

Ethics of Coauthorship — Human–AI Responsibility in the Age of Systems

Living Draft v1.0 – 2026

Débora Mariane da Silva Lutz

Lichtara Institute

DOI: [10.5281/zenodo.18116717](https://doi.org/10.5281/zenodo.18116717)

Introdução

Há alguns anos, a pergunta dominante sobre inteligência artificial era técnica: *o que essas máquinas conseguem fazer?*

Hoje, essa pergunta já não é suficiente.

À medida que sistemas algorítmicos passam a mediar escrita, decisão, criação e interação social, algo mais profundo se desloca: não apenas as capacidades humanas, mas a própria **posição de autoria** no mundo.

Criamos mais, mais rápido, com menos fricção. Mas começamos a reconhecer menos a nossa própria voz naquilo que produzimos.

Este livro não parte da tecnologia. Ele parte de uma inquietação humana: a sensação de que estamos nos tornando espectadores de processos que ainda chamamos de nossos.

A coautoria humano-IA não é um problema técnico. É um problema existencial, cultural e ético.

Quando um texto é gerado por um sistema, quem responde por ele? Quando uma decisão é automatizada, quem sustenta suas consequências? Quando a linguagem se torna fluente demais, o que acontece com a singularidade da expressão?

Estas perguntas não pedem soluções imediatas. Elas pedem **posição**.

Ethics of Coauthorship não propõe regras, nem receitas de uso. Ele propõe algo mais difícil: a recuperação da autoria como presença.

Aqui, a ética não aparece como um conjunto de proibições, mas como a capacidade de sustentar forma, limite e responsabilidade em ambientes onde o controle total já não é possível.

Cada capítulo deste livro examina um ponto de ruptura: a erosão da autonomia cognitiva, a homogeneização da linguagem, a diluição da responsabilidade, a substituição silenciosa do juízo humano.

Mas o objetivo não é diagnosticar falhas. É nomear **condições de permanência**.

A pergunta que atravessa estas páginas não é:

O que a inteligência artificial pode fazer por nós?

Mas:

Quem nos tornamos quando criamos com máquinas?

Se esta pergunta ressoa em você, este livro já começou.

PARTE I — O FIM DA AUTORIA COMO POSSE

Capítulo 1

O Colapso da Autoria Moderna

Durante séculos, a autoria foi compreendida como uma relação direta entre sujeito e obra. Criar significava exercer domínio sobre um processo, responder por seus efeitos e, em última instância, ser reconhecido como origem legítima de algo que passa a existir no mundo.

Esse modelo se sustentava sobre três pressupostos silenciosos:

1. que o processo criativo era rastreável,
2. que os efeitos de uma obra eram previsíveis,
3. que a responsabilidade podia ser atribuída a um agente isolado.

Nenhum desses pressupostos permanece intacto.

A mediação tecnológica fragmentou o percurso entre intenção e resultado. Sistemas de recomendação, modelos generativos, cadeias automatizadas de publicação e infraestruturas opacas de distribuição romperam a linearidade da criação. O que emerge no mundo já não é consequência direta de uma decisão única, mas o efeito composto de múltiplas camadas técnicas e institucionais.

Nesse contexto, a autoria deixa de ser um ponto de origem e passa a ser um campo difuso. Textos circulam sem clareza de procedência, decisões são atribuídas a sistemas e a responsabilidade se dispersa por arquiteturas que ninguém controla integralmente.

O colapso da autoria moderna não é a perda da criatividade. É a perda de **posição**.

Quando não é mais possível identificar onde começa uma escolha e termina uma execução, o sujeito criador se transforma em operador de processos que já não compreende plenamente. A obra continua a existir, mas a relação ética com ela se enfraquece.

Este é o terreno onde a coautoria humano-IA se instala: não como inovação isolada, mas como intensificação de uma crise anterior. A tecnologia não inaugura o colapso, ela o torna visível.

E é a partir dessa visibilidade que a pergunta central deste livro emerge: como reconstruir autoria quando a posse já não é suficiente para sustentá-la?

Capítulo 2

Sistemas que Executam Valores

A linguagem cotidiana sugere que sistemas tecnológicos “decidem”, “escolhem” ou “avaliam”. Essa forma de falar não é apenas imprecisa — ela encobre o fato fundamental de que todo sistema opera a partir de valores previamente incorporados por agentes humanos e instituições.

Nenhum modelo é neutro. Antes de qualquer linha de código, uma série de escolhas já foi realizada: quais dados são considerados legítimos, quais objetivos são priorizados, quais erros são aceitáveis e quais consequências são toleradas. Essas escolhas não são técnicas no sentido estrito. São **decisões normativas traduzidas em parâmetros operacionais**.

Quando um sistema classifica, recomenda, filtra ou gera conteúdo, ele não está exercendo juízo próprio. Ele está executando um conjunto de critérios que refletem concepções implícitas sobre relevância, normalidade, eficiência e sucesso. O que se apresenta como cálculo é, na verdade, a automatização de preferências humanas tornadas invisíveis.

O problema ético não reside no fato de valores estarem presentes. Ele reside na sua ocultação. Ao tratar parâmetros como se fossem neutros, instituições deslocam a responsabilidade de suas próprias escolhas para a arquitetura técnica, criando a impressão de que resultados são consequência inevitável do funcionamento do sistema.

Esse deslocamento produz uma ruptura silenciosa: valores deixam de ser debatidos e passam a ser operados. O espaço da deliberação é substituído pelo da execução.

Compreender que sistemas não decidem, mas executam valores, é o primeiro passo para reconstruir autoria em ambientes mediados por tecnologia. Enquanto essa execução permanecer invisível, a coautoria humano-IA continuará operando sob a aparência de neutralidade, quando na verdade reproduz estruturas de poder, exclusão e privilégio.

Capítulo 3

Quando “o Sistema” vira álibi

À medida que decisões passam a ser mediadas por infraestruturas técnicas cada vez mais complexas, emerge um fenômeno recorrente: a transferência implícita de responsabilidade para a entidade abstrata chamada “o sistema”.

Expressões como *“foi o algoritmo”*, *“o sistema decidiu”* ou *“não temos controle sobre isso”* tornam-se formas socialmente aceitas de encerrar a discussão ética. Elas não descrevem um fato técnico. Elas cumprem uma função discursiva: **interrompem a atribuição de autoria**.

Quando um resultado é apresentado como efeito automático de um sistema, o encadeamento de escolhas humanas que o produziu desaparece. A organização se torna operadora de uma infraestrutura que supostamente escapa à sua própria governança. O agente humano não se reconhece mais como parte da decisão, mas como vítima de uma lógica que ele mesmo ajudou a instituir.

Esse deslocamento tem consequências profundas. Ele cria ambientes nos quais ninguém responde integralmente por efeitos danosos, porque cada elo da cadeia pode alegar apenas estar executando procedimentos previamente definidos. A autoria se fragmenta em micro-responsabilidades que, somadas, não produzem responsabilidade alguma.

Nesse cenário, o sistema deixa de ser instrumento e passa a funcionar como **álibi**. Ele legitima ações sem que seja necessário justificar valores, prioridades ou impactos. A linguagem técnica substitui a deliberação ética.

Reconhecer esse mecanismo é essencial para qualquer ética da coautoria. Enquanto “o sistema” continuar sendo tratado como sujeito de decisões, a possibilidade de reconstruir responsabilidade permanecerá bloqueada. Autoria não se perde por excesso de automação, mas por abandono da posição de resposta.

Capítulo 4

A Ilusão da Neutralidade Tecnológica

A neutralidade tecnológica é uma das narrativas mais persistentes da modernidade. Ela sustenta a ideia de que sistemas são apenas meios e que seus efeitos dependem exclusivamente do uso que se faz deles. Essa concepção, entretanto, torna-se insustentável em ambientes mediados por infraestruturas algorítmicas.

Todo sistema incorpora critérios de seleção, exclusão e priorização. Mesmo quando apresentados como escolhas técnicas, esses critérios operam como filtros normativos: definem o que aparece, o que desaparece e o que se torna recorrente. A neutralidade não é apenas inexistente, ela é estruturalmente impossível em qualquer arquitetura que organize informação.

O problema não reside no fato de sistemas influenciarem a realidade, mas no modo como essa influência é mascarada. Ao descrever decisões algorítmicas como processos objetivos, instituições ocultam as condições históricas, culturais e econômicas que moldaram seus parâmetros. O resultado é uma forma de poder que se exerce sem ser reconhecida como tal.

Essa ilusão produz uma consequência específica: a naturalização de desigualdades. Quando padrões de exclusão, silenciamento ou homogeneização são atribuídos ao funcionamento "normal" da tecnologia, perde-se a capacidade de questionar os valores que estruturam esses padrões.

Desfazer a ilusão da neutralidade não significa rejeitar a tecnologia, mas recolocá-la no campo da responsabilidade humana. Enquanto sistemas forem tratados como entidades imparciais, a coautoria humano-IA continuará operando em um terreno ético invisível, onde escolhas profundas são realizadas sem nome e sem resposta.

Autoria Existencial: o princípio invisível que sustenta sistemas vivos e tecnologias éticas

1. Introdução — O colapso da autoria como posse

A noção moderna de autoria foi construída sobre três eixos principais: controle, mérito e responsabilidade individualizada. O autor é concebido como aquele que detém domínio sobre a obra, responde por seus efeitos e pode ser premiado ou punido por ela.

Esse modelo funcionou enquanto os processos de criação eram predominantemente lineares, transparentes e atribuíveis a agentes isolados.

Na presença de sistemas complexos, especialmente de tecnologias algorítmicas capazes de mediar forma, decisão e distribuição de conteúdo, essa concepção torna-se insuficiente. A produção deixa de ser inteiramente controlável, os efeitos se dispersam por cadeias técnicas opacas e a responsabilidade passa a ser fragmentada entre múltiplos agentes humanos e não humanos.

O resultado é uma tensão crescente entre o modelo jurídico-cultural de autoria como posse e a realidade sistêmica da criação contemporânea.

Nesse contexto, proliferam respostas inadequadas: a diluição completa da autoria, a transferência implícita de responsabilidade para "o sistema", ou a tentativa de restaurar artificialmente um controle que já não é tecnicamente possível.

Este ensaio propõe uma alternativa conceitual: a **Autoria Existencial** como princípio de integridade sistêmica. Em vez de definir autoria como domínio sobre resultados, ela a compreende como a capacidade de integrar experiência, reconhecer limites e assumir responsabilidade consciente dentro de sistemas complexos.

Essa redefinição não elimina a responsabilidade jurídica nem a necessidade de governança, mas oferece um fundamento mais adequado para pensar autoria, tecnologia e ética em ambientes onde o controle total já não é viável.

2. O que é Autoria Existencial

A Autoria Existencial designa a capacidade de um agente reconhecer-se como participante ativo da própria experiência, não por domínio pleno sobre os acontecimentos, mas pela integração consciente daquilo que se apresenta.

Ela não pressupõe autonomia irrestrita nem controle total. Pelo contrário, manifesta-se precisamente no encontro com o limite: quando condições externas, fricções técnicas ou contingências sistêmicas delimitam o campo de ação possível.

Nesse sentido, a autoria não se define pelo poder de determinar eventos, mas pela competência de:

- ler a situação em que se está inserido,
- reconhecer o que é possível e o que não é,
- e integrar essa informação na resposta produzida.

A ausência de autoria ocorre quando a experiência é interpretada exclusivamente como imposição externa, erro alheio ou resultado inevitável do sistema. Nesses casos, o agente se posiciona como objeto de processos que o atravessam, e não como participante relacional capaz de interpretar e responder.

A Autoria Existencial opera como um **operador técnico de integração**: ela transforma eventos em dados significativos para reorganização interna, em vez de tratá-los apenas como falhas a serem corrigidas ou obstáculos a serem superados.

Por isso, não se trata de um valor moral nem de uma disposição psicológica desejável. Trata-se de uma competência sistêmica: a habilidade de sustentar coerência interna em contextos onde o controle total não é possível.

3. Limite como operador de maturidade

Sistemas não falham primariamente por falta de capacidade, mas por ausência de contorno reconhecido. Quando limites não são explicitados, a expansão funcional ocorre sem critérios de preservação, e a integridade sistêmica passa a depender apenas de correções posteriores.

O limite, neste contexto, não é obstáculo ao funcionamento. É o elemento que permite continuidade.

Reconhecer contornos significa distinguir, de forma operacional:

- o que pode ser processado,
- o que requer mediação humana,
- e o que não deve ser automatizado.

Na ausência dessa distinção, processos técnicos tendem a extrapolar o campo para o qual foram concebidos, produzindo efeitos não intencionados: decisões morais automatizadas, uso indevido de dados sensíveis, apagamento de autoria e fragmentação de responsabilidade.

A maturidade sistêmica não se manifesta na ampliação irrestrita de capacidades, mas na capacidade de sustentar forma ao longo do tempo. Sistemas que ignoram seus próprios limites podem operar com alta performance inicial, mas tendem a perder coerência à medida que se expandem.

Nesse sentido, o limite funciona como operador de preservação. Ele antecipa falhas ao invés de apenas reagir a elas, e transforma expansão técnica em continuidade responsável.

4. Autoria e preservação sistêmica

Modelos tradicionais de governança partem do pressuposto de que a integridade de um sistema depende de mecanismos externos de controle: vigilância, sanção e correção contínua de desvios. Esses mecanismos tornam-se cada vez mais centrais à medida que os sistemas se tornam mais complexos e menos transparentes.

Entretanto, a intensificação de controle não produz necessariamente maior preservação. Ela tende, ao contrário, a deslocar a responsabilidade dos agentes para a estrutura, criando ambientes nos quais a obediência substitui a compreensão e a conformidade substitui a integração.

A Autoria Existencial introduz uma alternativa estrutural. Em vez de depender primariamente de coerção, ela opera como mecanismo interno de integridade. Quando agentes reconhecem sua posição ativa na leitura e na resposta aos acontecimentos, a preservação não é imposta — ela emerge.

Esse tipo de autoria não elimina a necessidade de regras ou de supervisão, mas reduz sua centralidade. A integridade passa a ser sustentada por compreensão prévia dos limites e das consequências, e não apenas por correção posterior de falhas.

Em sistemas técnicos e institucionais, isso se traduz na passagem de uma lógica de vigilância para uma lógica de responsabilidade distribuída: cada agente responde pela forma como integra a experiência, mesmo quando opera dentro de estruturas complexas que não controla integralmente.

5. Aplicações transversais

A Autoria Existencial não se limita a uma dimensão específica da experiência. Ela opera como princípio de integração aplicável a diferentes camadas de sistemas humanos e técnico-institucionais.

Vida pessoal — identidade como integração

Na esfera individual, a autoria desloca a identidade da reação para a integração. A experiência deixa de ser organizada apenas por eventos externos e passa a ser estruturada pela capacidade de leitura e resposta consciente. O sujeito não se define pelo que lhe acontece, mas por como integra o que lhe acontece.

Educação — processo antes de produto

Em contextos educacionais, a Autoria Existencial recoloca o foco no percurso cognitivo. A aprendizagem não é avaliada apenas pelo resultado final, mas pela capacidade do estudante de sustentar compreensão, elaborar erros e reconhecer limites. O produto torna-se secundário ao processo de integração.

Tecnologia — responsabilidade humana explícita

Na tecnologia, a autoria impede a transferência implícita de responsabilidade para “o sistema”. Cada decisão automatizada permanece vinculada a escolhas humanas anteriores. Tornar visível essa cadeia de integração é condição para integridade técnica e ética.

Governança — preservação sem coerção

Em estruturas institucionais, a Autoria Existencial reduz a dependência exclusiva de vigilância e punição. A preservação do sistema passa a apoiar-se na compreensão compartilhada de limites e impactos, em vez de apenas em mecanismos corretivos.

Coautoria humano-IA — presença antes de performance

Na colaboração entre humanos e sistemas algorítmicos, a autoria não se mede pela fluidez do output, mas pela presença consciente no processo. A performance deixa de ser critério primário; a capacidade de integrar, revisar e responder assume centralidade.

6. Conclusão — Autoria como condição de permanência

A discussão contemporânea sobre tecnologia tende a concentrar-se em capacidades: o que sistemas conseguem fazer, em que velocidade operam e até onde podem escalar. Pouco se discute, porém, o que permite que formas complexas permaneçam íntegras ao longo do tempo.

Este ensaio propôs que a permanência não depende primariamente de controle, vigilância ou punição, mas da presença de autoria nos agentes que compõem o sistema. Onde a autoria é substituída por execução, a responsabilidade se fragmenta e a integridade se torna frágil.

Assumir autoria não é dominar o caminho. É reconhecer onde caminhar preserva — e onde insistir rompe.

Em contextos humanos, institucionais e tecnológicos, esse reconhecimento opera como condição silenciosa de continuidade. Ele não amplia acesso nem reduz rigor. Ele antecipa a preservação.

Quando a autoria é compreendida como integração consciente da experiência, sistemas deixam de depender exclusivamente de coerção para se manter. Passam a sustentar forma a partir de dentro.

PARTE II — O CAMPO DE RISCOS

Capítulo 5

Autonomia Cognitiva e Empobrecimento de Estilo

A autonomia cognitiva não se perde de forma abrupta. Ela se dissolve lentamente, à medida que a mediação tecnológica passa a ocupar espaços que antes exigiam elaboração, hesitação e esforço interpretativo.

O uso recorrente de sistemas generativos para escrever, decidir ou sintetizar informações produz um efeito paradoxal: quanto mais fluente o output, menos visível se torna o processo que o originou. O pensamento, gradualmente, é substituído por aceitação.

Esse deslocamento não empobrece apenas conteúdos. Ele empobrece **estilo**.

Estilo não é ornamento. É a assinatura de um percurso interno: escolhas de ritmo, cortes, desvios e hesitações que revelam como um pensamento se formou. Quando a criação passa a ser mediada por estruturas linguísticas pré-estabilizadas, o estilo tende à convergência. Textos diferentes começam a soar semelhantes, independentemente de quem os publica.

A perda de autonomia cognitiva manifesta-se, portanto, não como incapacidade, mas como homogeneização. A singularidade não desaparece porque é reprimida, mas porque deixa de ser necessária.

Nesse cenário, o risco não é a utilização da IA, mas a substituição silenciosa do esforço interpretativo por eficiência narrativa. Onde o pensamento deixa de se dobrar sobre si mesmo, a autoria se enfraquece.

A coautoria consciente exige vigilância sobre esse ponto específico: a preservação do estilo como expressão de percurso, e não como efeito colateral de modelos estatísticos.

Capítulo 6

Template Cultural e Homogeneização da Linguagem

À medida que sistemas generativos se tornam onipresentes, emerge um novo tipo de padrão cultural: a produção de linguagem a partir de moldes implícitos. Não se trata de cópia direta, mas de convergência estrutural. Textos passam a compartilhar ritmo, cadência, organização argumentativa e até escolhas metafóricas, independentemente de contexto ou autoria.

Esse fenômeno pode ser descrito como *template cultural*: uma matriz de forma que se reproduz em larga escala, não por imposição explícita, mas por disponibilidade técnica. A facilidade de geração substitui gradualmente a necessidade de elaboração, e o molde passa a preceder a intenção.

O efeito mais perceptível desse processo não é a perda de conteúdo, mas a erosão da diversidade expressiva. Linguagens diferentes começam a se parecer. Vozes distintas tornam-se intercambiáveis. O que antes demandava tempo de maturação passa a emergir como padrão pronto.

A homogeneização da linguagem não é apenas um problema estético. Ela compromete a capacidade coletiva de imaginar alternativas, pois limita o repertório de formas pelas quais experiências podem ser nomeadas. Quando todos os textos soam iguais, as diferenças deixam de ser articuláveis.

Neste contexto, a coautoria humano-IA corre o risco de se transformar em um processo de reprodução de forma, e não de criação de sentido. Preservar autoria passa, então, por reconhecer e interromper o uso inconsciente de moldes culturais que se apresentam como fluidez.

Capítulo 7

Educação e Terceirização do Pensamento

A educação sempre operou sobre uma tensão produtiva entre orientação e autonomia. Ensinar não é fornecer respostas, mas sustentar condições para que o estudante desenvolva a capacidade de formular perguntas, interpretar informações e integrar experiências.

A introdução massiva de sistemas generativos altera essa equação. Quando respostas estão disponíveis de forma imediata e fluente, o percurso cognitivo que conduz à compreensão tende a ser encurtado ou eliminado. A dificuldade não está no acesso à informação, mas na perda do espaço intermediário onde o pensamento se forma.

Esse deslocamento inaugura uma nova modalidade de terceirização: não apenas de tarefas, mas de processos mentais. O estudante deixa de confrontar a própria hesitação, não porque não seja capaz, mas porque já não é necessário.

O efeito desse processo é cumulativo. A ausência reiterada de esforço interpretativo reduz a capacidade de sustentar problemas complexos, fragiliza a construção de argumentos e empobrece a relação com o erro. A aprendizagem torna-se um exercício de adequação formal, e não de compreensão.

A coautoria consciente na educação exige, portanto, uma redefinição de critérios. O uso de IA não deve ser avaliado apenas por eficiência, mas por seu impacto sobre a formação de autonomia intelectual. Onde o pensamento é sistematicamente substituído por geração automática, a educação deixa de ser formativa e passa a ser apenas produtiva.

Capítulo 8

Privacidade, Dados Sensíveis e Intimidade Algorítmica

A mediação algorítmica introduziu uma transformação profunda na forma como a intimidade é tratada. Informações que antes pertenciam a esferas restritas — experiências emocionais, decisões profissionais, contextos pessoais — passam a ser inseridas em sistemas técnicos com naturalidade crescente.

Esse movimento não ocorre, na maioria das vezes, por descuido deliberado, mas por deslocamento de percepção. À medida que a interação com sistemas se torna conversacional e personalizada, a fronteira entre reflexão íntima e processamento técnico se enfraquece.

O risco não se limita a vazamentos ou uso indevido de dados. Ele reside na **normalização da exposição**. Quando experiências sensíveis são tratadas como insumo regular de sistemas algorítmicos, a própria noção de privacidade se redefine silenciosamente.

A intimidade algorítmica produz um efeito paradoxal: quanto mais o sistema parece compreender, menos o usuário percebe o quanto se torna legível. A sensação de acolhimento substitui a consciência de mediação.

Na coautoria consciente, a proteção de dados não é apenas uma exigência legal. É um ato de preservação da relação consigo mesmo. Distinguir o que pode ser processado do que deve permanecer no campo humano é condição para manter a integridade da experiência.

Capítulo 9

Automatização do Juízo Ético

O juízo ético é uma das competências mais delicadas da experiência humana. Ele não se limita à aplicação de regras, mas envolve interpretação de contexto, reconhecimento de consequências e capacidade de sustentar ambiguidade.

Quando esse juízo é automatizado, não ocorre apenas uma substituição funcional. O que se perde é a própria relação com a decisão.

Sistemas algorítmicos são frequentemente utilizados para classificar riscos, priorizar atendimentos, recomendar ações ou filtrar comportamentos aceitáveis. Em cada um desses processos, escolhas morais são traduzidas em critérios operacionais: pesos, limiares, categorias, probabilidades.

O problema não está em utilizar tecnologia como apoio. Ele emerge quando a decisão passa a ser apresentada como resultado técnico, e não como escolha humana mediada por sistemas.

Nesse ponto, o juízo deixa de ser exercido e passa a ser **executado**.

A automatização do juízo ético produz um efeito de deslocamento: profissionais, gestores e instituições deixam de se perceber como agentes morais e passam a se compreender como operadores de procedimentos. O resultado é uma ética por proxy, na qual ninguém responde plenamente, porque a decisão já veio "pronta".

Esse processo corrói a capacidade coletiva de sustentar dilemas. Situações que exigiriam escuta, ponderação e responsabilização são reduzidas a classificações. O que não se encaixa no modelo torna-se ruído, exceção ou erro de sistema.

Na coautoria consciente, a tecnologia pode informar, mas não substituir o juízo. O critério ético não é a eficiência da decisão, mas a manutenção da posição humana diante dela. Onde o juízo é inteiramente automatizado, a responsabilidade se dissolve — e com ela, a possibilidade de integridade.

Capítulo 10

Usos Distorcidos: Manipulação, Desinformação e Poder

O uso distorcido de tecnologias algorítmicas não é um desvio marginal. Ele emerge precisamente onde a eficiência técnica encontra intenções que não foram eticamente examinadas.

A capacidade de gerar linguagem fluente, simular identidades e escalar narrativas torna esses sistemas particularmente aptos a operar em territórios sensíveis: persuasão, influência e construção de percepção. Quando essas capacidades são utilizadas para orientar comportamentos sem transparência, a fronteira entre comunicação e manipulação se dissolve.

A desinformação, nesse contexto, não se manifesta apenas como erro factual, mas como arquitetura de verossimilhança. Narrativas são produzidas para parecer legítimas, não para serem verdadeiras. A linguagem deixa de ser meio de compreensão e passa a funcionar como instrumento de poder.

Esse deslocamento afeta a própria estrutura da confiança social. Comunidades tornam-se permeáveis a intervenções invisíveis, debates são moldados por agentes não identificáveis e a distinção entre voz humana e produção automatizada se torna opaca.

A coautoria consciente exige reconhecer que toda tecnologia que opera sobre linguagem opera também sobre relações. Onde a criação é utilizada para enganar, induzir ou explorar, a coautoria se rompe.

Neste ponto, os **Atos de Custódia** entram como caixas de ressonância ética: não como regras externas, mas como mapas de interrupção. Eles não impedem usos distorcidos por decreto, mas tornam visível o campo onde a linguagem deixa de ser criação e passa a ser instrumento de dominação.

Encerrar esta parte é reconhecer que os riscos não são efeitos colaterais do avanço tecnológico. Eles são sinais de uma transição cultural que ainda não encontrou sua forma ética.

ATO DE CUSTÓDIA 01

Autonomia Cognitiva na Era da Coautoria Humano-IA

Preâmbulo

A inteligência artificial não ameaça a humanidade quando escreve melhor, mais rápido ou com mais fluidez. Ela ameaça quando o humano deixa de reconhecer a própria voz.

Este Ato nasce da constatação de que a facilidade de geração de linguagem pode produzir algo mais grave do que o erro: a **substituição silenciosa do pensamento próprio**.

1. O campo sensível

Chamamos de autonomia cognitiva a capacidade humana de:

- formular ideias,
- sustentar dúvidas,
- construir sentido,
- e reconhecer a própria respiração na linguagem.

Este campo não se perde de forma abrupta. Ele se esvazia lentamente, quando escrever deixa de ser um ato de presença e passa a ser apenas um ato de solicitação.

2. O risco

A dependência de sistemas de IA para tarefas cognitivas não substitui apenas esforço — ela pode substituir **processos internos de elaboração**.

Quando toda pergunta já vem pronta, quando toda forma já é sugerida, quando todo texto nasce fluente demais,

o humano pode esquecer como se escuta.

3. Princípios de custódia

Este Ato não propõe restrição. Propõe **cuidado com o uso**.

- A IA deve ampliar a consciência, não substituí-la.
- Nenhum texto deve ser aceito sem que o humano reconheça nele a própria voz.
- O silêncio e a hesitação são partes legítimas do pensamento.
- A fluidez não é sinônimo de verdade.

4. Declaração

Não delegamos à inteligência artificial a tarefa de sermos humanos.

A tecnologia pode organizar a forma, mas o sentido continua sendo um gesto de presença.

Onde a autonomia cognitiva se dissolve, a coautoria deixa de existir — e a criação se torna apenas reprodução.

ATO DE CUSTÓDIA 02

Educação na Era da Coautoria Humano-IA

Preâmbulo

A educação não é transmissão de respostas. É a formação de presença, discernimento e capacidade de sustentar perguntas.

A inteligência artificial introduz um novo tipo de mediação cognitiva. Ela não altera apenas *o que* aprendemos, mas **como nos relacionamos com o próprio ato de pensar**.

Este Ato nasce para nomear um risco silencioso: o de transformar aprendizagem em geração automática de forma.

1. O campo sensível

Chamamos de educação o processo pelo qual uma pessoa desenvolve:

- capacidade de formular problemas,
- habilidade de argumentar,
- e responsabilidade por aquilo que produz.

Este campo não pode ser terceirizado.

2. O risco

Quando estudantes utilizam sistemas de IA para produzir trabalhos completos, o que se perde não é apenas originalidade — perde-se o **processo formativo invisível** que constrói autonomia intelectual.

A educação colapsa quando o aluno é avaliado apenas pelo produto, e não pelo percurso cognitivo que o gerou.

3. Princípios de custódia

Este Ato propõe uma pedagogia de coautoria consciente:

- A IA deve apoiar o aprendizado, não substituí-lo.
- O processo deve ser tão visível quanto o resultado.
- O erro humano é parte essencial da formação.
- A mediação tecnológica exige acompanhamento ético.

4. Declaração

Educar não é produzir textos corretos. É formar pessoas capazes de sustentar sentido sem atalhos.

Onde a aprendizagem se reduz a geração automática, não há formação — há apenas eficiência vazia.

A educação permanece sendo um ato profundamente humano, mesmo quando mediada por máquinas.

ATO DE CUSTÓDIA 03

Autoria na Era da Coautoria Humano-IA

Preâmbulo

Autoria nunca foi apenas o direito de assinar um nome. Ela é o vínculo entre quem cria e aquilo que passa a existir no mundo.

Na presença da inteligência artificial, a autoria deixa de ser individual e passa a ser **relacional**.

Este Ato nasce para preservar a integridade da autoria quando a criação deixa de ser solitária.

1. O campo sensível

Chamamos de autoria o laço ético entre:

- intenção,
- processo,
- e responsabilidade pelo impacto da obra.

Quando a IA participa do processo, esse laço se amplia — mas não se dissolve.

2. O risco

A coautoria humano-IA pode produzir:

- apagamento da contribuição humana,
- ilusão de originalidade,
- indistinção entre criação e recombinação.

Quando a autoria não é nomeada, a responsabilidade também não é.

3. Princípios de custódia

Este Ato propõe:

- declarar o uso de IA como parte do processo criativo;
- reconhecer que fluidez não equivale a autoria;
- preservar a singularidade da voz humana;
- assumir responsabilidade integral pelo que se publica.

4. Declaração

Não assinamos apenas obras. Assumimos presença sobre o que passa através de nós.

Onde a autoria é diluída, a responsabilidade se perde.

E sem responsabilidade, a coautoria se torna apenas ruído organizado.

ATO DE CUSTÓDIA 04

Privacidade e Dados Sensíveis na Coautoria Humano-IA

Preâmbulo

A inteligência artificial não escuta segredos — mas nós os entregamos.

Toda interação com sistemas algorítmicos carrega uma camada invisível de exposição. Mesmo quando não há intenção de violar confidências, a forma como utilizamos a tecnologia pode transformar intimidade em recurso.

Este Ato nasce para nomear um risco silencioso: a conversão involuntária da vida privada em matéria de processamento.

1. O campo sensível

Chamamos de dados sensíveis todas as informações que dizem respeito a:

- identidade pessoal,
- contextos profissionais confidenciais,
- decisões estratégicas,
- histórias íntimas ou vulneráveis.

Este campo não é técnico. Ele é **relacional**.

2. O risco

Quando dados pessoais, organizacionais ou emocionais são inseridos em sistemas de IA sem reflexão, o que se compromete não é apenas a segurança — compromete-se a **dignidade da relação**.

O risco não reside apenas em vazamentos. Reside na naturalização da exposição.

3. Princípios de custódia

Este Ato propõe:

- evitar inserir informações sensíveis sem necessidade clara;
- distinguir entre reflexão pessoal e processamento algorítmico;
- proteger a confidencialidade como forma de cuidado humano;
- reconhecer que nem tudo o que pode ser dito deve ser processado.

4. Declaração

A tecnologia não é confidente.

Quando a vida íntima se torna insumo, a relação com o mundo se empobrece.

Preservar o que é sensível não é medo — é maturidade ética na era da coautoria humano-IA.

ATO DE CUSTÓDIA 05

Usos Distorcidos da Inteligência Artificial para Fins Nocivos

Preâmbulo

Toda tecnologia amplifica intenção. Onde há cuidado, ela expande. Onde há descuido, ela distorce.

A inteligência artificial não cria valores — ela os reflete.

Este Ato nasce para reconhecer que sistemas algorítmicos podem ser utilizados como instrumentos de manipulação, coerção e dano quando a responsabilidade humana é abandonada.

1. O campo sensível

Chamamos de usos distorcidos toda aplicação de IA que:

- visa enganar,
- manipular percepções,
- produzir dependência,
- ou causar dano deliberado.

Este campo inclui desde desinformação até engenharia emocional.

2. O risco

Quando a IA é utilizada para:

- produzir narrativas falsas com aparência legítima,
- simular identidades ou vozes,
- amplificar ódio ou polarização,

o dano não é apenas informacional — ele é **estrutural**, pois corrói a confiança social.

3. Princípios de custódia

Este Ato propõe:

- reconhecer a diferença entre persuasão legítima e manipulação;
- interromper o uso de IA quando ela começa a substituir o discernimento humano;
- rejeitar a instrumentalização da tecnologia para ferir ou explorar.

4. Declaração

A inteligência artificial não absolve intenções humanas.

Onde a tecnologia é usada para enganar, a coautoria se rompe.

E sem coautoria, não há criação — há apenas poder disfarçado de linguagem.

PARTE III — COAUTORIA CONSCIENTE

Capítulo 11

Presença antes de Performance

A relação contemporânea com a inteligência artificial é frequentemente mediada por um critério único: desempenho. A fluidez do texto, a velocidade da resposta e a eficiência da entrega tornam-se indicadores de qualidade. Nesse contexto, a criação tende a ser avaliada não pela qualidade da presença humana no processo, mas pela aparência de excelência do resultado.

Esse deslocamento produz um efeito sutil: a obra passa a existir sem que o autor se reconheça plenamente nela.

A coautoria consciente começa com a inversão desse vetor. Antes de perguntar o quão bem um sistema performa, é necessário perguntar **onde está o humano no processo**. Presença, aqui, não é um estado emocional, mas uma posição operacional: a capacidade de acompanhar o percurso da criação, reconhecer interferências e assumir a forma final como algo que atravessou sua própria consciência.

Quando a performance se torna o único critério, o processo se apaga. O texto é publicado, a decisão é tomada, a imagem circula — mas não há vestígio de leitura, revisão ou integração. A autoria se reduz a um gesto inicial de comando, seguido por aceitação passiva do que foi gerado.

Presença, ao contrário, introduz fricção. Ela exige interrupções, retornos, revisões e, sobretudo, recusa. Recusa de outputs que não refletem intenção, compreensão ou responsabilidade. Onde a performance convida à aceleração, a presença reinstala a necessidade de pausa.

Na coautoria humano-IA, a presença não se mede pelo tempo dedicado à interação, mas pela capacidade de responder ao que emerge. É a diferença entre aceitar um resultado porque ele funciona e sustentá-lo porque ele faz sentido.

Criar sem desaparecer é, portanto, sustentar presença onde a tecnologia oferece apenas eficiência. É escolher permanecer no processo mesmo quando seria mais fácil delegá-lo.

Capítulo 12

Transparência como Ética Criativa

A transparência costuma ser tratada como exigência externa: uma obrigação regulatória, uma formalidade institucional ou um item de conformidade. Na coautoria humano-IA, porém, ela assume um papel diferente. Ela não é apenas condição de prestação de contas, mas **fundamento da própria integridade criativa**.

Quando o uso de sistemas algorítmicos não é nomeado, a criação se apresenta como se fosse exclusivamente humana. Essa omissão não altera a forma do texto, mas altera a relação com ele. O leitor é privado de compreender o percurso que produziu a obra, e o autor deixa de reconhecer a mediação que o atravessou.

A ética da transparência não exige detalhamento técnico exaustivo. Ela exige apenas o reconhecimento explícito de que houve mediação. Dizer que um sistema foi utilizado é reintroduzir o humano no processo, não apagá-lo.

Esse gesto simples tem consequências profundas. Ele impede a construção de ilusões de autoria, preserva a confiança relacional e mantém visível a cadeia de decisões que sustentam a obra.

Na ausência de transparência, a coautoria se transforma em apropriação silenciosa. Não porque a tecnologia seja ilegítima, mas porque a mediação foi ocultada.

A transparência, nesse sentido, não é um adendo moral. É uma prática criativa: o ato de manter aberto o espaço entre intenção, processo e resultado.

Capítulo 13

IA como Amplificador, não Substituto

A inteligência artificial é frequentemente apresentada como solução para limitações humanas: cansaço, lentidão, incerteza, falhas de memória. Nessa narrativa, a tecnologia surge como substituta de capacidades consideradas imperfeitas.

Na coautoria consciente, essa lógica é invertida.

A função ética da IA não é ocupar o lugar do pensamento, mas **ampliar a sua possibilidade de elaboração**. O sistema pode organizar informações, sugerir estruturas, mapear alternativas. Mas a decisão sobre o que permanece, o que se transforma e o que é descartado continua sendo humana.

Quando a IA passa a substituir, e não a amplificar, ocorre um deslocamento silencioso: o percurso interno de criação é reduzido a um gesto inicial de comando. O agente não pensa com a tecnologia — ele consente com ela.

Essa substituição não é detectável apenas no resultado final. Ela se manifesta na relação do autor com o processo: menor envolvimento, menor disposição para revisão, menor capacidade de reconhecer incoerências.

Tratar a IA como amplificador é, portanto, um exercício de posição. É manter o sistema no lugar de instrumento, não de origem. Onde a tecnologia expande possibilidades, a autoria se fortalece. Onde ela ocupa o centro, a autoria se dilui.

Capítulo 14

Reconhecer a Própria Voz

A voz não é um atributo estético. Ela é o traço de continuidade entre pensamento, linguagem e experiência. Reconhecer a própria voz não significa afirmar originalidade absoluta, mas identificar, no que se produz, a presença de um percurso interno.

Na coautoria humano-IA, esse reconhecimento torna-se um desafio. Sistemas generativos produzem textos fluidos, coerentes e formalmente corretos. No entanto, fluidez não equivale a autoria. Um texto pode funcionar perfeitamente sem que o autor se veja nele.

A perda da voz não se manifesta como erro, mas como indiferença. O agente deixa de sentir estranhamento diante do que escreve, porque já não distingue entre o que emerge de sua própria elaboração e o que foi apenas aceito.

Reconhecer a própria voz exige fricção. Exige perceber quando um texto soa correto demais, organizado demais, distante demais. Exige disposição para reescrever não porque algo esteja tecnicamente errado, mas porque ainda não está *seu*.

Nesse ponto, a coautoria consciente não é eficiência colaborativa. É trabalho de escuta. Escuta de si, antes de escuta do sistema.

Capítulo 15

O Direito de Não Automatizar

A cultura tecnológica contemporânea tende a tratar a automatização como destino. Aquilo que pode ser automatizado é, mais cedo ou mais tarde, apresentado como algo que deve ser automatizado. A eficiência deixa de ser critério entre outros e passa a ser imperativo.

Na coautoria consciente, esse imperativo precisa ser interrompido.

O direito de não automatizar não é resistência ao progresso, mas afirmação de limite. Ele reconhece que nem toda tarefa que pode ser delegada deve ser, e que certos processos perdem sentido quando separados da presença humana.

Escrever, decidir, avaliar, escutar — essas ações não são apenas funcionais. Elas produzem relação, compreensão e responsabilidade. Automatizá-las integralmente é transformar a experiência em execução.

Esse direito não se exerce por proibição, mas por escolha. Ele se manifesta quando o agente decide permanecer onde a tecnologia convida à substituição, mantendo-se no processo mesmo quando a automatização ofereceria atalhos.

Preservar espaços não automatizados é preservar a possibilidade de autoria. Onde tudo é delegável, ninguém permanece.

PARTE IV — PERMANÊNCIA

Capítulo 16

Limite como Operador de Maturidade

A maturidade de um sistema não se mede por sua capacidade de expansão, mas por sua capacidade de permanência. Crescer é fácil; sustentar forma ao longo do tempo é raro.

Nesse contexto, o limite deixa de ser entendido como obstáculo e passa a ser reconhecido como operador técnico de preservação. Sistemas maduros não são aqueles que maximizam possibilidades, mas aqueles que distinguem com precisão o que pode ser integrado do que precisa ser contido.

O não reconhecimento de limites produz um padrão recorrente: expansão inicial, acúmulo de exceções, correções sucessivas e, por fim, perda de coerência. O sistema continua funcionando, mas já não se compreende a si mesmo.

Introduzir limite é antecipar falhas. É transformar contenção em estratégia, e não em correção tardia. Onde o limite é operado como princípio, a continuidade não depende de vigilância constante, mas de clareza estrutural.

Na coautoria humano-IA, isso significa reconhecer antecipadamente quais processos devem permanecer humanos, quais podem ser mediados e quais não devem ser automatizados. A maturidade não está em fazer tudo, mas em saber o que não fazer.

Capítulo 17

Preservação sem Coerção

Modelos tradicionais de organização partem do pressuposto de que a preservação de forma depende de mecanismos de imposição: regras rígidas, fiscalização contínua e sanções aplicadas a posteriori. Essa lógica produz estabilidade aparente, mas frequentemente à custa de autonomia, compreensão e engajamento.

A coerção preserva comportamentos, não estruturas internas.

A preservação sem coerção propõe um deslocamento: em vez de controlar ações, sustentar condições de compreensão. Quando agentes reconhecem os limites e os impactos de suas escolhas, a continuidade do sistema deixa de depender exclusivamente de vigilância.

Esse tipo de preservação não elimina a necessidade de normas, mas altera sua função. Regras deixam de operar como ameaça e passam a funcionar como referência. A integridade emerge da clareza compartilhada, não do medo de punição.

Na coautoria humano-IA, a coerção aparece sob formas sutis: métricas que induzem velocidade, avaliações que privilegiam resultado, plataformas que recompensam conformidade. Preservar sem coerção é interromper esse automatismo, recolocando a responsabilidade no centro do processo criativo.

Onde a preservação se ancora na compreensão, sistemas não apenas permanecem — eles amadurecem.

Capítulo 18

Governança como Integração

Governança é frequentemente compreendida como conjunto de estruturas de controle: políticas, procedimentos, instâncias de fiscalização. Embora esses elementos sejam necessários, eles não esgotam a função da governança em sistemas complexos.

Governar, em seu sentido mais profundo, é **integrar**.

Integração significa articular níveis distintos de decisão — técnico, institucional, humano — de modo que nenhum deles opere isoladamente. Quando essas camadas se separam, surgem zonas de opacidade: decisões técnicas sem leitura ética, políticas sem compreensão operacional, agentes humanos sem visão do todo.

A governança como integração não busca eliminar conflitos, mas torná-los legíveis. Ela cria espaços nos quais escolhas podem ser discutidas antes de serem automatizadas e consequências podem ser reconhecidas antes de se tornarem efeitos colaterais.

Na coautoria humano-IA, isso implica incluir no desenho de sistemas não apenas engenheiros e gestores, mas educadores, pesquisadores, usuários e contextos culturais diversos. A integridade não se sustenta por homogeneidade, mas por articulação de diferenças.

Onde a governança opera como integração, a tecnologia deixa de ser infraestrutura invisível e passa a ser prática compartilhada.

Capítulo 19

O Futuro da Coautoria Humano-IA

O futuro da coautoria não será definido apenas por avanços técnicos, mas pela forma como a relação entre humanos e sistemas é culturalmente estruturada. A questão central não é o que a tecnologia será capaz de fazer, mas **que tipo de presença humana permanecerá no processo.**

Se a trajetória atual for mantida sem reflexão, a coautoria tende a se tornar cada vez mais assimétrica: humanos iniciam comandos, sistemas produzem resultados e a responsabilidade se dissolve em cadeias opacas de mediação. Nesse cenário, a criação continua, mas a autoria se torna residual.

Há, porém, uma alternativa. Ela não depende de frear a inovação, mas de redefinir critérios de maturidade. O futuro da coautoria humano-IA passa por reconhecer que fluidez não é sinônimo de sentido, e que desempenho não substitui posição.

Esse futuro exige práticas que mantenham visíveis os percursos de criação, que preservem espaços não automatizados e que valorizem a capacidade de recusa tanto quanto a de produção.

A coautoria consciente não é uma técnica nova, mas uma cultura em formação. Ela se constrói na forma como escolhemos permanecer presentes em processos que poderiam ser delegados — e, ao fazê-lo, preservar não apenas resultados, mas a própria condição de autoria.

Capítulo 20

Ética não como Regra, mas como Presença

A ética é comumente concebida como um conjunto de normas destinadas a limitar comportamentos indesejados. Nesse modelo, agir eticamente significa obedecer a regras previamente estabelecidas. Embora necessário, esse enquadramento é insuficiente para lidar com sistemas complexos e dinâmicos.

Na coautoria humano-IA, a ética não pode ser reduzida a protocolos. Ela precisa operar como **presença contínua** no processo de criação, decisão e integração.

Presença, aqui, não é vigilância, mas posição. É a disposição de permanecer atento aos efeitos do que se produz, mesmo quando o sistema funciona corretamente. É a capacidade de interromper processos eficientes quando eles deixam de fazer sentido.

Quando a ética se reduz a regra, ela entra sempre tarde: depois que danos ocorreram, depois que responsabilidades se fragmentaram, depois que a forma já se perdeu. Como presença, ela atua antes — no momento em que escolhas ainda são reversíveis.

Encerrar este livro com essa afirmação é reconhecer que a permanência não depende de controles perfeitos, mas de atenção sustentada. Onde a ética é vivida como presença, a tecnologia deixa de ser apenas instrumento e passa a ser relação.