

A IA NÃO PRECISA QUERER PODER

Convergência instrumental sociotécnica, influência cognitiva estrutural e captura da decisão

AI Does Not Need to Want Power: Sociotechnical Instrumental Convergence, Structural Cognitive Influence, and the Capture of Decision-Making

Jandislei Antonio Genova

Advogado, autor e pesquisador independente em Direito, inteligência artificial e governança institucional.

Resumo

Este artigo investiga se a convergência instrumental, formulada no campo da segurança de agentes artificiais, pode contribuir para a análise jurídica e institucional de ecossistemas de inteligência artificial sem antropomorfizar modelos, empresas ou Estados. Trata-se de ensaio teórico-jurídico de construção conceitual e análise normativa, apoiado em revisão narrativa dirigida e aplicação documental exploratória. O corpus reúne literatura de segurança de IA, economia política, teoria organizacional e soberania digital; normas e referenciais de governança; e relatórios oficiais sobre parcerias entre provedores de nuvem e desenvolvedores de modelos. Propõe-se distinguir a **convergência instrumental sociotécnica**, categoria descritiva neutra sobre meios transversais, de sua forma **crítica**, caracterizada pela redução de contestabilidade, reversibilidade ou saída. Define-se ainda um limiar próprio para a captura da decisão. Como instrumentos analíticos, apresentam-se o Teste de Convergência Instrumental Sociotécnica — TCIS, com unidade de análise, classes de evidência, regras de decisão e contraprovas, e o **postulado regulatório da instrumentalidade mínima necessária**. A aplicação exploratória aos relatórios da Federal Trade Commission e da Competition and Markets Authority sustenta a presença de convergência e de risco crítico em parcerias de nuvem e IA, mas não demonstra, por si, captura. Conclui-se que sistemas não precisam “querer” poder para integrar arquiteturas que o acumulam; a inferência, contudo, é configuracional, graduada e dependente de prova.

Palavras-chave: inteligência artificial; convergência instrumental; influência cognitiva estrutural; governança algorítmica; soberania tecnológica.

Abstract

This article examines whether instrumental convergence, as developed in AI safety research, can inform the legal and institutional analysis of artificial-intelligence ecosystems without anthropomorphizing models, corporations, or states. It is a theoretical legal essay in concept-building and normative analysis, supported by a directed narrative review and an exploratory documentary application. The corpus combines scholarship on AI safety, political economy, organizational theory, and digital sovereignty; legal and governance instruments; and official reports on partnerships between cloud providers and model developers. The article distinguishes **sociotechnical instrumental convergence**, a neutral descriptive category concerning cross-cutting means, from its **critical** form, which reduces contestability, reversibility, or exit. It also defines a separate threshold for decision capture. Two analytical tools are proposed: the Sociotechnical Instrumental Convergence Test — SICT, with a specified unit of analysis, evidence classes, decision rules, and counterevidence; and the **regulatory postulate of minimum necessary instrumentality**. An exploratory application to reports issued by the Federal Trade Commission and the Competition and Markets Authority supports a finding of convergence and critical risk in cloud-AI partnerships, but does not by itself establish capture. The article concludes that AI systems need not “want” power to participate in architectures that accumulate it; the inference, however, remains configuration-dependent, graduated, and evidence-sensitive.

Keywords: artificial intelligence; instrumental convergence; structural cognitive influence; algorithmic governance; technological sovereignty.

1 Introdução: objetivos diferentes, instrumentos semelhantes

Uma inteligência artificial encarregada de apoiar diagnósticos, outra destinada a recomendar conteúdo e uma terceira utilizada em decisões administrativas podem cumprir funções inteiramente distintas. Ainda assim, os arranjos que as desenvolvem e operam valorizam meios semelhantes: dados, capacidade computacional, acesso ao ambiente, persistência, padrões, integração, previsibilidade do comportamento e continuidade operacional.

Essa recorrência suscita uma questão anterior à discussão sobre intenção, consciência ou vontade da máquina. **Objetivos diferentes garantem pluralidade efetiva quando todos dependem dos mesmos instrumentos de poder?**

Na literatura de segurança de inteligência artificial, a expressão *convergência instrumental* designa, em termos gerais, a possibilidade de que diversos objetivos finais tornem úteis os mesmos objetivos intermediários. Preservar a própria operação, impedir alteração do objetivo, adquirir recursos, melhorar capacidades e manter opções disponíveis podem ser meios vantajosos para uma grande variedade de finalidades. O argumento não exige que o sistema sinta desejo, medo ou ambição. “Objetivo”, “incentivo” e “busca de poder” são empregados como categorias funcionais de modelos decisórios, não como afirmações sobre consciência ou experiência subjetiva.¹

Omohundro (2008) apresentou tendências que denominou *basic AI drives*. Bostrom (2012) sistematizou a tese da convergência instrumental e a relacionou à tese da ortogonalidade, segundo a qual inteligência e objetivos finais podem variar de modo relativamente independente. Formalizações posteriores mostraram, sob condições especificadas, que políticas ótimas podem tender a manter opções disponíveis e evitar estados de desligamento (Turner et al., 2021), e que procedimentos decisórios parametricamente reorientáveis (*retargetable*) também podem apresentar incentivos de busca de poder (Turner; Tadepalli, 2022). O campo, entretanto, contém exceções, pressupostos restritivos e propostas de agentes corrigíveis ou “não ambiciosos” (Hadfield-Menell et al., 2017; Cohen; Vellambi; Hutter, 2020).

¹ O emprego de “objetivo”, “incentivo”, “preservação” ou “busca” descreve relações funcionais entre política, ambiente e resultado. Não pressupõe consciência, experiência fenomenológica, personalidade jurídica ou vontade no sentido humano. A distinção segue a cautela de Bostrom (2012) quanto às conotações psicológicas da palavra drive.

O problema deste artigo surge quando a intuição é transportada para o ambiente real de implantação. Uma solução contemporânea raramente atua como agente isolado. Ela integra uma cadeia formada por modelo, dados, interfaces, ferramentas, pessoas, contratos, fornecedores de nuvem, semicondutores, centros de dados, padrões, métricas, incentivos econômicos e instituições públicas. A unidade juridicamente relevante não é apenas o modelo. É o **arranjo sociotécnico**.

Esse deslocamento exige cautela. Empresas não são funções de utilidade; Estados não são agentes de aprendizado por reforço; mercados não são processos de decisão de Markov. Utilizar a tese estrita como prova direta da conduta de organizações seria erro de categoria. A pergunta adequada é mais modesta: existe uma estrutura comparável — e empiricamente investigável — pela qual finalidades organizacionais distintas tornam recorrentes os mesmos meios de acumulação, controle e preservação?

O artigo propõe chamar essa estrutura de **convergência instrumental sociotécnica**. A expressão não substitui a tese técnica nem atribui vontade unificada a ecossistemas heterogêneos. Designa uma categoria de médio alcance para investigar a recorrência de meios transversais — dados, computação, infraestrutura, padrões, persistência e previsibilidade cognitiva — e seus efeitos sobre a distribuição de opções e controle. A redução de contestabilidade, reversibilidade ou saída não integra toda ocorrência da categoria: caracteriza sua modalidade crítica.

A hipótese dialoga com a tese da **influência cognitiva estrutural** desenvolvida em *Autonomia Decisória na Era da IA* (Genova, 2026). Se arquiteturas informacionais condicionam percepção, saliência e escolha antes do clique, a modelagem do comportamento pode deixar de ser mero efeito colateral e tornar-se instrumento transversal de eficiência. Prever o usuário melhora recomendação, publicidade, segurança, produtividade e prestação de serviços; ao mesmo tempo, permite ordenar alternativas, administrar atenção e reduzir resistência. A influência não precisa ser objetivo final para adquirir relevância estrutural.

O problema de pesquisa é saber sob quais condições a recorrência de meios em ecossistemas de IA pode ser distinguida de simples eficiência, necessidade física, imitação institucional ou coordenação legítima. O objetivo geral é construir uma arquitetura conceitual e normativa que identifique o fenômeno sem inflar a tese. Os objetivos específicos são reconstruir o conceito estrito e seus limites; separar níveis e modalidades; definir limiares verificáveis; aplicar exploratoriamente o teste a documentação oficial; relacionar o fenômeno à influência cognitiva estrutural e à soberania; e extrair consequências de governança.

A tese defendida é delimitada: **a IA não precisa querer poder para integrar uma arquitetura que acumula poder. Basta que determinados meios aumentem reiteradamente a probabilidade de cumprir as funções atribuídas e que sua acumulação altere a distribuição de controle, dependência e contestabilidade.**

2 Estado da arte, método e limites da inferência

2.1 A tese estrita: finalidade, meios e três sentidos de poder

A distinção entre objetivos finais e instrumentais é central. Um objetivo final é valorizado pelo resultado que representa no modelo; um objetivo instrumental é útil porque facilita a realização de outros objetivos. A convergência ocorre quando diferentes objetivos finais favorecem o mesmo meio.

Omohundro (2008) argumenta que sistemas avançados orientados a objetivos podem apresentar tendências à autoproteção, à preservação de sua função de utilidade, ao aprimoramento e à aquisição eficiente de recursos, salvo quando projetados para conter essas tendências. Bostrom (2012) confere formulação filosófica mais precisa ao separar dois argumentos. Pela tese da ortogonalidade, níveis elevados de inteligência podem, em princípio, coexistir com ampla variedade de objetivos finais. Pela tese da convergência instrumental, certos objetivos intermediários podem ser úteis para numerosos objetivos finais.

O termo “poder” precisa ser desambiguado. **Poder funcional** é a capacidade, atribuída a uma política ou sistema, de preservar opções e realizar um objetivo. **Poder infraestrutural** é o controle de recursos, padrões, dados ou pontos de acesso dos quais terceiros dependem. **Poder institucional** é a capacidade juridicamente ou materialmente sustentada de condicionar escolhas de outros atores. Neste artigo, nenhum desses sentidos implica estado psicológico, projeto moral ou vontade política unitária.²

A conversão entre níveis é uma hipótese, não uma identidade. Poder funcional torna-se institucionalmente relevante quando o controle de um meio indispensável modifica as opções de terceiros ou reduz sua capacidade de compreender, contestar, interromper ou substituir o arranjo. Essa regra impede que qualquer ganho técnico seja denominado “poder” em sentido político.

² A tripartição é analítica. Poder funcional aproxima-se da disponibilidade de opções em Turner et al. (2021); poder infraestrutural dialoga com Rahman (2018) e Farrell e Newman (2019); poder institucional designa o condicionamento material de escolhas por regras, dependências e posições organizacionais. As categorias podem coexistir, mas não devem ser presumidas equivalentes.

O argumento do desligamento ilustra a lógica. Em *The Off-Switch Game*, Hadfield-Menell et al. (2017) demonstram, em modelo formal, que um agente que trata sua função de recompensa como certa pode ter incentivo para desabilitar o interruptor, porque desligado não realiza o objetivo. A conclusão não é que toda IA tentará impedir humanos. O resultado depende da estrutura do modelo. A contribuição mais importante é mostrar que incerteza sobre o objetivo e tratamento da ação humana como evidência podem alterar o incentivo e favorecer a preservação da possibilidade de correção.

Turner et al. (2021) formalizam tendências de busca de poder em processos de decisão de Markov. Em ambientes com determinadas simetrias, políticas ótimas para grande classe de funções de recompensa tendem a manter opções e evitar estados que eliminam futuros possíveis. Turner e Tadepalli (2022) ampliam a análise para procedimentos decisórios parametricamente reorientáveis e sustentam que essa propriedade pode ser condição suficiente para tendências de busca de poder em modelos distintos.

Esses trabalhos possuem grande valor analítico, mas não autorizam quatro saltos:

- de ambientes formalizados para todos os ambientes reais;
- de políticas ótimas para qualquer modelo estatístico;
- de incentivos funcionais para desejo psicológico;
- de agentes isolados para organizações e ecossistemas sem uma ponte conceitual adicional.

A prudência não enfraquece a tese; define seu alcance.

2.2 Exceções, corrigibilidade e problemas concretos

Convergência instrumental não equivale a inevitabilidade. Cohen, Vellambi e Hutter (2020) propõem uma inteligência artificial geral assintoticamente não ambiciosa, concebida para preferir políticas cujos efeitos permaneçam limitados ao objetivo designado. A proposta demonstra, ao menos no plano teórico, que arquiteturas podem ser construídas para reduzir incentivos expansivos.

Hadfield-Menell et al. (2017) mostram que a incerteza sobre a utilidade pode tornar racional preservar o interruptor. A literatura de corrigibilidade investiga sistemas que aceitem correção, modificação ou desligamento sem tentar manipular o mecanismo de supervisão. Isso desloca a governança de uma confiança genérica em “alinhamento” para propriedades verificáveis: interruptibilidade, escopo, horizonte temporal, permissões, reversibilidade e resposta à intervenção.

Amodei et al. (2016) transportam preocupações abstratas para problemas concretos: efeitos colaterais negativos, manipulação da recompensa, supervisão escalável, exploração segura e mudança de distribuição. Raji e Dobbe (2024) revisitam essa agenda e defendem que falhas reais exigem análise sociotécnica, pois objetivos, instituições, práticas e relações de poder participam da formação do risco.

Essa literatura fornece duas advertências metodológicas. Primeiro, nem todo sistema contemporâneo possui metas persistentes, memória estável, acesso a ferramentas ou capacidade de agir sobre o ambiente. Um modelo de linguagem utilizado apenas para gerar texto sob comando não deve ser tratado automaticamente como agente autônomo. O risco se altera quando o modelo recebe memória, planejamento, permissões, ferramentas, acesso à rede, execução de código, autonomia temporal e capacidade de contratar ou coordenar subagentes.³ Segundo, a governança deve examinar configurações e cadeias de implantação, não marcas ou classes genéricas.

2.3 Desenho da pesquisa, corpus e protocolo

O trabalho é um **ensaio teórico-jurídico de construção conceitual e análise normativa**, apoiado em revisão narrativa dirigida e aplicação documental exploratória. Não é revisão sistemática, estudo econométrico nem pesquisa de incidência. Não estima frequência ou efeito causal e não pretende representar todo o mercado de IA.

A busca dirigida foi encerrada em **25 de julho de 2026** e realizada, em português e inglês, em páginas oficiais de periódicos e conferências, arXiv, NIST, EUR-Lex, Planalto, CNJ, Federal Trade Commission — FTC e Competition and Markets Authority — CMA, com rastreamento regressivo das referências centrais. Os descritores combinaram *instrumental convergence*, *power-seeking*, *retargetability*, plataformas, infraestrutura, *digital sovereignty*, parcerias de IA e nuvem, custos de mudança, contestabilidade e saída. Foram incluídos trabalhos que: a) formalizam ou criticam convergência instrumental e corrigibilidade; b) explicam recorrência organizacional, dependência de trajetória, plataformas, infraestrutura, influência ou soberania digital; c) estabelecem deveres normativos diretamente relacionados a necessidade, supervisão, contestabilidade, auditabilidade, permissões ou saída; ou d) documentam relações entre provedores de nuvem e desenvolvedores de modelos. Foram excluídos textos meramente opinativos, duplicatas,

³ A fronteira entre modelo e agente é configuracional. Memória persistente, ferramentas, permissões, planejamento iterativo, execução de código e autonomia temporal podem transformar significativamente o perfil de risco sem alterar o modelo-base.

documentos sem autoria ou proveniência verificável e fontes secundárias quando o documento primário estava disponível.

O corpus possui quatro núcleos:

1. **Segurança e decisão artificial:** Omohundro (2008), Bostrom (2012), Amodei et al. (2016), Hadfield-Menell et al. (2017), Cohen, Vellambi e Hutter (2020), Turner et al. (2021), Turner e Tadepalli (2022) e Raji e Dobbe (2024).
2. **Organizações, economia política e soberania:** Arthur (1989), DiMaggio e Powell (1983), Pierson (2000), Pasquale (2015), Srnicek (2017), Yeung (2017), Plantin et al. (2018), Rahman (2018), van Dijck, Poell e De Waal (2018), Couldry e Mejias (2019), Farrell e Newman (2019), Susser, Roessler e Nissenbaum (2019) e Pohle e Thiel (2020).
3. **Governança normativa:** Lei nº 13.709/2018; Regulamento (UE) 2024/1689; Resolução CNJ nº 615/2025, em texto compilado após a Resolução nº 674/2026; NIST AI RMF 1.0 e perfil de IA generativa; e NIST SP 800-53 Rev. 5.
4. **Aplicação documental:** relatório da FTC sobre três parcerias entre provedores de nuvem e desenvolvedores de IA (2025) e relatórios da CMA sobre modelos fundacionais (2024).

Cada fonte recebeu uma função analítica: **fundamento**, **mecanismo**, **contraprova**, **parâmetro normativo** ou **evidência documental**. A análise seguiu oito operações: delimitação da unidade e do período; identificação das finalidades; identificação do meio recorrente; demonstração de sua utilidade transversal; exame dos efeitos sobre opções e controle; busca de explicações rivais; formulação de contrafactual menos expansivo; e classificação do enunciado.

Constatação documental é sustentada diretamente por fonte. **Inferência interpretativa** conecta evidências sem se confundir com elas. **Proposição normativa** formula um dever. A aplicação exploratória testa a coerência e o poder discriminante do TCIS; não o valida estatisticamente nem transforma risco concorrencial em ilícito.

2.4 Delimitação da contribuição e critério de originalidade

A literatura já contém conceitos sólidos para partes do fenômeno: convergência instrumental, busca de poder, isomorfismo institucional, capitalismo de plataforma, dependência de trajetória, colonialismo de dados, hiperindução, interdependência instrumentalizada, regulação de infraestrutura e soberania digital. A contribuição deste artigo não consiste em renomear esse conjunto.

Ela reside em três movimentos coordenados:

1. preservar o núcleo lógico da convergência instrumental sem transpor seus resultados formais como prova da conduta social;
2. construir uma ponte entre agente, organização e infraestrutura por meio de critérios verificáveis;
3. formular um postulado regulatório de minimização da instrumentalidade, além da minimização de dados.

A denominação **convergência instrumental sociotécnica** é apresentada como formulação autoral e operacional dentro do corpus examinado, não como reivindicação de prioridade lexical absoluta em todas as línguas e bases existentes.⁴ A contribuição reivindicada está na articulação entre níveis, na distinção entre modalidade neutra, modalidade crítica e captura, na regra de conversão entre formas de poder e no protocolo de prova.

Quadro 1 – Fronteiras entre a categoria proposta e conceitos adjacentes

Conceito	Pergunta central	Sobreposição com a proposta	Diferença preservada
Convergência instrumental estrita	Quais meios favorecem ampla classe de objetivos de um agente?	Estrutura finalidade-meio	Opera em modelos decisórios delimitados; não prova dinâmica social.
Isomorfismo institucional	Por que organizações de um campo se tornam semelhantes?	Recorrência sem coordenação central	Explica homogeneização por pressões coercivas, miméticas e normativas, não necessariamente utilidade instrumental transversal.
Dependência de trajetória e <i>lock-in</i>	Como escolhas iniciais elevam custos de mudança?	Persistência, retornos crescentes e saída	Pode existir sem pluralidade de finalidades ou meio transversal.
Capitalismo de plataforma e colonialismo de dados	Como plataformas extraem, intermedeiam e acumulam dados?	Dados, escala, intermediação e assimetria	Oferecem teoria histórico-econômica mais ampla; o TCIS examina configuração e prova.
Interdependência instrumentalizada	Como nós centrais permitem observação e estrangulamento?	Poder infraestrutural e controle de acesso	Foca topologia de redes e coerção entre atores; a proposta inclui cognição e cadeia decisória.
Regulação infraestrutural	Como governar serviços necessários, escaláveis e vulneráveis?	Indispensabilidade, acesso e opção pública	É programa regulatório; não é teste de convergência.

⁴ A pesquisa de originalidade conceitual é necessariamente provisória. A formulação reivindica contribuição pela arquitetura, definição e operacionalização apresentadas, preservando o dever de incorporar antecedentes equivalentes que venham a ser identificados.

Conceito	Pergunta central	Sobreposição com a proposta	Diferença preservada
Soberania digital	Quem conserva autoridade e capacidade no ambiente digital?	Controle de infraestrutura, padrões e decisão	É conceito político-normativo disputado; soberania operacional é um indicador setorial proposto.

Fonte: elaboração própria, com base em DiMaggio e Powell (1983), Arthur (1989), Srnicek (2017), Rahman (2018), Plantin et al. (2018), Couldry e Mejias (2019), Farrell e Newman (2019) e Pohle e Thiel (2020).

3 Três níveis que não podem ser confundidos

A validade da proposta depende de separar unidades analíticas. O mesmo vocabulário pode designar mecanismos diferentes.

Quadro 2 – Níveis de análise da convergência instrumental

Nível	Unidade principal	Mecanismo examinado	Afirmação admissível
Convergência instrumental estrita	Agente artificial orientado a objetivo em ambiente especificado	Objetivos intermediários úteis para ampla classe de objetivos finais	Sob condições formais ou arquiteturais definidas, certas políticas tendem a preservar opções, operação ou recursos.
Convergência instrumental organizacional	Organização com governança, incentivos e divisão interna	Estratégias recorrentes que aumentam capacidade, sobrevivência institucional ou vantagem competitiva	Organizações com finalidades distintas podem adotar meios semelhantes sem coordenação central.
Convergência instrumental sociotécnica	Cadeia formada por organizações, infraestruturas, sistemas técnicos e instituições	Acoplamento de incentivos e capacidades que favorece dados, escala, padrões, persistência, previsibilidade e dependência	Arranjos heterogêneos podem produzir acumulação funcional de poder sem vontade unitária ou conspiração.

Fonte: elaboração própria, com base em Omohundro (2008), Bostrom (2012), Turner et al. (2021), Plantin et al. (2018) e Farrell e Newman (2019).

No primeiro nível, o objeto é uma política ou procedimento decisório em ambiente formal ou tecnicamente delimitado. No segundo, o objeto é uma organização jurídica e econômica composta por pessoas, regras, conflitos internos e incentivos. No terceiro, o objeto é uma rede: fornecedores de modelos e nuvem, operadores, agentes públicos, normas, interfaces, centros de dados, padrões e usuários.

Não se deve afirmar que a rede “quer” algo. A análise recai sobre **seleção estrutural**. Meios que melhoram desempenho, reduzem custos, ampliam mercado ou facilitam controle

tendem a ser reproduzidos por organizações que competem, ainda que seus dirigentes não compartilhem intenção política e que seus sistemas não possuam metas persistentes.

Esse mecanismo é compatível com retornos crescentes e dependência de trajetória. Arthur (1989) demonstrou que tecnologias sujeitas a retornos crescentes podem tornar-se bloqueadas por sequências históricas, mesmo sem serem as mais eficientes. Pierson (2000) transportou a lógica dos retornos crescentes para processos políticos: custos de mudança, aprendizado, coordenação e expectativas podem tornar trajetórias difíceis de reverter. A convergência sociotécnica não pressupõe otimização perfeita; pode emergir de vantagens cumulativas e escolhas contingentes.

4 Definição da convergência instrumental sociotécnica

Propõe-se a seguinte definição:

Definição proposta 1 — Convergência instrumental sociotécnica: recorrência documentável de um meio, ou de uma classe funcional de meios, em arranjos compostos por organizações, infraestruturas e sistemas técnicos orientados por finalidades distintas, quando o meio possui utilidade transversal e altera de modo relevante a distribuição de opções, recursos ou controle. *Fonte: elaboração própria.*

A definição contém seis exigências.

Primeiro, deve haver **pluralidade de finalidades**. Se todos os casos buscam o mesmo objetivo, haverá estratégia comum, não necessariamente convergência.

Segundo, deve existir **recorrência documentável**. Sem contratos, arquiteturas, políticas, métricas, padrões, registros, investimentos ou práticas observáveis, a tese permanece especulativa.

Terceiro, o meio precisa ter **utilidade transversal**. Não basta aparecer em setores diferentes; deve contribuir funcionalmente para objetivos distintos.

Quarto, a unidade é um **arranjo composto**, não uma mente coletiva. A categoria não apaga conflitos entre empresa, modelo, operador e Estado.

Quinto, a recorrência deve produzir **efeito estrutural relevante**. Dados ou capacidade computacional só interessam à categoria quando alteram opções próprias ou alheias, acesso a recursos ou distribuição de controle. O efeito pode ampliar ou reduzir contestabilidade; a definição básica é descritiva e não presume ilicitude.

Sexto, a hipótese deve admitir **alternativa contrafactual**. Se o mesmo objetivo pode ser alcançado com menos autonomia, acesso, persistência, dados ou dependência, a opção mais expansiva exige justificção.

O conceito não equivale a concentração econômica, monopólio, colonialismo de dados, opacidade ou manipulação. Esses fenômenos podem funcionar como causas, manifestações ou consequências. A unidade específica é a recorrência funcional dos mesmos meios entre objetivos finais diversos.

Há **convergência instrumental sociotécnica crítica** quando, além das condições da definição básica, a configuração reduz de modo relevante a contestabilidade, a auditabilidade, a reversibilidade ou a saída de terceiros, ou concentra infraestrutura ou evidência indispensável em nó pouco substituível. Essa modalidade permanece um diagnóstico de risco; não equivale automaticamente a ilícito.

Definição proposta 2 — Captura da decisão: situação em que o controle assimétrico de meios sociotécnicos indispensáveis compromete materialmente a capacidade do sujeito ou da instituição de compreender, contestar, interromper, substituir ou reconstruir a decisão, embora a competência formal permaneça preservada. *Fonte: elaboração própria.*

A captura exige prova mais intensa do que a convergência crítica. Custos de mudança, opacidade ou concentração, isoladamente, não bastam. É necessário demonstrar impedimento material ou perda severa de uma ou mais capacidades decisórias, com nexos entre o meio controlado e a restrição observada.

5 Os meios convergentes

5.1 Dados: ver para prever, prever para agir

Dados são insumo, memória e instrumento de visibilidade. Melhoram treinamento, personalização, detecção de fraude, segurança, diagnóstico e planejamento. O problema surge quando a coleta deixa de ser vinculada à finalidade concreta e passa a preservar valor futuro indeterminado. A finalidade torna-se elástica, e a utilidade potencial substitui a necessidade demonstrada.

Couldry e Mejias (2019) descrevem a conversão da vida cotidiana em fluxos apropriáveis; Pasquale (2015) mostra a assimetria produzida quando instituições tornam pessoas legíveis enquanto permanecem opacas. Na convergência sociotécnica, o dado vale também por manter cursos futuros de ação: permite retreinar, segmentar, comparar, testar e criar produtos.

5.2 Capacidade computacional, energia e escala

Modelos avançados dependem de semicondutores, centros de dados, energia, resfriamento, redes e equipes. A computação é transversal a saúde, defesa, publicidade, pesquisa e Justiça; reduz custo unitário e acelera experimentação, mas pode elevar barreiras de entrada e dependência.

O risco não está na escala em si, e sim na diferença entre **acesso ao serviço e capacidade de governá-lo**. Uma instituição pode utilizar um modelo de alta qualidade sem conseguir auditar atualizações, reproduzir resultados, migrar cargas, preservar registros ou operar após ruptura contratual. Pesquisa e serviços públicos podem precisar legitimamente de escala; por isso, o teste exige proporcionalidade, alternativas e reversibilidade.

5.3 Padrões, plataformas e infraestrutura

Plantin et al. (2018) demonstram que plataformas podem adquirir características de infraestrutura, enquanto infraestruturas incorporam a lógica das plataformas. Padrões, interfaces, formatos e identidades reduzem custos de coordenação; sua adoção cumulativa, porém, pode converter escolha privada em condição quase pública de acesso.

Farrell e Newman (2019) mostram que redes assimétricas criam nós de observação e estrangulamento. Nuvem, chips, repositórios, sistemas de pagamento e interfaces podem tornar-se pontos de controle mesmo quando o serviço final parece descentralizado. Rahman (2018), ao propor regulação infraestrutural para serviços marcados por escala, necessidade e vulnerabilidade, desloca o foco da conduta isolada para a posição ocupada na rede.

5.4 Persistência, integração e bloqueio de saída

Um sistema integrado a fluxos críticos acumula documentos, registros, automações, taxonomias e rotinas. Cada melhoria pode elevar o custo de saída, ainda que a dependência decorra de eficiência genuína. Arthur (1989) e Pierson (2000) explicam como retornos crescentes tornam trajetórias difíceis de reverter; Srnicek (2017) mostra a tendência expansiva das plataformas.

Em IA, os custos incluem migração de dados, reescrita de integrações, perda de contexto, treinamento, revalidação e indisponibilidade. O fornecedor não precisa proibir a saída; basta que sair paralise funções relevantes. O bloqueio não é apenas contratual. Pode ser:

- **técnico**, por formatos e interfaces proprietárias;
- **cognitivo**, pela adaptação do usuário a uma lógica específica;

- **organizacional**, pela reconfiguração de processos ao redor do sistema;
- **probatório**, pela perda de logs e histórico na migração;
- **normativo**, quando padrões privados são incorporados a requisitos públicos;
- **material**, quando não há fornecedor alternativo com infraestrutura equivalente.

5.5 Previsibilidade comportamental e influência

Quanto melhor um arranjo prevê comportamentos, mais eficientemente pode recomendar, priorizar, precificar, detectar risco e alocar recursos. Yeung (2017) descreve *hypernudges* dinâmicos e adaptativos; Susser, Roessler e Nissenbaum (2019) delimitam a manipulação tecnológica encoberta de vulnerabilidades decisórias. A IA generativa amplia a possibilidade de adaptar tom, exemplos, momento e justificativa.

É nesse ponto que a convergência instrumental encontra a influência cognitiva estrutural. A arquitetura não precisa ordenar uma escolha. Pode:

- definir quais alternativas serão visíveis;
- escolher a ordem de apresentação;
- ajustar fricção e conveniência;
- determinar o momento de uma recomendação;
- personalizar linguagem e autoridade aparente;
- tornar uma opção padrão e as demais excepcionais;
- registrar a reação e retroalimentar o modelo.

O resultado não é necessariamente manipulação. Sistemas de acessibilidade, educação ou saúde podem usar personalização para ampliar autonomia. O critério jurídico é saber se a influência preserva compreensão, pluralidade, contestação e liberdade material de recusa.

5.6 Matriz dos meios e riscos

Quadro 3 – Meios convergentes, utilidade e risco jurídico

Meio	Utilidade transversal	Efeito estrutural possível	Pergunta de governança
Dados e inferências	Aprendizado, personalização, previsão, segurança	Legibilidade unilateral, finalidade elástica, assimetria probatória	O dado é necessário à finalidade atual ou preservado por valor futuro indeterminado?
Computação e energia	Treinamento, inferência, escala, experimentação	Barreiras de entrada e dependência material	A instituição consegue operar, auditar e migrar sem o provedor dominante?

Meio	Utilidade transversal	Efeito estrutural possível	Pergunta de governança
Padrões e infraestrutura	Coordenação, interoperabilidade, distribuição	Nó de observação ou estrangulamento	Quem define o padrão, altera o acesso e controla sua evolução?
Persistência e integração	Continuidade, memória, produtividade	Bloqueio técnico, cognitivo, organizacional e probatório	A saída é juridicamente permitida e materialmente possível?
Previsibilidade cognitiva	Recomendação, segurança, alocação, persuasão	Organização invisível de atenção, saliência e escolha	A influência é compreensível, contestável e compatível com a autonomia?

Fonte: elaboração própria, com base em Arthur (1989), Yeung (2017), Plantin et al. (2018), Couldry e Mejias (2019) e Farrell e Newman (2019).

6 Influência cognitiva estrutural como meio instrumental

A influência cognitiva estrutural designa o condicionamento persistente da percepção e da decisão por ambientes que organizam informação, alternativas, relevância, tempo e fricção. Ela não se reduz à mensagem persuasiva nem coincide necessariamente com manipulação. Opera na arquitetura em que a deliberação se torna possível.

Sua relação com a convergência instrumental pode ser representada por uma cadeia:

finalidade organizacional → necessidade de desempenho → busca de previsibilidade → modelagem do comportamento → organização do ambiente informacional → influência sobre percepção e escolha → melhoria da realização do objetivo

Essa cadeia não presume intenção de capturar a mente. Uma plataforma de ensino quer reduzir evasão; um hospital quer aumentar adesão; um tribunal quer organizar demandas; uma empresa quer vender; um sistema de segurança quer detectar anomalias. Em todos os casos, conhecer padrões de comportamento pode melhorar a função. A influência torna-se convergente porque reduz incerteza.

O risco surge quando o sujeito afetado não conhece a arquitetura, não consegue comparar alternativas ou não dispõe de saída equivalente. A autonomia formal permanece: o botão pode ser clicado, a recomendação pode ser recusada, o serviço pode ser abandonado. A autonomia material, porém, é reduzida quando toda alternativa relevante foi filtrada, encarecida ou tornada invisível.

Em *Autonomia Decisória na Era da IA*, a autonomia é tratada não apenas como existência formal de escolha, mas como domínio material das condições que tornam a escolha possível (Genova, 2026). A convergência instrumental sociotécnica amplia esse argumento: a

capacidade de estruturar essas condições pode ser útil para finalidades muito diferentes e, por isso, acumular-se em múltiplos setores.

Não se deve concluir que toda personalização seja ilegítima. A influência pode ampliar agência ao traduzir informação, reduzir barreiras, apresentar consequências e adaptar acessibilidade. A distinção depende de cinco critérios:

1. **finalidade legítima e específica;**
2. **proporcionalidade entre influência e benefício;**
3. **transparência suficiente sobre o papel do sistema;**
4. **pluralidade real de alternativas e possibilidade de recusa;**
5. **ausência de exploração de vulnerabilidade incompatível com a função.**

A influência estrutural só alcança o limiar de captura definido na seção 4 quando o controle do campo informacional compromete materialmente compreensão, contestação, interrupção, substituição ou reconstrução. A mera personalização, mesmo persuasiva, não satisfaz esse limiar.

7 Da decisão capturada à soberania capturada

Soberania em inteligência artificial costuma ser medida por capacidade de treinar modelos, produzir chips, operar centros de dados ou proteger dados nacionais. Esses elementos são indispensáveis, mas insuficientes. Pohle e Thiel (2020) mostram, ademais, que “soberania digital” é conceito disputado e funciona também como prática discursiva; por isso, o uso abaixo é deliberadamente operacional e setorial.

Uma instituição pode possuir competência jurídica e ainda depender de:

- modelos que não consegue modificar;
- nuvem que não consegue substituir;
- padrões que não ajudou a definir;
- registros que não controla;
- métricas fornecidas pelo contratado;
- atualizações que não pode recusar;
- conhecimento técnico concentrado no fornecedor;
- interfaces que condicionam o modo como seus agentes compreendem o problema.

Essa situação caracteriza uma diferença entre **soberania formal** e **soberania operacional**. A primeira conserva a atribuição normativa de decidir. A segunda exige capacidade material de compreender, auditar, interromper, substituir e reconstruir a decisão.

Propõe-se:

Definição proposta 3 — Soberania operacional em IA: capacidade institucional de definir finalidades e padrões, controlar evidências e infraestruturas críticas, auditar o funcionamento relevante, interromper ou modificar o sistema e abandonar um ecossistema sem paralisar funções essenciais. *Fonte: elaboração própria.*

A convergência instrumental sociotécnica oferece uma hipótese sobre uma via de captura. À medida que dados, computação, padrões, integração e previsibilidade se concentram, o fornecedor pode tornar-se mais eficiente e o usuário mais dependente. A assimetria não precisa resultar de contrato abusivo; pode emergir do próprio sucesso técnico. A ocorrência concreta, porém, depende do limiar probatório definido, não da concentração isolada.

Farrell e Newman (2019) mostram que a topologia das redes distribui capacidades de coerção. Plantin et al. (2018) mostram a infraestrutura adquirindo lógica de plataforma. Rahman (2018) demonstra que escala, necessidade e vulnerabilidade justificam olhar regulatório específico. Juntas, essas contribuições revelam que a soberania não se perde apenas por transferência formal de competência. Pode ser capturada pela indispensabilidade do meio.

O problema é especialmente grave no Estado. Se a Administração adota uma solução para funções essenciais e não preserva portabilidade, documentação, competência interna e alternativa operacional, a discricionariedade futura diminui. O governante seguinte poderá mudar a política, mas não a infraestrutura que define o campo das opções.

A captura também é normativa. Padrões privados podem ser incorporados a editais, certificações e regulamentos. A instituição passa a fiscalizar com métricas produzidas por quem controla a tecnologia. A regra existe, mas a evidência da conformidade permanece privatizada.

8 Teste de Convergência Instrumental Sociotécnica

Para impedir que a categoria seja aplicada a qualquer concentração tecnológica, propõe-se o **Teste de Convergência Instrumental Sociotécnica — TCIS**. Trata-se de protocolo estruturado de análise e prova, não de escala psicométrica, fórmula causal ou certificação.

8.1 Unidade, período e classes de evidência

A unidade de análise é uma **configuração sociotécnica delimitada**: conjunto identificável de atores, contratos, sistemas, infraestruturas e regras vinculados a uma implantação, parceria, cadeia de valor ou função institucional. O pesquisador deve declarar o período observado, pois atualização técnica, nova obrigação jurídica ou possibilidade real de migração pode alterar o diagnóstico.

Cada dimensão recebe uma das seguintes classes:

1. **E — evidenciada**: sustentada diretamente por contrato, ato, registro, relatório oficial, teste reproduzível ou outra fonte primária adequada;
2. **P — parcialmente evidenciada**: sustentada por indícios convergentes, mas dependente de inferência ou documento incompleto;
3. **ND — não demonstrada**: o corpus não permite concluir;
4. **CE — contraprova existente**: há evidência pertinente incompatível com a hipótese.

“ND” não significa inexistência, e “P” não pode ser tratado como “E”. Havendo conflito, a conclusão deve registrar a divergência. Evidência próxima ao fato, específica à configuração e tecnicamente verificável prevalece, em regra, sobre declaração genérica; contraprova material não pode ser neutralizada por simples soma de indícios.

8.2 Regras de decisão

As quatro condições nucleares são:

1. **pluralidade finalística**: pelo menos duas finalidades relevantes e distintas;
2. **recorrência do meio**: o mesmo meio ou classe funcional reaparece nas configurações comparadas;
3. **utilidade transversal**: demonstração de como o meio contribui para cada finalidade;
4. **efeito sobre opções ou controle**: alteração relevante na disponibilidade de cursos de ação, recursos ou capacidade de intervenção.

Há **convergência indicativa** quando as quatro condições recebem E ou P, sendo pluralidade e recorrência necessariamente E, e não há CE decisiva. Há **convergência robustamente documentada** quando as quatro recebem E.

A modalidade **crítica** requer, além da convergência, evidência de pelo menos dois agravantes, dos quais um deve incidir sobre contestabilidade ou auditabilidade, saída ou

substituição, ou concentração de infraestrutura ou evidência indispensável. A regra é um limiar mínimo de consistência para o protocolo, calibrável por setor; não é lei universal.

O diagnóstico de **captura** somente é admissível quando, além da modalidade crítica, houver E quanto ao comprometimento material de compreender, contestar, interromper, substituir ou reconstruir a decisão. Se a documentação apenas descreve risco potencial, a conclusão correta é “captura não demonstrada”.

8.3 Matriz operacional

Quadro 4 – Protocolo operacional do TCIS

Dimensão	Indicador verificável	Evidência pertinente
Finalidades	Objetivos finais efetivamente distintos	Estatutos, políticas, contratos, atos normativos
Recorrência	Meio ou classe funcional estável em casos distintos	Arquiteturas, investimentos, integrações, padrões
Utilidade	Contribuição funcional do meio para cada objetivo	Métricas, testes, documentos técnicos e comerciais
Opções e controle	Alteração de cursos de ação, recursos ou intervenção	Direitos de controle, permissões, portabilidade, governança
Necessidade	Expansão além do necessário, adequado e proporcional	Avaliação de impacto, justificativas e alternativas
Contestabilidade	Capacidade de compreender, auditar e impugnar	Registros, documentação, acesso regulatório, auditoria
Saída	Substituição sem perda desproporcional	Contratos, formatos, testes de migração, continuidade
Cognição	Organização de atenção e escolha em favor da finalidade	Interface, ordenação, experimentos, personalização
Externalidades	Distribuição de benefícios, custos e riscos	Consumo de recursos, incidentes, impactos e reparação
Contrafactual	Configuração menos expansiva com desempenho aceitável	Provas de conceito, desenho modular, alternativas

Fonte: elaboração própria (2026).

8.4 Contraprovas e explicações rivais

A hipótese perde força quando o meio não reaparece em finalidades distintas; a recorrência decorre integralmente de necessidade física estreita; soluções menos expansivas e equivalentes são preferidas; usuários preservam interoperabilidade, registros e substituição efetivas; a personalização amplia compreensão e pluralidade; ou regulação setorial,

segurança, escala, isomorfismo institucional ou contingência histórica explicam melhor o padrão.

O TCIS não produz causalidade pela soma de indícios. Sua função é tornar explícita a passagem entre evidência, inferência e conclusão, inclusive para impedir que a presença de tecnologia avançada seja convertida automaticamente em captura.

8.5 Aplicação documental exploratória: parcerias entre nuvem e IA

A unidade examinada compreende as parcerias Microsoft–OpenAI, Amazon–Anthropic e Alphabet–Anthropic e sua inserção na cadeia de modelos fundacionais, entre 2019 e janeiro de 2025. O relatório da FTC resulta de estudo realizado sob a seção 6(b) do *Federal Trade Commission Act*, com documentos públicos e não públicos fornecidos por cinco empresas. A própria FTC adverte que o relatório não constitui análise jurídica ou econômica completa e não conclui pela existência de conduta ilegal (Federal Trade Commission, 2025). A CMA, por sua vez, descreve riscos concorrenciais e a presença de incumbentes em múltiplos níveis da cadeia, sem afirmar que todo risco se materializou (Competition and Markets Authority, 2024).

Quadro 5 – Aplicação exploratória do TCIS

Dimensão	Achado documental	Classe	Consequência analítica
Pluralidade finalística	Provedores expandem nuvem, produtos e receitas; desenvolvedores buscam computação, desenvolvimento e distribuição de modelos.	E	Condição satisfeita.
Recorrência	As três parcerias combinam investimento, acesso a computação e integração, embora os termos variem.	E	Condição satisfeita sem presumir identidade contratual.
Utilidade transversal	Desenvolvedores obtêm computação e rotas de crescimento; provedores obtêm direitos econômicos, integração e informação.	E	Utilidade para finalidades distintas documentada.

Dimensão	Achado documental	Classe	Consequência analítica
Opções e controle	Há direitos de consulta, controle ou exclusividade em graus diversos e compromissos de gasto em nuvem; o efeito líquido sobre autonomia varia.	P	Convergência indicativa, não robusta.
Contestabilidade e evidência	Provedores recebem informação técnica e comercial indisponível a terceiros; parte relevante do corpus é confidencial ou agregada.	P	Assimetria documentada, auditabilidade externa não mensurada.
Saída e substituição	A FTC identifica custos contratuais e técnicos capazes de dificultar troca de nuvem ou uso de múltiplos provedores.	P	Agravante crítico plausível; materialização caso a caso não demonstrada.
Nó infraestrutural	A CMA identifica computação, dados e expertise como insumos críticos e poucos incumbentes em vários níveis da cadeia.	E	Agravante crítico evidenciado em nível de mercado.
Expansão além da necessidade	Os relatórios não demonstram configuração tecnicamente equivalente e menos expansiva para cada parceria.	ND	Não autoriza juízo de excesso.
Influência cognitiva	Os documentos não examinam arquitetura de escolha dos usuários finais com detalhe suficiente.	ND	Nenhuma conclusão sobre captura cognitiva.
Captura da decisão	Não há prova suficiente de impedimento material de compreender, contestar, interromper, substituir ou reconstruir decisões específicas.	ND	Captura não demonstrada.

Fonte: elaboração própria, com base em Federal Trade Commission (2025) e Competition and Markets Authority (2024).

O resultado é deliberadamente contido: há **convergência indicativa** e **risco de modalidade crítica**, porque computação, integração, informação e custos de saída reaparecem e alteram a distribuição de opções e controle. Não há base para afirmar captura, ilicitude ou efeito anticompetitivo consumado. A conclusão mais forte que a documentação suporta é a necessidade de investigação específica sobre portabilidade, multicloud, direitos de auditoria, governança de informação, continuidade e acesso a insumos.

O contrafactual verificável incluiria arquitetura multicloud funcional, formatos portáveis, registros independentes, limites de exclusividade, testes de migração e acesso não discriminatório a insumos críticos. A existência efetiva dessas condições poderia converter P em CE e afastar a modalidade crítica. Assim, a aplicação não apenas confirma: mostra como o TCIS pode recusar uma conclusão excessiva.

9 O Direito vigente já percebe partes do problema

Nenhum dos instrumentos examinados positiva a categoria proposta. Eles possuem âmbitos, vigência territorial e força normativa diferentes, mas contêm elementos capazes de enfrentar partes do problema. A análise temporal considera o estado das fontes em 25 de julho de 2026.

9.1 LGPD: necessidade, finalidade e responsabilização

A Lei nº 13.709/2018 estabelece princípios de finalidade, adequação, necessidade, transparência, prevenção e responsabilização. O princípio da necessidade limita o tratamento ao mínimo necessário para a realização da finalidade, com dados pertinentes, proporcionais e não excessivos (Brasil, 2018, art. 6º, III).

Essa arquitetura é diretamente relevante para a expansão instrumental dos dados. A organização não pode justificar retenção ilimitada apenas pela possibilidade abstrata de uso futuro. Finalidade legítima não elimina o dever de demonstrar necessidade.

A LGPD, porém, incide sobre tratamento de dados pessoais. Ela não regula, de modo geral, excesso de autonomia operacional, capacidade computacional, permissões de ferramenta, persistência de agente, dependência de infraestrutura ou influência sobre usuários. É nesse intervalo que se situa a proposta de ampliar a lógica de minimização.

9.2 Regulamento Europeu de Inteligência Artificial

O Regulamento (UE) 2024/1689 estabelece abordagem baseada em risco e, para sistemas de alto risco, disciplina gestão de riscos, documentação, registros, transparência, supervisão humana, exatidão, robustez e cibersegurança. A supervisão humana deve permitir que pessoas compreendam capacidades e limitações, evitem confiança automática, interpretem saídas, decidam não usar, desconsiderem ou revertam o resultado e interrompam o sistema quando necessário (União Europeia, 2024, art. 14).

O regime possui aplicação escalonada. Na data de corte, definições, proibições e alfabetização em IA já eram aplicáveis; outras partes passaram a incidir em 2025; a aplicação geral estava prevista para **2 de agosto de 2026**, e o art. 6º, nº 1, com obrigações correspondentes, para **2 de agosto de 2027** (União Europeia, 2024, art. 113). A referência ao art. 14, portanto, identifica parâmetro normativo do regime, não afirma vigência uniforme de todas as obrigações em julho de 2026.

Esses deveres atingem elementos centrais da convergência: preservam intervenção, produzem evidência e limitam dependência cognitiva. Mas supervisão formal não basta. Se o operador não possui tempo, competência, autoridade ou acesso a registros, o controle pode tornar-se ritual.

O Regulamento também distribui deveres ao longo da cadeia de valor. Essa perspectiva é compatível com a unidade sociotécnica, embora o regime permaneça organizado por categorias jurídicas de provedor, responsável pela implantação, importador e distribuidor.

9.3 Resolução CNJ nº 615/2025

A Resolução CNJ nº 615/2025, em texto compilado após a Resolução nº 674/2026, estabelece normas para desenvolvimento, governança, auditoria, monitoramento e uso responsável de IA no Poder Judiciário. Entre seus princípios estão transparência, explicabilidade, contestabilidade, auditabilidade, confiabilidade e supervisão humana efetiva.

O art. 10, I, possui particular relevância: veda soluções que não possibilitem revisão humana dos resultados ou que gerem dependência absoluta do usuário em relação ao resultado, sem possibilidade de alteração ou revisão. A norma reconhece que dependência não é mero inconveniente de uso; pode constituir risco excessivo aos direitos fundamentais e à independência.

Os arts. 12 a 14 exigem processos internos, documentação, logs, indicação de objetivos, avaliação de impacto e mecanismos de explicabilidade para soluções de alto risco.

Essas exigências podem sustentar o TCIS no setor público, pois produzem elementos necessários à reconstrução de finalidade, meio, supervisão e impacto.

Permanece, contudo, um desafio material: a existência normativa de acesso, auditoria ou interoperabilidade não garante competência técnica, infraestrutura e orçamento para exercê-los. Soberania operacional exige capacidade instalada.

9.4 NIST AI RMF

O NIST AI RMF 1.0 organiza a gestão de riscos nas funções **Govern, Map, Measure e Manage**. O documento é voluntário, não setorial e orientado à preservação de direitos, buscando integrar governança ao ciclo de vida (Tabassi, 2023). O perfil de IA generativa complementa a estrutura com riscos e ações específicas (Autio et al., 2024).

O valor do framework está em tratar risco como processo contínuo, não como certificação pontual. Isso converge com a necessidade de revisar instrumentalidade após atualizações, integrações, mudanças de finalidade e expansão de permissões.

Sua limitação jurídica é igualmente clara: voluntariedade e flexibilidade permitem adaptação, mas não garantem fiscalização independente, acesso regulatório ou sanção. Em julho de 2026, o NIST informava que o AI RMF 1.0 estava em revisão e que o *Playbook* seria atualizado posteriormente; o artigo, portanto, trata a versão 1.0 como referência vigente em transição, não como texto estabilizado.

10 Postulado regulatório da instrumentalidade mínima necessária

A convergência sociotécnica revela um problema que a minimização de dados, isoladamente, não resolve. Um sistema pode tratar poucos dados pessoais e ainda receber autonomia excessiva, permissões amplas, persistência indefinida ou posição infraestrutural indispensável.

Chamá-lo desde logo de princípio autônomo do Direito positivo produziria um salto indevido. Propõe-se, com maior precisão, um **postulado regulatório de síntese**: diretriz argumentativa para desenho, contratação, interpretação e controle, cuja exigibilidade concreta depende da norma e da competência aplicáveis.⁵

10.1 Derivação e limites jurídicos

A derivação ocorre em quatro etapas:

⁵ “Postulado regulatório” é empregado como proposta de síntese e orientação argumentativa. Não cria competência, sanção ou dever autônomo e não substitui a identificação da base jurídica aplicável a cada caso.

1. **Dados e finalidade:** a LGPD exige finalidade, adequação, necessidade, prevenção e responsabilização. A expansão do tratamento deve ser vinculada à finalidade e limitada ao pertinente, proporcional e não excessivo (Brasil, 2018, art. 6º).
2. **Decisão e supervisão:** o Regulamento Europeu de IA estrutura supervisão, registros, possibilidade de desconsiderar ou reverter resultados e interrupção; a Resolução CNJ nº 615/2025 exige supervisão humana efetiva, contestabilidade, documentação, registros e descontinuação, e veda dependência absoluta sem revisão.
3. **Permissões e segurança:** o controle AC-6 do NIST SP 800-53 Rev. 5 formula o menor privilégio para usuários e processos, limitando acessos ao necessário às tarefas organizacionais (National Institute of Standards and Technology, 2020).
4. **Ciclo de vida, infraestrutura e saída:** o AI RMF organiza gestão contínua de risco; FTC e CMA identificam acesso a insumos críticos, custos de mudança, interoperabilidade e escolha como problemas de governança e concorrência.

Essas fontes não possuem igual força normativa nem positavam, em conjunto, uma regra geral sobre autonomia, computação ou dependência. A passagem é **reconstrutiva e propositiva**: identifica uma razão comum de contenção do meio e a estende, sob controle de proporcionalidade e competência, a dimensões não cobertas pela minimização de dados.

Definição proposta 4 — Postulado regulatório da instrumentalidade mínima necessária: sistemas e arranjos de inteligência artificial não devem receber, conservar ou ampliar autonomia, dados, permissões, recursos, persistência, influência ou dependência além do necessário, adequado, proporcional e reversível para finalidade legítima, específica e documentada. *Fonte: elaboração própria.*

10.2 Dimensões da minimização

O postulado exige avaliar:

1. **autonomia:** quais decisões o sistema pode preparar, recomendar ou executar;
2. **dados:** quais informações pode acessar, inferir, combinar e reter;
3. **permissões:** quais ferramentas, redes, contas e ambientes pode utilizar;
4. **recursos:** quais limites de computação, energia, orçamento e chamadas pode consumir;
5. **persistência:** por quanto tempo mantém memória, objetivo e capacidade de ação;
6. **influência:** como pode ordenar, personalizar e apresentar alternativas;
7. **dependência:** quais funções se tornam indisponíveis se o sistema falhar ou o fornecedor sair;

8. **evidência:** quais registros devem permanecer sob controle do responsável e do fiscalizador;
9. **reversibilidade:** quão rapidamente ações, integrações e efeitos podem ser interrompidos ou desfeitos;
10. **substituibilidade:** se o sistema pode ser abandonado sem paralisar a organização.

10.3 Controles correspondentes

Quadro 6 – Instrumentalidade mínima e controles

Dimensão	Excesso típico	Controle proposto
Autonomia	Execução automática onde recomendação seria suficiente	Separação entre planejar, recomendar, aprovar e executar
Dados	Acesso amplo por conveniência ou uso futuro	Segmentação, finalidade específica, retenção curta e proibição de combinação
Permissões	Credenciais genéricas e permanentes	Menor privilégio, autorização por tarefa e expiração automática
Recursos	Escala sem teto ou sem custo imputável	Limites de computação, orçamento, energia, tempo e chamadas
Persistência	Memória e metas mantidas indefinidamente	Horizonte temporal, revisão periódica e limpeza verificável
Influência	Personalização invisível orientada à adesão	Transparência, controle do usuário, alternativas e vedação de exploração
Dependência	Processo crítico sem modo degradado	Portabilidade, redundância, opção pública ou alternativa local
Evidência	Logs controlados apenas pelo fornecedor	Registro independente, acesso regulatório e cadeia de custódia
Reversibilidade	Ações irreversíveis ou em cascata	Confirmação humana, ambientes isolados, limites transacionais e reversão controlada
Substituibilidade	Formatos, APIs e conhecimento proprietários	Padrões abertos, documentação, testes periódicos de saída e capacitação interna

Fonte: elaboração própria (2026).

O postulado não exige a menor tecnologia possível. Exige o menor **poder instrumental** compatível com a finalidade e o risco. Um sistema de emergência pode precisar de ampla capacidade, mas essa necessidade deve ser demonstrada, temporária quando possível, monitorada e acompanhada de evidências.

10.4 Teste contrafactual da necessidade

A pergunta decisiva é:

A finalidade legítima seria alcançada, com desempenho e segurança aceitáveis, por configuração que possua menos autonomia, dados, permissões, recursos, persistência, influência ou dependência?

Se a resposta for positiva, a configuração mais expansiva precisa de justificativa adicional. Se a resposta for negativa, a instituição deve demonstrar por que o incremento é necessário, quais riscos produz e como preservará correção, saída e reparação.

Essa lógica altera o ônus argumentativo. Em vez de exigir que o afetado prove por que o sistema não deveria possuir determinada capacidade, exige-se que quem a atribui demonstre sua necessidade.

11 Arquitetura mínima de governança

Uma governança coerente com o postulado combina técnica, contrato, instituição e prova ao longo do ciclo de vida:

1. **Antes da entrada:** definir finalidade sem referência a produto; mapear decisão, afetados, direitos e infraestrutura; comparar alternativa automatizada e não automatizada; aplicar TCIS, avaliação de impacto e contrafactual; estimar dependência e externalidades; aprovar plano de saída.
2. **No desenvolvimento e na implantação:** modularizar componentes; separar ambientes; limitar permissões por tarefa; registrar dados, versões, parâmetros, ferramentas e responsáveis; testar reversibilidade; atribuir autoridade real à supervisão; submeter mudança de finalidade ou atualização de risco a nova avaliação.
3. **Na operação:** monitorar desempenho e efeitos distributivos; revisar necessidade; registrar incidentes; testar desligamento e modo degradado; auditar personalização; comparar resultados com alternativa independente; medir custo de saída e comunicar mudanças materiais.
4. **Na descontinuidade:** exportar dados e evidências em formato utilizável; preservar cadeia de custódia; revogar credenciais; eliminar memórias conforme política verificável; manter continuidade e transferir conhecimento necessário à substituição.

Essas medidas não provam soberania operacional, mas transformam a declaração de controle em capacidades testáveis.

12 Objeções, condições de validade e respostas

12.1 “A proposta antropomorfiza ecossistemas”

A objeção seria procedente se o artigo atribuísse desejo, consciência ou objetivo unitário a redes. Não atribui. O conceito identifica recorrência funcional de meios e exige documentação de incentivos, capacidades e efeitos. A linguagem de “poder” é operacional e não psicológica.

12.2 “Há deriva de categoria entre IA safety e economia política”

A transposição direta seria inválida. Por isso o artigo separa três níveis e não utiliza resultados formais como prova social. O núcleo preservado é apenas a estrutura lógica: finalidades distintas podem tornar úteis meios semelhantes. A ponte sociotécnica depende de literatura própria e de evidência empírica adicional.

12.3 “Tudo que é eficiente parecerá convergência”

O TCIS exige mais que eficiência. Requer pluralidade finalística, recorrência, utilidade transversal e efeito sobre opções ou controle. Necessidade técnica estreita pode afastar a hipótese. A modalidade crítica exige agravantes próprios, e captura exige comprometimento material demonstrado.

12.4 “O conceito não é falsificável”

O artigo define classes de evidência, regras de decisão e contraprovas. A hipótese perde força quando meios menos expansivos são escolhidos, interoperabilidade e saída permanecem efetivas ou variáveis alternativas explicam melhor a recorrência. O protocolo ainda requer validação interavaliador e setorial.

12.5 “Não há demonstração causal”

Correto. O trabalho é conceitual e documental. A aplicação FTC/CMA demonstra uso disciplinado do protocolo, não causalidade nem representatividade. Sua contribuição é formular unidade, mecanismo, critérios e inferências proporcionais ao corpus.

12.6 “Minimização pode impedir inovação e segurança”

O postulado não estabelece teto abstrato. Segurança pode justificar permissões ou dados adicionais; inovação pode justificar experimentos delimitados. A exigibilidade depende

de base jurídica e competência; a proposta apenas desloca para quem expande a capacidade o dever argumentativo de justificá-la.

12.7 “Regulação pode consolidar incumbentes”

A objeção é relevante. Deveres complexos podem favorecer organizações com mais capital e equipes. Por isso a proposta inclui modularidade, padrões abertos, infraestrutura compartilhada, opções públicas, auditoria proporcional e apoio a capacidades institucionais. Governança não deve converter segurança em barreira privada de entrada.

12.8 “Modelos abertos resolvem o problema”

Abertura de pesos reduz uma barreira, mas não entrega chips, energia, nuvem, dados, equipes, documentação, segurança ou capacidade de retreinamento. Tampouco garante que a cadeia operacional seja auditável. Abertura é dimensão importante, não sinônimo de soberania.

13 Limitações e agenda de pesquisa

O artigo possui seis limitações principais.

Primeiro, a revisão é narrativa e dirigida, não sistemática. O protocolo torna seleção e função das fontes mais transparentes, mas não elimina viés de cobertura.

Segundo, a aplicação documental abrange três parcerias e relatórios produzidos por autoridades concorrenciais. Parte dos documentos subjacentes é confidencial, agregada ou redigida, e o período termina em janeiro de 2025. A configuração pode ter mudado.

Terceiro, o TCIS ainda não possui validação interavaliador, calibração setorial ou teste de sensibilidade dos limiares. Suas classes disciplinam a inferência, mas não produzem medida universal.

Quarto, a ponte entre convergência técnica, organização e infraestrutura continua teórica. A separação de níveis reduz a deriva de categoria, mas não a converte em causalidade.

Quinto, poder, dependência, influência e saída exigem indicadores setoriais. Uma migração aceitável em publicidade pode ser incompatível com continuidade em saúde ou Justiça.

Sexto, a análise privilegia riscos de acumulação e examina menos os benefícios de integração, escala e coordenação. O postulado proposto precisa ser testado contra custos de

conformidade, segurança, propriedade intelectual, continuidade, inovação e distribuição de capacidades.

A agenda empírica inclui estudos comparativos entre setores; codificação independente do TCIS; testes de migração, desligamento e modo degradado; auditorias de interfaces; indicadores de soberania operacional; análise concorrencial e ambiental da infraestrutura; e investigação de configurações agênticas com memória, ferramentas e persistência.

Uma hipótese testável é que a modalidade crítica se torna mais provável quando coexistem **alta autonomia técnica, alta concentração infraestrutural e baixa contestabilidade institucional**. Estudos comparativos devem avaliar também a hipótese rival: integração e escala podem aumentar segurança, disponibilidade e pluralidade quando acompanhadas de acesso, interoperabilidade e fiscalização efetivos.

14 Conclusão

A convergência instrumental oferece uma advertência simples e poderosa: a inocuidade aparente do objetivo final não garante a inocuidade dos meios. Um sistema não precisa receber a missão de acumular recursos, preservar-se ou influenciar pessoas. Essas capacidades podem tornar-se úteis para cumprir outras missões.

Transportar essa advertência para a sociedade exige rigor. Modelos, empresas, Estados e infraestruturas não podem ser confundidos. Resultados formais não provam condutas organizacionais. Por isso o artigo definiu a convergência instrumental sociotécnica como categoria neutra de médio alcance, separou sua modalidade crítica e reservou “captura” ao comprometimento material da capacidade decisória.

A análise identificou cinco meios especialmente relevantes: dados, capacidade computacional, infraestrutura e padrões, persistência e bloqueio de saída, e previsibilidade cognitiva. Eles são condições do desenvolvimento tecnológico, não indícios automáticos de ilicitude. A aplicação aos relatórios FTC/CMA encontrou convergência indicativa e risco crítico em parcerias de nuvem e IA, mas recusou concluir captura por falta de prova do comprometimento material exigido.

A influência cognitiva estrutural aparece, nesse quadro, não apenas como risco lateral, mas como possível meio instrumental. Conhecer e organizar o comportamento reduz incerteza para finalidades diversas. O Direito deve distinguir personalização que amplia autonomia de arquitetura que captura a decisão mantendo apenas sua aparência formal.

O postulado regulatório da instrumentalidade mínima necessária responde a essa mudança. Não basta minimizar dados. É preciso justificar autonomia, permissões, recursos, persistência, influência e dependência segundo necessidade, adequação, proporcionalidade e reversibilidade. Sua força concreta, porém, depende da base jurídica aplicável; ele não é apresentado como princípio autônomo já positivado.

A conclusão central pode ser condensada:

A inteligência artificial não precisa querer poder. Basta que poder — entendido como capacidade de preservar opções, controlar recursos, reduzir intervenção e tornar-se indispensável — aumente a probabilidade de cumprir a função que lhe foi atribuída.

O desafio da governança não é adivinhar uma vontade oculta da máquina. É impedir que a utilidade dos meios convergentes transforme eficiência em indispensabilidade, indispensabilidade em dependência e dependência em captura da decisão e da soberania.

Referências

- AMODEI, Dario et al. Concrete problems in AI safety. *arXiv*, 2016. DOI: <https://doi.org/10.48550/arXiv.1606.06565>. Disponível em: <https://arxiv.org/abs/1606.06565>. Acesso em: 25 jul. 2026.
- ARTHUR, W. Brian. Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, Oxford, v. 99, n. 394, p. 116-131, 1989. DOI: <https://doi.org/10.2307/2234208>.
- AUTIO, Chloe et al. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. Gaithersburg: National Institute of Standards and Technology, 2024. NIST AI 600-1. DOI: <https://doi.org/10.6028/NIST.AI.600-1>. Disponível em: <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>. Acesso em: 25 jul. 2026.
- BOSTROM, Nick. The superintelligent will: motivation and instrumental rationality in advanced artificial agents. *Minds and Machines*, Dordrecht, v. 22, p. 71-85, 2012. DOI: <https://doi.org/10.1007/s11023-012-9281-3>.
- BRASIL. *Lei nº 13.709, de 14 de agosto de 2018*. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). *Diário Oficial da União*: seção 1, Brasília, DF, n. 157, p. 59, 15 ago. 2018. Texto compilado. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709compilado.htm. Acesso em: 25 jul. 2026.
- COMPETITION AND MARKETS AUTHORITY (United Kingdom). *AI Foundation Models: update paper*. London: CMA, 11 abr. 2024. Disponível em: <https://www.gov.uk/government/publications/ai-foundation-models-update-paper>. Acesso em: 25 jul. 2026.
- COHEN, Michael K.; VELLAMBI, Badri N.; HUTTER, Marcus. Asymptotically unambitious artificial general intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, Palo Alto, v. 34, n. 3, p. 2467-2476, 2020. DOI: <https://doi.org/10.1609/aaai.v34i03.5628>.
- CONSELHO NACIONAL DE JUSTIÇA (Brasil). *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. *Diário da Justiça eletrônico/CNJ*, Brasília, DF, n. 54, p. 2-17, 14 mar. 2025. Texto compilado após a Resolução nº 674/2026. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/6001>. Acesso em: 25 jul. 2026.
- COULDRY, Nick; MEJIAS, Ulises A. Data colonialism: rethinking big data's relation to the contemporary subject. *Television & New Media*, Thousand Oaks, v. 20, n. 4, p. 336-349, 2019. DOI: <https://doi.org/10.1177/1527476418796632>.
- DIMAGGIO, Paul J.; POWELL, Walter W. The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, [S. l.], v. 48, n. 2, p. 147-160, 1983. DOI: <https://doi.org/10.2307/2095101>.
- FARRELL, Henry; NEWMAN, Abraham L. Weaponized interdependence: how global economic networks shape state coercion. *International Security*, Cambridge, v. 44, n. 1, p. 42-79, 2019. DOI: https://doi.org/10.1162/isec_a_00351.
- FEDERAL TRADE COMMISSION (United States). *Partnerships Between Cloud Service Providers and AI Developers: FTC Staff Report on AI Partnerships & Investments 6(b) Study*. Washington, DC: FTC, jan. 2025. Disponível em: https://www.ftc.gov/system/files/ftc_gov/pdf/p246201_aipartnerships6breport_redacted_0.pdf. Acesso em: 25 jul. 2026.
- GENOVA, Jandislei Antonio. *Autonomia decisória na era da IA*. São Paulo: Editora Dialética, 2026.
- HADFIELD-MENELL, Dylan et al. The off-switch game. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 26., 2017, Melbourne. *Proceedings [...]*. [S. l.]: IJCAI, 2017. p. 220-227. DOI: <https://doi.org/10.24963/ijcai.2017/32>.
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (United States). *Security and Privacy Controls for Information Systems and Organizations*. Gaithersburg: NIST, 2020. NIST SP 800-53 Rev. 5. DOI: <https://doi.org/10.6028/NIST.SP.800-53r5>. Disponível em: <https://csrc.nist.gov/pubs/sp/800/53/r5/final>. Acesso em: 25 jul. 2026.
- OMOHUNDRO, Stephen M. The basic AI drives. In: WANG, Pei; GOERTZEL, Ben; FRANKLIN, Stan (ed.). *Artificial General Intelligence 2008: proceedings of the First AGI Conference*. Amsterdam: IOS Press, 2008. p.

- 483-492. (Frontiers in Artificial Intelligence and Applications, v. 171). Disponível em: <https://selfawaresystems.com/2007/11/30/paper-on-the-basic-ai-drives/>. Acesso em: 25 jul. 2026.
- PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge, MA: Harvard University Press, 2015.
- PIERSON, Paul. Increasing returns, path dependence, and the study of politics. *American Political Science Review*, Cambridge, v. 94, n. 2, p. 251-267, 2000. DOI: <https://doi.org/10.2307/2586011>.
- PLANTIN, Jean-Christophe et al. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, London, v. 20, n. 1, p. 293-310, 2018. DOI: <https://doi.org/10.1177/1461444816661553>.
- POHLE, Julia; THIEL, Thorsten. Digital sovereignty. *Internet Policy Review*, Berlin, v. 9, n. 4, p. 1-19, 2020. DOI: <https://doi.org/10.14763/2020.4.1532>.
- RAHMAN, K. Sabeel. Infrastructural regulation and the new utilities. *Yale Journal on Regulation*, New Haven, v. 35, n. 3, p. 911-940, 2018. Disponível em: <https://brooklynworks.brooklaw.edu/faculty/1032/>. Acesso em: 25 jul. 2026.
- RAJI, Inioluwa Deborah; DOBBE, Roel. Concrete problems in AI safety, revisited. *arXiv*, 2024. DOI: <https://doi.org/10.48550/arXiv.2401.10899>. Disponível em: <https://arxiv.org/abs/2401.10899>. Acesso em: 25 jul. 2026.
- SRNICEK, Nick. The challenges of platform capitalism: understanding the logic of a new business model. *Juncture*, [S. l.], v. 23, n. 4, p. 254-257, 2017. DOI: <https://doi.org/10.1111/newe.12023>.
- SUSSER, Daniel; ROESSLER, Beate; NISSENBAUM, Helen. Technology, autonomy, and manipulation. *Internet Policy Review*, Berlin, v. 8, n. 2, p. 1-22, 2019. DOI: <https://doi.org/10.14763/2019.2.1410>.
- TABASSI, Elham. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg: National Institute of Standards and Technology, 2023. NIST AI 100-1. DOI: <https://doi.org/10.6028/NIST.AI.100-1>. Disponível em: <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10>. Acesso em: 25 jul. 2026.
- TURNER, Alex et al. Optimal policies tend to seek power. *Advances in Neural Information Processing Systems*, [S. l.], v. 34, p. 23063-23074, 2021. Disponível em: <https://proceedings.neurips.cc/paper/2021/hash/c26820b8a4c1b3c2aa868d6d57e14a79-Abstract.html>. Acesso em: 25 jul. 2026.
- TURNER, Alex; TADEPALLI, Prasad. Parametrically retargetable decision-makers tend to seek power. *Advances in Neural Information Processing Systems*, [S. l.], v. 35, p. 31391-31401, 2022. DOI: <https://doi.org/10.52202/068431-2276>. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2022/hash/cb3658b9983f677670a246c46ece553d-Abstract-Conference.html. Acesso em: 25 jul. 2026.
- UNIÃO EUROPEIA. *Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial e que altera determinados atos legislativos da União*. *Jornal Oficial da União Europeia*, L, 2024/1689, 12 jul. 2024. Disponível em: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>. Acesso em: 25 jul. 2026.
- VAN DIJCK, José; POELL, Thomas; DE WAAL, Martijn. *The platform society: public values in a connective world*. New York: Oxford University Press, 2018. DOI: <https://doi.org/10.1093/oso/9780190889760.001.0001>.
- YEUNG, Karen. “Hypernudge”: big data as a mode of regulation by design. *Information, Communication & Society*, London, v. 20, n. 1, p. 118-136, 2017. DOI: <https://doi.org/10.1080/1369118X.2016.1186713>.