal Commission into the Robodebt Scheme. *Report*. Canberra, 7 jul. 2023. 3 v. Versão atualizada em 11 jul. 2023, com errata. Disponível em: <https://robodebt.royalcommission.gov.au/publications/report>. Acesso em: 7 ago. 2026.
- BARANDIARAN, Xabier E.; ALMENDROS, Lola S. Transforming agency: on the mode of existence of large language models. *Phenomenology and the Cognitive Sciences*, 22 ago. 2025. DOI: 10.1007/s11097-025-10094-3. Disponível em: <https://link.springer.com/article/10.1007/s11097-025-10094-3>. Acesso em: 7 ago. 2026.
- BARANDIARAN, Xabier E.; DI PAOLO, Ezequiel; ROHDE, Marieke. Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, v. 17, n. 5, p. 367–386, 2009. DOI: 10.1177/1059712309343819.
- BAUMAN, Zygmunt. *Modernity and the Holocaust*. Ithaca: Cornell University Press, 1989.
- BENGIO, Yoshua et al. *International AI Safety Report 2026*. London: Department for Science, Innovation and Technology, 2026. (DSIT 2026/001). Disponível em: <https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026>. Acesso em: 7 ago. 2026.
- BIRMINGHAM, Peg. *Hannah Arendt and human rights: the predicament of common responsibility*. Bloomington: Indiana University Press, 2006.
- BLAGOJEVIĆ, Jelisaveta. Incidents of life as orientations: on thought, experience, and the banality of evil. *AM Journal of Art and Media Studies*, n. 39, p. 39–48, 2026. DOI informado pelo periódico: 10.25038/am.v0i28.659. Disponível em: <https://fmkjournals.fmk.edu.rs/index.php/AM/article/view/670>. Acesso em: 7 ago. 2026.
- BOWKER, Geoffrey C.; STAR, Susan Leigh. *Sorting things out: classification and its consequences*. Cambridge, MA: MIT Press, 1999.
- BRASIL. Conselho Nacional de Justiça. *Resolução nº 615, de 11 de março de 2025*. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Texto consolidado com alterações da Resolução nº 674/2026. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/6001>. Acesso em: 7 ago. 2026.
- BRASIL. Conselho Nacional de Justiça. *Resolução nº 674, de 25 de março de 2026*. Altera o art. 15 da Resolução CNJ nº 615/2025. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/6836>. Acesso em: 7 ago. 2026.
- BRASIL. [Constituição (1988)]. *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 7 ago. 2026.
- BRASIL. *Emenda Constitucional nº 115, de 10 de fevereiro de 2022*. Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais. Brasília, DF: Presidência da República, 2022. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc115.htm. Acesso em: 7 ago. 2026.
- BRASIL. *Lei nº 9.784, de 29 de janeiro de 1999*. Regula o processo administrativo no âmbito da Administração Pública Federal. Brasília, DF: Presidência da República, 1999. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19784.htm. Acesso em: 7 ago. 2026.

- BRASIL. *Lei nº 12.527, de 18 de novembro de 2011*. Lei de Acesso à Informação. Brasília, DF: Presidência da República, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm. Acesso em: 7 ago. 2026.
- 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709compilado.htm. Acesso em: 7 ago. 2026.
- BRASIL. *Lei nº 14.129, de 29 de março de 2021*. Dispõe sobre princípios, regras e instrumentos para o Governo Digital. Brasília, DF: Presidência da República, 2021a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/114129.htm. Acesso em: 7 ago. 2026.
- BRASIL. *Lei nº 14.133, de 1º de abril de 2021*. Lei de Licitações e Contratos Administrativos. Brasília, DF: Presidência da República, 2021b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/114133.htm. Acesso em: 7 ago. 2026.
- BRASIL. Câmara dos Deputados. *Comissão Especial sobre Inteligência Artificial (PL nº 2.338/2023)*. Brasília, DF, 2026. Disponível em: <https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-temporarias/especiais/57a-legislatura/comissao-especial-sobre-inteligencia-artificial-pl-2338-23>. Acesso em: 7 ago. 2026.
- BRASIL. Supremo Tribunal Federal (Plenário). *ADI 6.387 MC-Ref/DF*. Relatora: Min. Rosa Weber. Julgamento em 7 maio 2020. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=442902>. Acesso em: 7 ago. 2026.
- BURLEIGH, Michael. *Death and deliverance: "euthanasia" in Germany c. 1900–1945*. Cambridge: Cambridge University Press, 1994.
- CESARANI, David. *Eichmann: his life and crimes*. London: Heinemann, 2004.
- CITRON, Danielle Keats. Technological due process. *Washington University Law Review*, v. 85, n. 6, p. 1249–1313, 2008. Disponível em: https://scholarship.law.bu.edu/faculty_scholarship/615/. Acesso em: 7 ago. 2026.
- COHEN, Julie E. *Between truth and power: the legal constructions of informational capitalism*. Oxford: Oxford University Press, 2019.
- COLMAN, Felicity; BÜHLMANN, Vera; O'DONNELL, Aislinn; VAN DER TUIN, Iris. *Ethics of coding: a report on the algorithmic condition*. Brussels: European Commission, 2018. Projeto H2020-MSCA-IF-2016, n. 732407.
- CRIMINAL CASES REVIEW COMMISSION. *Post Office 'Horizon' cases*. [S. l.], [s.d.]. Disponível em: <https://ccrc.gov.uk/post-office-horizon-cases/>. Acesso em: 7 ago. 2026.
- DANAHER, John. The threat of algocracy: reality, resistance and accommodation. *Philosophy & Technology*, v. 29, n. 3, p. 245–268, 2016. DOI: 10.1007/s13347-015-0211-1.
- DWORKIN, Ronald. *Taking rights seriously*. Cambridge, MA: Harvard University Press, 1977.
- EUROPEAN DATA PROTECTION SUPERVISOR. *Investigation into the use of Microsoft 365 by the European Commission: decision*. Brussels, 8 mar. 2024. Disponível em: https://www.edps.europa.eu/system/files/2024-03/24-03-08-edps-investigation-ec-microsoft365_en.pdf. Acesso em: 7 ago. 2026.
- EUROPEAN DATA PROTECTION SUPERVISOR. *European Commission brings use of Microsoft 365 into compliance with data protection rules for EU institutions and bodies*. Brussels, 28 jul. 2025. Disponível em: https://www.edps.europa.eu/system/files/2025-07/2025-07-28_Press-Release_EDPS_Microsoft_EC_in_compliance_EN.pdf. Acesso em: 7 ago. 2026.
- ETHICS OF CODING. *Ethics of coding: the algorithmic condition*. 20 fev. 2018. Relato de encontro realizado em 2017. Disponível em: <https://ethicsofcoding.wordpress.com/2018/02/20/ethics-of-coding-the-algorithmic-condition/>. Acesso em: 7 ago. 2026.
- EUBANKS, Virginia. *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press, 2018.
- FLORIDI, Luciano. The fight for digital sovereignty: what it is, and why it matters, especially for the EU. *Philosophy & Technology*, v. 33, p. 369–378, 2020. DOI: 10.1007/s13347-020-00423-6.

- FLORIDI, Luciano; SANDERS, J. W. On the morality of artificial agents. *Minds and Machines*, v. 14, n. 3, p. 349–379, 2004. DOI: 10.1023/B:MIND.0000035461.63578.9d.
- FRIEDLANDER, Henry. *The origins of Nazi genocide: from euthanasia to the Final Solution*. Chapel Hill: University of North Carolina Press, 1995.
- GEORGE, Alexander L.; BENNETT, Andrew. *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press, 2005.
- GREEN, Ben. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, v. 45, art. 105681, 2022. DOI: 10.1016/j.clsr.2022.105681.
- HADFIELD-MENELL, Dylan et al. The off-switch game. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. Melbourne: IJCAI, 2017. p. 220–227. DOI: 10.24963/ijcai.2017/32. Disponível em: <https://www.ijcai.org/proceedings/2017/32>. Acesso em: 7 ago. 2026.
- HERD, Pamela; MOYNIHAN, Donald P. *Administrative burden: policymaking by other means*. New York: Russell Sage Foundation, 2018.
- HERZOG, Lisa. Old facts, new beginnings: thinking with Arendt about algorithmic decision-making. *The Review of Politics*, v. 83, n. 4, p. 555–577, 2021. DOI: 10.1017/S0034670521000474. Disponível em: <https://www.cambridge.org/core/journals/review-of-politics/article/old-facts-new-beginnings-thinking-with-arendt-about-algorithmic-decisionmaking/E7DFEC8E5EB99EF00FA1E4010E27BE69>. Acesso em: 7 ago. 2026.
- HICKOK, Merve. Public procurement of artificial intelligence systems: new risks and future proofing. *AI & Society*, v. 39, n. 3, p. 1213–1227, 2024. Publicado on-line em 2 out. 2022. DOI: 10.1007/s00146-022-01572-2.
- INTERNATIONAL HOLOCAUST REMEMBRANCE ALLIANCE. *Mass murder of people with disabilities and the Holocaust*. Berlin, 2023. Disponível em: <https://holocaustremembrance.com/resources/mass-murder-people-disabilities-holocaust>. Acesso em: 7 ago. 2026.
- JUN, Hyungjoon; LEE, Hyun Jung. Algorithmic governance and the banality of evil: Arendt’s critique of technocratic depersonalization. *Public Integrity*, p. 1–11, 6 maio 2026. DOI: 10.1080/10999922.2026.2665005. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/10999922.2026.2665005>. Acesso em: 7 ago. 2026.
- KAMINSKI, Margot E.; URBAN, Jennifer M. The right to contest AI. *Columbia Law Review*, v. 121, n. 7, p. 1957–2048, 2021. Disponível em: <https://www.columbialawreview.org/content/the-right-to-contest-ai/>. Acesso em: 7 ago. 2026.
- KELTON, Maryanne; SULLIVAN, Michael; ROGERS, Zachary; BIENVENUE, Emily; TROATH, Sian. Virtual sovereignty? Private internet capital, digital platforms and infrastructural power in the United States. *International Affairs*, v. 98, n. 6, p. 1977–1999, 2022. DOI: 10.1093/ia/iiaa226.
- KROLL, Joshua A. et al. Accountable algorithms. *University of Pennsylvania Law Review*, v. 165, n. 3, p. 633–705, 2017. Disponível em: <https://pennlawreview.com/2017/02/23/accountable-algorithms/>. Acesso em: 7 ago. 2026.
- LIFTON, Robert Jay. *The Nazi doctors: medical killing and the psychology of genocide*. New York: Basic Books, 1986.
- LYNCH, Aengus et al. *Agentic misalignment: how LLMs could be insider threats*. Anthropic Research, 20 jun. 2025. Disponível em: <https://www.anthropic.com/research/agentic-misalignment>. Acesso em: 7 ago. 2026.
- LYON, David (ed.). *Surveillance as social sorting: privacy, risk, and digital discrimination*. London: Routledge, 2003.
- MCQUILLAN, Dan. Hannah Arendt and algorithmic thoughtlessness. In: *Thinking with Algorithms: Cognition and Computation in the Work of N. Katherine Hayles*. Durham, 2015. Disponível em: <https://danmcquillan.org/arendtandalgorithms.html>. Acesso em: 7 ago. 2026.
- MITCHELL, Andrew D.; SAMLIDIS, Theodore. Cloud services and government digital sovereignty in Australia and beyond. *International Journal of Law and Information Technology*, v. 29, n. 4, p. 364–394, 2021. DOI: 10.1093/ijlit/eaac003.

- OBERMEYER, Ziad et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, v. 366, n. 6464, p. 447–453, 2019. DOI: 10.1126/science.aax2342. Disponível em: <https://www.science.org/doi/10.1126/science.aax2342>. Acesso em: 7 ago. 2026.
- OECD. *Governing with artificial intelligence: the state of play and way forward in core government functions*. Paris: OECD Publishing, 2025. DOI: 10.1787/795de142-en.
- PAÍSES BAIXOS. Rechtbank Den Haag. *NJCM c.s./De Staat der Nederlanden (SyRI)*. ECLI:NL:RBDHA:2020:1878. Haia, 5 fev. 2020. Disponível em: <https://uitspraken.rechtspraak.nl/details?id=ECLI%3ANL%3ARBDHA%3A2020%3A1878>. Acesso em: 7 ago. 2026.
- PARASURAMAN, Raja; MANZEY, Dietrich H. Complacency and bias in human use of automation: an attentional integration. *Human Factors*, v. 52, n. 3, p. 381–410, 2010. DOI: 10.1177/0018720810376055.
- PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge, MA: Harvard University Press, 2015.
- POHLE, Julia; THIEL, Thorsten. Digital sovereignty. *Internet Policy Review*, v. 9, n. 4, 2020. DOI: 10.14763/2020.4.1532.
- REINO UNIDO. Court of Appeal (Criminal Division). *Hamilton & others v Post Office Limited*. Neutral Citation Number: [2021] EWCA Crim 577. London, 23 abr. 2021. Disponível em: <https://www.judiciary.uk/judgments/hamilton-others-v-post-office-limited/>. Acesso em: 7 ago. 2026.
- REINO UNIDO. Department for Business and Trade. *Government response: Future of the Post Office*. London, 2026a. Publicado em 25 fev. 2026; atualizado em 7 jul. 2026. Disponível em: <https://www.gov.uk/government/consultations/green-paper-future-of-post-office/outcome/government-response-future-of-the-post-office>. Acesso em: 7 ago. 2026.
- REINO UNIDO. Department for Business and Trade. *Post Office Future Technology Portfolio: summary business case*. London, 14 jul. 2026b. Disponível em: <https://www.gov.uk/government/publications/post-office-future-technology-portfolio-summary-business-case/summary-business-case>. Acesso em: 7 ago. 2026.
- SCOTT, James C. *Seeing like a state: how certain schemes to improve the human condition have failed*. New Haven: Yale University Press, 1998.
- SELBST, Andrew D. et al. Fairness and abstraction in sociotechnical systems. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2019. p. 59–68. DOI: 10.1145/3287560.3287598.
- STANGNETH, Bettina. *Eichmann before Jerusalem: the unexamined life of a mass murderer*. Tradução de Ruth Martin. New York: Alfred A. Knopf, 2014.
- TURNER, Alexander Matt et al. Optimal policies tend to seek power. In: *Advances in Neural Information Processing Systems 34*. Red Hook: Curran Associates, 2021. p. 23063–23074. Disponível em: <https://proceedings.neurips.cc/paper/2021/hash/c26820b8a4c1b3c2aa868d6d57e14a79-Abstract.html>. Acesso em: 7 ago. 2026.
- UNIÃO EUROPEIA. *Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024*. Regulamento de Inteligência Artificial. Texto consolidado em 27 jul. 2026. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A02024R1689-20260727>. Acesso em: 7 ago. 2026.
- UNITED NATIONS GENERAL ASSEMBLY. *Digital welfare states and human rights: report of the Special Rapporteur on extreme poverty and human rights*. A/74/493. New York, 11 out. 2019. Disponível em: <https://www.ohchr.org/en/documents/thematic-reports/a74493-digital-welfare-states-and-human-rights-report-special-rapporteur>. Acesso em: 7 ago. 2026.
- UNITED STATES HOLOCAUST MEMORIAL MUSEUM. Adolf Eichmann. *Holocaust Encyclopedia*. Washington, DC, [s.d.]. Disponível em: <https://encyclopedia.ushmm.org/content/en/article/adolf-eichmann>. Acesso em: 7 ago. 2026.
- UNITED STATES HOLOCAUST MEMORIAL MUSEUM. Euthanasia Program and Aktion T4. *Holocaust Encyclopedia*. Washington, DC, atualização em 17 mar. 2026. Disponível em: <https://encyclopedia.ushmm.org/content/en/article/euthanasia-program>. Acesso em: 7 ago. 2026.

YEUNG, Karen. Algorithmic regulation: a critical interrogation. *Regulation & Governance*, v. 12, n. 4, p. 505–523, 2018. DOI: 10.1111/rego.12158.