

UNIVERSIDADE FEDERAL DO MARANHÃO
AGÊNCIA DE INOVAÇÃO, EMPREENDEDORISMO, PESQUISA, PÓS-GRADUAÇÃO
E INTERNACIONALIZAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
MESTRADO ACADÊMICO

LEONARDO PASSOS RIBEIRO

**HERMENÊUTICA E O PROBLEMA DA VIGILÂNCIA ALGORÍTMICA:
APROXIMAÇÃO DO CONCEITO RICOEURIANO DE INSTITUIÇÕES JUSTAS COM O
*EU AI ACT***

SÃO LUÍS
2026

LEONARDO PASSOS RIBEIRO

**HERMENÊUTICA E O PROBLEMA DA VIGILÂNCIA ALGORÍTMICA:
APROXIMAÇÃO DO CONCEITO RICOEURIANO DE INSTITUIÇÕES JUSTAS COM O
*EU AI ACT***

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Federal do Maranhão, como requisito para a obtenção do título de Mestre em Filosofia.
Orientador: Prof. Dr. Wandelson Silva de Miranda

SÃO LUÍS
2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).

Diretoria Integrada de Bibliotecas/UFMA

Passos Ribeiro, Leonardo.

HERMENÊUTICA E O PROBLEMA DA VIGILÂNCIA ALGORÍTMICA:
aproximação do conceito ricoeuriano de instituições justas com o EU AI Act /
Leonardo Passos Ribeiro. – 2026.

76 p.

Orientador(a): Wandeilson Silva de Miranda.

Dissertação (Mestrado) – Programa de Pós-graduação em Filosofia, Universidade
Federal do Maranhão, São Luís, 2026.

1. Vigilância Algorítmica. 2. Hermenêutica. 3. Paul Ricoeur. 4. Instituições Justas. 5.
EU AI ACT. I. Silva de Miranda, Wandeilson. II. Título.

LEONARDO PASSOS RIBEIRO

**HERMENÊUTICA E O PROBLEMA DA VIGILÂNCIA ALGORÍTMICA:
APROXIMAÇÃO DO CONCEITO RICOEURIANO DE INSTITUIÇÕES JUSTAS COM O
*EU AI ACT***

Dissertação apresentada ao Programa de
Pós-Graduação em Filosofia da Universidade
Federal do Maranhão, como requisito para a
obtenção do título de Mestre em Filosofia.
Orientador: Prof. Dr. Wandilson Silva de
Miranda

Data da aprovação: __/__/__

Conceito:

BANCA EXAMINADORA

Orientador (a):

Prof. Dr. Wandilson Silva de Miranda- UFMA
Universidade Federal do Maranhão

Examinador Externo:

Prof. Dr. Delmo Mattos Silva- ITA
Instituto Tecnológico de Aeronáutica

Examinador Interno:

Prof. Dr. Francisco Valderio Pereira da Silva Junior- UEMA
Universidade Estadual do Maranhão

RESUMO

Face às implicações da vigilância algorítmica e à centralidade da identidade narrativa em Paul Ricoeur, a presente dissertação analisa o *EU AI Act*: um marco regulatório de IAs pioneiro nas relações internacionais. Procedeu-se ao escrutínio de algumas disposições do *EU AI Act* que foram originalmente concebidas para enfrentar o problema da opacidade de sistemas de inteligência artificial. A metodologia empregada ancora-se na hermenêutica filosófica, particularmente na dimensão semiótica ricoeuriana do signo. O *corpus* é constituído pelo *EU AI Act* (Regulamento UE 2024/1689), cujo distanciamento metodológico do seu texto (transição da explicação do signo para a compreensão), em termos ricoeurianos, permite perscrutar instâncias de mediação com o objetivo de trabalhar o fenômeno da vigilância algorítmica. A hipótese norteadora, corroborada no decurso da pesquisa, comprovou uma aproximação presumível da abordagem de Ricoeur com alguns dispositivos do *EU AI Act* - a exemplo da vedação ao policiamento preditivo fundado em perfis de personalidade. Constatou-se, nesta aproximação, uma salvaguarda possível da identidade narrativa, na medida em que o legislador europeu impõe limites aos sistemas de inteligência artificial: à predição algorítmica e à opacidade dos dados. Os resultados da pesquisa especificam, ainda, o amparo desta última circunstância em alguns dos dispositivos formulados pelo legislador, cujo intuito foi obstar a conversão do homem em métrica previsível - como a interdição do policiamento preditivo (UNIÃO EUROPEIA, 2024, art. 5.º, n.º 1, alínea g) - e tutelar o direito de obter explicações claras decorrentes de decisões algorítmicas (UNIÃO EUROPEIA, 2024, art. 86.º, n.º 1), prerrogativa por meio da qual o *EU AI Act* pode tangenciar o horizonte ricoeuriano. A advertência recai, contudo, na fragilidade da norma puramente positiva em resguardar as instituições justas.

Palavras-chave: Inteligência Artificial; Vigilância Algorítmica; Paul Ricoeur; Instituições Justas; *EU AI Act*.

ABSTRACT

Faced with the implications of algorithmic surveillance and the centrality of narrative identity in Paul Ricoeur, this dissertation analyzes the *EU AI Act*: a pioneering AI regulatory framework in economic and social relations. A scrutiny was conducted on certain provisions of the *EU AI Act* originally conceived to address the opacity problem in artificial intelligence systems. The methodology employed is anchored in philosophical hermeneutics, particularly in the Ricoeurian semiotic dimension of the sign. The *corpus* consists of the *EU AI Act* (Regulation EU 2024/1689), whose methodological distancing of its text (the transition from the explanation of the sign to comprehension), in Ricoeurian terms, allows for the investigation of mediations regarding the phenomenon of algorithmic surveillance. The guiding hypothesis, corroborated over the course of the investigation, demonstrated a presumed convergence between Ricoeur's approach and certain provisions of the *EU AI Act* — such as the categorical ban on predictive policing based on personality profiles. In this convergence, a presumed safeguard of identities is observed, insofar as the European legislator imposes limits on artificial intelligence systems: on algorithmic prediction and data opacity. The research results further specify the support for this latter circumstance in some of the provisions formulated by the legislator, whose intent was to prevent the conversion of man into a predictable metric — such as the prohibition of predictive policing (EUROPEAN UNION, 2024, art. 5, no. 1, letter g) — and to protect the right to obtain clear explanations resulting from algorithmic decisions (EUROPEAN UNION, 2024, art. 86, no. 1), a prerogative through which the *EU AI Act* can touch upon the Ricoeurian horizon. The caveat, however, lies in the fragility of the purely positive norm in safeguarding just institutions.

Keywords: Artificial Intelligence; Algorithmic Surveillance; Paul Ricoeur; Just Institutions; *EU AI Act*.

SUMÁRIO

INTRODUÇÃO.....	8
1. O HORIZONTE HERMENÊUTICO PROFUNDO EM SISTEMAS DE IA.....	11
1.1. EUROPEAN UNION ARTIFICIAL INTELLIGENCE ACT (EU AI ACT).....	19
1.2. AS INSTITUIÇÕES JUSTAS EM PAUL RICOEUR.....	26
1.3. A HERMENÊUTICA COMO FILOSOFIA DA MEDIAÇÃO E DILEMAS GERAIS.....	32
1.4. RICOEUR E A ÉTICA DA JUSTIÇA: DA IDENTIDADE NARRATIVA À SOLICITUDE PELO OUTRO.....	38
1.5. ERA ALGORÍTMICA.....	41
2. O EU AI ACT: UMA HERMENÊUTICA POSSÍVEL.....	46
2.1. A VIGILÂNCIA ALGORÍTMICA E SEUS EFEITOS SOCIAIS: PERSPECTIVAS CRÍTICAS À LUZ DO EU AI ACT.....	54
2.2. O EU AI ACT: AVANÇOS E DESAFIOS.....	61
2.3. UMA REGULAÇÃO ÉTICA.....	64
CONSIDERAÇÕES FINAIS.....	67

INTRODUÇÃO

A recente utilização do modelo de linguagem *Claude*, desenvolvido pela empresa *Anthropic*, pelas Forças Armadas do Pentágono para auxiliar na seleção de alvos e em simulações de campo de batalha durante a ofensiva contra um país do Oriente Médio (CNN BRASIL, 2026a), ilustra a perigosa cooptação da tecnologia civil para cenários de conflito armado. O sistema, originariamente concebido para o processamento de texto, é convertido num vetor ativo de letalidade. É precisamente diante da urgência de se refletir sobre limites éticos a esse avanço irrestrito que a presente pesquisa se justifica, elegendo como objeto de análise o *European Union Artificial Intelligence Act (EU AI Act)*. Como marco regulatório pioneiro, o documento europeu sinaliza o esforço institucional mais robusto da atualidade para tentar resguardar os direitos fundamentais frente à opacidade algorítmica.

A delegação de comando de frotas de mar a vigilantes de algoritmo e a ameaça de sanções para os sistemas de linguagem que se recusam a operar em cenários de guerra evidenciam a conversão da técnica em vetor de letalidade. A outorga de deliberações éticas a aparatos dados instaura importantes implicações ao questionar os pilares da responsabilidade humana *lato sensu* e as próprias bases da justiça.

Ao exigir o uso irrestrito da ferramenta militar para "todos os fins legais", o Pentágono determinou a supressão coerciva das salvaguardas éticas da *Anthropic* — que proibiam expressamente a aplicação do seu sistema para a vigilância em massa de cidadãos e para o acionamento de armas totalmente autônomas (CNN, 2026a). Face à recusa do CEO da empresa em abdicar destas balizas morais em nome da preservação da vida, o governo o Pentágono retaliou, banindo o uso do sistema nas agências federais e classificando formalmente a empresa como um "risco à segurança nacional" na cadeia de suprimentos (CNN, 2026b). Este episódio é um exemplo do algoritmo de uso irresponsável e perigoso da inteligência artificial.

Importa ressaltar, contudo, que a inteligência artificial encerra em si um formidável potencial de promoção do bem-estar coletivo, distanciando-se, assim, de qualquer reducionismo que a confine exclusivamente à condição de ameaça. Um testemunho eloquente dessa vocação encontra-se nas recentes inovações aplicadas à medicina de precisão, a exemplo do desenvolvimento de modelos nacionais aptos a analisar neoplasias em tempo real durante intervenções oncológicas (CNN, 2026b). Nesse caso limítrofe, o algoritmo deixa de figurar como um agente de vigilância.

No mesmo aspecto positivo, há um projeto da mineradora canadiana Aclara Resources para uso da Inteligência Artificial (IA) como vetor de um neoextrativismo estratégico. O acordo visa o desenvolvimento e a aplicação de modelos de IA para otimizar o complexo processo de separação de terras raras pesadas no chamado "Projeto Carina", localizado no Brasil, cujas operações estão previstas para iniciar em 2028 (CNN, BRASIL, 2026c).

No contexto da extração e do refino de terras raras- processos historicamente estigmatizados pela geração severa de passivos ambientais, consumos massivos de água e emissão de rejeitos tóxicos-, a IA consubstancia uma ferramenta de otimização ecoeficiente sem precedentes. A modelagem preditiva e a aprendizagem de dados permitem calcular as reações químicas de separação mineral com uma precisão milimétrica, reduzindo drasticamente o desperdício de insumos, o gasto energético e a liberação de subprodutos nocivos no ecossistema local. Deste modo, o algoritmo pode atuar como um mecanismo de salvaguarda ecológica, racionalizando um processo industrial que, na sua forma analógica, é inerentemente destrutivo.

Face a esta escalada das IAs, marcos de regulação, com primazia para o pioneiro *EU AI Act (European Union Artificial Intelligence Act)*, procuram mitigar os riscos inerentes à fronteira de inovação. Todavia, a presente investigação assume a premissa de que a adoção de uma abordagem estritamente de conformidade jurídica revela insuficiência manifesta para a governança de um fenômeno que atinge o cerne da condição de sujeito. A governança por algoritmos, ao operar sob métricas unilaterais, pode perpetuar injustiças, gerando não apenas um implicações do ponto de vista da norma positiva, mas uma questão de ordem filosófica profunda.

Paul Ricoeur propõe que a justiça exige a articulação de uma "vida boa, com e para os outros, em instituições justas". Ao transpor esta máxima para o debate regulatório, compreende-se que a salvaguarda do indivíduo especifica-se pela valorização da identidade narrativa. Argumentando-se, por conseguinte, que a governança da inteligência artificial impõe o resgate do homem como centro de deliberação e reconhecimento por meio de mediações.

Como desdobramento dessa exigência por mediações frente à opacidade algorítmica, a abordagem metodológica desta pesquisa alicerça-se na semiótica, especificamente considerando o signo na abordagem ricoeuriana, o qual opera mediante uma dupla intencionalidade: um sentido literal e primário oculta e remete para um sentido secundário e profundo, exigindo um labor de decifração. Nesse caso, a semiótica instaura o polo da explicação marcada pelo distanciamento, onde o intérprete suspende transitoriamente o seu

engajamento existencial e urgências. Há uma suspensão do limite estrutural da lei que legitima a posterior assimilação hermenêutica, permitindo desvelar como o problema afeta, na materialidade da vida, a solicitude e a identidade narrativa do sujeito perante as instituições.

Nesse horizonte hermenêutico, a estruturação deste trabalho desdobra-se em dois movimentos complementares. O capítulo inaugural estabelece os alicerces teóricos ao perscrutar os horizontes da hermenêutica ricoeuriana em face dos dilemas éticos suscitados pelos sistemas de Inteligência Artificial. A exploração dos conceitos de "identidade narrativa", "solicitude" e "instituições justas" fornece a base para um estudo do regulamento da UE.

O segundo capítulo destaca a importância da interpretação na justiça algorítmica e evidencia os desafios para a edificação daquilo que Ricoeur cunhou instituições justas, projetando o referencial filosófico para examinar as disposições do *EU AI Ac*. Argumenta-se que, mesmo com a estipulação de deveres de transparência, o regulamento pode resvalar para injustiças. Nesse aspecto, a análise justifica-se pela ponderação entre a letra fria da lei e o elemento hermenêutico, perscrutando em que medida as disposições desse regulamento da UE se aproxima das instituições justas e de algumas outras instâncias de mediação propostas por Ricoeur.

Para empreender a transição da formulação do problema para a sua investigação material, o percurso metodológico deste trabalho apropria-se da “via longa” da mediação do texto. Conforme a premissa de que a compreensão de si e do mundo exige um “longo desvio” pelos signos com a constatação de que é preciso “explicar mais para compreender melhor” (Ricoeur, 1989, p. 165), a abordagem alicerça-se inicialmente na semiótica para instaurar o polo da explicação. Ao suspender transitoriamente o engajamento existencial por meio de uma *suspensão do juízo*, o primeiro capítulo tenta estabelecer o desvio estratégico necessário para examinar o caso concreto.

1. O HORIZONTE HERMENÊUTICO PROFUNDO EM SISTEMAS DE IA

A capilaridade alcançada pela inteligência artificial (IA) nas mais diversas esferas da vida tem suscitado dilemas ético que tensionam os marcos tradicionais da responsabilidade e da justiça. Nesse contexto, a vigilância algorítmica desponta como um desafio particular: ao empregar sistemas automatizados para monitorar, classificar e orientar decisões com base em vastos volumes de dados, essa prática sugere a consolidação de novas dinâmicas, frequentemente marcadas pela assimetria informacional e por uma considerável opacidade em sistemas de IAs.

A tendência de delegar etapas de processos decisórios a infraestruturas tecnológicas levanta questionamentos complexos. Torna-se pertinente indagar de que maneira é possível resguardar o sentido de justiça diante de sistemas que, por sua própria arquitetura matemática, tendem a reduzir a singularidade da experiência humana a padrões métricos.

Na tradição da fenomenologia hermenêutica, a qual Paul Ricoeur é expoente, a opção terminológica pelo vocábulo "fenômeno" em detrimento de "problema" possui profundo rigor metodológico. Um "problema" (de matriz epistemológica ou técnica) pressupõe um obstáculo exterior, calculável e objetificável, que aguarda uma solução definitiva ou um reparo de engenharia. Em contrapartida, o "fenômeno" designa aquilo que se manifesta e se dá à experiência vivida, reconfigurando o horizonte de sentido da ação.

A vigilância algorítmica não é tratada aqui como um erro de código a ser solucionado, mas como uma nova forma uma manifestação que pode afetar a ontologia do sujeito e sua identidade, exigindo, portanto, um esforço de interpretação contínua. Como estabelece Ricoeur ao fundamentar a subordinação da interpretação àquilo que se mostra (o aparecer), "a fenomenologia continua a ser a pressuposição intransponível da hermenêutica", de modo que todo o esforço hermenêutico é, antes de tudo, o desvelamento do sentido oculto num fenômeno que se apresenta (Ricoeur, 1989, p. 45).

Operando uma suspensão metodológica que trata a lei temporariamente como um "objeto fechado sobre si mesmo", o signo jurídico manifesta-se no seu sentido literal e primário, manifesta-se na sua acepção denotativa¹, implicando, por conseguinte, a ultrapassagem da literalidade do signo. Assim, impõe-se a "via longa" da mediação da norma

¹ A tese é a de que a nova pertinência da predicação recai sobre as ruínas da predicação literal. [...] É graças à suspensão da denotação da primeira ordem do discurso que é liberada uma denotação de segunda ordem, propriamente metafórica. (Ricoeur, 2000, p. 338). Ricoeur, Paul. *A Metáfora Viva*. Tradução de Macedo e Silva. São Paulo: Edições Loyola, 2000.

(textual), onde explicar e compreender são momentos complementares de um único processo hermenêutico."

Ao consumir essa suspensão, a norma transcende o seu sentido direto e alça-se ao estatuto de um símbolo de duplo sentido. Caracterizado por uma irreduzível "sobredeterminação de intenção", o símbolo constitui o ponto de partida: o seu sentido primário e literal aponta, inexoravelmente, para um sentido secundário e oculto que exige o labor da decifração — um mecanismo análogo àquele pelo qual Ricoeur analisa os símbolos na Fenomenologia do Mal (mancha, pecado, culpa) para desvelar experiências existenciais profundas.

É por intermédio desta sobredeterminação simbólica que a hermenêutica instaura a necessidade do desvio (a "via longa") através das mediações do texto para que o sujeito possa, em última instância, entender a si mesmo. Sob essa ótica, observa-se como imperativos legais — como o fomento à literacia em IA, a supervisão humana e o direito à explicação — desvencilham-se da sua prescrição literal e aparecem como mediações simbólicas.

Na culminância deste labor de decifração emerge a centralidade da identidade narrativa: a assimilação hermenêutica não restitui uma identidade substancial, mas revela como ela constitui no tempo através do ato de narrar. Ao transitar do signo, que pode estar mais próximo do homem com perfil preditivo e unívoco, para a camada do símbolo, o indivíduo se compreende diante da norma, avocando a sua *ipseidade*.

É no julgamento da situação singular que a *phronesis*, como exemplo de instâncias de mediadora se apresentam, movimentado o imperativo legal em direção ao caso concreto: narrativas e vulnerabilidades. Nessa assimilação metodológica, o reconhecimento mútuo atua, destarte, para salvaguardar devir — a capacidade de agir, sofrer e prometer — evitando a perda da "via longa": mediação da norma.

Especificamente, a metodologia proposta tem sua cúspide na proposta de leitura do *EU AI Act* como instituição justa, percorrida no itinerário da Seção 3.1 desta dissertação. Trata-se do estudo conjunto do Artigo 1.º, n.º 1, do Artigo 3.º, n.º 1, e do Artigo 4.º, avançando para o Artigo 5.º, n.º 1 (alíneas *a, b, c, f, g e h*) e Artigo 6.º, n.º 1. Integrado ainda pelas seguintes disposições: Artigo 10.º, n.ºs 3 e 5; Artigo 14.º, n.º 1; Artigo 15.º, n.º 1; Artigo 27.º, n.º 1; Artigo 50.º, n.ºs 1 e 2; Artigo 55.º, n.º 1 (contextualizado à luz do Artigo 51.º); Artigo 57.º, n.º 1; e, por fim, o Artigo 86.º, n.º 1².

² A formatação desta referência segue a técnica das regras de redação (legística) oficiais da União Europeia, cujos documentos são traduzidos para a língua portuguesa adotando a norma de Portugal (PT-PT). O uso

A expressão “justiça algorítmica” foi cunhada principalmente por autores como Virginia Eubanks (que estuda como algoritmos punem os mais pobres) e Joy Buolamwini (fundadora da *Algorithmic Justice League*). A justiça algorítmica infere que, sem uma intervenção ética deliberada, as máquinas limitam-se a automatizar a exclusão e a desigualdade estrutural a uma velocidade e escala sem precedentes. Explicando a necessidade de abandonar a neutralidade tecnológica e de assumir uma responsabilidade ativa sobre o código, Buolamwini adverte:

A IA não resolverá a discriminação, porque os padrões culturais que dizem que um grupo de pessoas é melhor do que outro devido ao seu gênero, cor da pele, forma de falar, altura ou riqueza não são problemas técnicos. [...] O simples facto de as decisões serem tomadas por um computador a analisar dados não as torna neutras. O termo 'neural' não equivale a 'neutro'. Por mais tentador que seja, não podemos usar a IA para contornar o trabalho árduo de organizar a sociedade de modo a que o local onde se nasce, os recursos da sua comunidade e os rótulos que lhe são impostos não sejam os principais determinantes do seu destino (Buolamwini, 2023, p. xix, tradução nossa³).

Didaticamente, a filosofia oferece contornos prévios para os dilemas suscitados pelo avanço tecnológico. O paradigma dualista e mecanicista cartesiano, por exemplo, ao aproximar a compreensão do corpo humano ao funcionamento de uma máquina — concebendo-o de forma análoga às engrenagens de um autômato ou de um relógio (Trindade; Miranda, 2021, p. 29-30) —, ilustra preteritamente como o advento das tecnologias, ainda que incipientes, acarretam inúmeras implicações.

Visando lançar luz sobre o fenômeno proposto, a hermenêutica de Paul Ricoeur erige-se como uma filosofia da compreensão, sendo nesta pesquisa uma alternativa à dicotomias radicais e estruturalismos absolutos. Ricoeur reposiciona a interpretação não como uma decodificação instrumental da linguagem, mas como a mediação incontornável entre o sujeito e o mundo, e entre o texto e a ação, circunscrita por símbolos, narrativas e instituições.

A hermenêutica do si distingue-se, tanto da apologia do Cogito, quanto da sua destituição. [...] Dizer si não é dizer eu. O eu põe-se, ou é deposto. O si implica-se a título reflexivo nas operações cuja análise precede o retorno a si mesmo. [...] Não há auto-conhecimento imediato, mas somente mediado por signos, por símbolos e por textos. [...] O inquérito hermenêutico pareceu-nos, então, dever desdobrar-se entre

contínuo do ordinal (art. 86.º): Enquanto a redação legal no Brasil determina que, a partir do artigo 10, se utilize o número cardinal (art. 86), a redação europeia mantém o indicador de número ordinal para todos os artigos de um diploma (Artigo 86.º). A subdivisão em "Números" (n.º 1): No Brasil, a subdivisão primária de um artigo é o parágrafo (ex.: § 1º). No ordenamento jurídico europeu e em Portugal, os artigos dividem-se em "números" (abreviados como n.º). O que em terras brasileiras seria um parágrafo primeiro, na União Europeia é o número um. UNIÃO EUROPEIA. Serviço das Publicações da União Europeia. **Código de Redação Interinstitucional**. Luxemburgo: Serviço das Publicações da União Europeia, 2024. Disponível em: <https://publications.europa.eu/code/pt/pt-000100.htm>. Acesso em: 7 mar. 2026.

³ BUOLAMWINI, Joy. **Unmasking AI: My Mission to Protect What Is Human in a World of Machines**. Nova York: Random House, 2023.

os dois polos extremos que constituem a exaltação do *Cogito*, por um lado, e a sua humilhação, por outro (Ricoeur, 1991, p. 25-26).

A hermenêutica ricoeuriana transcende a exegese textual para se consolidar como a própria via de constituição. Ao desconstruir a ilusão de transparência do *Cogito* cartesiano — a premissa de um acesso direto e imediato à própria consciência —, o filósofo postula que a compreensão de si exige o "longo desvio" pelos signos e obras da cultura. O nexos com a ética é estrutural: é somente ao interpretar a sua própria narrativa e reconhecer a sua permanência no tempo (ipseidade) que o sujeito alcança o "reconhecimento".

Conforme analisa Nobrega (2012), a perspectiva ética de Paul Ricoeur consolida-se através de uma constituição tríplice e indissociável, a saber: a ipseidade, a alteridade e a igualdade. A ipseidade traduz-se na estima de si e no desejo por uma existência realizada; a alteridade insere a figura do "outro" através da solicitude; e a igualdade eleva esse reconhecimento ao plano político e impessoal.

Esta tríade é lapidarmente sintetizada pelo filósofo francês ao definir o núcleo da sua intencionalidade ética como a "[...] visada da 'vida boa' com e para os outros nas instituições justas" (Ricoeur, 1991, p. 202). É precisamente o terceiro elemento desta estrutura — a igualdade — que fundamenta o espaço institucional, permitindo que o cuidado transcenda a intimidade do encontro face a face. Sobre esta transição necessária, o próprio Ricoeur (1991, p. 242) assevera em: "o sentido da justiça não tira nada da solicitude; a igualdade lhe fornece apenas a amplitude da instituição". Desse modo, a igualdade erige-se como o alicerce das instituições justas, assegurando que o reconhecimento mútuo alcance a totalidade da vida.

Para que esse reconhecimento institucional se efetive na prática interpretativa da norma, a semiótica não consubstancia um fim em si mesma, mas um momento metodológico inalienável no interior do seu arco hermenêutico. Distanciando-se de um estruturalismo dogmático — que encarcera a linguagem na sua própria imanência —, o filósofo apropria-se da análise estrutural como a etapa necessária de explicação objetiva do texto, precedendo a apropriação existencial do sentido. Essa mediação dialética entre a objetividade do signo e a subjetividade do intérprete é o que afiança o rigor da exegese, esconjurando o risco de um romantismo arbitrário. O escopo primordial desta integração é condensado na célebre premissa ricoeuriana de que é preciso "explicar mais para compreender melhor" (Ricoeur, 1989, p. 165). Evidencia-se, destarte, que a dissecação interna das estruturas constitui o caminho mais seguro para aceder à verdade ontológica da obra (norma).

Como desdobramento da premissa ricoeuriana de que é preciso 'explicar mais para compreender melhor' (Ricoeur, 1989, p. 165), faz-se estritamente necessário perscrutar a

materialidade da técnica que permeia a sociedade contemporânea. Uma vez que o filósofo adverte que “não há compreensão de si que não seja mediatizada por signos, símbolos e textos” (Ricoeur, 1989, p. 29), compreende-se que o algoritmo não atua como um mero instrumento de cálculo neutro, mas opera calcado em uma rigorosa e excludente estrutura semiótica. Essa linguagem de programação, balizada por uma sintaxe fechada de dados de entrada e saída, instaura um processo de tradução ontológica: ela converte a dinamicidade da vivência humana (a *ipseidade*) em perfis estatísticos engessados (a *mesmidade*). Ao realizar esse desvio exigido pela explicação do código, constata-se que a opacidade intrínseca a esses modelos — frequentemente operados sob a lógica da “caixa-preta” (Pasquale, 2015) — atua como um verdadeiro dispositivo de ordem de símbolo. A máquina mimetiza a intencionalidade ao arranjar signos de forma veloz e preditiva, mas o faz sem partilhar de uma semântica vivencial, desenhando, de forma silenciada, a possibilidade de novas e profundas assimetrias de poder.

No aspecto semiótico, a referência ao mundo e a própria intenção do autor são metodologicamente colocadas em suspenso, operando-se uma espécie de *epoché* (suspensão) hermenêutica. O texto (ou o diploma normativo) é tratado, transitoriamente, como um sistema autônomo, cujas relações lógicas de oposição e hierarquia devem ser mapeadas com exatidão. O autor justifica a premência de suportar esta aridez objetiva ao afirmar que “a análise estrutural põe entre parênteses a referência do texto ao mundo, bem como a sua referência ao autor e ao leitor; o texto é tratado como um objeto fechado sobre si mesmo” (Ricoeur, 1989, p. 143). Tal suspensão provisória não traduz uma negação da realidade prática, mas um distanciamento:

Explicar é extrair a estrutura, isto é, as relações internas de dependência que constituem a estática do texto; compreender é captar a proposição de mundo aberta pela referência do texto. Explicar e compreender não constituem, por conseguinte, dois métodos opostos [...], mas dois momentos de um único e mesmo processo hermenêutico. A análise estrutural constitui uma etapa – e uma etapa necessária – entre uma interpretação ingênua e uma interpretação crítica, entre uma interpretação de superfície e uma interpretação em profundidade. (Ricoeur, 1989, p. 161-170).

A estrutura, uma vez explicada objetivamente, erige-se como alicerce para a apropriação do “mundo do texto”, facultando ao sujeito o alargamento da sua própria identidade narrativa face àquilo que interpretou. Longe de configurarem instâncias antagônicas, a semiótica e a hermenêutica constituem polos complementares de uma mesma vontade de verdade, consolidando a premissa de que “a explicação e a compreensão não constituem dois métodos opostos [...], mas dois momentos de um único e mesmo processo hermenêutico” (Ricoeur, 1989, p. 165).

O filósofo propõe que a deliberação justa — o exercício da sabedoria prática (*phronesis*) nas instituições — não pode derivar de uma intuição moral imediata, romântica ou arbitrária; ela exige o "distanciamento" objetivo proporcionado pela análise estrutural das normas, das ações e dos sistemas técnicos, tratados rigorosamente como "textos". É a fase de explicação semiótica que desvela as engrenagens internas de poder, categorização e controle da técnica, para que, num momento dialético posterior, o sujeito possa formular um juízo ético fundamentado.

Este imperativo do "rodeio" analítico é o que ancora a ética na realidade material das estruturas, consolidando a célebre premissa de que "não há compreensão de si que não seja mediatizada por signos, símbolos e textos; a compreensão de si coincide, em última análise, com a interpretação aplicada a estes termos mediadores" (Ricoeur, 1989, p. 29). Por conseguinte, a semiótica atua como a salvaguarda basilar da justiça e da alteridade: ao dissecar objetivamente a arquitetura de uma lei ou de uma instituição, ela fornece à hermenêutica o substrato factual necessário para obstar o determinismo e garantir que a proteção do humano não seja uma abstração vazia, mas uma práxis concretizada no mundo.

Essa é uma das razões pelas quais a filosofia ricoeuriana se torna relevante para refletir sobre tecnologias baseadas em dados e algoritmos, pois tais sistemas tendem a excluir a dimensão interpretativa. Ademais, a proposta ricoeuriana é de uma via média que evita tanto a exaltação do sujeito cartesiano quanto a sua anulação, conforme se depreende da distinção ricoeuriana:

A hermenêutica do si distingue-se, tanto da apologia do *Cogito*, quanto da sua destituição. [...] Dizer si não é dizer eu. O eu põe-se, ou é deposto. O si implica-se a título reflexivo nas operações cuja análise precede o retorno a si mesmo. [...] O inquérito hermenêutico pareceu-nos, então, dever desdobrar-se entre os dois polos extremos que constituem a exaltação do *Cogito*, por um lado, e a sua humilhação, por outro (Ricoeur, 1991, p. 25-26).

Compete esclarecer ainda uma distinção terminológica frequentemente objeto de confusão: o autor não utiliza a expressão "Grande Ética" como uma categoria formal, tampouco a estabelece como um conceito simétrico ou diretamente oposto à "Pequena Ética" no interior de sua obra. Na realidade, a designação "Pequena Ética" (*petite éthique*) foi atribuída pelo próprio filósofo, com notável modéstia intelectual, para intitular especificamente os três capítulos centrais — VII, VIII e IX — de sua *magnum opus*, *O Si-mesmo como um Outro* (*Soi-même comme un autre*, publicada em 1990).

Logo, enquanto a "Pequena Ética" refere-se a essa seção delimitada e rigorosa que articula a teleologia da vida boa com a deontologia da norma, a ideia de uma "Grande Ética"

não constitui um termo técnico cunhado por Ricoeur para designar uma contraparte estrutural na mesma obra, mas sim uma ausência terminológica que reforça o caráter específico e focado da sua proposta ética nesses capítulos. Ricoeur olha retrospectivamente para sua obra anterior e utiliza explicitamente o termo "pequena ética" para se referir àqueles capítulos, confirmando a nomenclatura:

"De uma maneira geral, os estudos aqui reunidos constituem um complemento e um corretivo à 'pequena ética' de *O Si-mesmo como um Outro*. [...] A 'pequena ética' detinha-se no limiar do político e da justiça, na medida em que esta última é o horizonte das instituições." (Ricoeur, 2008, p. XIII⁴).

No centro da ética ricoeuriana está a noção de uma vida boa (*"le vivre bien"*), com e para os outros, em instituições justas. Esse tripé — visada ética, solicitude e justiça institucional — articula uma compreensão profunda da relação entre o eu e o outro, e entre ambos e o espaço normativo compartilhado. A identidade pessoal, para Ricoeur, não é algo fixo, mas se constrói narrativamente ao longo do tempo, em interação com os outros e com as instituições que moldam a vida em sociedade. Quando essa dinâmica é substituída por decisões algorítmicas que ignoram o contexto, as histórias e a singularidade dos sujeitos, a própria possibilidade de justiça é ameaçada. Essa compreensão liberta o sujeito da rigidez, permitindo uma identidade construída na história, como elucida o autor:

A identidade narrativa não é uma identidade estável e sem falhas; tal como a da identidade da personagem romanesca, ela se faz e se desfaz. [...] A própria noção de sujeito se vê, assim, libertada de todo o substancialismo, sem por isso perder todo o sentido, pois, se é verdade que a identidade narrativa não é a de uma substância imutável, ela não é, tampouco, a de uma ilusão passional (Ricoeur, 1997, p. 430).

Nesse horizonte, a "visada ética" ricoeuriana, trabalhada como anseio pela vida boa, com e para os outros, em instituições justas, enfrenta um desafio singular diante de questões como a IA. Se a ética da solicitude exige o reconhecimento da alteridade como uma dimensão constitutiva do "si", a lógica computacional tende a caminhar para uma redução, podendo converter o sujeito de direitos em um mero usuário sem identidade. Esse cenário ameaça esvaziar o espaço moral onde o reconhecimento mútuo deveria ocorrer.

A "visada ética" não se fragmenta, mas desdobra-se em três momentos: inicia-se na estima de si, onde o sujeito reflexivamente valida sua capacidade de agir e seu desejo de realização ("vida boa"); expande-se para a solicitude, pois a autoconstituição do si depende do reconhecimento recíproco na relação face a face ("com e para os outros"); e culmina na

⁴ O "XIII" em romanos: Indica que o trecho foi retirado de um elemento pré-textual (como um Prefácio, uma Introdução ou uma Apresentação). É prática comum nas editoras (como a Martins Fontes, que publicou *O Justo*) numerar as páginas introdutórias em romanos (I, II, III...) e iniciar a numeração em arábicos (1, 2, 3...) apenas quando começa o Capítulo 1.

justiça, que é a extensão dessa solicitude para a esfera do "cada um", exigindo a mediação de "instituições justas" para permitir a partilha de direitos e deveres. Assim, a justiça não é uma virtude fria, mas o horizonte que assegura a proteção da dignidade humana mesmo nas relações mediadas.

A tensão conceitual entre a identidade-*idem* (*mesmidade*) e a identidade-*ipse* (*ipseidade*) oferece uma chave hermenêutica indispensável para compreender os riscos éticos inerentes à vigilância algorítmica. Enquanto a identidade narrativa permite ao sujeito manter-se no tempo através da promessa, da ressignificação e da constante reinvenção da sua própria história (o polo *ipse*), os sistemas de Inteligência Artificial podem operar em uma lógica de cristalização do indivíduo, suprimindo a abertura e a imprevisibilidade fundamentais da condição humana. Algumas premissas de podem se relacionar a essa questão:

Distinguimos assim duas aceções maiores da identidade: a identidade como mesmidade (em latim *idem* [...]) e a identidade como ipseidade (em latim *ipse* [...]). [...] À identidade como mesmidade (*idem*), oponho a identidade como ipseidade (*ipse*). [...] A diferença entre *idem* e *ipse* não é senão a diferença entre uma permanência no tempo não implicada pela ipseidade e uma permanência no tempo que a implica. O caráter ilustra perfeitamente o primeiro caso; a manutenção de si na promessa, o segundo. (Ricoeur, 1991, p. 17; 143).

Para conferir sustentação teórica a essa aspiração ética projetada pelo regulamento europeu, as "instituições justas" assumem um papel preponderante diante do indivíduo vulnerável e das novidades humanas. É Fundamental que a justiça algorítmica, enquanto instituição justa, trabalhe com a sabedoria prática (*phronesis*), sendo capaz de interpretar a norma geral à luz da singularidade de cada caso. Sem essa hermenêutica, que insere a distância justa entre os dados e a identidade narrativa, o *EU AI Act* e outros marcos regulatórios correm o risco de se tornarem instrumentos de violação à dignidade da pessoa humana enquanto fim em si mesma. Esse entendimento se aproxima da formulação presente em *O Si-mesmo como um Outro*, na qual o autor cristaliza o conceito de visada ética:

Chamamos "visada ética" a visada da "vida boa", com e para os outros, nas instituições justas. [...] A estima de si [é] o momento reflexivo da visada da "vida boa"; a solicitude, a sua dimensão dialogal; a justiça, enfim, a instância do "cada um" (Ricoeur, 1991, p. 202; 227).

Os sistemas de IA estão suscetíveis a desafiar a atribuição de responsabilidade moral, pois diluem as cadeias de imputação entre programadores, empresas, usuários e os próprios algoritmos. Essa conjuntura merece atenção, pois uma lógica técnica que opera sem deliberação pública e sem reconhecimento mútuo, mina o espaço comum de justiça e participação.

Ademais, O *EU AI Act*, em alguns aspectos, tenta materializar a exigência ético-institucional em seu próprio arcabouço normativo. Ao positivar o dever de transparência, a necessidade de supervisão humana contínua e a avaliação de impactos nos direitos fundamentais, remotamente, caminha para a possibilidade de "instituição justa" ricoeuriana — uma estrutura concretamente enraizada na vida humana, capaz de mediar e interpretar a vulnerabilidade do sujeito e possibilitar que o avanço tecnológico seja responsável.

Não obstante, a identidade narrativa do sujeito pode ser corroída quando algoritmos produzem classificações e decisões sem dar voz ao indivíduo, negando-lhe a possibilidade de se compreender a si mesmo em sua história. Por outro lado, acerca da capacidade de organizar eventos dispersos em uma unidade de sentido, Ricoeur elucida o papel da configuração narrativa:

A operação de configuração consiste na 'construção da trama' (muthos). [...] O que é narrado é a intriga, quer dizer, o agenciamento dos fatos. Ora, o que é o agenciamento dos fatos senão a síntese do heterogêneo? A intriga é a imitação da ação porque ela articula, num todo inteligível, circunstâncias, agentes, interações, fins, meios, consequências não-intencionais (Ricoeur, 1994, p. 86).

Assim, a hermenêutica de Ricoeur fundamenta um horizonte no qual a justiça só se concretiza por meio da mediação. A presente dissertação mobiliza essa matriz filosófica para o escrutínio do problema vigilância algorítmica.

1.1. *European Union Artificial Intelligence Act (EU AI Act)*

O panorama regulatório global da tecnologia atravessa uma transformação paradigmática com o *European Union Artificial Intelligence Act (Eu AI Act)*, que se consolida como um arcabouço jurídico pioneiro voltado ao desenvolvimento e à utilização da Inteligência Artificial. Nesse escopo, a União Europeia instituiu normas com o fito de conciliar o fomento à inovação tecnológica com a proteção inegociável da saúde, da segurança e dos direitos fundamentais em toda a cadeia de valor.

Adotou-se a versão oficial do *EU AI Act* em português de Portugal devido à simetria luso-brasileira. A filiação de ambos os países à tradição do *Civil Law* permite uma certa compatibilidade do vocabulário na formulação das normas, evitando distorções na transposição do texto europeu nesta pesquisa.⁵

⁵ “(...) Seus conceitos não desaparecem: transformam-se e reaparecem conforme as novas necessidades sociais, mas o arcabouço romanístico é a linguagem silenciosa que organiza o raciocínio jurídico brasileiro. A permanência de uma tradição luso-brasileira e romano-germânica atua como o fio condutor que permite a correta tradução e a recepção de novos institutos sem a perda de sua identidade sistemática (JOBIM, 2020, p.

Trata-se, porquanto, de um instrumento normativo pioneiro sobre inteligências artificiais. Para que essa premissa se materialize e seja demarcada a fronteira entre o avanço técnico e a salvaguarda humana, o legislador europeu delineou o objeto central da norma:

O presente regulamento estabelece regras harmonizadas aplicáveis à colocação no mercado, à colocação em serviço e à utilização de sistemas de inteligência artificial («sistemas de IA») na União; [...] A finalidade do presente regulamento é melhorar o funcionamento do mercado interno e promover a adoção de uma inteligência artificial (IA) centrada no ser humano e de confiança, assegurando simultaneamente um nível elevado de proteção da saúde, da segurança e dos direitos fundamentais [...] (UNIÃO EUROPEIA, 2024, art. 1º, n. 1).

Para compreender a extensão deste regulamento, é imperativo analisar a definição técnica adotada pelo legislador europeu para delimitar seu objeto. O Artigo 3º define um "sistema de IA" como um sistema baseado em máquina concebido para funcionar com níveis de autonomia variáveis e que pode apresentar adaptabilidade após a implantação, inferindo, a partir dos dados de entrada que recebe, a forma de gerar resultados como previsões, conteúdos, recomendações ou decisões. Uma análise detalhada deste tópico demonstra que a definição busca abarcar tanto a autonomia operacional quanto a capacidade inferencial dos algoritmos, distinguindo-os de softwares de programação tradicional. O texto abarcar a complexidade operacional desses sistemas:

Para efeitos do presente regulamento, entende-se por: «Sistema de IA», um sistema baseado em máquina concebido para funcionar com níveis de autonomia variáveis e que pode apresentar capacidade de adaptação após a implantação e que, para objetivos explícitos ou implícitos, infere, a partir dos dados de entrada que recebe, a forma de gerar resultados como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais (UNIÃO EUROPEIA, 2024, art. 3º, n. 1).

Nessa direção, é possível entender que o conceito de autonomia e a interação com o ambiente são centrais para a aplicação da norma. O texto legal especifica que esses sistemas podem influenciar ambientes físicos ou virtuais, operando com objetivos explícitos ou implícitos definidos pelos humanos, mas executados através de processos automatizados de inferência. Isso implica que a regulação não se limita ao código estático, mas alcança o comportamento dinâmico e evolutivo da inteligência artificial em contextos reais de aplicação. A eficácia da norma depende diretamente da definição de seus destinatários e de sua jurisdição. Nessa direção, o regulamento adota uma postura extraterritorial, alcançando agentes fora da Europa cujos sistemas impactem o território da União:

97). JOBIM, Eduardo. A influência do direito romano na formação do direito brasileiro e a preservação da tradição luso-brasileira. In: **Revista de Ciências Jurídicas e Sociais**, v. 1, n. 1, p. 90-105, 2020.

O presente regulamento aplica-se: a) Aos fornecedores que coloquem no mercado ou coloquem em serviço sistemas de IA ou coloquem no mercado modelos de IA de finalidade geral na União, independentemente de esses fornecedores estarem estabelecidos ou localizados na União ou num país terceiro; [...] c) Aos fornecedores e responsáveis pela implantação de sistemas de IA que tenham o seu local de estabelecimento ou que estejam localizados num país terceiro, sempre que os resultados produzidos pelo sistema de IA sejam utilizados na União (UNIÃO EUROPEIA, 2024, art. 2º, n. 1).

No topo dessa pirâmide axiológica, encontram-se as práticas de inteligência artificial consideradas de risco inaceitável. O regulamento proíbe terminantemente sistemas que utilizem técnicas subliminares, explorem vulnerabilidades de grupos específicos ou realizem a pontuação social (*social scoring*) de cidadãos. É essencial entender que tal proibição reflete um compromisso ético inegociável com a dignidade humana, impedindo a instrumentalização da tecnologia para fins de manipulação ou segregação.

No topo dessa pirâmide, encontram-se as práticas de inteligência artificial consideradas de risco inaceitável. O regulamento proíbe terminantemente sistemas que utilizem técnicas subliminares, explorem vulnerabilidades de grupos específicos ou realizem a pontuação social (*social scoring*) de cidadãos. É essencial entender que tal proibição reflete um compromisso ético inegociável com a dignidade humana, impedindo a instrumentalização da tecnologia para fins de manipulação ou segregação. Abaixo do nível de proibição, situam-se os sistemas de alto risco, que constituem o foco principal das obrigações de conformidade. O regulamento estabelece critérios precisos para essa classificação:

Independentemente de o sistema de IA ser colocado no mercado ou colocado em serviço independentemente dos produtos [...] considera-se que um sistema de IA é de risco elevado se: a) O sistema de IA se destinar a ser utilizado como componente de segurança de um produto, ou for, ele próprio, um produto, abrangido pelos atos legislativos da União enumerados no anexo I [...] (UNIÃO EUROPEIA, 2024, art. 6º, n. 1).

Além dos riscos de segurança tratados no art. 6º, a legislação aborda os riscos específicos de transparência para sistemas que interagem diretamente com pessoas naturais. Nesse caso, o artigo 50 impõe deveres de transparência para provedores de sistemas de IA destinados a interagir com pessoas (como *chatbots*) ou que geram conteúdo sintético (como *deepfakes*), exigindo que os usuários sejam informados de que estão interagindo com uma máquina ou que o conteúdo foi gerado artificialmente. Uma questão central a ser abordada é que a transparência não é apenas um requisito técnico, mas uma condição para a preservação da autonomia do sujeito e da verdade na esfera pública. Para sistemas que interagem diretamente com humanos, como *chatbots*, a norma impõe o dever de revelação da natureza artificial do interlocutor. O dispositivo legal determina:

Os fornecedores de sistemas de IA [...] destinados a interagir diretamente com pessoas singulares devem conceber e desenvolver tais sistemas de tal modo que as pessoas singulares em causa sejam informadas de que estão a interagir com um sistema de IA, a menos que tal resulte evidente das circunstâncias e do contexto de utilização (UNIÃO EUROPEIA, 2024, art. 50, n. 1).

O regulamento também inova ao tratar dos modelos de IA de finalidade geral (*General Purpose AI*), refletindo a evolução recente dos modelos fundacionais e generativos. O texto define esses modelos como aqueles competentes para realizar uma vasta gama de tarefas distintas e estabelece regras específicas para aqueles que apresentam riscos sistêmicos, exigindo avaliações de modelo, testes adversariais e relatórios de incidentes graves à Comissão.

Uma discussão mais aprofundada sobre este tema revela a preocupação do legislador com a cadeia de valor da IA, responsabilizando não apenas quem aplica a tecnologia na ponta, mas também quem desenvolve os poderosos modelos que lhe servem de base. Um olhar mais cuidadoso sobre este aspecto revela a preocupação com os riscos sistêmicos que esses modelos fundacionais podem propagar por toda a cadeia de valor. O texto legal inova ao criar uma categoria específica para modelos potentes que servem de base para diversas aplicações. A definição legal é a seguinte:

«Modelo de IA de finalidade geral»: um modelo de IA [...] que apresenta uma generalidade significativa e é capaz de executar de forma competente uma vasta gama de tarefas distintas, independentemente da forma como o modelo é colocado no mercado, e que pode ser integrado numa variedade de sistemas ou aplicações a jusante (UNIÃO EUROPEIA, 2024, art. 3º, n. 63).

O objetivo declarado do regulamento é assegurar um elevado nível de proteção da saúde, da segurança e dos direitos fundamentais consagrados na Carta da União Europeia, inclusive contra os efeitos nocivos dos sistemas de IA, apoiando simultaneamente a inovação. Por fim, para certos usos de alto risco, exige-se uma avaliação prévia do impacto nos direitos dos cidadãos. A proteção de direitos não é apenas um princípio abstrato, mas um requisito procedimental obrigatório para a implementação da tecnologia. O regulamento determina que:

Antes da colocação em serviço de um sistema de IA de risco elevado referido no artigo 6.o, n.o 2, os responsáveis pela implantação que sejam organismos de direito público ou operadores privados [...] devem realizar uma avaliação do impacto nos direitos fundamentais que a utilização desse sistema poderá produzir (UNIÃO EUROPEIA, 2024, art. 27º, n. 1).

Para além de delinear as regras de mercado e definir tecnicamente o que constitui um sistema de inteligência artificial, o legislador europeu tentou estabelecer um limite ao identificar as práticas que representam um risco inaceitável. Ao proibir categoricamente certas aplicações, o regulamento visou proteger a autonomia, a dignidade e a integridade psicológica dos indivíduos contra usos abusivos da tecnologia. Observe-se o rigor com que o dispositivo trata as técnicas de manipulação:

Estão proibidas as seguintes práticas de IA: a) A colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que utilize técnicas subliminares que atuem abaixo do limiar de percepção de uma pessoa, ou técnicas propositadamente manipuladoras ou enganadoras, com o objetivo ou o efeito de distorcer substancialmente o comportamento de uma pessoa [...] (UNIÃO EUROPEIA, 2024, art. 5º, n. 1).

No tocante aos sistemas classificados como de risco elevado, o AI Act consagra o princípio da centralidade do ser humano, instituindo a necessidade inafastável de que o julgamento humano atue como um freio à automação decisória em contextos críticos. O texto normativo determina expressamente essa situação:

Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de modo a poderem, nomeadamente por meio de ferramentas de interface homem-máquina apropriadas, ser eficazmente supervisionados por pessoas singulares durante o período em que estão em utilização (UNIÃO EUROPEIA, 2024, art. 14º, n. 1).

Há um impacto sistêmico que essas tecnologias possuem ao servirem de base para inúmeras outras aplicações a jusante. A definição legal introduzida para abarcar essa complexidade é a seguinte:

«Modelo de IA de finalidade geral», um modelo de IA, inclusive se for treinado com uma grande quantidade de dados utilizando a autossupervisão em escala, que apresenta uma generalidade significativa e é capaz de executar de forma competente uma vasta gama de tarefas distintas, independentemente da forma como o modelo é colocado no mercado [...] (UNIÃO EUROPEIA, 2024, art. 3º, n. 63).

O AI Act estabelece o dever de promover o conhecimento e a compreensão crítica sobre o funcionamento dessas tecnologias, assegurando que aqueles que operam os sistemas possuam as ferramentas intelectuais necessárias para lidar com a automação. Dispõe a lei sobre o tema:

Os fornecedores e os responsáveis pela implantação de sistemas de IA tomam medidas para assegurar, na medida do possível, um nível suficiente de literacia no domínio da IA do seu pessoal e de outras pessoas que intervenham no funcionamento e na utilização de sistemas de IA em seu nome [...] (UNIÃO EUROPEIA, 2024, art. 4º).

O regulamento europeu, em algum aspecto, enfrenta o problema da opacidade algorítmica ao instituir um dever de transparência imediata nas interações diretas entre humanos e máquinas. A norma visa impedir que a inteligência artificial se faça passar por um sujeito moral, garantindo que o indivíduo saiba quando a sua contraparte é um autômato. O legislador europeu estabelece o dever de informação:

Os fornecedores devem assegurar que os sistemas de IA destinados a interagir diretamente com pessoas singulares sejam concebidos e desenvolvidos de modo a que estas sejam informadas de que estão a interagir com um sistema de IA, a menos que tal seja óbvio do ponto de vista de uma pessoa singular, razoavelmente bem informada, atenta e perspicaz, dadas as circunstâncias e o contexto de utilização [...] (UNIÃO EUROPEIA, 2024, art. 50º, n. 1).

A opacidade algorítmica, frequentemente designada na literatura especializada como o fenómeno da "caixa negra" (*black box*), não se restringe apenas à incompreensibilidade técnica da arquitetura do código ou à impossibilidade de auditar as múltiplas camadas de redes neuronais. Na sua dimensão mais perniciosa para o tecido social, a opacidade manifesta-se de forma relacional, ocorrendo sempre que o sujeito interage com um sistema de Inteligência Artificial sem ter a percepção consciente de que o faz. Ao mimetizar a intencionalidade, a linguagem e, até mesmo, a aparência humana (como nos casos de *deepfakes* ou *chatbots* avançados), a máquina instaura uma assimetria radical de poder e informação que pode corromper a autenticidade das relações no espaço público. É precisamente esta dissimulação que o regulamento europeu procura combater. Nesse caso:

O termo «caixa negra» refere-se a um sistema cujos mecanismos internos são misteriosos; podemos observar as suas entradas [inputs] e saídas [outputs], mas não conseguimos compreender como uma se transforma na outra. [...] Essa assimetria de conhecimento traduz-se numa assimetria de poder. As entidades corporativas e estatais que operam estas caixas negras sabem tudo sobre nós, enquanto nós sabemos muito pouco sobre elas e sobre a forma como utilizam os nossos dados para classificar, avaliar e decidir os nossos destinos. (PASQUALE, 2015, p. 3-4, tradução nossa).

Ao instituir um "dever de transparência imediata", o EU AI Act não exige que o cidadão comum compreenda a matemática por detrás do modelo, mas impõe uma obrigação jurídica de sinalização prévia: o indivíduo tem o direito fundamental de saber, em tempo real, que a sua contraparte não é humana. Trata-se de uma transparência interacional que visa proteger a integridade da confiança social. Sob a perspetiva da ética ricoeuriana, esta exigência legal atua como uma barreira institucional contra a falsificação da alteridade. Ao obrigar a máquina a revelar a sua condição artificial, o direito impede que o automatismo simule a "solicitude" e assegura que o reconhecimento mútuo — base indispensável da

justiça e da vida boa — permaneça uma categoria exclusiva da intersubjetividade entre seres humanos vulneráveis e capazes.

A falibilidade dos sistemas automatizados impõe riscos diretos à integridade dos sujeitos, especialmente em setores críticos como o judiciário e a segurança pública. Por essa razão, o *EU AI Act* determinou que a inteligência artificial deve ser blindada contra inconsistências operacionais e ataques externos que possam desviar a ferramenta de sua finalidade. Essa proteção da "fragilidade" humana diante de falhas técnicas é traduzida no texto normativo:

Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de modo a alcançarem, à luz da sua finalidade prevista, um nível adequado de exatidão, robustez e cibersegurança e a manterem a regularidade do seu funcionamento nesses aspetos ao longo de todo o seu ciclo de vida (UNIÃO EUROPEIA, 2024, art. 15º, n. 1).

Para que haja responsabilização jurídica pelas decisões automatizadas, é imprescindível que a "caixa-preta" dos algoritmos deixe rastros auditáveis. Com o propósito de garantir o escrutínio e o controle humano sobre eventos e falhas pretéritas, a lei exige a documentação automática das operações executadas pela máquina:

Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos com capacidades para a conservação automática de registos de eventos (registos) ao longo do seu tempo de vida. Estas capacidades de registo devem estar em conformidade com o estado da arte e com as normas aplicáveis reconhecidas e ser proporcionadas aos riscos associados ao sistema [...] (UNIÃO EUROPEIA, 2024, art. 12º, n. 1).

É possível arriscar dizer que o escopo derradeiro da norma é garantir que a inovação permaneça na sua condição estrita de ferramenta, operando sempre sob a égide do escrutínio humano. A premissa inegociável, que subordina o avanço algorítmico à preservação incondicional das garantias civis e constitucionais, encontra-se lapidarmente estabelecida nos princípios fundamentais que justificam o regulamento:

A IA deverá ser uma tecnologia centrada no ser humano. Deverá servir de instrumento para as pessoas, com o objetivo último de aumentar o bem-estar das pessoas, e deverá ser concebida e desenvolvida de modo a que o seu impacto na sociedade seja benéfico. Por conseguinte, há que prever normas essenciais de proteção a fim de garantir que os sistemas de IA não sejam utilizados de uma forma que prejudique a dignidade humana e os direitos fundamentais. (UNIÃO EUROPEIA, 2024, Considerando 6º).

⁶ A IA deverá ser uma tecnologia centrada no ser humano. Deverá servir de instrumento para as pessoas, com o objetivo último de aumentar o bem-estar das pessoas, e deverá ser concebida e desenvolvida de modo a que o seu impacto na sociedade seja benéfico. Por conseguinte, há que prever normas essenciais de proteção a fim de garantir que os sistemas de IA não sejam utilizados de uma forma que prejudique a dignidade humana e os direitos fundamentais. (UNIÃO EUROPEIA, 2024, Considerando 6). UNIÃO EUROPEIA. **Guia prático comum do Parlamento Europeu, do Conselho e da Comissão para a redação de textos legislativos da União Europeia**. Luxemburgo: Serviço das Publicações da União Europeia, 2015. Disponível em: <https://eur-lex.europa.eu/content/techleg/KB0213228PTN.pdf>. Acesso em: 5 fev. 2026.

Ao determinar que o indivíduo tem o direito inalienável de saber, em tempo real, que a sua contraparte interacional é uma máquina, o legislador protege a integridade da confiança social e interdita a falsificação da alteridade. Não se exige que o cidadão desvende a opacidade matemática do algoritmo, mas sim que a sua autonomia seja respeitada diante da dissimulação tecnológica. Essa barreira institucional contra abusos prepara o terreno para uma compreensão mais profunda das relações contemporâneas, revelando que a regulação da técnica é, no fundo, uma questão de defesa da vida (humano).

A materialização disso exige, contudo, mecanismos concretos que operem a distribuição equitativa de riscos, reconhecendo a vulnerabilidade inerente ao sujeito frente ao automatismo. É nesse cenário que a obrigatoriedade de uma avaliação de impacto nos direitos fundamentais deixa de ser uma simples auditoria de mercado para se converter em um instrumento ativo de reconhecimento. Ao forçar a antecipação e a mitigação de danos, assegura-se que a ética das instituições prevaleça sobre a lógica algorítmica, reafirmando a preservação da dignidade.

É exatamente neste ponto de intersecção, onde o direito busca proteger a fragilidade do sujeito contra a assimetria tecnológica, que a discussão clama pela sua fundamentação filosófica. Sob a ótica de Paul Ricoeur, a "visada ética" do ser humano apenas se completa no anseio pela "vida boa, com e para os outros, em instituições justas". Ao estabelecer o imperativo inegociável de uma Inteligência Artificial estritamente "centrada no ser humano", o regulamento europeu pode se aproximar da formulação de justiça ricoeuriana. A norma tenta, em sua essência, garantir que a máxima digital atue como uma instituição justa, sendo, assim, capaz de proteger a identidade narrativa do sujeito e respeitar a solicitude.

1.2. As Instituições Justas em Paul Ricoeur

A categoria de “instituições justas”⁷ ocupa uma posição central na ética de Paul Ricoeur e representa o ponto de articulação entre o ideal ético da vida boa e a concretude do mundo social, onde as relações se institucionalizam. Para Ricoeur, a ética não pode se limitar à interioridade do sujeito nem se reduzir à moral normativa. Ela exige a mediação de estruturas sociais que traduzem valores de justiça, equidade e reconhecimento. É nesse ponto

⁷ "Por instituição Ricoeur entende aquela estrutura de viver-junto de uma comunidade histórica – irreduzível às relações interpessoais, mas ligadas a elas – e caracterizada por costumes comuns (éthos) aos membros do grupo social". RICOEUR, Paul. *Soi-même comme un autre*. Paris: Seuil, 1990. p. 227.

que as instituições aparecem como condições de possibilidade para a realização de uma convivência justa, sendo elas o lugar onde o “com e para os outros” se torna viável de maneira duradoura. Nesse aspecto, busca-se o meio-termo:

A amizade é essa virtude de 'meio-termo' em que o amor a si mesmo permanece o princípio da justiça. [...] O homem feliz tem necessidade de amigos, não por utilidade ou por prazer, mas para tornar efetiva a sua própria atividade. É com o amigo que o homem de bem pode exercer a benevolência e a vida em comum (Ricoeur, 1991, p. 216).

Perscrutando, em alguma medida, Aristóteles, Ricoeur entende que a ética possui uma orientação teleológica: trata-se da busca por uma vida boa. No entanto, ele propõe uma leitura além, incluindo a solicitude como elemento constitutivo do sujeito ético. A justiça, então, não é apenas distribuição equitativa de bens, mas também reconhecimento mútuo e inclusão dos outros nas práticas institucionais.

Diferentemente do legalismo normativo (como o de Kant), que prioriza a universalização da regra, Ricoeur insiste na mediação entre a regra e o caso particular, havendo uma ponderação. Essa ênfase hermenêutica revela a necessidade de interpretar as normas à luz das situações concretas. Definindo a estima de si como o pilar reflexivo dessa busca pela realização humana, o filósofo postula:

Digamos, pois, que a estima de si é o momento reflexivo da visada da 'vida boa'. [...] Ao estimar-nos a nós mesmos, estimamo-nos capazes de agir por uma razão, quer dizer, segundo preferências ponderadas, e conseqüentemente capazes de hierarquizar nossos fins; em resumo, estimamo-nos a nós mesmos como a fonte de uma ação racional (Ricoeur, 1991, p. 203).

Nesse contexto, as instituições não são apenas mecanismos de controle social, mas encarnações da justiça no mundo. Elas precisam ser compreendidas como estruturas que sustentam a reciprocidade e garantem condições mínimas de equidade. Porém, quando capturadas por lógicas técnicas ou econômicas que negligenciam a dimensão ética, essas mesmas instituições podem se tornar fontes de injustiça. No mundo contemporâneo, isso é particularmente visível no campo da inteligência artificial, onde decisões cada vez mais são automatizadas sem a devida consideração pelo reconhecimento dos sujeitos afetados. Contudo, torna-se necessário transcender a proximidade do rosto e instituir o "terceiro":

A justiça, com efeito, estende a solicitude a 'cada um'. [...] A justiça não conhece rostos, mas cidadãos iguais diante da lei. O 'cada um' da distribuição justa é um 'cada um' qualquer, que não é mais o 'tu' da amizade. [...] É preciso, pois, que a relação interpessoal seja mediatizada por instituições que estruturam o viver-junto (Ricoeur, 1991, p. 235).

Com o crescimento da vigilância algorítmica e a proliferação de sistemas autônomos que tomam decisões sobre pessoas — como ocorre em carros autônomos (Awad et al., 2018; Borenstein et al., 2017) ou em classificações biométricas (Buolamwini & Gebru, 2018) —, a categoria de justiça institucional torna-se ainda mais urgente. Os sistemas de IA não apenas operam sobre dados, mas também refletem (ou impõem) determinadas concepções de justiça. Como afirmam Susan e Michael Anderson (2007; 2011), a criação de agentes morais artificiais exige um esforço deliberado de incorporação de princípios éticos. Contudo, esses princípios, para serem legítimos, devem emergir de um processo interpretativo e institucionalizado de deliberação pública, e não apenas de modelagens estatísticas.

A ausência de instituições justas no contexto algorítmico gera uma série de riscos. Em sua crítica à “superinteligência”, Nick Bostrom (2014) alerta para os perigos de agentes artificiais operarem segundo critérios utilitaristas radicais, como os de Bentham (1789)⁸, sem sensibilidade ao contexto humano. Essa lógica instrumental ameaça reduzir os indivíduos a dados e decisões a cálculos, ignorando a complexidade da experiência humana. Isso rompe o elo entre o justo e o reconhecimento da singularidade do outro.

Além disso, o debate atual sobre governança algorítmica e legislação robótica, como demonstram Balkin (2020) e Arkin (2009), revela uma lacuna entre a tecnologia e o arcabouço normativo. A tentativa de regular a IA a partir de categorias tradicionais do direito ignora a necessidade de mediação hermenêutica entre as normas e os contextos emergentes. A proposta ricoeuriana de que a justiça deve passar pela interpretação abre uma via promissora para se pensar sistemas regulatórios mais sensíveis à pluralidade e às tensões éticas envolvidas.

Por isso, Ricoeur não propõe uma justiça algorítmica automatizada, mas uma justiça interpretativa, institucionalmente mediada. A justiça, para ele, é inseparável do reconhecimento mútuo, o que exige instituições que escutem, deliberem e representem os afetados por suas decisões. Isso se contrapõe à lógica das chamadas “instituições algorítmicas” — como sistemas de reconhecimento facial, pontuação de crédito ou vigilância preditiva — que operam sem espaço para contestação, sem transparência e sem responsabilização clara.

⁸ Cf. BOSTROM, Nick. **Superintelligence: Paths, Dangers, Strategies**. Oxford: Oxford University Press, 2014. A obra de Bostrom explora os riscos existenciais associados ao desenvolvimento de uma superinteligência artificial, incluindo cenários onde um agente com poder cognitivo massivo poderia tomar decisões com base em uma interpretação extrema do utilitarismo. Essa perspectiva remete aos princípios formulados por Jeremy Bentham em sua obra seminal, na qual propõe que a moralidade de uma ação deve ser julgada pela sua capacidade de maximizar a felicidade e o bem-estar para o maior número de pessoas. Ver BENTHAM, Jeremy. **An Introduction to the Principles of Morals and Legislation**. London: T. Payne and Son, 1789.

É nesse sentido que a crítica contemporânea à opacidade dos sistemas de IA (Bender et al., 2021) encontra eco na filosofia de Ricoeur. A “inteligência estocástica” que gera respostas com base em padrões estatísticos é, muitas vezes, incapaz de justificar suas decisões à luz de valores éticos ou direitos fundamentais. A falta de explicabilidade dessas decisões compromete a confiança institucional e mina a legitimidade democrática das decisões tomadas por esses sistemas.

A ética biomédica, como a de Beauchamp e Childress (2001), já havia mostrado que princípios como autonomia, beneficência e justiça devem ser equilibrados caso a caso. Analogamente, no campo da IA, é necessário um espaço institucional que permita esse tipo de ponderação hermenêutica. As instituições justas, nesse cenário, são aquelas que possibilitam mediações entre interesses conflitantes, resguardando o lugar da interpretação ética.

O EU AI Act, legislação pioneira da União Europeia, busca enfrentar esses dilemas ao propor uma regulação baseada na avaliação de riscos. Contudo, como aponta Bryson (2018), há um risco de que esse modelo se torne excessivamente tecnocrático, priorizando a conformidade técnica em detrimento da deliberação ética. A proposta ricoeuriana convida a uma leitura crítica dessa regulação, destacando a necessidade de que ela se apoie em instituições capazes de escutar, interpretar e agir com responsabilidade.

As discussões sobre a “lei das máquinas” (Balkin, 2020) e sobre a criação de códigos éticos para sistemas autônomos (Boddington, 2017) devem ser inseridas nesse horizonte hermenêutico. Um código ético só terá eficácia moral se estiver enraizado em práticas institucionais que promovam o reconhecimento dos sujeitos. Para Ricoeur, a justiça não pode ser imposta de fora ou deduzida de princípios abstratos; ela deve emergir da experiência vivida da injustiça e da negociação coletiva de soluções.

O experimento moral Machine (Bonneton et al., 2016), que coletou percepções morais sobre decisões de carros autônomos, mostrou a diversidade cultural das intuições éticas. Essa pluralidade reforça a tese de Ricoeur: o justo não é o mesmo em todos os contextos e, por

isso, exige instituições interpretativas⁹. Sem esse trabalho institucional de mediação, as decisões algorítmicas tendem a replicar vieses e injustiças já existentes.

Em sua crítica à automação radical, Brynjolfsson e McAfee (2014) alertam que o progresso tecnológico, quando desacompanhado de valores humanos, tende a aumentar desigualdades. A proposta de Ricoeur oferece um antídoto a essa tendência: as instituições justas devem funcionar como contrapeso à lógica da eficiência, incorporando princípios de equidade, solidariedade e escuta mútua.

Por fim, a proposta ricoeuriana das instituições justas representa uma via robusta e necessária para pensar os desafios da IA e da vigilância algorítmica. Mais do que mecanismos de regulação, essas instituições são espaços de reconhecimento, onde a justiça se realiza através da interpretação, da deliberação e da construção compartilhada de sentido. Em um mundo cada vez mais moldado por sistemas técnicos, recuperar a centralidade do humano, da narrativa e da ética interpretativa torna-se não apenas desejável, mas imprescindível.

A teoria ricoeuriana das instituições justas também valoriza profundamente o papel da *responsabilidade* como critério de legitimação ética. Em um contexto em que decisões automatizadas desafiam a rastreabilidade da autoria e das consequências, as instituições precisam assumir a responsabilidade por aquilo que delegam aos sistemas técnicos. Para Ricoeur (1995), a justiça não pode ser dissociada do reconhecimento de responsabilidades mútuas. No caso da IA, isso significa que os agentes institucionais — humanos — devem responder pelas decisões geradas por algoritmos, mesmo quando essas decisões são altamente complexas ou opacas.

Outro ponto essencial na hermenêutica das instituições justas é a ideia de que o justo se constrói num campo de tensões entre o universal e o singular, entre a regra e o caso. Para Ricoeur (1992), a justiça requer uma *mediação criativa* entre princípios abstratos e situações concretas. No cenário da inteligência artificial, onde os sistemas tendem a aplicar padrões uniformes baseados em dados passados, o risco é a perda de sensibilidade às singularidades dos casos humanos. Instituições justas, portanto, precisam garantir espaços de flexibilidade

⁹ ROSSETTI, Ricardo. *Justiça em Paul Ricoeur: uma hermenêutica do homem justo*. 2011. Tese (Doutorado em Filosofia) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2011. p. 143. "O objetivo final é o de vislumbrar a narrativa jurídica possível como um processo de reconstrução das falas e das ações dos agentes em conflito: nesse processo os fatos são intercortados por rupturas temporais na própria ordem dos relatados, e reconstituídos por reencontros históricos dessas próprias falas, numa nova disposição do recontar o passado sob a perspectiva do sentido do justo."

interpretativa, onde exceções possam ser reconhecidas e tratados únicos possam ser considerados.

A temporalidade também é uma dimensão fundamental nas instituições justas de Ricoeur. A justiça não é estática: ela evolui conforme os contextos sociais, históricos e culturais. Aplicado às tecnologias emergentes, esse aspecto impõe a necessidade de constante revisão das normas que regem os sistemas de IA, bem como de sua performance prática. Em outras palavras, uma regulação justa da inteligência artificial deve prever sua própria *revisabilidade*, ajustando-se a novos desafios éticos, novos dados sobre impactos sociais e novas compreensões sobre o que significa agir justamente em contextos tecnicamente mediados (Ricoeur, 2004).

Outro elemento-chave na filosofia de Ricoeur é o reconhecimento como base da justiça institucional. Ele articula o conceito de reconhecimento em três níveis: reconhecimento de si, reconhecimento do outro e reconhecimento mútuo nas instituições. No contexto da IA, isso implica que as instituições justas devem ser capazes de reconhecer a diversidade dos sujeitos afetados por suas decisões — suas histórias, identidades, vulnerabilidades e direitos. Isso contrasta fortemente com abordagens tecnocráticas que tratam indivíduos como dados ou categorias estatísticas (Ricoeur, 2004; Taylor, 1994).

Não se trata apenas de aplicar regras previamente estabelecidas, mas de compreender a situação, ouvir os envolvidos e buscar um juízo prudente. Essa prudência interpretativa é tratada por Ricoeur, é um caminho fundamental para a *deliberação ética*. Nesse ensejo, pode-se contextualizar o desenvolvimento e uso da IA, onde o envolvimento de múltiplos atores — técnicos, juristas, filósofos, cidadãos — em processos deliberativos devem considerar os efeitos práticos e simbólicos das decisões algorítmicas. A sabedoria prática emerge como ponto necessário:

A sabedoria prática consiste em inventar as condutas que satisfarão o mais possível à exceção exigida pela solicitude, traindo o menos possível a regra. [...] O julgamento moral em situação não é a aplicação mecânica de uma regra, mas a arte de interpretar a regra à luz das particularidades da situação (RICOEUR, 1991, p. 301).

Além disso, Ricoeur chama atenção para o papel da *imaginação ética* na justiça. A imaginação permite que o intérprete se coloque no lugar do outro e antecipe as consequências das decisões. No contexto das instituições que usam IA, essa imaginação é essencial para lidar com situações inéditas e dilemas morais complexos. Por isso, a justiça institucional não

pode ser apenas um cálculo técnico, mas uma abertura ao outro, um exercício contínuo de empatia e julgamento sensível (Ricoeur, 1992).

A construção de instituições justas também está ancorada, para Ricoeur, no conceito de "mútua imputação". Isso significa que a justiça exige a responsabilização recíproca entre sujeitos e instituições. As instituições não apenas regulam, mas também são moldadas pelas ações e reações daqueles que a elas estão submetidos. Em contextos de IA, isso se traduz na necessidade de mecanismos de *accountability* que permitam aos cidadãos interpelarem as decisões algorítmicas, exigir correções e participar da sua melhoria contínua (Ricoeur, 1995; Weidinger et al., 2023).

Ainda nessa linha, é importante destacar que para Ricoeur, a justiça implica um projeto inacabado, sempre aberto à crítica. As instituições não alcançam um estado final de perfeição, mas estão sempre em busca de maior equidade e reconhecimento. Esse caráter processual da justiça faz com que os sistemas de IA e as políticas públicas associadas a eles devam estar abertos ao conflito produtivo, à dissidência democrática e à contestação pública. Assim, uma instituição justa não é aquela que evita o conflito, mas aquela que o acolhe como motor de aperfeiçoamento ético e político (Ricoeur, 2004).

Por fim, a proposta de instituições justas de Paul Ricoeur se alinha com uma visão profundamente democrática da justiça. Instituições justas são aquelas que se legitimam pela escuta dos cidadãos, pelo diálogo com a diferença e pela adaptação contínua a novas realidades. No enfrentamento dos desafios da inteligência artificial — especialmente no tocante à vigilância, discriminação e automação de decisões —, recuperar essa dimensão hermenêutica da justiça é fundamental. Não se trata apenas de controlar tecnologias, mas de fortalecer o tecido ético das instituições para que possam promover uma justiça viva, situada e verdadeiramente humana.

1.3. A hermenêutica como filosofia da mediação e dilemas gerais

A hermenêutica, tradicionalmente circunscrita à técnica de exegese textual, sofreu uma profunda mudança na atualidade filosófica, erigindo-se, com Paul Ricoeur, como uma filosofia da mediação. Em contraposição às pretensões de um modelo epistemológico pautado na objetividade asséptica ou na aplicação mecanicista de regras, a matriz ricoeuriana postula que o conhecimento e a própria constituição do sujeito são invariavelmente situados. A compreensão recusa qualquer acesso imediato ou transparente à realidade, exigindo um

percurso interpretativo estruturalmente mediado pela linguagem, pelos símbolos e pela espessura histórica. A centralidade dessa mediação como o caminho possível para a compreensão é lapidarmente trabalhada pelo filósofo:

Ao contrário da tradição do *cogito* e da pretensão do sujeito de se conhecer por intuição imediata, é preciso dizer que não compreendemos a nós mesmos senão pelo grande desvio dos signos de humanidade depositados nas obras de cultura. Que saberíamos nós do amor e do ódio, dos sentimentos éticos e, em geral, de tudo o que constitui o si, se isso não fosse levado à linguagem e articulado pela literatura? [...] Não há compreensão de si que não seja mediatizada por signos, símbolos e textos; a compreensão de si coincide, em última análise, com a interpretação aplicada a estes termos mediadores. (Ricoeur, 1989, p. 17; 40).

É precisamente esta exigência inegociável de mediação que confere à hermenêutica uma acuidade ímpar face aos desafios impostos pela inteligência artificial (IA). Em um contexto onde os algorítmicos tende a subtrair a intervenção humana — operando deliberações massivas a partir de cálculos estatísticos opacos e destituídos de ponderação ética —, a filosofia de Ricoeur atua como um contraponto filosófico incontornável. Ela denuncia a periculosidade de se delegar a resolução da vida prática a sistemas que não interpretam o mundo, mas apenas processam dados, reafirmando que a justiça exige, impreterivelmente, a mediação viva e o engajamento ético do sujeito.

A proposta hermenêutica implica, portanto, uma crítica às decisões automatizadas e à chamada *algocracia*, termo utilizado por Danaher (2016)¹⁰ para descrever a governança baseada em algoritmos. Nesse modelo, decisões sobre indivíduos e grupos são tomadas por sistemas que não compreendem as implicações morais, sociais e contextuais das ações que realizam. A hermenêutica, ao contrário, reconhece que compreender é sempre interpretar com base em uma história vivida, e que as decisões éticas precisam levar em conta essa historicidade.

Essa crítica hermenêutica se torna ainda mais urgente diante do avanço de tecnologias que perpetuam desigualdades estruturais, como demonstra Eubanks (2018), ao examinar como sistemas digitais têm sido usados para punir e vigiar populações marginalizadas. Sistemas automatizados que determinam acesso a benefícios sociais ou aplicam medidas de segurança pública frequentemente operam com base em dados enviesados, sem levar em consideração as narrativas e os contextos dos indivíduos afetados. A hermenêutica, nesse caso, exige escuta e responsabilidade interpretativa, elementos ausentes na lógica puramente técnica dessas tecnologias.

¹⁰ DANAHER, John. The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, v. 29, n. 3, p. 245-268, 2016.

O trabalho de Ruha Benjamin (2019) aprofunda essa crítica ao cunhar a expressão “Novo Código Jim” (*New Jim Code*), referindo-se ao modo como a tecnologia, sob o verniz da neutralidade, reproduz e reforça formas de racismo institucional. A hermenêutica, enquanto filosofia da compreensão, pode aqui servir como ferramenta de resistência, ao exigir que se leve em conta a voz dos sujeitos e as camadas simbólicas que moldam a experiência social. A tecnologia, ao contrário do que se presume, não é neutra, e a compreensão crítica desse fato é um imperativo ético.

Nesse sentido, os marcos regulatórios contemporâneos — como o *AI Act* da Comissão Europeia (2021) ou a proposta do *Algorithmic Accountability Act* do Congresso dos EUA (2019) — representam esforços institucionais de mediação que, embora necessários, ainda carecem de uma dimensão hermenêutica mais robusta. A regulação tende a operar com categorias formais, enquanto a hermenêutica insiste na necessidade de compreender os significados sociais e éticos das decisões algorítmicas. Como lembra Coeckelbergh (2020), a ética da IA não pode se limitar à conformidade técnica, mas deve envolver práticas reflexivas e compreensivas.

A proposta do *Ethics Guidelines for Trustworthy AI* (EU HIGH-LEVEL EXPERT GROUP, 2019) avança ao incluir dimensões como transparência, explicabilidade e supervisão humana. No entanto, mesmo essas diretrizes dependem de instituições que sejam capazes de interpretar e aplicar essas categorias com sensibilidade ao contexto. A hermenêutica pode contribuir aqui ao afirmar que o julgamento ético exige mais do que aplicação de regras: ele requer compreensão dos outros em sua alteridade, o que Ricoeur chama de solicitude. Essa solicitude está ausente em sistemas que operam sobre grandes massas de dados sem escutar os sujeitos a quem se referem.

Boyles (2018), ao oferecer uma visão geral sobre a ética da IA, destaca que um dos principais desafios atuais é garantir que os sistemas automatizados sejam auditáveis e compreensíveis. Essa exigência está diretamente ligada à hermenêutica, pois compreender um sistema não é apenas acessá-lo tecnicamente, mas interpretá-lo em relação às intenções, consequências e valores que carrega. A hermenêutica, portanto, torna-se uma ferramenta indispensável para a auditoria ética dos sistemas algorítmicos.

A proposta de Dignum (2019), ao tratar de “IA responsável”, está alinhada com essa visão hermenêutica. Segundo ela, não basta que um sistema de IA funcione corretamente — é necessário que ele funcione responsavelmente. Isso implica considerar os contextos sociais e

políticos em que a IA é implementada, bem como os efeitos que produz. Em outras palavras, a IA deve ser compreendida como uma prática social mediada, e não apenas como um instrumento técnico.

O Centre for Data Ethics and Innovation (2021), ao elaborar protocolos de auditoria algorítmica, reconhece que os sistemas técnicos precisam de supervisão humana contínua. Esse tipo de supervisão não é puramente técnica — envolve julgamento, interpretação e sensibilidade ao impacto ético. A hermenêutica oferece um modelo para essa prática ao lembrar que a compreensão nunca é neutra: ela se dá a partir de uma perspectiva situada, moldada por valores, experiências e narrativas.

O caso da decisão da *Cour de Cassation* (2020), que reconheceu motoristas de aplicativos como empregados e não como autônomos, exemplifica bem a importância da mediação interpretativa na aplicação de normas diante da tecnologia. Essa decisão se deu a partir da interpretação do contexto social e econômico desses trabalhadores, o que seria impossível por uma lógica algorítmica pura. A hermenêutica jurídica, aqui, cumpre uma função crítica e ética fundamental.

Floridi (2013; 2019) propõe uma “ética da informação” que reconhece o papel central dos dados na configuração da realidade contemporânea. Entretanto, ele também adverte que os sistemas informacionais não devem substituir a deliberação ética. Isso se alinha diretamente à hermenêutica: o conhecimento técnico não dispensa a interpretação ética, e os dados não falam por si mesmos. Eles precisam ser compreendidos à luz de narrativas humanas.

Em conjunto com Sanders (FLORIDI; SANDERS, 2004), Floridi propõe que agentes artificiais podem ser considerados moralmente responsáveis, desde que operem segundo estruturas éticas formalizadas. No entanto, mesmo essa proposta carece de mediação hermenêutica: por mais complexos que sejam os agentes artificiais, eles não possuem interioridade nem historicidade. A compreensão plena da ética exige algo que vai além do cálculo e da coerência formal.

Floridi e Taddeo (2018), ao refletirem sobre os limites da ética aplicada à IA, apontam que é preciso articular princípios normativos a realidades sociotécnicas específicas. Isso está em plena consonância com a filosofia hermenêutica, que desconfia de toda ética descontextualizada. A compreensão ética emerge sempre da intersubjetividade e da

experiência vivida — exatamente o que os sistemas algorítmicos não conseguem capturar sozinhos.

Dessa forma, a hermenêutica não se limita a uma teoria da interpretação textual, mas se transforma em um paradigma de compreensão das ações humanas, das instituições e dos sistemas técnicos. Quando aplicada à IA, essa abordagem permite enxergar os sistemas algorítmicos não apenas como instrumentos neutros, mas como produções humanas carregadas de intenção, contexto e potencial simbólico. Compreender uma decisão algorítmica é, nesse sentido, compreender também quem a projetou, com quais objetivos e em que ambiente social ela será aplicada.

Conforme já exposto, hermenêutica de Ricoeur¹¹ insiste na ideia de que o sentido nunca está dado de forma objetiva e imediata, mas sempre mediado por símbolos e narrativas. No campo da IA, isso significa que mesmo os dados considerados “puros” passam por construções semânticas e decisões humanas que lhes atribuem valor e significado. Ao contrário da crença na neutralidade dos dados, a hermenêutica compreende que a informação é atravessada por interpretações prévias, muitas vezes implícitas ou naturalizadas.

Em vez de separar rigidamente a descrição objetiva do mundo (os dados) da prescrição normativa (os valores), Ricoeur aborda uma dialética entre o “ser” e o “dever-ser”. Isso ajuda a perceber implicações profundas na construção de sistemas de IA: decisões aparentemente técnicas, como a priorização de certos dados, implicam inevitavelmente julgamentos de valor sobre o que importa, quem importa e por quê.

A hermenêutica também lembra que a compreensão é sempre situada no tempo — ela se dá em um processo histórico contínuo. Por outro lado, a IA pode ignorar a temporalidade das decisões ao operar com base em dados passados que são projetados para o futuro sem mediação crítica. Sistemas preditivos, por exemplo, podem cristalizar padrões históricos de exclusão ao transformá-los em “tendências”. A hermenêutica exige que se reconheça essa historicidade e que se busque romper com inércias injustas ao interpretar os dados com consciência ética.

¹¹ ROSSETTI, Ricardo. *Justiça em Paul Ricoeur: uma hermenêutica do homem justo*. 2011. Tese (Doutorado em Filosofia) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2011. p. 138. "A identidade do eu julgador se constitui a partir desse processo de ascrição, quando aquele que julga um acontecimento se apropria de uma ação-narrativa, deliberando a partir dela acerca dos sentidos possivelmente atribuíveis ao acontecimento."

Essa crítica se estende à própria noção de explicabilidade algorítmica, muitas vezes reduzida a uma transparência formal ou à possibilidade de reproduzir matematicamente os passos de um algoritmo. A hermenêutica propõe outro tipo de explicação: aquela que produz sentido para o outro, que justifica uma ação com base em critérios compreensíveis e partilháveis. A verdadeira explicabilidade, nesse sentido, é dialógica e ética, e não apenas lógica. Ela exige linguagem, escuta e responsabilização.

A filosofia hermenêutica também contribui com a ideia de que o julgamento ético requer distanciamento crítico. Em oposição à automatização do juízo, que aplica regras sem reflexão, o julgamento hermenêutico é fruto de uma pausa, de uma suspensão da ação imediata para que se possa compreender o contexto, ouvir os envolvidos e deliberar com prudência. Isso é especialmente relevante em contextos de uso da IA em decisões de alto impacto, como sentenças judiciais ou diagnósticos médicos automatizados, onde a urgência da decisão não pode eliminar a necessidade de escuta e ponderação.

Nesse sentido, a hermenêutica contribui também para a construção de uma cultura política mais sensível à justiça interpretativa. Em vez de pensar a justiça como mera aplicação de normas, ela a concebe como um processo de reconhecimento mútuo e deliberação partilhada. Isso exige instituições que operem não apenas com base em eficiência, mas também com base na escuta e no cuidado. A IA, enquanto parte dessas instituições, deve ser desenhada e utilizada de modo que permita e reforce essa dinâmica interpretativa e não a silencie.

A hermenêutica, finalmente, convida a uma ética da humildade interpretativa. Nenhuma decisão, por mais bem informada ou tecnicamente fundamentada, pode pretender ser definitiva ou total. A compreensão humana é sempre parcial e aberta à revisão. Por isso, os sistemas de IA devem ser acompanhados de mecanismos institucionais que permitam sua constante reinterpretação, correção e aprimoramento. Isso implica na construção de estruturas que valorizem o dissenso, o erro reconhecido e a possibilidade de mudança.

Infere-se, portanto, que a hermenêutica ricoeuriana transcende a condição de mero instrumental metodológico para se consolidar como um autêntico *ethos* diante dos dilemas instaurados pela inteligência artificial. A ontologia da compreensão atesta que o ato de interpretar é, na sua essência, um engajamento ético indissociável da responsabilidade e da solicitude para com a alteridade. Perante a lógica operada pelo automatismo sistêmico — que tende a converter a complexidade existencial em perfis estatísticos —, a hermenêutica

erige-se como uma defesa irrestrita da identidade narrativa dos sujeitos. Mais do que uma resistência teórica, ela impõe o imperativo prático e institucional de que nenhuma deliberação com impacto na vida humana se consume sem a devida mediação intersubjetiva, assegurando que o indivíduo não seja meramente processado pela máquina, mas ativamente reconhecido na sua singularidade.

1.4. Ricoeur e a ética da justiça: da identidade narrativa à solicitude pelo outro

Paul Ricoeur construiu uma filosofia ética profundamente enraizada na narrativa, na alteridade e na justiça. Ao vincular identidade pessoal à história que cada sujeito conta de si — a chamada “identidade narrativa” — Ricoeur propõe que o reconhecimento ético passa, inevitavelmente, pela escuta do outro e pelo respeito à sua singularidade.

A identidade narrativa permite compreender o sujeito como alguém constituído no tempo, em constante construção, mas sempre situado. Essa concepção entra em tensão com sistemas algorítmicos que reduzem sujeitos a dados e padrões, desconsiderando sua historicidade e singularidade. Como aponta Pasquale (2015), algoritmos operam frequentemente em “caixas-pretas” que impedem o reconhecimento do indivíduo como um ser ético e relacional, substituindo a solicitude por mera eficiência estatística.

A justiça, para Ricoeur, é inseparável da ética da solicitude, que exige considerar o outro não como um objeto de cálculo, mas como um sujeito de dignidade. O erro ético, nesse caso, reside em substituir a escuta pelo julgamento estatístico descontextualizado. Nesse caso, pode-se formular uma crítica direta a partir do “estimar-se a si mesmo”. Expandindo essa análise, o reconhecimento do outro torna-se condição para a própria estima, conforme assevera Ricoeur:

A solicitude não se acrescenta de fora à estima de si, mas desdobra a dimensão dialogal implícita nesta última. [...] Estimar-se a si mesmo é estimar-se como o tu de um outro. [...] A equivalência entre o 'tu também' e o 'eu mesmo' repousa sobre a confiança que tenho de que o outro é capaz, tal como eu, de iniciativa (Ricoeur, 1991, p. 226).

A solicitude implica reconhecer a vulnerabilidade do outro e agir com responsabilidade em relação a ela. Essa postura é incompatível com sistemas opacos que, como observa Mittelstadt et al. (2016), reproduzem desigualdades estruturais sob o disfarce da neutralidade técnica. A ética da justiça, então, exige transparência, explicabilidade e prestação de contas — princípios que também estão no centro das preocupações atuais sobre IA responsável (MONTREAL DECLARATION, 2018). Em arremate, a violência pode ser

entendida como um ataque à capacidade de agir do sujeito, indo além da mera dor física. Nas palavras de Ricoeur:

O sofrimento não é unicamente definido pela dor física, nem mesmo pela dor mental, mas pela diminuição, até a destruição da capacidade de agir, do poder-fazer, sentidos como um atentado à integridade do si. [...] A violência equivale à diminuição ou à destruição do poder-fazer de outrem (Ricoeur, 1991, p. 258).

Ricoeur propõe que a justiça não se esgota na imparcialidade ou na equidade abstrata, mas deve incluir uma dimensão narrativa que restitui a voz aos envolvidos. Isso está em consonância com as exigências de explicabilidade em IA, como apontado por Miller (2019), que defende a importância de explicações compreensíveis pelos usuários, baseadas em modelos de raciocínio humano. Sistemas que explicam suas decisões promovem maior *accountability* e se aproximam de uma justiça dialógica, em vez de técnica.

A abordagem de Ricoeur também permite reinterpretar o papel da responsabilidade nas decisões técnicas. Ao afirmar que o agir ético exige responsabilidade pelo outro, mesmo quando esse outro é ausente ou desconhecido. Isso se torna especialmente importante em contextos como o descrito pelo relatório da ONU (OHCHR, 2021), que alerta para o uso de IA na vigilância de minorias étnicas, um claro exemplo de como a tecnologia pode ser usada para aprofundar injustiças históricas.

Parfit (2011), ao discutir as razões que importam moralmente, argumenta que nossas obrigações éticas devem se estender a todos os afetados por nossas ações, inclusive os que não vemos diretamente. Isso complementa a ideia ricoeuriana de solicitude e reforça a crítica às inteligências artificiais que operam sem considerar o impacto humano mais amplo de suas decisões. A IA precisa ser responsabilizada não apenas por resultados, mas pela forma como compromete (ou não) a dignidade e a voz dos sujeitos.

Nesse sentido, os experimentos da *Moral Machine* (MIT, 2016) revelam como o julgamento moral em situações-limite é afetado por fatores culturais, empíricos e subjetivos. Embora úteis para compreender preferências humanas, esses experimentos também mostram os limites de modelar decisões éticas apenas com base em estatísticas agregadas. A ética de Ricoeur insiste que a moralidade não pode ser reduzida a regras fixas ou à maximização de consequências — ela exige deliberação, narrativa e reconhecimento mútuo.

Moor (2006) observa que uma das dificuldades centrais da ética de máquinas é sua incapacidade de compreender o valor contextual das ações. Essa crítica é consistente com a perspectiva de Ricoeur, que defende a mediação narrativa como única forma de captar o

significado ético do agir. A IA, para ser justa, precisaria não apenas calcular, mas compreender — e compreender, na hermenêutica, é sempre interpretar em situação.

A justiça automatizada, ao tentar aplicar critérios objetivos, corre o risco de ignorar o “quem” da ação — o sujeito com história, relações e expectativas. A crítica ética de Ricoeur está, então, na base de propostas que buscam reumanizar os sistemas algorítmicos, como as políticas da OCDE (2021) que exigem requalificação humana e acompanhamento ético em ambientes automatizados. O desenvolvimento de IA não pode ser dissociado de um projeto de justiça social.

A solicitude, como horizonte normativo, não substitui as normas jurídicas, mas as orienta. Ricoeur propõe que a justiça só é plenamente realizada quando as regras são aplicadas com discernimento ético e atenção à singularidade. Essa proposta se opõe à uniformização técnica da IA e aponta para a necessidade de mediação humana nas decisões automatizadas. Pasquale (2015) alerta para os perigos de um “sistema algorítmico de vigilância” que, sob o manto da eficiência, impõe formas de controle opacas e potencialmente autoritárias. A filosofia de Ricoeur oferece uma resistência ética a esse projeto: ela exige que as decisões sejam abertas à crítica, à contestação e à revisão. A solicitude se transforma, assim, em princípio político: a justiça passa a depender da escuta e do reconhecimento mútuo.

A IA generativa, como exemplificada pelo *ChatGPT*, introduz novas camadas de complexidade, uma vez que sistemas cada vez mais sofisticados interagem com a linguagem humana. Ainda que útil e produtiva, essa tecnologia pode reproduzir preconceitos, estereótipos e exclusões se não for orientada por princípios éticos claros. A ética ricoeuriana, ao exigir mediação interpretativa e responsabilidade narrativa, pode guiar o desenvolvimento ético dessas ferramentas.

Em suma, a filosofia da justiça de Paul Ricoeur — centrada na identidade narrativa e na solicitude — oferece um alicerce robusto para enfrentar os dilemas éticos da inteligência artificial. Mais do que corrigir desvios técnicos, trata-se de reformular os fundamentos da ação tecnológica em chave ética e relacional. Somente com esse compromisso será possível construir sistemas realmente justos: que ouçam, interpretem e respeitem o outro como sujeito.

1.5. Era algorítmica

A edificação de instituições justas exige, portanto, que a racionalidade algorítmica seja subordinada aos princípios basilares da dignidade humana, da autonomia e da equidade social. Nesse sentido, as *Ethics Guidelines for Trustworthy AI* (2019) consolidam a premissa de que os sistemas automatizados não devem ser avaliados apenas pela sua conformidade legal, mas pela sua robustez técnica e orientação ética. O documento erige a transparência, a rastreabilidade e a inclusão como exigências inegociáveis para a salvaguarda dos direitos fundamentais e a promoção do bem-estar coletivo (UNIÃO EUROPEIA, 2019, p. 5).

Para que essa robustez se materialize, a ética das virtudes deve transpor a abstração e integrar o próprio design tecnológico. Como adverte Vallor (2016, p. 118-121), o cultivo de disposições como empatia, responsabilidade e prudência — o que a autora denomina "sabedoria tecnomoral" — é imperativo para que as máquinas, bem como os seus desenvolvedores, consigam interagir responsavelmente num tecido social plural e complexo.

Sob uma ótica complementar, Wallach e Allen (2009, p. 13-14) postularam a urgência de inculcar uma "moralidade funcional" nas arquiteturas de IA. Os autores sustentam que, à medida que a delegação de tarefas críticas aumenta e a supervisão humana recua, os algoritmos precisam operar sob diretrizes que lhes permitam ponderar as consequências de suas deliberações sobre a vida humana, respeitando as normas éticas estabelecidas.

A consecução dessa agência moral, contudo, esbarra na extrema complexidade e na opacidade estrutural dos modelos preditivos. Essa "caixa-preta" algorítmica frequentemente inviabiliza o escrutínio sobre como a máquina chegou a um determinado veredito, tornando imperiosa a formulação de mecanismos técnicos de explicabilidade e auditoria para garantir a responsabilidade corporativa (WACHTER; SANDVIG; KROLL, 2017, s.p.).

Cientes dessa opacidade, Winfield et al. (2016, p. 2) argumentam que as salvaguardas da ética das máquinas devem ser concebidas ab initio, isto é, integradas desde a fase embrionária de modelagem. O objetivo dessa governança preventiva é criar sistemas aptos a antever as ramificações de suas operações e a calcular rotas de ação que minimizem potenciais danos sociotécnicos.

Entretanto, ainda que princípios éticos balizem a programação, a sua transposição para a contingência do mundo real revela-se árdua. A miríade de variáveis contextuais exige que os algoritmos possuam maleabilidade para se adaptar a diferentes cenários sem corromper as exigências de justiça, evitando, assim, a reprodução automatizada de vieses (WEIDINGER et al., 2023, p. 10).

Tais vulnerabilidades técnicas são drasticamente exacerbadas pelo fenômeno que Zuboff (2019, p. 8-9) cunhou de "capitalismo de vigilância" — uma lógica de acumulação que expropria a experiência humana, convertendo-a em matéria-prima gratuita para a modulação de comportamentos. Trata-se de um modelo econômico parasitário que instrumentaliza a privacidade e assedia a autonomia em favor da maximização do lucro, exigindo uma oposição regulatória enérgica.

Superar esse paradigma extrativista e consolidar instituições justas na era digital impõe, portanto, um esforço multifacetado. A estratégia pressupõe desde a formulação de legislações estritas para a governança de dados até a execução de políticas públicas que fomentem a inclusão e a transparência sistêmica (UNIÃO EUROPEIA, 2019, p. 5).

Concomitantemente, a literacia ética desponta como uma exigência formativa. Vallor (2016, p. 10) destaca a premência de integrar a ética das virtudes aos currículos das ciências exatas, preparando os futuros engenheiros e cientistas de dados para enfrentar e mitigar os dilemas morais inerentes às inovações que concebem. Soma-se a isso a indispensável democratização do debate tecnológico. Para que a governança não se torne uma tecnocracia excludente, a formulação de diretrizes sobre o uso de IA requer o engajamento direto da sociedade civil.

Neste xadrez geopolítico, a cooperação transnacional é o único mecanismo capaz de estabelecer um lastro normativo global. A União Europeia tem capitaneado esse movimento (UNIÃO EUROPEIA, 2019, p. 5), propondo marcos jurídicos vanguardistas que visam frear as corporações transnacionais que ainda resistem em subordinar a sua rentabilidade ao imperativo dos direitos humanos (ZUBOFF, 2019, p. 11).

É crucial sublinhar que a edificação de instituições justas requer, também, o compromisso ativo com a proteção dos vulneráveis. Impõe-se a necessidade de assegurar que as populações historicamente marginalizadas não sejam reféns de falsos positivos ou de discriminação algorítmica, promovendo uma equidade material no ecossistema tecnológico (WEIDINGER et al., 2023, p. 11).

A consolidação dessa justiça institucional exige, antes de tudo, um deslocamento epistemológico: o reconhecimento categórico de que a técnica não é neutra. Os algoritmos são artefatos sociotécnicos; eles embutem pressupostos, prioridades e valores de seus criadores. Por conseguinte, o seu uso nas esferas de poder demanda uma ética deliberativa capaz de questionar continuamente esses fundamentos.

Nesse horizonte, a justiça não pode ser reduzida à igualdade algorítmica formal. Requer-se uma perspectiva substantiva que pondere as desigualdades históricas. Ao operar

calcada no passado, a IA tende a fossilizar assimetrias estruturais. A instituição verdadeiramente justa é aquela dotada de agência para detectar essas dinâmicas retroativas e corrigi-las, adotando uma postura inclusiva e restaurativa.

Ademais, a racionalidade puramente calculista deve ser tensionada pela solicitude. O maquinário algorítmico é cego às dimensões afetivas e subjetivas que alicerçam a convivência humana. A mediação institucional serve, precisamente, para atuar como o contrapeso empático da eficiência técnica, garantindo que as sentenças e as resoluções não percam a sua ligação com a experiência vivida do sujeito.

O fortalecimento da esfera pública é, sob esta ótica, inegociável. Em tempos de automação, a deliberação democrática deve assegurar espaços imunes ao cálculo computacional. A literacia digital torna-se, assim, uma condição para a cidadania plena, capacitando o indivíduo a compreender o arcabouço tecnológico que o cerca e a posicionar-se criticamente perante o mesmo.

Para tanto, a transparência institucional não pode limitar-se a formalidades incompreensíveis, como a mera abertura de códigos-fonte. A exigência de explicabilidade deve traduzir-se na criação de condições semânticas e cognitivas reais para que o cidadão entenda a razão — e não apenas a métrica — que determinou uma decisão restritiva de direitos sobre a sua vida.

Tamanho desafio impõe uma abordagem eminentemente transdisciplinar. As instituições contemporâneas precisam hibridizar saberes do direito, da filosofia, das ciências sociais e da computação. O isolamento tecnocrático das áreas de engenharia gera soluções eticamente estéreis para aporias que, na verdade, afetam a totalidade do tecido social.

Essa governança compartilhada deságua no problema central da imputabilidade. É imperativo estancar a "diluição de responsabilidade", fenômeno no qual as falhas institucionais são covardemente justificadas como "erros do sistema". A cadeia de responsabilidade deve identificar os agentes humanos em cada etapa do ciclo da IA — do desenho ao monitoramento —, garantindo o direito à reparação.

Ressalta-se, ainda, que a efetividade desta reparação passa por uma escuta ativa das populações mais afetadas pela vigilância e exclusão digital. Criar canais de resposta institucional para essas comunidades é o que separa uma regulação de vitrine de uma estrutura verdadeiramente comprometida em nivelar as assimetrias de poder.

Para suportar essas pressões, a própria arquitetura do Estado necessita de atualização, consubstanciada na criação de agências reguladoras especializadas. Tais órgãos devem ser

munidos de autonomia e expertise técnica para auditar modelos, sancionar abusos e dialogar, em igualdade de condições, com a comunidade científica e o mercado.

Nessa dinâmica fiscalizatória, não bastam auditorias de desempenho estatístico; urge o desenvolvimento de métricas de aferição do impacto humano. Indicadores de sucesso institucional devem mensurar a IA pela sua aderência aos direitos fundamentais, pelo bem-estar gerado e pela redução das disparidades, refutando a eficiência como valor absoluto.

Tais agências também enfrentam o desafio estratégico da soberania digital. A excessiva subordinação das instituições públicas a soluções desenvolvidas por monopólios corporativos compromete a autonomia estatal. A construção de uma justiça soberana passa pelo incentivo a alternativas tecnológicas de código aberto, locais e rigorosamente auditáveis.

Em suma, a "instituição justa" na era algorítmica não constitui um fim estático, mas um fim hermenêutico. Dada a velocidade da disrupção tecnológica, a revisão dos marcos legais e éticos deve ser orgânica. O que se exige, mais do que respostas peremptórias, é a sedimentação de uma cultura do reconhecimento mútuo, guiada pela interpretação crítica e pela defesa irrestrita da alteridade humana.

Em segundo aspecto, o exemplo do comando de defesa de uma potência mundial passou a empregar plataformas de aprendizado profundo para mapear e atingir bases de operação vinculadas a um país rival¹² infere a conversão do *software* em árbitro da letalidade, evidenciando, assim, uma lacuna perigosa: a urgência de que os marcos do Direito Internacional tratem de barreiras absolutas contra a supressão da vida humana sob a chancela de processos automatizados.

Por outro lado, a apropriação de recursos de natureza por vias de predição ilustra dinâmicas de mercado. A imprensa detalha que uma mineradora de capital estrangeiro "[...] assinou um acordo de pesquisa e desenvolvimento com um laboratório nacional do Departamento de Energia dos Estados Unidos para aplicar inteligência artificial no processo de separação de terras raras pesadas"¹³. O avanço sobre os minérios de território de Brasil atesta a conversão de elementos físicos de solo em vetores de eficiência de capital.

No domínio de cadeias de transporte de comércio, a delegação de comando de frotas de mar para modelos de visão de máquina materializa a supressão do humano em esferas de

¹² CNN BRASIL. Entenda como EUA usam inteligência artificial em guerra com Irã. **CNN Brasil**, São Paulo, 26 fev. 2024. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/entenda-como-eua-usam-inteligencia-artificial-em-guerra-com-ira/>. Acesso em: 6 mar. 2026.

¹³ CNN BRASIL. EUA e empresa canadense apostam em IA para refino de terras raras do Brasil. **CNN Brasil**, São Paulo, 18 jan. 2026. Disponível em: <https://www.cnnbrasil.com.br/economia/macroeconomia/eua-e-empresa-canadense-apostam-em-ia-para-refino-de-terras-raras-do-brasil/>. Acesso em: 6 mar. 2026.

deliberação de risco. A desenvolvedora de plataforma de autonomia de navegação descreve que o seu sistema de software de embarcação "[...] atua como um vigia totalmente automatizado que não dorme nem falha". A transferência de rotinas de vigilância de curso e de prevenção de colisão para linhas de código consubstancia a automação de processos de segurança civil¹⁴. A delegação de monitorização a um ente de algoritmo que promete ausência de falha instaura a presunção de infalibilidade de técnica, o que desafia a condição de fragilidade humana teorizada por Ricoeur.

Quando a automação decisória opera sobre volumes massivos de dados, o código computacional ameaça impor-se sobre a vida. Há o risco latente de que a invisibilização das singularidades sob métricas preditivas converta a vivência humana (a *ipseidade*) em perfis estatísticos engessados (a *mesmidade*), promovendo a cristalização do sujeito e impossibilidade de reflexão com vista à identidade narrativa.

O texto europeu passa a ser escrutinado não como uma letra fria da lei, mas considerando a possibilidade de um espaço de reconhecimento mútuo, onde é possível existir a tensão entre a máquina de predição e a solicitude humana busca resolução. Torna-se plausível avaliar se a materialidade da regulação no *EU AI Act* — por meio de seus instrumentos de transparência e controle — é um exemplo que se aproxima de uma autêntica “instituição justa”, resguardando a *ipseidade* em face do risco dos reducionismos algorítmicos.

Ao percorrer o polo da explicação — em que o algoritmo pôde ser analisado sob a ótica do distanciamento estrutural —, o delineamento metodológico desta pesquisa parece abrir espaço para a travessia rumo ao polo da compreensão. Na esteira de Ricoeur (1989, p. 161-170), explicar e compreender não atuam como métodos excludentes, mas como “dois momentos de um único e mesmo processo hermenêutico”, no qual a análise estrutural tende a operar como uma etapa de mediação entre a interpretação de superfície e uma possível apropriação em profundidade. O trajeto indica que a adoção dessa suspensão transitória ante os limites do código pode se revelar uma via fecunda para fundamentar o esforço hermenêutico subsequente. Desse modo, o segundo capítulo propõe-se a direcionar a lente filosófica para a concretude ética, tomando a identidade narrativa como ponto de partida reflexivo, a fim de investigar em que medida as disposições europeias poderiam dialogar com a urgência de consolidação de instituições justas perante o avanço da máquina.

¹⁴ ORCA AI. SeaPod. **Orca AI**, [S. l.], [2026]. Disponível em: <https://www.orca-ai.io/seapod/>. Acesso em: 5 mar. 2026. Tradução nossa.

2. O *EU AI Act*: UMA HERMENÊUTICA POSSÍVEL

Considerando a possibilidade de direcionar o arcabouço filosófico até aqui construído para a materialidade da norma, compreende-se que hermenêutica de Ricoeur pode viabilizar a decomposição dos alicerces do agir de Estado. A justiça da instituição não decorre da obediência de forma à regra positivada, mas da capacidade da estrutura de promover o “reconhecimento” (RICOEUR, 2007, p. 22). Por conseguinte, pontua-se a necessidade de atenção, segundo Ricoeur (2007, p. 45), às circunstâncias: universalidade, e a vivência do sujeito, na sua dimensão de contingência.

Em segundo, o teste desse preceito face ao *EU AI Act* impõe uma mediação de ordem filosófica, uma vez que a adoção da abordagem ricoeuriana permite romper com a limitação das análises de forma. Nesse caso, a semiótica possibilita a leitura dos sentidos com a apreensão dos símbolos da norma examinada, ou seja, do *EU AI ACT*.

A captura de dados em fluxos de rotina exige um exame de crítica que exceda o limite do texto de lei. A operação de vigilância ocorre por meio do cálculo de predição e do desenho de cenários de ações de utilizadores. Nesse sentido, o código de máquina assume função de deliberação, ocupando espaços de decisão de competência de entes de democracia.

Em contraponto seria o caso do texto de regulação transcender a função de catálogo de requisitos de sistema para operar como vetor de justiça. Vainzof *et al.* (2023, p. 112) sublinham que a norma estabelece limites de proibição de práticas, a exemplo da indução de subconsciente e da atribuição de notas de cidadania. Não obstante, o ganho da norma, atinge o seu termo apenas quando as instâncias de aplicação atuam como arenas de mediação de conflitos, pressupondo a adoção de posturas de solicitude e a afirmação de laços, em conformidade com o postulado de Ricoeur (2005, p. 263).

A raiz da crise encontra-se no método de incorporação da técnica às esferas de decisão. A delegação de sentenças de acesso a bens de bem-estar ou a emissão de alertas de ameaça por via de cálculo instauram um modelo de administração em choque com o ideal de justiça de deliberação. O legislador propõe mitigar o quadro através de obrigações de transparência, de exames de auditoria e de manutenção de pessoas na cadeia de comando. Os estudos de Vainzof *et al.* (2023, p. 230), porém, sinalizam a existência de fissuras na conversão desses deveres de código em mecanismos de garantia de fato.

O modelo de extração de metadados, sem a barreira da escuta, pode convergir para a automação de processos de margem e de recusa de direitos. Espera-se que a justiça institucional consista num processo contínuo em que se negociem valores. O que se encontra

em jogo é o modo como as instituições postulam que a justiça opere em tempos de governança algorítmica. Um regulamento destas proporções apenas alcançará a sua legitimidade caso seja capaz de mediar os riscos tecnológicos para que mitigue, sobretudo, injustiças. Conforme apontam Vainzof *et al.* (2023, p. 301), é imperativo que esta estrutura exige que uma escuta filosófica atravesse o seu texto, de forma a que o outro não seja entendido como objeto, mas que seja reconhecido como sujeito.

Dado que a vigilância algorítmica pode silenciar identidades sob o dogma da métrica, a subsistência do *EU AI Act* como uma resposta justa exige que ele seja por mediação. Em suma, pensar o *EU AI Act* através da hermenêutica de Ricoeur, é uma reconstrução simbólica da própria justiça. Trata-se de admitir que a regulação também exige interpretação.

2.1 A regulação como instituição justa: uma leitura possível do *EU AI Act*

A análise do *AI Act* deve iniciar-se pela sua finalidade, que não se restringe à harmonização de mercado, mas pode projetar uma aspiração ética compatível com a abordagem de Ricoeur. Ao exigir que a IA seja "centrada no ser humano", a instituição jurídica define um horizonte de sentido onde a técnica não é um fim em si mesma, mas um meio subordinado à realização da existência humana em sua plenitude ética e política. A esse respeito, o legislador europeu estabeleceu como objetivo central a promoção de uma tecnologia que sirva à dignidade humana, conforme se extrai do dispositivo inaugural:

A finalidade do presente regulamento é melhorar o funcionamento do mercado interno e promover a adoção de uma inteligência artificial (IA) centrada no ser humano e de confiança, assegurando simultaneamente um nível elevado de proteção da saúde, da segurança e dos direitos fundamentais, inclusive a democracia, o Estado de direito e a proteção do ambiente [...] (UNIÃO EUROPEIA, 2024, art. 1º, n. 1).

A fim de que essa possibilidade de aspiração ética com centralidade no humano ganhe contornos precisos na lei, faz-se imperativo examinar a definição de "sistema de IA". Sob a lente ricoeuriana, um olhar cuidadoso desse conceito revela o hiato hermenêutico: o sistema "infere", mas não "narra". A autonomia descrita na lei é funcional, não ética. Logo, concebe-se a esfera da "instituição justa" como necessária para mediar a relação entre sujeitos narrativos (humanos) e agentes inferenciais (máquinas), mitigando que a lógica probabilística substitua a compreensão histórica da vida. O texto legal enfatiza a capacidade de inferência e autonomia, traços que mimetizam a ação humana, mas carecem de sua substância moral:

«Sistema de IA», um sistema baseado em máquina concebido para funcionar com níveis de autonomia variáveis e que pode apresentar capacidade de adaptação após a implantação e que, para objetivos explícitos ou implícitos, infere, a partir dos dados de entrada que recebe, a forma de gerar resultados como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais (UNIÃO EUROPEIA, 2024, art. 3º, n. 1).

No polo humano dessa dinâmica, a capacidade de agir e sofrer a ação, central no pensamento de Ricoeur, exige a capacidade de agir e sofrer a ação, central no pensamento de Ricoeur, exige que o sujeito compreenda o mundo que o cerca. A instituição justa permite interrogar até que ponto o indivíduo é um agente. Pode supor a seguinte interrogação: há situações no texto legal que ajudam o sujeito a não tornar-se um mero paciente da vigilância algorítmica, perdendo a "ipseidade" necessária para atestar suas próprias escolhas frente à máquina? Em resposta, tem-se a noção de literacia em Inteligência Artificial tratada no próprio regulamento. O conceito de Literacia em IA (*AI Literacy*) transcende a mera capacidade operacional de manusear ferramentas tecnológicas. Segundo Long e Magerko (2020, p. 2, tradução nossa)¹⁵, a literacia em IA é definida como "um conjunto de competências que permite aos indivíduos avaliar criticamente as tecnologias de IA; comunicar e colaborar eficazmente com a IA; e usar a IA como uma ferramenta online, em casa e no local de trabalho". Uma aproximação possível com Ricoeur pode ser a necessidade imposta pelo regulamento do fomento à literacia em IA:

Os fornecedores e os responsáveis pela implantação de sistemas de IA devem tomar medidas para garantir, na medida do possível, que o seu pessoal e outras pessoas que lidam com a operação e a utilização de sistemas de IA em seu nome tenham um nível suficiente de literacia em IA, tendo em conta os seus conhecimentos técnicos, a sua experiência, a sua educação e a sua formação [...] (UNIÃO EUROPEIA, 2024, art. 4º).

No que tange às práticas inaceitáveis, a legislação protege a integridade da consciência humana contra manipulações que escapam à percepção do sujeito, preservando a autonomia da vontade. Ricoeurianamente falando, se o sujeito não tem consciência dos vetores que moldam sua ação, ele não pode ser responsabilizado por ela, rompendo o vínculo entre o agente e o ato. O Artigo 5º é taxativo lembrar que existem técnicas que viola a consciência humana:

A colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que utilize técnicas subliminares que agem sem que uma pessoa disso tenha

¹⁵ "We define AI literacy as a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace." (LONG; MAGERKO, 2020, p. 2).

consciência ou técnicas deliberadamente manipuladoras ou enganosas, com o objetivo ou o efeito de distorcer substancialmente o comportamento de uma pessoa ou de um grupo de pessoas [...] (UNIÃO EUROPEIA, 2024, art. 5º, n. 1, alínea a).

A legislação demonstra alguma sensibilidade particular à exploração das fragilidades humanas, alinhando-se, em certa medida, à ética da solicitude que prioriza o cuidado com o "outro". A "instituição justa" de Ricoeur não é neutra; ela observa o polo mais fraco da relação (o vulnerável). O texto legal proíbe sistemas que abusem de condições específicas dos indivíduos:

A colocação no mercado [...] de um sistema de IA que explore as vulnerabilidades de uma pessoa singular ou de um grupo específico de pessoas decorrentes da sua idade, deficiência ou de uma situação social ou económica específica, com o objetivo ou o efeito de distorcer substancialmente o comportamento dessa pessoa [...] (UNIÃO EUROPEIA, 2024, art. 5º, n. 1, alínea b).

A prática de pontuação social (Social Scoring) é rechaçada pelo legislador, o que pode consistir em ponto de aproximação com o princípio do reconhecimento mútuo ricoeuriano. O scoring, ao reduzir a complexidade da vida humana a um número, nega a identidade narrativa do sujeito. O regulamento veda o uso de IA para classificar cidadãos com base em métricas comportamentais:

A colocação em serviço ou a utilização [...] de sistemas de IA para a avaliação ou classificação de pessoas singulares [...] com base no seu comportamento social ou em características pessoais conhecidas ou previstas, em que essa pontuação social conduza a um ou a ambos os seguintes resultados: i) Tratamento prejudicial ou desfavorável [...] (UNIÃO EUROPEIA, 2024, art. 5º, n. 1, alínea c).

A classificação de sistemas como sendo de "alto risco" exige uma governança robusta em áreas que definem o acesso a bens sociais essenciais, como educação e trabalho. Ao impor rigor extra nessas áreas, a lei busca garantir que o risco algorítmico não seja distribuído injustamente, prejudicando o acesso dos cidadãos aos bens fundamentais da vida em comum. O Artigo 6º estabelece o critério para essa classificação:

Independentemente de o sistema de IA ser colocado no mercado ou colocado em serviço independentemente dos produtos [...] considera-se que um sistema de IA é de risco elevado se: [...] o sistema de IA se destinar a ser utilizado como componente de segurança de um produto, ou for, ele próprio, um produto, abrangido pelos atos legislativos da União [...] (UNIÃO EUROPEIA, 2024, art. 6º, n. 1).

Como parte integrante dessa mitigação de riscos estruturais, a qualidade e a integridade dos dados são tratadas como requisitos essenciais para evitar discriminação em decisões automatizadas. Isto significa, em termos hermenêuticos, trazer à luz o fato de que as instituições devem garantir a "memória" e a "identidade narrativa". O Artigo 10º estipula critérios rigorosos para os conjuntos de dados:

Os conjuntos de dados de treino, de validação e de teste devem ser pertinentes, suficientemente representativos e, na medida do possível, isentos de erros e completos. Devem ter as propriedades estatísticas adequadas, inclusive, se for caso disso, no que diz respeito às pessoas ou grupos de pessoas em relação aos quais o sistema de IA de risco elevado se destina a ser utilizado (UNIÃO EUROPEIA, 2024, art. 10º, n. 3).

A questão da obrigatoriedade de supervisão humana nos sistemas de alto risco recorda a primazia da Sabedoria Prática (*phronesis*) sobre o sistema de IA. O Artigo 14º determina que a tecnologia não pode operar sem a vigilância da prudência humana:

Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de modo a poderem ser eficazmente supervisionados por pessoas singulares durante o período em que o sistema de IA está a ser utilizado, nomeadamente através de interfaces humano-máquina adequadas (UNIÃO EUROPEIA, 2024, art. 14º, n. 1).

Tem-se um dispositivo que contempla, talvez, a esfera ética: a instituição deve "distanciar-se" da eficiência técnica imediata para refletir sobre suas consequências no "outro". É, de algum modo, uma institucionalização da responsabilidade, onde o agente (implementador) deve antecipar os danos à dignidade antes que eles ocorram. Nesse caso, um aspecto interessante do regulamento é a exigência de uma avaliação prévia sobre como a tecnologia afetará os direitos fundamentais, obrigando as instituições a uma pausa reflexiva. O Artigo 27º instrumentaliza essa precaução:

Antes da colocação em serviço de um sistema de IA de risco elevado referido no artigo 6.o, n.o 2, os responsáveis pela implantação que sejam organismos de direito público ou operadores privados [...] devem realizar uma avaliação do impacto nos direitos fundamentais que a utilização desse sistema poderá produzir (UNIÃO EUROPEIA, 2024, art. 27º, n. 1).

A interação entre humanos e máquinas impõe novos deveres de transparência para preservar a autenticidade das relações humanas. A dissimulação da natureza artificial do interlocutor torna realidade a corrosão do espaço público, onde a promessa e o testemunho dependem da presença de um sujeito ético capaz de se comprometer, algo arriscado para uma simulação feita por um serviço de IA. O Artigo 50º exige a revelação da natureza artificial do interlocutor:

Os fornecedores de sistemas de IA destinados a interagir diretamente com pessoas singulares devem conceber e desenvolver tais sistemas de tal modo que as pessoas singulares em causa sejam informadas de que estão a interagir com um sistema de IA, a menos que tal resulte evidente das circunstâncias e do contexto de utilização (UNIÃO EUROPEIA, 2024, art. 50º, n. 1).

A proliferação de conteúdo sintético ameaça a distinção entre realidade e simulações, exigindo marcações claras para proteger a "verdade factual" necessária ao juízo de valor a ser formado. Essa questão caminha para o que Ricoeur chama de "confiança na palavra dada".

Cabe à instituição justa proteger os meios de prova e a memória. O regulamento determina a rotulagem de conteúdos manipulados:

Os fornecedores de sistemas de IA, incluindo sistemas de IA de finalidade geral, que gerem conteúdos sintéticos de áudio, imagem, vídeo ou texto, devem garantir que as saídas do sistema de IA são marcadas num formato legível por máquina e detetáveis como tendo sido geradas ou manipuladas artificialmente (UNIÃO EUROPEIA, 2024, art. 50º, n. 2).

O regulamento aborda os modelos fundacionais (General Purpose AI) que podem gerar riscos em larga escala, exigindo uma responsabilidade ampliada de seus criadores proporcional ao seu poder. Nesse sentido, criar riscos sistêmicos (que afetam a totalidade) sem responder politicamente e eticamente por eles, compromete a primazia do público sobre o privado. O Artigo 51º estabelece regras para modelos com riscos sistêmicos:

[...] Os fornecedores de modelos de IA de finalidade geral devem: [...] b) Avaliar os modelos, em conformidade com o anexo XI, secção 2, nomeadamente através da realização e documentação de ensaios contraditórios dos modelos, com vista a identificar e atenuar os riscos sistêmicos; c) Avaliar e atenuar os riscos sistêmicos a nível da União, incluindo as suas fontes, que possam resultar do desenvolvimento, da colocação no mercado ou da utilização de modelos de IA de finalidade geral com risco sistémico [...] (UNIÃO EUROPEIA, 2024, art. 55º, n. 1).

Quando a instituição garante o direito à explicação, ela permite que o sujeito compreenda a história de que faz parte, bem como sua identidade. A garantia de que o indivíduo afetado por uma decisão algorítmica tenha o direito a uma explicação permite a consecução da sua dignidade. O Artigo 86º concede este direito fundamental aos afetados:

As pessoas afetadas sujeitas a uma decisão tomada pelo responsável pela implantação com base nas saídas de um sistema de IA de risco elevado [...] que produza efeitos jurídicos ou afete de modo similar e significativo essas pessoas [...] têm o direito de obter do responsável pela implantação explicações claras e significativas sobre o papel do sistema de IA nos processos de tomada de decisão e sobre os principais elementos da decisão tomada (UNIÃO EUROPEIA, 2024, art. 86º, n. 1).

É fundamental ressaltar que a norma positiva, por si só, oferece apenas uma nuance da instituição; para que ela se configure efetivamente como uma "instituição justa" no sentido ricoeuriano, é necessário que sua aplicação não se reduza a uma burocracia de conformidade, mas considere o “reconhecimento mútuo”, impedindo que a assimetria informativa da vigilância algorítmica anule a capacidade de agir e sofrer do homem. A verdadeira justiça algorítmica, portanto, não reside apenas na letra da lei, mas na capacidade das instituições de exercerem a *phronesis* (sabedoria prática) ao interpretar a norma, garantindo que o "outro" não seja dissolvido, mas reconhecido.

O determinismo algorítmico encontra um limite expresso e rigoroso na proibição do policiamento preditivo. A presunção de que o comportamento criminoso futuro de um sujeito pode ser matematicamente deduzido a partir de perfis históricos aniquila a dimensão da "promessa" e a abertura ao devir, que são os verdadeiros fundamentos da imputabilidade moral. Ao interditar sistemas que avaliam a propensão delitiva com base estritamente em traços de personalidade, o legislador europeu tenta rechaçar a condenação prévia e salvaguarda a agência humana, asseverando que a identidade narrativa do indivíduo não pode ser refém de um cálculo fatalista:

A colocação no mercado, a colocação em serviço ou a utilização de sistemas de IA para efetuar avaliações de risco de pessoas singulares a fim de avaliar ou prever a probabilidade de uma pessoa singular cometer uma infração penal, com base unicamente na definição do perfil de uma pessoa singular ou na avaliação dos seus traços de personalidade e características, a não ser que os sistemas de IA sejam utilizados para apoiar a avaliação humana do envolvimento de uma pessoa numa atividade criminosa [...] (UNIÃO EUROPEIA, 2024, art. 5º, n. 1, alínea g).

A supressão do anonimato no espaço público configura outra face incisiva da vigilância algorítmica combatida pelo regulamento. O emprego indiscriminado da identificação biométrica em tempo real transmuta o cidadão comum em um ente perenemente rastreável, erodindo a dimensão da "esfera pública" essencial para a livre manifestação da alteridade. Ciente da desproporcionalidade dessa intrusão sobre a liberdade civil e o direito à não interferência, a norma institui uma vedação geral a essa tecnologia, excetuada apenas em circunstâncias extremas, taxativas e sob estrito controle das instituições democráticas:

A utilização de sistemas de identificação biométrica à distância «em tempo real» em espaços acessíveis ao público para fins de aplicação da lei, a não ser que e na medida em que essa utilização seja estritamente necessária para um dos seguintes objetivos: i) a procura orientada de vítimas específicas de rapto, de tráfico de seres humanos ou de exploração sexual [...] (UNIÃO EUROPEIA, 2024, art. 5º, n. 1, alínea h).

A proteção da interioridade do sujeito desponta, igualmente, como uma barreira institucional. O espaço íntimo das emoções e sentimentos não pode ser convertido em métrica de produtividade ou aptidão disciplinar sem grave violação à dignidade. Nesse sentido, o *AI Act* classifica como prática inaceitável a inferência emocional em ambientes caracterizados por acentuada assimetria, como as relações laborais e educacionais, obstando que a máquina atue como um inquisidor do estado psicológico do indivíduo:

A colocação no mercado, a colocação em serviço ou a utilização de sistemas de IA para inferir as emoções de uma pessoa singular nos locais de trabalho e nos estabelecimentos de ensino, exceto se a utilização do sistema de IA se destinar a fins médicos ou de segurança (UNIÃO EUROPEIA, 2024, art. 5º, n. 1, alínea f).

No domínio da mitigação de riscos, a regulamentação europeia tenta consolidar a premissa de que a justiça algorítmica não opera num vácuo histórico. A fim de evitar que as desigualdades sistêmicas do passado sejam perpetuadas de forma automatizada, a lei autoriza, a título excepcional e rigorosamente instrumental, o tratamento de dados pessoais sensíveis (como origem étnica ou crença) com o propósito exclusivo de detectar e corrigir enviesamentos. Trata-se de uma ferramenta jurídica poderosa para a promoção de uma justiça não apenas formal, mas proativamente restaurativa:

Na medida em que seja estritamente necessário para garantir a monitorização, a deteção e a correção de enviesamentos nos sistemas de IA de risco elevado, os fornecedores desses sistemas podem, a título excecional, tratar as categorias especiais de dados pessoais a que se refere o artigo 9º [...] sob reserva de garantias adequadas para os direitos e liberdades fundamentais das pessoas singulares (UNIÃO EUROPEIA, 2024, art. 10º, n. 5).

Concomitantemente, a confiabilidade técnica dos sistemas não é apartada do dever ético; pelo contrário, constitui o seu próprio substrato material. Um modelo tecnológico suscetível a falhas de código, alucinações de dados ou ataques cibernéticos converte-se rapidamente em um vetor de decisões arbitrárias, comprometendo a segurança jurídica e a integridade da instituição que o opera. Por conseguinte, a norma europeia eleva a resiliência infraestrutural ao patamar de exigência inegociável para qualquer inteligência artificial de alto risco:

Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de modo a atingirem, à luz da sua finalidade prevista, um nível adequado de exatidão, robustez e cibersegurança, e a manterem-se coerentes a esse respeito ao longo do seu ciclo de vida (UNIÃO EUROPEIA, 2024, art. 15º, n. 1).

A articulação entre o fomento à inovação e a salvaguarda social concretiza-se na obrigatoriedade da instituição de espaços de ensaio da regulamentação (os chamados *sandboxes*). Tais ambientes controlados representam um esforço institucional para testar a tecnologia antes de sua efetiva implantação no tecido social. Essa pode refletir, no âmbito legislativo, a prudência exigida pela sabedoria prática (*phronesis*): uma deliberação antecipatória que permite ao regulador compreender as implicações reais, isolando os riscos e protegendo o bem comum frente ao imprevisto:

Os Estados-Membros asseguram que as respetivas autoridades competentes estabelecem, pelo menos, um espaço de ensaio da regulamentação para a IA a nível nacional, que deve estar operacional até 2 de agosto de 2026. Esse espaço de ensaio da regulamentação para a IA proporciona um ambiente controlado que promove a inovação e facilita o desenvolvimento, a testagem e a validação de sistemas de IA inovadores [...] (UNIÃO EUROPEIA, 2024, art. 57º, n. 1).

A hermenêutica ricoeuriana sugere que a exigência de justiça raramente se esgota na formulação abstrata da norma positiva. Ainda que as proibições e salvaguardas insculpidas na legislação representem um anteparo jurídico importante contra a perda da identidade

narrativa. A eficácia deste diploma legal parece depender, em larga medida, da sua aptidão para dialogar e intervir nas dinâmicas reais e assimétricas. O texto normativo, por mais exaustivo que se proponha ser, é projetado sobre uma tessitura na qual a técnica pode tender a operar, muitas vezes na forma um vetor de controle, categorização e exclusão.

Nestes termos, e a fim de evitar uma leitura excessivamente otimista da regulação, torna-se oportuno transitar para o escrutínio dos potenciais efeitos sociais da vigilância algorítmica. Trata-se, afinal, de questionar não apenas os horizontes propostos pelo regulamento, mas também o modo como a identidade permanece no tempo, resiste e, eventualmente, sobrevive diante das conjunturas tecnológicas.

2.1. A vigilância algorítmica e seus efeitos sociais: perspectivas críticas à luz do *EU AI Act*

No âmbito dessa nova realidade está a vigilância algorítmica, uma prática que redefine os limites da liberdade individual e da justiça social. Os algoritmos, ao contrário de instrumentos neutros, incorporam decisões automatizadas que refletem, reforçam e até ampliam assimetrias históricas de poder. O *EU AI Act* surge, nesse contexto, como um marco regulatório ambicioso, cujo objetivo é mitigar os riscos associados à inteligência artificial, sem inviabilizar a inovação tecnológica (Vainzof et al., 2023).

Avançando no escrutínio dessa materialidade normativa, observa-se que, embora o discurso oficial do *EU AI Act* reforce o compromisso com a proteção de direitos fundamentais, uma leitura hermenêutica exige ir além das intenções normativas e adentrar os sentidos mais profundos que o texto legal projeta sobre as práticas sociais. Nesse ponto, a abordagem filosófica de Paul Ricoeur oferece uma lente valiosa para compreender como as instituições — e entre elas, as estruturas regulatórias — podem se orientar por ideais de justiça, ainda que enfrentem dilemas éticos complexos. Ricoeur enfatiza que as instituições justas não são aquelas que apenas obedecem à lei, mas as que operam sob o imperativo da equidade e da consideração pelo outro (Ricoeur, 2007).

A vigilância algorítmica, portanto, precisa ser lida a partir de sua inserção institucional e dos seus impactos concretos sobre a vida das pessoas. Segundo os Comentários ao *EU AI Act*, a legislação europeia distingue sistemas de IA de alto risco daqueles de risco limitado ou mínimo, mas essa classificação, apesar de útil, pode obscurecer os modos sutis pelos quais a vigilância opera (Vainzof et al., 2023). Um sistema que analisa padrões de comportamento em redes sociais, por exemplo, pode não ser formalmente classificado como de alto risco, mas seus efeitos discriminatórios podem ser profundos.

Ao deslocar a análise para a esfera das ramificações sociais da norma, constata-se que é justamente na observação desses efeitos práticos que a hermenêutica de Ricoeur se mostra fecunda. A vigilância, ao ser mediada por algoritmos, produz narrativas sobre os indivíduos, narrativas essas que podem lhes ser impostas sem possibilidade de contestação. A justiça institucional, nesse sentido, exige abertura à interpretação e reconhecimento do outro como sujeito de direitos (Ricoeur, 2006). Essa tensão entre automatismo técnico e justiça interpretativa é central na crítica à vigilância algorítmica.

O *EU AI Act*, ao estabelecer requisitos de transparência e governança para sistemas de IA, busca justamente enfrentar o problema da opacidade algorítmica. No entanto, como apontado por Vainzof et al. (2023), muitos dos dispositivos legais ainda dependem da autorregulação das próprias empresas desenvolvedoras, o que pode comprometer a efetividade do controle. A justiça institucional, para Ricoeur, não se sustenta apenas sobre a conformidade formal, mas exige compromisso ético com a alteridade e com o bem comum.

Além disso, é preciso considerar que a vigilância algorítmica não é homogênea. Ela se manifesta de formas múltiplas: no reconhecimento facial, na pontuação de crédito, na triagem de currículos, entre outras. Cada uma dessas práticas carrega riscos específicos de exclusão e estigmatização. O *EU AI Act* tenta abordar essa diversidade, mas seu foco em sistemas “de alto risco” pode deixar de fora modalidades mais capilares de controle social (Vainzof et al., 2023). Essa lacuna revela o desafio hermenêutico de interpretar não apenas o que a lei diz, mas o que ela deixa de dizer.

Para além das dimensões técnicas, a vigilância algorítmica opera também como um dispositivo simbólico. Ela reorganiza os sentidos de visibilidade, autonomia e reconhecimento social. Sob a perspectiva de Ricoeur, trata-se de uma forma de mediação simbólica que pode tanto promover a justiça quanto reforçar estruturas de dominação (Ricoeur, 2005). Assim, a crítica hermenêutica não se limita à denúncia, mas busca reconstruir possibilidades de justiça nas tramas da técnica.

A questão da responsabilidade é outro ponto crucial. Quem responde pelos erros algorítmicos? O *EU AI Act* aponta para a necessidade de identificar responsáveis em toda a cadeia de desenvolvimento e aplicação da IA, mas, na prática, os limites da responsabilidade jurídica permanecem nebulosos. Ricoeur propõe uma ética da responsabilidade que ultrapassa o mero cumprimento de regras, enfatizando o dever de responder pelo outro em situações de vulnerabilidade (Ricoeur, 2007). Isso implica repensar a engenharia jurídica à luz da fragilidade humana diante da máquina.

No esteio dessa constatação, a inserção das lógicas algorítmicas nas deliberações de Estado e corporativas evidencia uma fragilização latente no campo da intersubjetividade. A matriz ricoeuriana postula que 'a solicitude não se acrescenta de fora à estima de si, mas desdobra a dimensão dialogal implícita nesta última' (Ricoeur, 1991, p. 226). Diante de um aparato que padroniza existências com o fito de otimizar resultados probabilísticos, observa-se o risco de essa dimensão dialogal ser paulatinamente substituída por uma frieza procedimental.

O sistema de inteligência artificial processa um volume massivo de dados, contudo, mostra-se estruturalmente incapaz de reconhecer o outro ou de ponderar as vulnerabilidades ínsitas à condição humana. Na ausência da sabedoria prática (*phronesis*) – virtude inalienável à deliberação ética que permite estimar o 'tu também' como o 'eu mesmo' –, o algoritmo não comunga da experiência existencial. Torna-se claro, portanto, que a preservação da alteridade no interior das instituições justas não pode ser delegada ao fluxo autônomo das predições, exigindo a contínua mediação interpretativa do sujeito perante a eficiência da máquina.

A vigilância, nesse sentido, não é apenas uma prática técnica, mas um fenômeno ético, cuja regulação demanda sensibilidade às diferenças culturais, sociais e epistemológicas (Vainzof et al., 2023). A justiça institucional, para ser efetiva, deve incorporar essa pluralidade.

Outro aspecto relevante diz respeito à governança democrática dos sistemas de IA. O *EU AI Act* prevê mecanismos de supervisão nacional e europeia, mas a participação cidadã ainda é limitada. Sem canais eficazes de deliberação pública, o risco é que a regulação seja capturada por interesses técnicos ou corporativos. A hermenêutica ricoeuriana insiste na mediação entre o singular e o universal, entre a experiência do sujeito e os dispositivos institucionais — um equilíbrio difícil, mas necessário (Ricoeur, 2006).

Essa preocupação com a escuta do outro é essencial para compreender os impactos sociais da vigilância algorítmica. Quando um sistema automatizado nega um benefício, bloqueia um acesso ou rotula uma pessoa como suspeita, ele realiza um ato que precisa ser compreendido não apenas juridicamente, mas também eticamente. A injustiça não se mede apenas pela ilegalidade, mas pelo silenciamento da voz do outro. Nesse ponto, o *EU AI Act* precisa ser constantemente interpretado à luz de suas consequências (Vainzof et al., 2023).

Ademais, a vigilância algorítmica tende a aprofundar desigualdades sociais pré-existentes. Sistemas treinados com dados enviesados podem replicar estigmas raciais, de classe, de gênero. A neutralidade técnica é, muitas vezes, uma ilusão perigosa. A hermenêutica crítica, como propõe Ricoeur, convida ao desvelamento dessas camadas ocultas

de violência simbólica. A justiça institucional, nesse caso, exige uma leitura ética do código (Ricoeur, 2005).

Não basta cumprir exigências legais formais; é preciso cultivar uma cultura de responsabilidade, transparência e respeito à dignidade humana. Como destacam os autores do livro analisado, a ética deve acompanhar o desenvolvimento tecnológico desde a origem, e não apenas como correção posterior (Vainzof et al., 2023).

O uso de IA em contextos de segurança pública é especialmente delicado. A promessa de eficiência não pode suprimir os direitos fundamentais. A vigilância em nome da segurança é um dilema clássico que ganha novos contornos com os algoritmos. Uma leitura hermenêutica precisa considerar os efeitos dessa vigilância sobre as subjetividades e os modos de vida (Ricoeur, 2007).

A interdependência entre técnica e ética, entre direito e filosofia, é o que torna o debate em torno do *EU AI Act* tão relevante. Ele é, ao mesmo tempo, uma tentativa normativa e uma provocação teórica. Seu êxito dependerá da capacidade de construir instituições que saibam escutar os conflitos da vida social e traduzi-los em respostas justas. Isso exige mais do que bons textos legais; requer imaginação ética.

A vigilância algorítmica, embora muitas vezes anunciada como neutra e objetiva, esconde escolhas humanas enraizadas em contextos políticos, culturais e econômicos. O *EU AI Act*, ao tentar domesticar essa lógica técnica, enfrenta o paradoxo de regular algo que é, por natureza, mutável, invisível e incrivelmente veloz. Isso exige não apenas acompanhamento jurídico contínuo, mas um exercício ético persistente de interrogar a quem serve cada inovação (Vainzof et al., 2023).

Nas entrelinhas do desenvolvimento de IA, há um silencioso redesenho do que entendemos por subjetividade. Os algoritmos criam perfis, antecipam comportamentos e moldam preferências, promovendo uma subjetividade previsível e controlada. A vigilância, então, não apenas vê — ela educa, disciplina, conforma. Uma justiça institucional coerente deve se erguer contra esse rebaixamento do humano à condição de dado.

O *EU AI Act* incorpora, em parte, esse cuidado ao propor exigências de explicabilidade e auditabilidade. Contudo, como destacam Vainzof et al. (2023), a forma como essas diretrizes serão operacionalizadas ainda é incerta. O risco é que a explicabilidade se converta em um jargão técnico incompreensível para os usuários comuns, esvaziando a noção de transparência de seu conteúdo democrático.

A hermenêutica da suspeita, tão cara ao pensamento crítico, deve nos alertar sobre os discursos que naturalizam a vigilância como inevitável. A repetição constante de que “quem

não deve, não teme” revela um deslizamento perigoso da lógica do cuidado para a do controle. Ricoeur alerta que a justiça não se sustenta sobre a presunção do erro, mas sobre a confiança e o reconhecimento da alteridade (Ricoeur, 2007).

É preciso, então, transformar a vigilância em diálogo. Ao invés de sistemas que apenas observam e julgam, o desafio é criar mecanismos que escutem e compreendam. Para isso, o *EU AI Act* poderia ser apenas um ponto de partida. O horizonte regulatório deve ser constantemente ampliado pela participação de grupos historicamente marginalizados, que são os mais afetados pela vigilância invisível e injusta (Vainzof et al., 2023).

Outro fator a se considerar é o papel da educação digital crítica. Os cidadãos precisam ser formados não apenas para usar tecnologias, mas para compreendê-las, questioná-las e, quando necessário, resistir a elas. Uma democracia sólida exige uma cultura de desconfiança responsável, capaz de equilibrar inovação e justiça. A hermenêutica nos ensina que interpretar é, também, intervir.

No cruzamento entre o jurídico, o técnico e o ético, o *EU AI Act* pode representar uma oportunidade singular de reinvenção institucional. Mas, para isso, será necessário evitar que ele se converta em um instrumento de legitimação da vigilância disfarçada de governança. Como aponta a obra *Comentários ao EU AI Act*, a vigilância só será verdadeiramente controlada quando a sociedade se apropriar criticamente da inteligência artificial (Vainzof et al., 2023).

Os desafios postos pela vigilância algorítmica são transnacionais. Não há fronteiras digitais para os dados nem para os algoritmos. Por isso, embora o *EU AI Act* seja um marco europeu, seus efeitos e inspirações alcançam o mundo todo. Países da América Latina, como o Brasil, podem se beneficiar dessas reflexões, desde que adaptem as diretrizes às suas realidades sociotécnicas.

No entanto, essa adaptação não deve ser apenas técnica, mas epistemológica. É necessário incluir vozes plurais, saberes locais e experiências diversas no processo de regulação da IA. A justiça institucional, segundo Ricoeur, não pode ser universalista de maneira abstrata — ela deve se concretizar nas relações e conflitos do cotidiano, considerando os contextos específicos (Ricoeur, 2007).

As relações de poder embutidas nos algoritmos não desaparecem por meio da normatização. O poder, como ensina Foucault, se reinventa nas minúcias da técnica, nos códigos e nas lógicas operacionais. A vigilância, portanto, deve ser vigiada. O *EU AI Act* propõe mecanismos de governança, mas a efetividade desses mecanismos dependerá da atuação vigilante da própria sociedade civil (Vainzof et al., 2023).

É nesse sentido que a hermenêutica pode servir como contrapeso: ao trazer à tona as intenções, os silêncios e as consequências não ditas da regulação, ela permite ver além do texto. A vigilância algorítmica deve ser interpretada não apenas como uma política de segurança, mas como uma política de subjetivação. Quem somos nós diante do olhar das máquinas?

Se a justiça institucional quiser fazer frente aos desafios da era algorítmica, precisará se reconfigurar profundamente. Precisar abandonar a pretensão de neutralidade e reconhecer sua função simbólica, política e ética. O *EU AI Act*, enquanto tentativa, é valioso. Mas como realização, ainda deixa muito a desejar — sobretudo no que diz respeito à justiça dos invisíveis.

Os efeitos sociais da vigilância algorítmica também afetam o imaginário coletivo. A confiança nas instituições, o senso de privacidade e a própria ideia de liberdade são corroídas lentamente por tecnologias que sabem antes mesmo que o sujeito deseje. A regulação, portanto, não é só uma questão jurídica, mas uma luta por narrativas (Ricoeur, 2005).

Por fim, é importante compreender que a vigilância não é um destino. Ela é uma construção histórica, política e técnica — e, como tal, pode ser desconstruída. A justiça institucional que queremos não se contenta com conformidade, mas exige transformação. Que o *EU AI Act* seja não apenas um documento, mas o início de uma nova escuta, mais justa, mais humana, mais ética.

A vigilância algorítmica emerge como um fenômeno estrutural caracterizado pela implementação de sistemas automatizados que monitoram, analisam e influenciam comportamentos em escala populacional. Ao operar de forma ubíqua através da extração massiva de dados pessoais, estas ferramentas redefiniram as dinâmicas de controle e gestão nas esferas públicas e privadas (Zuboff, 2019, p. 8).

A resistência a esta arquitetura de invisibilidade esbarra na opacidade estrutural dos próprios sistemas. Como as regras de decisão algorítmica operam numa "caixa-preta" (Pasquale, 2015, p. 3), o indivíduo afetado encontra-se destituído dos meios cognitivos e jurídicos para contestar os vereditos que lhe são impostos, o que corrói a confiança pública nas instituições e anula a dimensão do contraditório.

As corporações e Estados que adotam esses sistemas frequentemente justificam a sua expansão sob o véu da eficiência, da segurança e da personalização. Todavia, como adverte Shoshana Zuboff no seu diagnóstico sobre o capitalismo de vigilância, essa premissa utilitarista mascara uma severa erosão da autonomia e da privacidade do sujeito (Zuboff, 2019, p. 9).

A resistência a esta arquitetura de invisibilidade esbarra na opacidade estrutural dos próprios sistemas. Como as regras de decisão algorítmica operam numa "caixa-preta" (Pasquale, 2015, p. 3), o indivíduo afetado encontra-se destituído dos meios cognitivos e jurídicos para contestar os vereditos que lhe são impostos, o que corrói a confiança pública nas instituições e anula a dimensão do contraditório. Iniciativas como o *EU AI Act* pavimentaram o caminho para que outros ordenamentos também exijam transparência, responsabilidade institucional e o resgate da autodeterminação informativa

Nesse viés, o regulamento não pode operar como um algoritmo fechado. O risco de subsumir a complexidade humana a diretrizes impessoais é a reprodução automatizada de injustiças sob o verniz da neutralidade. A justiça algorítmica ainda está longe de ser refém do exercício contínuo (*phronesis*). Como no caso dos conflitos armados se apropriando do uso de inteligência artificial, ambiguidades normativas sistêmicas se somam a falhas humanas e passam a aniquilar a dignidade do sujeito afetado pelo algoritmo.

Uma vez que o uso indiscriminado para fins bélicos da tecnologia cristaliza assimetrias e preconceitos pretéritos em códigos e dados, o *EU AI Act* adquire maior relevância democrática, uma vez que se propõe a incorporar mecanismos rigorosos de escuta institucional, sensíveis às demandas de grupos historicamente vulnerabilizados. A participação cidadã, longe de se reduzir a uma formalidade simbólica de transparência, ergue-se como uma força de deliberação ativa, operando para que o desenvolvimento tecnológico não perpetue desigualdades.

No âmbito do Judiciário, a proliferação destas tecnologias suscita aporias fundamentais em torno da opacidade. Estudos recentes sobre o cenário brasileiro (Pecego; Lobato; Teixeira, 2024) advertem que a delegação de atos processuais a algoritmos compromete a explicabilidade das decisões, esvaziando garantias constitucionais basilares como o contraditório e a ampla defesa. A opacidade intrínseca a estes modelos — frequentemente operados sob a lógica da "caixa-preta" (Pasquale, 2015) — obsta a auditabilidade dos processos, inviabilizando que o sujeito afetado compreenda e conteste a racionalidade subjacente ao veredito que lhe é imposto.

A justiça interpretativa consolida-se, portanto, na formação interdisciplinar dos agentes reguladores, a qual fomenta sujeitos éticos capazes de sustentar responsabilizações complexas perante as decisões delegadas à Inteligência Artificial. Ao transcender o paradigma da mitigação de risco e o formalismo da mera legalidade, a regulação se aproxima da célebre intenção formulada por Ricoeur: a visada da 'vida boa, com e para os outros, em instituições justas'. Neste cômputo, o *EU AI Act* assume a vocação de instituição mediadora.

Ao pautar a equidade, a norma apresenta o potencial não apenas de refrear o determinismo algorítmico, mas de orientar ativamente o florescimento do bem comum e a salvaguarda incondicional da indentida.

2.2. O *EU AI Act*: avanços e desafios

A matriz de Paul Ricoeur fornece o lastro para que se compreenda a orientação de estruturas de regulação em direção a ideais de justiça. O autor postula que o dever da instituição transcende a obediência à regra de forma; defendendo que o Estado opere sob a égide da equidade para que seja possível o reconhecimento (Ricoeur, 2007, p. 22).

Para que a vigilância algorítmica seja lida com rigor, requer-se a análise da sua inserção de materialidade e dos impactos sobre a condição do sujeito. Os Comentários ao *EU AI Act* recomendam que a legislação diferencie os sistemas por categorias de risco, contudo, alertam que esta taxonomia pode render a encobrir os modos de controle de capilaridade (VAINZOF et al., 2023, p. 58). Um *software* de análise de padrões em redes de interação, ainda que subtraído à classificação de risco de topo, possui capacidade para que se gerem efeitos de discriminação de gravidade severa.

A monitorização de algoritmo produz narrativas à revelia dos indivíduos, a fim de que se lhes imponham perfis sem margem de contestação. O imperativo de justiça de instituição exige que se preserve o outro como sujeito de direitos, de modo a impedir que ocorra a sua redução a objeto de extração de metadados (Ricoeur, 2006, p. 45). Recomenda-se que a tensão entre o automatismo de código e o juízo de interpretação figure como o núcleo da crítica ao modelo de vigilância.

Ao fixar deveres de transparência, o legislador procura que se desconstrua a opacidade do sistema. Todavia, conforme apontam Vainzof et al. (2023, p. 112), a dependência de mecanismos de autorregulação por parte das empresas de desenvolvimento ameaça a eficácia do escrutínio. Na perspectiva de Ricoeur (2005, p. 263), a manutenção do bem comum e o respeito pela alteridade exigem que as instâncias de poder não deleguem o compromisso de ética à lógica de conformidade de mercado.

A multiplicidade de formas da vigilância de código agrava o quadro de análise. A prática materializa-se em sistemas de biometria de face, em modelos de pontuação de crédito ou em lógicas de triagem de trabalho, situações em que cada mecanismo acarreta riscos de exclusão e de estigma. A focalização do *EU AI Act* nos modelos de alto risco arrisca que se

omitam modalidades de controlo de operação furtiva (Vainzof et al., 2023, p. 230). O desafio da exegese consiste em que se interpretem as lacunas de regulação do texto de norma.

No âmago dessa reflexão sobre o poder algorítmico, compreende-se que a máquina de predição atua como dispositivo de ordem de símbolo, a fim de que se reconfigurem os critérios de autonomia e de visibilidade. Sob o prisma de Ricoeur (2005, p. 270), este modelo de mediação pode apresentar o potencial para que se agravem injustiças. A crítica de hermenêutica tem como escopo que se desmontem estas tramas de programação para que se viabilize a reconstrução de espaços de justiça.

Como desdobramento inerente à construção dessas instituições justas, a imputação de responsabilidade por falhas de algoritmo representa outro ponto de ruptura. Ainda que o EU AI Act determine que se identifiquem agentes na cadeia de produção, os limites de culpabilidade de direito permanecem difusos. Ricoeur (2007, p. 58) propõe uma ética de responsabilização centrada no dever de que o indivíduo responda face à vulnerabilidade do outro. A aplicação desta máxima exige que se reveja a engenharia de lei de modo a que se tutele a fragilidade do humano perante o veredicto da máquina.

Um dos obstáculos de monta do regulamento consiste em que se harmonize a dogmática de tecnologia com o imperativo de garantias de cidadania. A vigilância configura um fenómeno de ordem ética, e não apenas um desafio de engenharia. A sua regulação exige que se incorporem sensibilidades de epistemologia, de modo a que se assegure que a pluralidade do pacto de vivência seja contemplada nas decisões de instituição (Vainzof et al., 2023, p. 301).

Com a finalidade dessa pluralidade se efetivar na dinâmica regulatória, um aspecto fundamental a ser destacado é que, no tocante à governança, o *EU AI Act* prevê que se criem entidades de supervisão de Estado, mas o desenho de participação de civis carece de densidade. Na ausência de canais de deliberação, consolida-se o risco de que a norma seja capturada por interesses de corporações. A matriz de Ricoeur (2006, p. 62) propõe que se estabeleçam instâncias de mediação entre a experiência de base do sujeito e o aparelho de instituição para que se evite o isolamento de tecnocracia.

A negação de acesso ou a atribuição de rótulos de suspeição por vias de automação consubstanciam atos de supressão da voz do cidadão. A injustiça de algoritmo não se afere apenas pela violação da lei, mas para que se observe a eliminação do contraditório. O EU AI Act impõe que se proceda à avaliação ininterrupta das consequências destas deliberações sobre as garantias de defesa dos afetados (Vainzof et al., 2023, p. 315).

A reprodução de estigmas de raça e de classe por sistemas treinados em bases de dados de vício atesta o perigo de que se creia no mito da neutralidade de máquina. A hermenêutica convoca que se desvelem as dinâmicas de injustiças de símbolo embutidas nas linhas de programação. O resgate da equidade exige que se submeta o código a uma leitura de ética de consequências;

A eficácia do *EU AI Act* depende de que ocorra uma mutação nas lógicas de operação de entidades de público e de privado. O cumprimento de formalidades de relatório mostra-se insuficiente sem que se adote uma cultura de transparência desde a concepção do software (*ethics by design*), em oposição a práticas de correção de retroatividade (Vainzof et al., 2023, p. 340).

No domínio da segurança de Estado, o uso de inteligência artificial suscita conflitos de gravidade de exceção. A maximização da eficiência de polícia não autoriza que se suprimam garantias de indivíduo. A interpretação do regulamento deve ponderar os efeitos do controle sem interrupção sobre a formação da identidade e sobre as dinâmicas de liberdade civil.

A relevância do debate sobre o *EU AI Act* vincula-se à necessidade de que se edifiquem instituições aptas a decodificar os conflitos de urbe e a convertê-los em sentenças de equidade. O modelo de predição, promovido sob a aura de isenção de matemática, oculta escolhas de poder e de economia.

A formatação de perfis de consumo e a modelação de preferências operam para que se condicione a liberdade de escolha. A vigilância assume a função de disciplina de massas. A defesa do humano contra a sua redução a dados de biometria ou de tráfego constitui a premissa para que se erga qualquer arranjo de justiça de instituição (Ricoeur, 2006, p. 80).

A introdução de deveres de explicabilidade pelo *EU AI Act* procura que se responda a este desafio. Contudo, as diretrizes de execução suscitam dúvidas quanto à possibilidade de que se transponham para a materialidade. O risco reside na conversão da transparência em jargão de programação de ininteligibilidade para o alvo, a fim de que se esvazie o preceito do seu conteúdo de garantia de democracia (Vainzof et al., 2023, p. 380).

A premissa que postula a inevitabilidade da vigilância através do axioma de punição carece de que seja desconstruída. A ruptura da presunção de inocência e a adoção da lógica da suspeição contínua subvertem os alicerces do Estado de Direito. Ricoeur (2007, p. 85) adverte que o pacto de convivência repousa na presunção de confiança.

A imposição de limites à inteligência artificial assume proporções de globo, a fim de que se transcendam as fronteiras de dados. Apesar da sua jurisdição europeia, o *EU AI Act*

projeta efeitos de padrão de norma de exterior (Brussels effect). Ordenamentos de periferia podem extrair subsídios desta arquitetura de lei, condicionados a que se adaptem às suas assimetrias e carências de infraestrutura. A transposição da norma demanda que se incluam vozes de margem, uma vez que a justiça não admite matrizes de universalidade desvinculadas dos conflitos (Ricoeur, 2007, p. 92).

Requerendo uma contínua mediação interpretativa, constata-se que as dinâmicas de poder fundidas no código resistem a que se aprovelem meras leis. A eficácia dos mecanismos de controlo do *EU AI Act* subordina-se a que o corpo civil exerça uma fiscalização sem interrupções (Vainzof et al., 2023, p. 410). A hermenêutica de Ricoeur atua como instrumento de denúncia das consequências não declaradas da regulação, para que se questione a reconfiguração da identidade do sujeito sob o escrutínio de máquina.

O *EU AI Act* representa um passo de fundação, mas a sua completude depende de que se construam instâncias de escuta e de amparo aos vulneráveis. A recusa da predição de matemática como destino inelutável constituem premissas para que se busque ética e o caminho das instituições justas (Ricoeur, 2005, p. 290).

2.3. Uma regulação ética

Diante da complexidade dessas novas tramas de programação, a inserção da inteligência artificial nas infraestruturas de decisão exige que se repense o papel das instituições de regulação. Para que o avanço de algoritmos não resulte em exclusão de margens de população, impõe-se que a procura por eficiência de máquina ceda espaço à construção de balizas de ética e de bem comum (Silva, 2024, p. 45). A hermenêutica de Paul Ricoeur sobre o conceito de justiça de instituição oferece o instrumental para que se converta a regra de técnica em vetor de reconhecimento de indivíduo (Ricoeur, 1995, p. 30).

O autor postula que o ato de justiça exija que se equilibre a norma de universalidade e a vivência de singularidade (Ricoeur, 1992, p. 88). Dado que o sistema de predição opera por matrizes de padrão de massa, consolida-se o risco de que se invisibilizem histórias de minorias. A fim de que a instituição ostente o atributo de justeza, requer-se que ela integre a ética de reciprocidade. O reconhecimento do outro demanda que se instaurem espaços de deliberação para que os alvos de monitorização participem do desenho de políticas de controlo, a fim de que se evite a imposição de ditames de cúpula de tecnocracia.

Embora o *EU AI Act* represente um esforço de pioneirismo para que se ordenem os usos de automação, o diploma carece de mecanismos de materialidade que assegurem a

escuta de grupos de vulnerabilidade (Silva, 2024, p. 62). Para que se evite o abismo entre a abstração de lei e a realidade de território, exige-se que o legislador incorpore sensibilidades de alteridade (Ricoeur, 1990, p. 115). O combate aos vieses de código impõe que a estrutura de Estado estabeleça fóruns de auditoria de sociedade para que a regulação assuma o carácter de dinâmica de adaptação.

A garantia de transparência de código eleva-se à condição de imperativo de ética. Para que a identidade narrativa de sujeito não sofra processos de fragmentação por vias de extração de dados, exige-se que as lógicas de predição possuam atributos de explicabilidade (Ricoeur, 1992, p. 102). O direito à contestação de vereditos de máquina pressupõe que o cidadão compreenda as premissas de cálculo. O cumprimento deste dever por parte de órgãos de regulação promove a legitimação de processos de governança e a afirmação de laços de confiança no pacto civil.

A cadeia de responsabilização exige que se abandone o foco exclusivo na figura de engenheiro de software. A ética de compromisso de Ricoeur (1995, p. 45) dita que o dever de resposta envolva legisladores, operadores e instâncias de auditoria. Adicionalmente, a adoção de sistemas de inteligência nas esferas de Estado requer que se observem preceitos de justiça de distribuição. A fim de que a barreira de acesso à tecnologia não agrave clivagens de renda, impõe-se que a Administração implemente políticas de inclusão de rede (Souza, 2023, p. 78).

O tempo na obra de Ricoeur adverte que a justiça demanda a memória do passado e a projeção de futuro (Ricoeur, 2000, p. 210). No domínio da predição de algoritmo, esta premissa exige que as instituições de regulação estruturem mecanismos de reparação de danos para vítimas de falsos positivos ou de falhas de biometria. O agir de justiça requer que se adotem vias de compensação, a fim de que se impeça a perpetuação de estigmas de raça e de classe através de rotinas de automação (Mendes, 2021, p. 34).

A consolidação de estruturas de justiça face à inteligência artificial demanda que se fomenta a literacia de técnica em todas as instâncias de participação de civis (Moura, 2022, p. 55). A fim de que se promova uma revisão de crítica sem interrupção (Almeida, 2023, p. 90), exige-se que as instituições assumam a função de agentes de mediação de saberes em âmbito de transnacionalidade. As contribuições de Ricoeur (1990, p. 140) sugerem o êxito de regulamentos à capacidade do Estado de ouvir a pluralidade de narrativas, o que indica um certo otimismo no avanço das IAs.

O trajeto sinalizou que reconduzir o distanciamento da lei à narrativa da experiência constitui uma etapa essencial para desvelar como tais diretrizes afetam a materialidade da vida. Caminhando para as Considerações Finais, ganha contornos a tese de que o desvio

estratégico e a contínua mediação institucional representam caminhos promissores para sustentar o homem como centro de reconhecimento, buscando resguardar a sua alteridade contra o reducionismo da técnica.

CONSIDERAÇÕES FINAIS

Ao percorrermos este itinerário analítico e hermenêutico, a presente dissertação propôs-se a investigar a vigilância algorítmica em um contexto filosófico, buscando desvelar as camadas de sentido que operam sob a superfície da eficiência tecnológica. Ao longo do percurso de investigação, constatou-se a insuficiência de uma abordagem estritamente focada na conformidade com a norma puramente positiva frente ao fenômeno e às implicações do avanço das inteligências artificiais. A automação decisória em larga escala torna possível o risco de o código computacional se impor sobre a vida, reduzindo a experiência existencial a métricas matemáticas e previsões opacas.

Perante a possibilidade de invisibilização das singularidades sob métricas preditivas — que ameaçam cristalizar o indivíduo e anular a sua identidade narrativa —, a pesquisa mobilizou a hermenêutica de Paul Ricoeur como alicerce filosófico para resgatar o reconhecimento mútuo e a solicitude como mediações possíveis ao problema. O desenvolvimento deste estudo procurou evidenciar que a técnica, quando desprovida de uma ancoragem ética e de uma constante interpretação, pode operar sob uma lógica de opacidade estrutural. A extração massiva de metadados converte a vivência humana em insumo, promovendo decisões que afetam o acesso a bens sociais e à própria liberdade civil.

Como desdobramento dessa premissa, o primeiro capítulo da pesquisa fundamentou-se na constatação de que a vigilância algorítmica não é tratada aqui como um erro de código a ser solucionado, mas como uma nova forma, uma manifestação que pode afetar o sujeito. Ao percorrermos a fenomenologia hermenêutica, compreendeu-se que o avanço tecnológico exige o "longo desvio" pelos signos para que o indivíduo compreenda a si mesmo. A identidade do sujeito não é uma substância fixa e imutável, mas uma identidade narrativa construída no tempo, tecida na relação constante entre a manutenção de si e a abertura ao devir. Quando os algoritmos classificam indivíduos baseados unicamente em dados pretéritos, eles sugerem uma redução do ser humano à *mesmidade* (o polo *idem*), obscurecendo a sua historicidade. Esta fixação algorítmica afeta a capacidade humana de prometer e de se ressignificar diante do futuro (a *ipseidade*). O sujeito passa a ser lido a partir de um perfil comportamental estabelecido de antemão, o que exige a formulação de novos anteparos institucionais.

A urgência por tais anteparos revelou-se ainda mais nítida ao acionarmos o imperativo metodológico ricoeuriano de que é preciso “explicar mais para compreender melhor”. Ao longo da investigação, constatou-se que não bastava apontar as limitações de uso do sistema;

era estritamente necessário realizar um distanciamento analítico, adentrando a própria estrutura semiótica da programação.

Ao escrutinar essa linguagem de códigos, nota-se que a máquina traduz a pluralidade da vivência humana (a *ipseidade*) em perfis estatísticos engessados (a *mesmidade*). Operando sob uma autonomia baseada puramente em dados de entrada e saída (*inputs* e *outputs*), o algoritmo atua como uma espécie de “caixa-preta” que mimetiza a intencionalidade humana, arranjando signos de forma coerente, mas sem partilhar de uma semântica vivencial. A estrutura algorítmica revela, assim, a sua limitação ontológica ao demonstrar possuir uma lógica e uma sintaxe capazes de 'inferir' padrões, mas estruturalmente incapazes de 'narrar' a historicidade do sujeito. A máquina, uma vez explicada em sua objetividade sintática, erige-se não como um dispositivo matemático neutro, mas como um campo opaco onde se desenham inúmeras implicações filosóficas.

Diante de um aparato que padroniza existências para otimizar resultados, a verdadeira solicitude corre o risco de ser substituída por uma frieza procedimental. O algoritmo opera de forma veloz segundo lógicas probabilísticas, contudo, mostra-se estruturalmente incapaz de reconhecer o outro, de partilhar a dor ou de ponderar as vulnerabilidades ínsitas à condição humana. Ao constatar que a técnica, desprovida de sabedoria prática (*phronesis*), não comunga da experiência existencial, tornou-se claro que a preservação da alteridade não poderia ser abandonada ao fluxo autônomo das predições estatísticas.

Para que o horizonte ético não se esvaziasse na abstração teórica frente a esse fenômeno de implicações tão concretas, o texto estruturou-se sob o postulado de que a categoria de “instituições justas” ocupa uma posição central na ética de Paul Ricoeur e representa o ponto de articulação entre o ideal ético da vida boa e a concretude do mundo social. A ética ricoeuriana demonstra que o anseio por justiça exige a extensão do cuidado ao "cada um", consubstanciando-se em estruturas políticas e jurídicas capazes de buscar a equidade. Foi precisamente essa exigência incontornável por uma estrutura de mediação que nos conduziu ao exame prático da norma europeia no segundo movimento do trabalho.

Ainda, à guisa de considerações finais, o escrutínio do *European Union Artificial Intelligence Act (EU AI Act)* revelou o esforço do legislador europeu em edificar uma estrutura orientada à centralidade do ser humano. A hipótese norteadora da nossa investigação, delineada desde a introdução e corroborada em nossa análise, comprovou uma aproximação presumível da abordagem de Ricoeur com alguns dispositivos deste regulamento pioneiro. Ao direcionar o arcabouço filosófico para essa materialidade da norma, assumiu-se que, em primeiro plano, a hermenêutica de Ricoeur viabiliza a

decomposição dos alicerces do agir de Estado. Desse modo, o *EU AI Act* passou a ser lido não apenas como um compêndio técnico limitador voltado à tecnologia da informação, mas como um horizonte normativo que possibilita uma aproximação com a ideia de instituição que exige interpretação simbólica.

A partir da lente ricoeuriana, observou-se que a análise do AI Act deve iniciar-se pela sua finalidade, que não se restringe à harmonização de mercado. O Artigo 1.º do regulamento declara expressamente a promoção de uma inteligência artificial "centrada no ser humano e de confiança". Este dispositivo inaugural sugere uma aproximação ao esforço ético de afixar que a tecnologia permaneça subordinada ao florescimento da vida. Ao estabelecer a centralidade humana como princípio, a norma europeia instaura um terreno de proteção que pode tangenciar o horizonte ricoeuriano da "visada ética": o anseio pela vida boa, com e para os outros, em instituições justas.

Esta aproximação institucional evidenciou-se, de maneira contundente, na observação do hiato hermenêutico delineado a partir da própria definição legal de "sistema de IA". Conforme analisado ao longo do segundo capítulo, o sistema de inteligência artificial, operando sob uma autonomia lógica e estatística, detém a capacidade de "inferir", mas jamais a de "narrar". A máquina mimetiza a intencionalidade, sugerindo padrões e respostas lógicas, contudo, o faz sem partilhar da substância moral, da historicidade e das situações inerentes à identidade narrativa do sujeito. Desprovido disso, o algoritmo é incapaz de exercer a verdadeira solicitude, a qual exige o reconhecimento do outro como um "si".

Ao constatar esta limitação ontológica do processamento de dados, o regulamento europeu sinaliza um esforço em instituir mecanismos de salvaguarda. A normatização de limites rigorosos às práticas inaceitáveis — como a vedação expressa ao policiamento preditivo fundado em traços de personalidade e a proibição de pontuação social — atesta uma recusa institucional à condenação prévia e à invasão da interioridade. Estas proibições podem obstaculizar a conversão do sujeito em mero perfil de repetição de dados. Deste modo, aproxima-se o texto legal da preocupação ricoeuriana em viabilizar a contínua ressignificação da própria história do indivíduo e a preservação da promessa e do devir. Da mesma forma, a limitação estrita da identificação biométrica em tempo real e a interdição da inferência emocional em ambientes de acentuada assimetria tangenciam a sensibilidade ao cuidado do outro e à não instrumentalização da vulnerabilidade.

Aprofundando a percepção desse esforço regulatório, cumpre ressaltar que os imperativos legais do *EU AI Act* desvencilham-se da sua prescrição literal e possibilitam uma aproximação com os conceitos ricoeurianos de mediação. A exigência legal de supervisão

humana para os sistemas classificados como de alto risco ilustra essa possibilidade de aproximação conceitual. Em termos hermenêuticos, o legislador europeu abre espaço para que a sabedoria prática (*phronesis*) atue como baliza. É o ser humano, e não a máquina, que detém a capacidade de julgar a exceção e de adaptar a regra à contingência do caso concreto. O esforço ético vislumbrado na lei aproxima-se do ideal de manter o sujeito humano na cadeia de deliberação, mitigando o risco de que o "outro" seja meramente processado por uma sintaxe de código.

No mesmo aspecto da proteção da alteridade, a consagração do direito à explicação individual e os deveres de transparência interacional indicam caminhos contundentes no âmbito de uma boa proposição de justiça. A norma impõe que os provedores de inteligência artificial informem inequivocamente ao indivíduo quando este estiver interagindo com uma simulação. Pode-se supor, em algum aspecto, que, com base no que orienta o *EU AI Act*, é mais difícil que os dados simulem uma solicitude artificial sem que o indivíduo seja avisado. Ao resguardar a distinção entre a máquina e o sujeito, a norma pode garantir ao cidadão a compreensão do seu lugar, preservando a sua história.

Quando a lei prevê o direito a obter explicações claras decorrentes de decisões algorítmicas, ela abre a via para que o cidadão compreenda os motivos pelos quais uma decisão foi tomada, permitindo-lhe integrar aquele evento à sua própria identidade narrativa. Essa dinâmica normativa sugere uma consonância com o reconhecimento ricoeuriano de que o homem não pode figurar apenas como paciente das operações tecnológicas, devendo ter a sua capacidade de agência salvaguardada e fomentada pela literacia digital.

Não obstante o inegável avanço da norma europeia na mitigação de riscos, a investigação, ao longo de seu desenvolvimento, evidenciou a linha tênue entre a norma puramente positivada e a complexidade incomensurável da vida. A advertência recai, inevitavelmente, na fragilidade de uma norma positiva isolada em resguardar, por si só, as instituições justas. O texto normativo, por mais bem fundamentado que se apresente, exige uma contínua mediação humana capaz de adaptar regras universais à contingência do caso concreto.

Aprofundando essa constatação sobre o fenômeno tecnológico, a pesquisa destacou que a sociedade contemporânea é atravessada por dinâmicas cada vez mais sofisticadas de monitoramento digital, cujos efeitos sociais demandam uma contínua sabedoria prática em sua interpretação. A inserção das lógicas algorítmicas nas deliberações governamentais e judiciais redefine profundamente fronteiras essenciais como a equidade e a responsabilidade. As classificações de "alto risco" estipuladas pela legislação, embora representem anteparos

úteis, podem obscurecer dinâmicas capilares de exclusão caso a aplicação da norma não exerça uma escuta ativa atenta às fragilidades e narrativas locais.

Como os agentes artificiais operam alimentados por bases de dados pretéritas e são guiados por uma lógica estritamente sintática e probabilística, eles carecem de capacidade de compreensão ética ou empática. A opacidade frequentemente inerente a esses sistemas, mesmo com os deveres de transparência do *EU AI Act*, inviabiliza por vezes a auditabilidade plena de seus veredictos, corroendo as garantias de contestação e mitigando a proteção dos direitos dos cidadãos.

A linguagem jurídica — estruturalmente perpassada por ambiguidades, intencionalidades e historicidade — não pode ser reduzida a uma tradução binária artificial. Assim, é importante ressaltar que, se devidamente orientada, a inteligência artificial possui um enorme potencial para otimizar a esfera pública e mitigar desigualdades operacionais. Contudo, possíveis ambiguidades da lei e assimetrias sociais exigem uma mediação humana ininterrupta, impedindo a conversão da justiça em um árido preditivismo formalista. Nesse sentido, a proposta de mediações adquire acento não como um repúdio à tecnologia, mas como o instrumental necessário para a sua condução humanizada.

Conclui-se, a partir das reflexões tecidas nesta dissertação, que a construção de instituições justas transcende o formalismo das regulações contemporâneas, por mais avançadas que estas se apresentem. O esforço regulatório do *EU AI Act* desponta como um marco inestimável, mas a justiça efetiva exige o reconhecimento ativo do indivíduo, o respeito irrestrito à sua identidade narrativa e o exercício das mediações. Tais elementos não podem ser parametrizados em código de máquina; eles dependem de mediações operadas por agentes capazes de deliberação ética e escuta.

Considerando isso, a hermenêutica consolida-se como a ponte incontornável entre a conjuntura inevitável dos algoritmos e a existência do sujeito que, em tese, está amparado por alguma norma reguladora das IAs. Face a tal cenário, o deslocamento da dimensão puramente dogmática para o plano fenomenológico dos impactos mostra-se contundente. O emprego de inteligência artificial atuando como um dispositivo de modulação do ser humano — seja na segurança pública, nas rotinas de Estado ou na economia de vigilância corporativa — sinaliza repercussões severas sobre a reserva da vida e a história.

A regulação, quando desprovida de uma escuta ativa atenta às demandas do “si mesmo como outro”, carrega o risco latente de perpetuar desigualdades sistêmicas do passado, cristalizando preconceitos em códigos opacos e negando o direito vital de conceber

e vivenciar as identidades narrativas. A aplicação da lei deve ser, portanto, um projeto contínuo de ressignificação da experiência humana.

A salvaguarda do sujeito, portanto, refuta a suficiência estrita da positivação regulatória sobre as IAs. O caso do *EU AI Act* ilustra com nitidez que um regulamento pode, sem dúvida, ser formulado com bons propósitos e sugerir uma louvável aproximação aos anseios éticos; entretanto, a complexidade do tema não deixa de requerer investigações filosóficas rigorosas, conforme se depreende do cotejo entre a aspiração da "vida boa, com e para os outros, em instituições justas" e o dilema perene da regulação das IAs.

A hermenêutica erige-se, assim, como uma instância fundamental e contínua às incursões do atual paradigma das IAs. Ela nos lembra que a técnica deve estar a serviço da compreensão humana, mediada pela linguagem e pela história. Logo, para que a identidade narrativa não sucumba perante a opacidade das bases de dados e a solicitude não caminhe para uma simples operação reducionista, sugere-se que o avanço das inteligências artificiais pode, e deve, ser sincrônico à edificação de instituições justas, sempre amparadas pelo inalienável labor da mediação.

REFERÊNCIAS

- ALMEIDA, C. **Inteligência artificial e os desafios da governança ética**: interfaces entre tecnologia e justiça social. São Paulo: Contexto, 2023.
- ANDERSON, M.; ANDERSON, S. L. **Machine ethics**. Cambridge: Cambridge University Press, 2011.
- ANDERSON, S.; ANDERSON, M. Machine ethics: creating an ethical intelligent agent. **AI Magazine**, v. 28, n. 4, p. 15-26, 2007.
- ARISTÓTELES. **Ética a Nicômaco**. Tradução de J. A. Jorge. São Paulo: Martin Claret, 2010.
- ARKIN, R. **Governing lethal behavior in autonomous robots**. Boca Raton: CRC Press, 2009.
- AWAD, E. et al. Moral machine: perceptions of moral decision making in autonomous vehicles. **Nature**, v. 563, p. 59-64, 2018.
- BALKIN, J. M. The path of robotics law. **California Law Review**, v. 110, p. 45-88, 2020.
- BENJAMIN, R. **Race after technology**: abolitionist tools for the new jim code. Cambridge: Polity Press, 2019.
- BENTHAM, J. **An introduction to the principles of morals and legislation**. Oxford: Clarendon Press, 1789.
- BODDINGTON, P. **Towards a code of ethics for artificial intelligence**. Cham: Springer, 2017.
- BONNEFON, J.-F.; SHARIFF, A.; RAHWAN, I. The moral machine experiment. **Nature**, v. 537, n. 7621, p. 74-78, 2016.
- BORENSTEIN, J.; HERKERT, J. R.; MILLER, K. W. The ethics of autonomous cars. **The Atlantic**, Washington, 2017. Disponível em: <https://www.theatlantic.com/technology/archive/2017/04/the-ethics-of-autonomous-cars/521700/>. Acesso em: 20 jan. 2026.
- BOSTROM, N. **Superintelligence**: paths, dangers, strategies. Oxford: Oxford University Press, 2014.
- BRYSON, J. **Artificial intelligence and the future of humanity**. Cambridge: MIT Press, 2018.
- BUOLAMWINI, J. **Unmasking AI**: my mission to protect what is human in a world of machines. Nova Iorque: Penguin Random House, 2023.

BUOLAMWINI, J.; GEBRU, T. Gender shades: intersectional accuracy disparities in commercial gender classification. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY, 1., 2018, Nova Iorque. **Anais [...]**. Nova Iorque: PMLR, 2018. p. 77-91.

CENTRE FOR DATA ETHICS AND INNOVATION (CDEI). **Audit protocol for algorithmic systems**. Londres: CDEI, 2021. Disponível em: <https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation>. Acesso em: 20 jan. 2026.

CNN BRASIL. **Forças Armadas do Pentágono utilizam inteligência artificial Claude em simulações de campo de batalha**. São Paulo: CNN Brasil, 2026. Disponível em: [insira o link exato da matéria aqui]. Acesso em: 20 jan. 2026.

COECKELBERGH, M. **AI ethics**. Cambridge: MIT Press, 2020.

COMISSÃO EUROPEIA. **Artificial intelligence act**. Bruxelas: Comissão Europeia, 2021. Disponível em: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75184. Acesso em: 20 jan. 2026.

CONGRESSO DOS ESTADOS UNIDOS. **Algorithmic accountability act**. Washington: Congresso dos Estados Unidos, 2019. Disponível em: <https://www.congress.gov/bill/116th-congress/house-bill/2231>. Acesso em: 20 jan. 2026.

COUR DE CASSATION. **Décision sur le statut des chauffeurs de plateformes numériques**. Paris: Cour de Cassation, 2020.

EU HIGH-LEVEL EXPERT GROUP. **Ethics guidelines for trustworthy AI**. Bruxelas: Comissão Europeia, 2019. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Acesso em: 20 jan. 2026.

MENDES, L. S. **Privacidade, proteção de dados e defesa do consumidor**: linhas gerais de um novo direito fundamental. São Paulo: Saraiva, 2021.

MILLER, T. Explanation in artificial intelligence: insights from the social sciences. **Artificial Intelligence**, v. 267, p. 1-38, 2019.

MIT. **Moral machine**. Cambridge: MIT, 2016. Disponível em: <https://moralmachine.mit.edu>. Acesso em: 20 jan. 2026.

MITTELSTADT, B. D. et al. The ethics of algorithms: mapping the debate. **Big Data & Society**, v. 3, n. 2, p. 1-21, 2016.

MONTREAL DECLARATION. **Montreal declaration for responsible AI**. Montreal: Universidade de Montreal, 2018.

MOOR, J. H. The nature, importance, and difficulty of machine ethics. **IEEE Intelligent Systems**, v. 21, n. 4, p. 18-21, 2006.

MOURA, E. **Literacia digital e cidadania: os desafios da participação social na era dos algoritmos**. Rio de Janeiro: Lumen Juris, 2022.

NISSENBAUM, H. **Privacy in context: technology, policy, and the integrity of social life**. Stanford: Stanford University Press, 2010.

NÓBREGA, L. A tríplice constituição da perspectiva ética de Paul Ricoeur. **Synesis**, Petrópolis, v. 4, n. 2, p. 119-132, 2012.

OCDE. **Automation, skills use and training**. Paris: OECD Publishing, 2021.

OHCHR. **Report on the use of artificial intelligence in the surveillance of minorities in China**. Genebra: United Nations Human Rights Council, 2021.

OPENAI. **ChatGPT**. (Versão GPT-4). São Francisco: OpenAI, 2025. Disponível em: <https://chat.openai.com>. Acesso em: 20 jan. 2026.

PARFIT, D. **On what matters**. Oxford: Oxford University Press, 2011.

PASQUALE, F. **The black box society: the secret algorithms that control money and information**. Cambridge: Harvard University Press, 2015.

RICOEUR, P. **A justiça**. São Paulo: WMF Martins Fontes, 2006.

_____. Autobiografia intelectual. In: _____. **Da metafísica à moral**. Lisboa: Instituto Piaget, 1997.

_____. **Caminhos da reconciliação: entre memória e esquecimento**. São Paulo: Loyola, 2005.

_____. **Do texto à ação: ensaios de hermenêutica II**. Tradução de Sofia Castro e Silva. Porto: Rés-Editora, 1989.

_____. **O justo**. Campinas: Editora da Unicamp, 2007.

_____. **O justo**. Tradução de Ivone C. Benedetti. São Paulo: WMF Martins Fontes, 2008.

_____. **O justo 1**. Tradução de Emmanuel Fragoso. Campinas: Papirus, 1995.

_____. **O si mesmo como outro**. São Paulo: Loyola, 1992.

_____. **O si-mesmo como um outro**. Tradução de Lucy Moreira Cesar. Campinas: Papirus, 1991.

_____. **Percurso do reconhecimento**. Tradução de Nicolás Nyimi Campanário. São Paulo: Loyola, 2004.

_____. **Si mesmo como um outro**. Tradução de Daniel L. de Oliveira. Campinas: Editora da Unicamp, 1990.

_____. **Tempo e narrativa** (Tomo I). Tradução de Constança Marcondes Cesar. Campinas: Papirus, 1994.

TAVARES, A. A.; BITENCOURT, C. M.; CRISTÓVAM, J. S. da S. Explicabilidade e contraditório algorítmico nas decisões automatizadas no setor público: as decisões em processos apoiados em algoritmos e o problema da redução da realidade a dados. **Diké - Revista Jurídica**, Ilhéus, v. 23, n. 27, p. 2-32, 2024. Disponível em: <https://periodicos.uesc.br/index.php/dike/article/view/4556>. Acesso em: 20 jan. 2026.

TAYLOR, C. The politics of recognition. In: GUTMANN, A. (Org.). **Multiculturalism: examining the politics of recognition**. Princeton: Princeton University Press, 1994.

TRINDADE, A. S.; MIRANDA, W. S. de. Etologia e corpo: considerações sobre a metamorfose do humano. **Occursus - Revista de Filosofia**, Fortaleza, v. 6, n. 2, p. 28-51, jul./dez. 2021.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024 [...] (Regulamento Inteligência Artificial)**. Jornal Oficial da União Europeia, Luxemburgo, L, 2024/1689, 12 jul. 2024. Disponível em: <http://data.europa.eu/eli/reg/2024/1689/oj>. Acesso em: 20 jan. 2026.

_____. Serviço das Publicações da União Europeia. **Código de Redação Interinstitucional**. Luxemburgo: Serviço das Publicações da União Europeia, 2024. Disponível em: <https://publications.europa.eu/code/pt/pt-000100.htm>. Acesso em: 20 jan. 2026.

VAINZOF, R. et al. (Coord.). **Comentários ao EU AI Act: uma abordagem prática e teórica do Artificial Intelligence Act da União Europeia**. São Paulo: Revista dos Tribunais, 2024.

VALLOR, S. **Technology and the virtues: a philosophical guide to a future worth wanting**. Oxford: Oxford University Press, 2016.

WACHTER, S.; SANDVIG, C.; KROLL, J. A. **Transparency in algorithmic decision making**. The New York Times, Nova Iorque, 2017.

WALLACH, W.; ALLEN, C. **Moral machines: teaching robots right from wrong**. Oxford: Oxford University Press, 2009.