

Agência e Integridade Epistêmica na Escrita Distante

Agency and Epistemic Integrity in Distant Writing

Tiegue Vieira RODRIGUES*^{id}

RESUMO: Neste artigo, examinamos as dimensões epistêmicas e éticas da “*escrita distante*”, uma forma de composição assistida por Inteligência Artificial (IA), no contexto da produção filosófica e acadêmica. Perante o desafio que os Modelos de Linguagem Amplia (LLMs) colocam às noções tradicionais de autoria, o texto desenvolve uma visão positiva para o seu uso responsável. Argumentamos que, em vez de diminuir a agência humana, a escrita distante reconfigura o papel do autor para o de um agente epistêmico curador, cuja responsabilidade intelectual se manifesta através de uma nova forma de trabalho cognitivo. Com base em conceitos da epistemologia da virtude, distinguimos rigorosamente a agência epistêmica intencional do filósofo da mera instrumentalidade epistêmica da máquina. Enquanto o agente humano opera no espaço normativo das razões, respondendo a compromissos epistêmicos como verdade e justificação, os LLMs funcionam como instrumentos que facilitam processos cognitivos sem possuir estados mentais intencionais. Neste contexto, enquadrados a curadoria como um ato cognitivo exigente que envolve três dimensões de conhecimento: o conhecimento processual (saber-como) manifestado no design sofisticado de prompts; o conhecimento avaliativo necessário para a avaliação crítica dos outputs; e o conhecimento proposicional (saber-que) que fundamenta a integração argumentativa coerente. O artigo analisa também os riscos epistêmicos inerentes a esta prática, particularmente a geração de simulacros epistêmicos, textos que exibem a forma superficial do discurso filosófico sem a substância conceitual genuína. Propomos que o cultivo de virtudes intelectuais específicas, como a vigilância epistêmica, a humildade intelectual e o compromisso com a compreensão profunda, é indispensável para a mitigação destes riscos. Por fim, defendemos que a integridade autoral já não pode ser governada exclusivamente por noções tradicionais de plágio como apropriação indevida, mas exige um novo imperativo de transparência, para o qual desenvolvemos uma taxonomia gradual de divulgação ética que reconhece diferentes níveis de assistência tecnológica na produção textual.

PALAVRAS-CHAVE: Escrita Distante. Agência Epistêmica. Curadoria Epistêmica. Inteligência Artificial. Autoria.

ABSTRACT: In this article, we examine the epistemic and ethical dimensions of "distant writing," a form of AI-assisted composition, in the context of philosophical and academic production. Given the challenge that Large Language Models (LLMs) pose to traditional notions of authorship, the text develops a positive vision for their responsible use. We argue that, rather than diminishing human agency, distant writing reconfigures the author's role into

* Doutor em filosofia pela Pontifícia Universidade Católica (PUCRS), Professor Adjunto de Filosofia do Departamento de Filosofia da Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.
tiegue.rodrigues@gmail.com

that of an epistemic curatorial agent, whose intellectual responsibility manifests through a new form of cognitive labor. Drawing on concepts from virtue epistemology, we rigorously distinguish the intentional epistemic agency of the philosopher from the mere epistemic instrumentality of the machine. While the human agent operates in the normative space of reasons, responding to epistemic commitments such as truth and justification, LLMs function as instruments that facilitate cognitive processes without possessing intentional mental states. In this context, we frame curation as a demanding cognitive act involving three dimensions of knowledge: procedural knowledge (know-how) manifested in sophisticated prompt design; evaluative knowledge necessary for critical assessment of outputs; and propositional knowledge (know-that) that grounds coherent argumentative integration. The article also analyzes the epistemic risks inherent in this practice, particularly the generation of epistemic simulacra, texts that exhibit the superficial form of philosophical discourse without genuine conceptual substance. We propose that the cultivation of specific intellectual virtues, such as epistemic vigilance, intellectual humility, and commitment to deep understanding, is indispensable for mitigating these risks. Finally, we argue that authorial integrity can no longer be governed exclusively by traditional notions of plagiarism as misappropriation, but requires a new imperative of transparency, for which we develop a graded taxonomy of ethical disclosure that recognizes different levels of technological assistance in textual production.

KEYWORDS: Distant Writing. Epistemic Agency. Epistemic Curation. Artificial Intelligence. Authorship.

Artigo recebido em: 11.10.2025

Artigo aprovado em: 14.11.2025

1 Introdução

A recente e rápida proliferação de Modelos de Linguagem Ampla (LLMs)¹ antecipa uma reconfiguração significativa nos fundamentos epistêmicos da autoria, desafiando pressupostos estabelecidos sobre criatividade, intencionalidade e trabalho intelectual. Essa transformação, que afeta todas as áreas da produção textual, assume um caráter particularmente provocador no domínio da escrita filosófica, uma prática historicamente associada ao pensamento rigoroso, à clareza conceitual e ao esforço reflexivo individual.

Nesse cenário de transformações, o conceito de “*escrita distante*” (*distant writing*), conforme proposto por Luciano Floridi (2025a), oferece um quadro teórico robusto

¹ Doravante, empregaremos a sigla *LLM* (ou *LLMs*, no plural), derivada do original *Large Language Models*, para abreviar os termos “*Modelo de Linguagem Ampla*” e “*Modelos de Linguagem Ampla*”, respectivamente.

para compreender esse novo modo de produção textual. Diferentemente da “*leitura distante*”, proposta por Franco Moretti (2013), que utiliza ferramentas computacionais para analisar retrospectivamente grandes corpora de textos², a escrita distante é fundamentalmente prospectiva e generativa, aproveitando os LLMs para criar novas obras. Nela, o autor humano transita de um compositor direto para o de um *designer* ou curador, que guia um sistema de IA através da elaboração de *prompts* e do refinamento iterativo. O autor passa, assim, da escrita no texto para a escrita através da máquina, transformando a natureza do ato criativo.

A emergência dessa prática levanta questões epistemológicas urgentes, especialmente no contexto filosófico: se um autor utiliza um LLM para gerar partes de um artigo, quem é o responsável pelo status epistêmico do produto final? Perante essas questões, e reconhecendo que o uso dessa tecnologia parece ter vindo para ficar, o presente texto, ao mesmo tempo em que problematiza a escrita distante, procura construir uma visão positiva para a sua aplicação apropriada no contexto da escrita filosófica e acadêmica³. Ao invés de adotarmos uma postura simplesmente crítica ou proibitiva, defenderemos a tese de que, na escrita filosófica distante, o autor não abre mão da sua agência epistêmica, mas antes a *reconfigura*, assumindo o papel crucial de um agente epistêmico curador⁴. A agência epistêmica, conforme pensada a partir de

² O termo “*leitura distante*,” proposto por Moretti (2013), não se refere a uma leitura desatenta, mas a uma abordagem metodológica das humanidades digitais que utiliza métodos computacionais (como a modelagem de tópicos, a análise de redes e a mineração de dados) para analisar vastos arquivos literários. O objetivo é identificar padrões de larga escala, tendências e estruturas que não seriam visíveis através da leitura atenta e individual de textos (*close reading*). Para uma exploração aprofundada, ver Moretti (2013), *Distant Reading*.

³ Neste artigo, focamos apenas em dimensões epistêmicas e éticas da autoria. Para uma discussão mais ampla sobre as implicações estéticas e o debate sobre se a IA pode ser genuinamente “criativa”, ver a obra seminal de Boden (2004) sobre a criatividade, bem como análises mais recentes sobre a arte gerada por IA, como a de Zylinska (2020).

⁴ A escolha do termo “curador” em vez de “editor” ou “supervisor” é deliberada. Enquanto um editor tipicamente refina um trabalho já concebido por um autor, e um supervisor orienta o trabalho de outro agente, o curador (no sentido museológico) seleciona, contextualiza e organiza elementos para criar uma nova exposição com um significado próprio. O termo “curador” capta melhor a agência de alto nível do autor na escrita distante, que envolve não apenas a correção, mas a seleção estratégica e a construção de um todo coerente a partir de materiais gerados externamente.

uma perspectiva epistêmica, refere-se à capacidade de um sujeito ser um convededor ativo e responsável pelas suas crenças e afirmações (Zagzebski, 1996). Argumentamos que o uso de LLMs na filosofia não elimina esse tipo de agência, mas desloca o seu exercício para um tipo de curadoria, um processo que não é um ato passivo de seleção, mas uma atividade epistêmica cognitivamente exigente.

Para desenvolver essa tese construtiva, o artigo irá, em primeiro lugar, desvendar a arquitetura epistêmica da escrita filosófica distante, distinguindo a agência epistêmica do autor da instrumentalidade epistêmica do LLM. Em seguida, analisará a divisão do trabalho cognitivo e os riscos inerentes a essa prática, como a geração de *simulacros epistêmicos*, textos que simulam a forma do conhecimento sem a sua substância⁵. Argumentar-se-á que virtudes intelectuais como a vigilância, a humildade epistêmica e a autorreflexão crítica são indispensáveis para mitigar esses riscos (Battaglia, 2022; Kidd; Battaglia, 2023). Por fim, o artigo abordará as implicações para a autoria, a originalidade e a integridade acadêmica na filosofia, propondo que a transparência sobre a mediação da IA se torne uma norma de responsabilidade intelectual. O objetivo final é fornecer um quadro teórico que preserve a centralidade da agência epistêmica, oferecendo um caminho para o uso ético e produtivo dessa nova tecnologia.

2 A Arquitetura Epistêmica da Escrita Filosófica Distante

Para que seja possível acessar adequadamente as implicações que surgem a partir do uso da escrita distante na filosofia, é preciso desvendar a sua arquitetura epistêmica, ou seja, a estrutura subjacente de conhecimento, agência e justificação que a governa. Esse novo modo de produção textual não representa uma simples

⁵ O conceito de “simulacros epistêmicos” inspira-se na teoria de Jean Baudrillard (1994), que descreve a progressão dos simulacros em quatro estágios, culminando na “simulação pura”, em que a imagem não tem qualquer relação com uma realidade subjacente, tornando-se a sua própria realidade. Um “simulacro epistêmico” funciona de forma análoga: é um texto que imita perfeitamente a forma do conhecimento (argumentos, provas, citações) sem ter por base a substância do conhecimento (compreensão, justificação, verdade), tornando-se uma simulação de um discurso acadêmico.

atualização das ferramentas do filósofo, como se trocasse a máquina de escrever por um computador. Pelo contrário, ela fundamentalmente reconfigura a distribuição do trabalho cognitivo, obrigando-nos a dividir, de maneira clara, a contribuição do agente humano e a do sistema artificial. Pode-se notar que, no cerne dessa arquitetura, encontramos uma distinção filosófica crucial: a diferença entre a agência epistêmica, que permanece firmemente com o autor humano, e a instrumentalidade epistêmica, que caracteriza a função do LLM. Ignorar essa distinção leva a uma compreensão equivocada da autoria e da responsabilidade, enquanto a sua elucidação revela a natureza exigente do novo papel do filósofo como um curador epistêmico.

2.1 Agência Epistêmica *versus* Instrumentalidade Epistêmica

Quando consideramos a tradição analítica da epistemologia, especialmente no tocante à epistemologia da virtude⁶, identificamos que o termo “agência epistêmica” é definido como a capacidade de um agente não apenas de possuir crenças, mas de formá-las, avaliá-las e assumir responsabilidade por elas de maneira ativa e racional (Greco, 2010; Zagzebski, 1996). Nesse sentido, um agente epistêmico é aquele capaz de deliberar, responder a razões e justificar as suas afirmações. Trata-se alguém que se posiciona no espaço normativo do conhecimento, sendo digno de crédito pelas suas crenças verdadeiras e responsável pelos seus erros. Essa concepção de agência necessariamente implica intencionalidade, consciência e uma capacidade para a avaliação reflexiva, atributos que normalmente são tomados como sendo constitutivos da prática filosófica. O filósofo não apenas afirma uma tese, mas compromete-se com a sua defesa, antecipa objeções e situa-a dentro de uma tradição de pensamento.

⁶ A epistemologia da virtude é uma área ampla que se pode dividir, grosso modo, em duas vertentes principais. A primeira, por exemplo, o confiabilismo da virtude (associado a teóricos como Ernest Sosa e John Greco), foca nas faculdades cognitivas fiáveis (como a visão ou a memória) como “virtudes”. A segunda, o responsabilismo da virtude (associado a teóricas como Linda Zagzebski), foca nos traços de caráter intelectual (como a mente aberta, a curiosidade e a coragem intelectual) como as virtudes centrais. Neste texto, assumimos principalmente a segunda vertente, ao enfatizar o caráter e as disposições do agente. Para uma visão geral, ver Baehr (2011).

Os LLMs, por outro lado, operam inteiramente fora desse espaço normativo. Conforme argumentado por Floridi (2025b), os sistemas de IA podem ser vistos como uma forma de agência sem inteligência (agência não-epistêmica), capazes de executar tarefas complexas de forma autônoma, mas desprovidos de compreensão, consciência ou estados mentais intencionais. O seu funcionamento não é guiado pelo raciocínio inferencial, mas pela otimização matemática e pela previsão da próxima palavra com base em correlações estatísticas identificadas em vastos conjuntos de dados de treino (Buckner, 2019). Consequentemente, um LLM não sabe ou crê em nada no sentido proposicional. Quando gera um texto, não está expressando uma crença justificada, mas sim compondo uma sequência linguisticamente plausível; análoga a um “papagaio” que pode imitar a fala humana sem qualquer compreensão do seu significado (Bender *et al.*, 2021).

Nesse sentido, o LLM é, na melhor das hipóteses, um instrumento epistêmico. Podemos considerar como um instrumento epistêmico algo que se constitui enquanto uma ferramenta capaz de facilitar ou expandir as nossas capacidades cognitivas, como um telescópio, um microscópio ou uma calculadora⁷. Essas ferramentas claramente têm um papel causal na aquisição de conhecimento, mas ninguém estaria disposto a atribuir agência epistêmica ao telescópio pela descoberta de uma nova lua, ou à calculadora pela solução de uma determinada equação. A responsabilidade e o crédito pertencem ao astrônomo e ao matemático que os utilizam. Os LLMs são, sem dúvida, instrumentos de uma sofisticação sem precedentes, cuja capacidade geradora ultrapassa, em muito, a de ferramentas anteriores. No entanto, do ponto de vista epistêmico, ele permanece na categoria de instrumento. A sua complexidade não lhe

⁷ A noção de “instrumento epistêmico” aqui discutida correlaciona-se com a tese da mente estendida de Clark e Chalmers (1998). Nessa perspectiva, um objeto externo (como um caderno ou um LLM) pode tornar-se parte constitutiva de um sistema cognitivo, funcionando como uma extensão da mente do agente. É crucial notar, no entanto, que mesmo como parte de uma “mente estendida”, o objeto não adquire agência. Ele permanece um instrumento, e a agência epistêmica de todo o sistema cognitivo híbrido continua a ser exclusivamente do utilizador humano.

confere agência; apenas amplia o escopo das tarefas que podem ser delegadas pelo agente humano.

Uma analogia pode ser traçada com a crítica de Platão à escrita no diálogo *Fedro*. Sócrates argumenta que a palavra escrita é como uma pintura: parece viva, mas se lhe fizermos uma pergunta, ela mantém um silêncio majestoso. O texto escrito não pode se defender, clarificar as suas ambiguidades ou se adaptar a diferentes interlocutores; é um instrumento inerte, desprovido da alma viva do discurso dialético. O LLM é, em muitos aspectos, a manifestação última desse texto platônico. Ele pode simular uma resposta, gerando mais texto em reação a um novo *prompt*, mas, tal como o texto de Platão, não possui uma compreensão genuína para se “defender” com base em razões. A sua resposta não nasce de uma reflexão sobre o significado, mas de um novo cálculo probabilístico.

Assim como a escrita, para Platão, era um instrumento que ameaçava a memória e o verdadeiro conhecimento (a dialética), os LLMs são instrumentos que, se mal compreendidos, ameaçam a atribuição correta da agência e da responsabilidade epistêmica. Elas não são um interlocutor no diálogo filosófico, mas sim um sofisticado autômato que organiza símbolos, ecoando o famoso argumento do *Quarto Chinês* de Searle (1980)⁸, segundo o qual a manipulação sintática de símbolos não constitui, por si só, uma compreensão semântica.

Considere um exemplo prático: um filósofo que utiliza um LLM com o *prompt*: “Escreva um diálogo no estilo de Platão em que Sócrates discute a ética da inteligência artificial com um jovem programador chamado Technos”. O modelo pode gerar um texto impressionante, imitando a estrutura maiêutica, o tom interrogativo de Sócrates

⁸ O Argumento do Quarto Chinês é uma experiência mental proposta por John Searle (1980) para contestar a tese da IA forte, segundo a qual um computador devidamente programado pode ter uma mente e compreensão genuínas. Searle imagina-se dentro de um quarto, a manipular símbolos chineses seguindo um livro de regras em inglês, de tal forma que, para um observador externo, parece que ele comprehende chinês. No entanto Searle argumenta que ele não comprehende uma única palavra de chinês. A sua conclusão é que a manipulação puramente sintática de símbolos não é suficiente para a semântica, ou seja, para a compreensão. Para uma discussão aprofundada do argumento e das suas numerosas críticas, ver a entrada da *Stanford Encyclopedia of Philosophy* sobre o tema (Cole, 2020).

e a deferência do seu interlocutor. Pode até articular argumentos que parecem coerentes, como a preocupação de que a IA, ao focar na eficiência, possa negligenciar a virtude. No entanto o LLM não comprehende o que é a ética, a virtude, o método socrático ou a própria IA. O texto é uma colagem estilística e temática baseada em padrões. A coerência é aparente, a profundidade é simulada. O valor filosófico do texto não reside no resultado bruto da máquina, mas sim no que o filósofo (agente) faz a seguir: a sua avaliação, a sua crítica, a sua decisão de o descartar, refinar ou integrar no seu próprio argumento. A máquina fornece o mármore sintático; o filósofo, como agente epistêmico, é o escultor que lhe impõe (ou não) uma forma conceitual com significado e validade.

2.2 O Trabalho Epistêmico da Curadoria

Se o LLM é um instrumento, então o trabalho do filósofo na escrita distante não pode ser reduzido à simples operação de uma ferramenta. Em vez disso, a agência epistêmica do autor se manifesta através de um exigente trabalho de curadoria. Esse termo, emprestado do mundo da arte, é aqui redimensionado para descrever um processo intelectual multifacetado que vai muito além da mera seleção. O curador de um museu não se limita a escolher obras; ele as interpreta, contextualiza e organiza para construir uma narrativa coerente que gera conhecimento e provoca uma resposta informada no espectador (Elgin, 2017). De forma análoga, o escritor-curador (filósofo) não apenas seleciona trechos de texto gerados pela IA, mas orquestraativamente todo o processo dialógico com a máquina para construir um argumento filosófico robusto e original. Esse trabalho curatorial é epistemicamente generativo e pode ser decomposto em, pelo menos, três tipos de conhecimento interdependentes, que o autor humano deve possuir e aplicar.

O primeiro é o conhecimento processual, ou o “saber-como” (Ryle, 1949). Refere-se à competência prática de interagir eficazmente com o LLM. Na filosofia, isso se traduz na arte de construir *prompts* que sejam conceitualmente precisos e

filosoficamente férteis. Um *prompt* vago como “fale sobre o livre-arbítrio” produzirá um resumo genérico e inútil. Um *prompt* de um especialista, como “Desenvolva um argumento compatibilista que responda à *Objeção da Manipulação* de Derek Pereboom, utilizando uma analogia com a programação de um agente autônomo em IA”, é uma forma de *design* epistêmico. Exige um conhecimento profundo do debate para guiar a máquina para um espaço latente de possibilidades textuais que seja relevante.

O segundo é o que podemos chamar de *conhecimento avaliativo*, a capacidade de julgar a qualidade, a coerência e a validade do resultado gerado. É aqui que a perícia filosófica é mais crítica. Um LLM pode produzir um parágrafo sobre o imperativo categórico de Kant que é gramaticalmente perfeito e usa a terminologia correta. No entanto apenas o agente-curador pode avaliar: a formulação gerada do imperativo categórico é uma representação fiel ou uma simplificação enganadora? O exemplo fornecido pela IA para ilustrar o conceito (por exemplo, sobre uma promessa mentirosa) captura a essência do teste de universalização ou falha sutilmente? O texto distingue adequadamente entre deveres perfeitos e imperfeitos? Esses são julgamentos que exigem uma sensibilidade filosófica apurada, uma capacidade de detectar nuances, ambiguidades e erros conceituais que a máquina, focada na plausibilidade superficial, é incapaz de reconhecer (Grasso, 2023).

O terceiro, e mais fundamental, é o conhecimento semântico ou proposicional, o “*saber-que*”. Trata-se do corpo de conhecimento filosófico que o autor possui sobre o tema em questão. Na escrita tradicional, esse conhecimento é a fonte direta do conteúdo. Na escrita distante, funciona como o critério final de verificação e o alicerce para o trabalho de curadoria. O autor humano continua sendo o portador da responsabilidade pela veracidade das afirmações, pela solidez das inferências e pela integridade das fontes (dado que os LLMs são notoriamente conhecidos por “alucinar” ou fabricar) (Strzelecki, 2023). Sem esse conhecimento prévio, o autor se torna um mero transcritor de uma caixa-preta, correndo o risco de propagar falsidades ou absurdos sob um verniz de eloquência, uma crítica que ecoa o conceito de “*bullshit*”, de

Frankfurt (2005), em que as afirmações são produzidas sem qualquer consideração pela sua verdade.

Uma analogia útil para capturar a dinâmica desse processo é a de um compositor que trabalha com um pianista improvisador genial, mas tematicamente amnésico. O compositor (o filósofo) estabelece o tema, a estrutura harmônica e o objetivo emocional da peça. Ele pode pedir ao pianista (o LLM) para improvisar uma variação sobre um motivo específico. O pianista, com a sua técnica prodigiosa, pode produzir uma cascata de melodias e harmonias complexas e belas. No entanto ele não tem memória do tema inicial nem uma visão global da composição. É o compositor que tem de ouvir atentamente, selecionar os fragmentos que servem a sua visão, rejeitar os que se desviam e tecer os elementos escolhidos numa obra coesa e com significado. O ato criativo reside tanto na instrução inicial como, e talvez mais importante, no julgamento avaliativo e integrador do compositor. A agência criativa e epistêmica é sua. O LLM, tal como o pianista amnésico, é um poderoso gerador de variações, mas é o escritor-curador que lhe confere direção e propósito, exercendo o que pode ser chamado de metacontrole epistêmico. Esse processo iterativo de instruir, gerar, avaliar e refinar reflete um ciclo de calibração epistêmica, em que o agente humano ajusta continuamente a sua interação com o instrumento para garantir que os resultados se alinhem com os seus padrões de verdade e coerência (Pritchard, 2016). Em suma, a arquitetura epistêmica da escrita filosófica distante não marginaliza o autor humano; pelo contrário, exige que ele opere em um nível mais elevado de supervisão, exercendo a sua agência não através da produção direta de cada palavra, mas através da orquestração e curadoria crítica de todo o processo de criação de significado.

3. A Divisão do Trabalho Cognitivo e os Riscos Epistêmicos

A caracterização da escrita distante como uma interação entre um agente epistêmico humano e um instrumento epistêmico artificial estabelece as bases para

uma análise mais profunda da sua dinâmica operacional. Essa interação constitui uma nova forma de divisão do trabalho cognitivo, um processo em que as tarefas de geração de ideias, articulação linguística e refinamento argumentativo são distribuídas através dos limites do cérebro humano e do sistema computacional. Embora essa colaboração humano-máquina possa desbloquear uma notável eficiência e criatividade, ela é intrinsecamente carregada de vulnerabilidades. A fluidez e a capacidade de simulação do LLM criam um terreno fértil para riscos epistêmicos sutis e, por vezes, perigosos.

Navegar nesse novo ecossistema exige, portanto, não apenas competência técnica, mas uma consciência aguçada das suas armadilhas e a mobilização de um conjunto robusto de virtudes intelectuais para salvaguardar a integridade do trabalho filosófico.

3.1 A Natureza Distribuída da Escrita Filosófica com IA

A interação entre um filósofo e um LLM, na prática da escrita distante, inaugura uma nova e complexa forma de cognição distribuída. O conceito de “mente estendida”, introduzido originalmente por Andy Clark e David Chalmers (1998), desafia a visão tradicional segundo a qual a cognição seria um processo que ocorreria única e exclusivamente dentro dos limites do cérebro e do corpo humano. Para melhor compreender essa noção de “mente estendida”, considere um exemplo. Imagine que Otto, um homem com Alzheimer, usa um caderno para registrar informações importantes sobre sua vida cotidiana. Em sua visita ao museu, Otto faz questão de anotar em seu caderno o nome do seu guia e de todas as obras com as quais teve contato, para que não corra o risco de esquecê-las. Clark e Chalmers argumentam que o caderno, nesse caso, não é apenas uma ferramenta externa, mas parte do sistema cognitivo de Otto; o processo de consultar o caderno desempenha a mesma função que

a memória biológica, conforme o “*Princípio de Paridade*”⁹. O sistema cognitivo relevante não é apenas “Otto”, mas a unidade acoplada e funcionalmente integrada “Otto-caderno”.

A escrita distante opera segundo um princípio semelhante, mas com uma ordem de magnitude de complexidade e interatividade muito superior. O LLM não é um repositório passivo como o caderno de Otto; ele é um parceiro dinâmico e gerador no processo de pensamento. Ele funciona como um *andaime epistêmico*. “Andaime epistêmico” é um termo definido como uma estrutura externa ao agente cognitivo que permite ao agente a realização de tarefas cognitivas que, de outro modo, seriam muito demandantes ou até impraticáveis (Clark; Chalmers, 1998). Imaginemos um filósofo trabalhando sobre a relação entre a fenomenologia de Husserl e a de Merleau-Ponty. Tradicionalmente, esse trabalho exigiria horas de leitura e síntese para mapear os pontos de convergência e divergência. Utilizando um LLM de forma estratégica, o filósofo pode delegar a tarefa inicial de resumir as posições de cada autor sobre um conceito específico, como a intencionalidade ou o *mundo da vida*. O LLM pode gerar rapidamente resumos estruturados que servem de base, o andaime, para a análise. Isso não substitui o pensamento filosófico; pelo contrário, liberta recursos cognitivos. Ao externalizar o trabalho mais laborioso de recolha e organização inicial, o filósofo pode concentrar a sua energia mental nas tarefas de ordem superior: a análise crítica, a descoberta de uma nova síntese, a formulação de uma objeção original. O LLM torna-se parte de um ciclo de *feedback* cognitivo, em que o autor solicita, avalia o andaime gerado, refina a sua própria compreensão e formula novas questões, num processo iterativo de construção de conhecimento.

⁹ O “*Princípio da Paridade*”, formulado por Clark e Chalmers (1998), é o critério central para determinar se um processo externo conta como parte de um sistema cognitivo. Segundo esse princípio, se uma parte do mundo funciona como um processo que, se fosse realizado na cabeça, não hesitaríamos em reconhecer como parte do processo cognitivo, então essa parte do mundo é (assim o afirmamos) parte do processo cognitivo. É esse princípio que nos permite argumentar que o caderno de Otto não é apenas uma ferramenta, mas uma extensão da sua memória.

Essa dinâmica estabelece uma clara, ainda que inédita, divisão do trabalho cognitivo. As tarefas são distribuídas de acordo com a natureza da capacidade de cada componente do sistema híbrido. Ao LLM (o instrumento) são delegadas as funções que dependem de processamento de dados em larga escala, reconhecimento de padrões e geração sintática, tais como: o rascunho inicial de seções de texto, a exploração de diferentes formulações estilísticas para aumentar a clareza, a sumarização de literatura acadêmica relevante ou a articulação de objeções e contraexemplos padronizados. No entanto as funções que exigem agência epistêmica genuína permanecem como domínio exclusivo do filósofo humano. Tais tarefas incluem: a definição do propósito e da tese do argumento (o *telos* da investigação), a formulação de hipóteses e *prompts* criativos, a avaliação da solidez e validade lógica da relevância filosófica do conteúdo gerado, o julgamento sobre a nuance e a precisão conceitual, a verificação da veracidade das fontes e, de forma mais crucial, a assunção de total responsabilidade intelectual pelo produto final.

A divisão do trabalho encontra um paralelo esclarecedor na forma como a “Big Science” opera¹⁰. Considere um físico de partículas no CERN (a Organização Europeia para a Pesquisa Nuclear) analisando os dados de uma colisão no *Grande Colisor de Hádrons* (LHC). Nenhum cientista comprehende, por si só, cada detalhe dos milhões de componentes do detector ou do *software* que processa os *peta bytes* de dados. O conhecimento é necessariamente distribuído por uma vasta colaboração de especialistas e instrumentos computacionais complexos. O papel do cientista principal não é realizar os cálculos, mas sim projetar a experiência, interpretar os resultados fornecidos pelos computadores à luz da teoria e, finalmente, construir um argumento que justifique uma descoberta.

¹⁰ A análise da ciência como uma forma de conhecimento distribuído é um tópico central na sociologia e filosofia da ciência. Autores como Peter Galison, em obras como *Image and Logic* (1997), exploraram como o conhecimento em campos como a física de partículas é produzido através de vastas colaborações entre teóricos, experimentalistas e complexos aparatos técnicos, em que nenhum indivíduo detém todo o conhecimento do sistema. Essa perspectiva oferece um modelo útil para pensar sobre a colaboração humano-IA na escrita.

Da mesma forma, o filósofo na escrita distante pode não compreender os mecanismos internos do LLM, mas a sua perícia reside em saber como “operar” a ferramenta de forma a produzir resultados significativos e, mais importante, em saber como interpretar, validar e contextualizar esses resultados dentro de um quadro filosófico rigoroso. A autoridade epistêmica do artigo sobre o *Bóson de Higgs*¹¹ não pertence aos computadores, mas aos cientistas que o assinam. Analogamente, a autoria de um artigo filosófico produzido através da escrita distante pertence inequivocamente ao filósofo que o concebeu, o guiou e se responsabiliza por ele.

3.2 Vulnerabilidades Cognitivas e "Simulacros Epistêmicos"

A profunda integração entre o pensamento humano e a geração maquinária, embora potente, abre a porta a um conjunto de vulnerabilidades cognitivas que são particularmente agudas na escrita filosófica. O risco mais insidioso dos LLMs não reside nos seus erros óbvios, mas na sua fluência enganadora. Os modelos modernos são excepcionalmente proficientes a imitar as convenções da escrita acadêmica, produzindo textos que são gramaticalmente impecáveis e estilisticamente polidos. Essa qualidade, por si só, pode mascarar lacunas no conteúdo e na lógica, apresentando um verniz de autoridade que desarma a vigilância crítica.

Essa fluência cria uma perigosa ilusão epistêmica, um “*efeito de halo*”, em que a forma polida do texto confere uma falsa aura de rigor ao seu conteúdo. Na filosofia, em que a nuance da linguagem e a estrutura retórica são frequentemente inseparáveis da própria substância do argumento, essa capacidade de imitação constitui uma armadilha de primeira ordem. Um parágrafo bem estruturado e com uma terminologia técnica precisa pode, facilmente, contornar as defesas críticas iniciais de

¹¹ O bóson de Higgs é a partícula elementar associada ao campo de Higgs, cuja função é explicar a origem da massa das partículas fundamentais no Modelo Padrão da física. Sua existência foi proposta nos anos 1960 e confirmada experimentalmente em 2012 pelos detectores ATLAS e CMS no Large Hadron Collider.

um leitor ou, mais perigosamente, de um autor apressado ou excessivamente confiante na ferramenta.

Essa característica dá origem ao que pode ser designado, de forma apropriada, como simulacros epistêmicos: resultados que simulam perfeitamente a forma do conhecimento sem possuir a sua substância. O conceito, que ecoa a noção de “simulacro”, de Jean Baudrillard (1994), como uma cópia sem original, descreve um texto que possui todos os marcadores externos de um discurso filosófico legítimo (conceitos precisos, estrutura argumentativa, conectores lógicos), mas que é vazio no seu núcleo epistêmico¹². É, na sua essência, um zumbi epistêmico: parece e comporta-se como um argumento filosófico vivo, mas falta-lhe a mente, isto é, a compreensão, a intencionalidade e a justificação, que lhe daria vida intelectual.

Considere, como exemplo, um filósofo que solicita a um LLM “uma análise da relação do *Dasein* com a temporalidade na obra ‘Ser e Tempo’, de Heidegger”. O modelo pode gerar um texto convincente, utilizando corretamente termos como “*ser-para-a-morte*”, “*autenticidade*” e “*cuidado*”. No entanto o LLM não tem qualquer compreensão fenomenológica de como esses conceitos se interligam na complexa ontologia de Heidegger. O texto é uma montagem sintática baseada em padrões extraídos de milhares de textos sobre o tema, uma imitação perfeita de uma explicação filosófica, mas desprovida do pensamento que deveria animá-la. Para um autor que não seja um especialista profundo em Heidegger, esse simulacro pode ser indistinguível de uma análise genuína, levando à sua incorporação acrítica no trabalho.

Uma analogia filosófica poderosa pode ser traçada com os Sofistas da Grécia Antiga. Figuras como Górgias e Protágoras eram mestres da retórica, capazes de

¹² Embora tenhamos usado o termo “simulacro” de forma geral, a teoria de Baudrillard (1994) é mais granular, descrevendo quatro estágios de representação: (1) o reflexo de uma realidade básica; (2) a máscara e perversão de uma realidade básica; (3) a máscara da ausência de uma realidade básica; e (4) o simulacro puro, que não tem qualquer relação com qualquer realidade. Os simulacros epistêmicos gerados por LLMs podem variar ao longo desse espectro, desde a paráfrase (estágio 2) até à fabricação completa de factos e argumentos (estágio 4).

construir discursos eloquentes sobre qualquer tema com o objetivo de persuadir uma audiência, independentemente da verdade. O seu foco era a opinião (*doxa*), não o conhecimento (*episteme*). Os LLMs funcionam como os Sofistas supremos da era digital. A sua arquitetura é otimizada não para a veracidade, mas para a plausibilidade textual, o correlato moderno da persuasão. Quando solicitado a defender uma tese, o LLM monta a sequência de palavras mais provável que, nos seus dados de treino, corresponde a essa defesa. É, na sua essência, uma máquina de Górgias, capaz de gerar um discurso com igual convicção aparente, mas sem qualquer compromisso com a verdade. O filósofo que utiliza essa ferramenta sem uma vigilância implacável se arrisca a se tornar um aprendiz de sofista.

Para além desses riscos conceituais, existem perigos práticos e imediatos, como o fenômeno bem documentado das “*alucinações*” da IA. Estas ocorrem quando um modelo gera informações factualmente incorretas, mas apresenta-as com total confiança, como atribuir uma citação a um filósofo errado ou inventar artigos acadêmicos inexistentes para suportar um determinado argumento (Strzelecki, 2023). Na escrita filosófica, a integridade das fontes e a genealogia correta de uma ideia não são meras formalidades; são centrais para a honestidade intelectual. Uma referência fabricada não é simplesmente um erro, é uma corrupção do registro acadêmico. A responsabilidade pela verificação meticolosa de cada afirmação e fonte gerada pela IA recai, de forma intransigente, sobre o autor humano.

Finalmente, os LLMs podem perpetuar e amplificar preconceitos sistêmicos presentes nos seus vastos dados de treino (Friedman; Nissenbaum, 1996). Um modelo treinado predominantemente no cânone filosófico ocidental pode, ao ser solicitado a comparar teorias éticas, marginalizar ou apresentar de forma caricatural as perspectivas da filosofia africana ou asiática. Isso não acontece por malícia, mas por um reflexo da assimetria estatística dos textos com que aprendeu. O sistema cognitivo distribuído se torna, assim, uma zona de vulnerabilidade, em que a delegação do

trabalho cognitivo acarreta o risco de uma correspondente, e perigosa, delegação da vigilância epistêmica e ética.

3.3 Virtudes Intelectuais como Antídoto

Se a escrita distante, praticada sem o devido cuidado, é uma zona de vulnerabilidade epistêmica, a solução para os seus perigos não reside numa rejeição tecnofóbica nem em listas de verificação puramente processuais. Acreditamos que a resposta mais robusta e duradoura se encontra no domínio da ética e da epistemologia, especificamente no caráter do próprio agente. A epistemologia da virtude, com o seu foco nas disposições, traços de caráter e excelências intelectuais do conhecedor, oferece um quadro normativo ideal para guiar o filósofo nessa nova paisagem (Baehr, 2011; Zagzebski, 1996). Para navegar de forma responsável na colaboração com os LLMs, o autor deve cultivar e exercer ativamente um conjunto de virtudes intelectuais que funcionam como o antídoto necessário para as vulnerabilidades cognitivas inerentes a esse processo.

A primeira e mais fundamental dessas disposições é a vigilância epistêmica. Essa virtude transcende a mera verificação de fatos; é uma postura ativa e habitual de ceticismo metódico. O filósofo vigilante aborda cada fragmento de texto gerado por um LLM não como uma afirmação a ser aceita, mas como uma hipótese a ser rigorosamente testada. Numa analogia com o método cartesiano, o autor deve se manter inicialmente cético frente aos resultados da máquina. Cada argumento, cada definição e cada exemplo, por mais plausível que pareça, deve ser submetido ao tribunal da razão e do conhecimento prévio do autor. A vigilância exige que se questione incessantemente: a cadeia de inferências é logicamente sólida ou apenas retoricamente persuasiva? A analogia oferecida pela IA é verdadeiramente esclarecedora ou apenas uma simplificação enganadora? A fonte citada é real e o seu conteúdo foi representado com precisão? Essa virtude transforma o autor de um

consumidor passivo de texto a um interrogador incansável, garantindo a higiene epistêmica do seu trabalho.

Intimamente ligada à vigilância está a humildade epistêmica, uma virtude de dupla face. Por um lado, implica uma consciência clara e constante das limitações do LLM. O filósofo epistemicamente humilde nunca deve esquecer que está interagindo com uma ferramenta não senciente, uma “caixa-preta” estatística propensa a erros, preconceitos e à fabricação de absurdos convincentes. Essa atitude previne a confiança cega e a tendência para a delegação de tarefas intelectuais críticas.

Por outro lado, a humildade manifesta-se no reconhecimento das limitações do próprio conhecimento do autor. Nenhum filósofo é especialista em todos os domínios. A humildade leva o autor a se perguntar: “tenho a profundidade de conhecimento necessária para avaliar criticamente esta análise sobre a lógica modal que o LLM acabou de gerar?” Essa consciência fomenta a honestidade intelectual e incentiva a busca por verificação externa junto de outros especialistas humanos, funcionando como uma salvaguarda essencial contra a arrogância que a aparente onisciência da IA poderia induzir (Nyholm, 2023).

A terceira virtude indispensável é a *autorreflexão crítica*, que funciona como uma espécie de “*meta-virtude*” que supervisiona o processo cognitivo do próprio autor. O seu foco não está no resultado da máquina, mas na própria interação do autor com ela. O filósofo autorreflexivo questiona as suas próprias motivações e a suscetibilidade a preconceitos cognitivos¹³. Ele deve estar atento ao *viés de confirmação*: a tendência para aceitar, de forma acrítica, os resultados da IA que reforçam as suas próprias teses, enquanto submete os resultados dissonantes a um escrutínio muito mais severo. Essa virtude implica perguntar: “Estou utilizando o LLM para explorar genuinamente um

¹³ O estudo dos vieses cognitivos (*cognitive bias*) é uma área de investigação prolífica na psicologia e na economia comportamental. O viés de confirmação, a tendência para procurar, interpretar e recordar informações de uma forma que confirma as crenças pré-existentes, é um dos mais bem documentados. Para uma introdução acessível aos principais vieses que afetam o raciocínio humano, ver a obra de Daniel Kahneman (2011), *Thinking, Fast and Slow*.

problema ou para encontrar atalhos que evitem o trabalho árduo do pensamento original? A minha satisfação com este parágrafo gerado deve-se à sua qualidade intrínseca ou ao fato de ele articular convenientemente algo que eu tinha dificuldade em expressar?”. Esse tipo de introspecção é o antídoto para a destituição da agência humana que alguns críticos temem (Clowes, 2023), pois reafirma o controle consciente e deliberado do autor sobre o seu próprio processo intelectual.

Essas três disposições centrais são complementadas por outras virtudes igualmente importantes. A coragem intelectual é necessária para descartar um trecho de texto gerado pela IA que seja eloquente e bem escrito, mas filosoficamente impreciso ou superficial, resistindo à tentação de o manter devido ao esforço já investido. A minuciosidade intelectual manifesta-se na paciência para verificar meticulosamente cada detalhe, resistindo à pressão por produtividade que essas ferramentas podem criar.

Em conjunto, esse mosaico de virtudes constitui o núcleo da responsabilidade epistêmica na era da IA. Elas não são qualidades opcionais, mas sim condições necessárias para a prática da escrita distante de uma forma que seja intelectualmente honesta e filosoficamente valiosa. O produto final de uma colaboração humano-máquina não reflete apenas a capacidade da tecnologia, mas, e de forma mais decisiva, o caráter intelectual e a integridade do agente humano que a conduziu.

4 Implicações para a Autoria, Originalidade e Integridade Filosófica

A reconfiguração da escrita como um ato de curadoria e *design*, mediado por poderosos instrumentos generativos, transcende as questões de processo e adentra o coração normativo da prática acadêmica. A emergência da escrita distante obriga a uma reavaliação crítica dos conceitos fundamentais que sustentam a integridade da filosofia e de outras disciplinas: autoria, originalidade e plágio. Essa nova realidade do trabalho intelectual imposta pela escrita distante exige que repensem e reforcemos as estruturas epistêmicas, éticas e normativas que tradicionalmente se baseiam na

suposição de um autor humano único, intencional e singular.

Nosso interesse não é defender o abandono desses princípios, mas de adaptá-los a um novo contexto. Isso implica necessariamente uma mudança de foco, afastando-se da noção de plágio como “roubo” e aproximando-se de uma ética da transparência, da responsabilidade e da confiança epistêmica.

4.1 Repensando o Plágio e a Atribuição

Além de transformar nossos modos de produção intelectual, a introdução da escrita distante na prática filosófica também desestabiliza os fundamentos éticos que orientam essa atividade, exigindo uma revisão profunda das noções de autoria, plágio e atribuição.

De acordo com a tradição, a noção de plágio está fundamentalmente focada na ideia de apropriação indevida de um produto (texto, ideia etc.) que pertence a outro autor. Em linhas gerais, entende-se o plágio como uma forma de roubo intelectual: um ato que retira de alguém o reconhecimento, a autoria e o resultado de seu esforço, apresentando-os como se fossem do plagiador (Scanlon, 2003). Esse modelo parte da premissa de que existe um autor portador de direitos morais e epistêmicos sobre a sua obra. Contudo, essa estrutura normativa entra em colapso quando a origem do texto é um Modelo de Linguagem Amplo (LLM), cuja natureza não se encaixa nas categorias tradicionais de autoria humana.

Como foi estabelecido, um LLM não é um agente no sentido filosoficamente relevante. É um instrumento sofisticado, uma máquina estatística desprovida de intencionalidade, consciência, direitos ou da capacidade de “possuir” o conteúdo que gera. Nesse contexto, a acusação de “plagiar de uma IA” torna-se um erro categórico¹⁴.

¹⁴ Essa posição filosófica está cada vez mais a ser refletida no domínio jurídico. O U.S. Copyright Office, por exemplo, emitiu orientações em 2023 afirmando que uma obra gerada inteiramente por IA sem qualquer intervenção criativa humana não pode ser protegida por direitos de autor, uma vez que a lei de direitos de autorais dos EUA protege apenas os frutos do trabalho intelectual que se baseiam nos poderes criativos da mente humana. A ausência de autoria humana na fonte do texto gerado pela IA

Não se pode roubar propriedade intelectual de uma entidade que não pode, em primeiro lugar, possuí-la. Tentar aplicar o modelo tradicional de plágio a essa nova realidade é como tentar aplicar as leis de trânsito a um cardume de peixes: as categorias simplesmente não se correspondem. Essa constatação, contudo, não nos absolve do problema ético. Pelo contrário, obriga-nos a localizá-lo com maior precisão.

Se o erro não reside na violação dos direitos de um inexistente autor-máquina, então qual é a natureza da transgressão? A resposta reside numa mudança de enquadramento: o erro ético fundamental da utilização não declarada de LLMs não é a apropriação, mas sim a decepção e a deturpação. O dano não é infligido à máquina, mas à comunidade de leitores e à própria integridade do discurso filosófico.

A prática acadêmica assenta num pacto de confiança implícito entre o autor e o leitor. O leitor confia que o trabalho apresentado sob o nome de um autor é um reflexo honesto do seu esforço cognitivo, da sua perícia e da sua voz intelectual. A não divulgação do uso substancial de um LLM viola esse pacto de forma fundamental. O leitor é levado a acreditar que um argumento complexo ou uma prosa eloquente nasceram da mente do autor, quando, na verdade, podem ter sido gerados por um sistema computacional. Essa omissão constitui uma deturpação epistêmica grave, pois impede que a comunidade avalie corretamente a origem, o mérito e a autoridade do trabalho (Roig, 2006).

Aqui podemos trazer uma analogia esclarecedora: a prática do *ghostwriting* acadêmico¹⁵. Considere certas áreas “sensíveis” do conhecimento¹⁶. Por exemplo, tomemos o campo da medicina. Nesse sentido, a publicação de um artigo científico

reforça a ideia de que o plágio, como violação de um direito de autor, é uma categoria inaplicável. Ver U.S. Copyright Office (2023).

¹⁵ O *ghostwriting* (ou escrita fantasma) é a prática de contratar alguém para escrever um texto que será oficialmente creditado a outra pessoa. No contexto acadêmico, é considerado uma grave violação da integridade, pois o autor nomeado (quem assina o trabalho) recebe o crédito e a autoridade por um esforço intelectual que não realizou. Essa prática engana a comunidade de leitores e revisores, deturpando a verdadeira origem do trabalho.

¹⁶ O termo “sensíveis”, remete àquelas áreas do conhecimento que lidam diretamente com o bem-estar e a vida humana e seu impacto é de alto custo.

escrito por um médico renomado, mas secretamente escrito por uma empresa de comunicação contratada por uma empresa farmacêutica, constituiria uma grave violação ética (Ross *et al.*, 2008). O problema central não é exatamente a delegação da escrita a outro agente, mas os fatos dessa delegação serem mantidos em segredo. A ocultação cria uma falsa atribuição de autoridade, engana os revisores e leitores, e mina a credibilidade do processo científico. A utilização não declarada de um LLM na filosofia é uma forma de *ghostwriting* tecnológico. De modo similar, o filósofo, ao apresentar o texto gerado como seu, empresta a sua credibilidade e o seu nome a um produto cuja origem cognitiva é, na melhor das hipóteses, híbrida e, na pior, inteiramente delegada à máquina. Dessa maneira, o leitor é privado do conhecimento de que está dialogando com o pensamento de um ser humano ou com o resultado estatístico de uma máquina.

Com o intuito de aprofundarmos a nossa compreensão do termo “autoria” nesse novo contexto, podemos recorrer a uma analogia: a obra *Fonte* (1917), produzida por Marcel Duchamp¹⁷. Duchamp não é tomado como o fabricante do urinol em nenhum sentido relevante; tudo o que ele fez foi selecionar esse objeto (que normalmente é produzido em massa), removeu-o do seu contexto funcional original, deu-lhe um novo contexto e, ao fazê-lo, assinou-o e apresentou-o como arte. O seu ato autoral foi um gesto de seleção, recontextualização e enquadramento conceitual, não de manufatura. A originalidade da obra reside nesse ato de *design* e não na criação do objeto físico.

A prática do filósofo na escrita distante é, em muitos aspectos, análoga. O seu ato autoral pode residir não na geração da prosa, mas na formulação de um *prompt* genial, na seleção criteriosa do resultado da IA e na sua integração engenhosa num

¹⁷ A discussão sobre a autoria na arte conceitual é vasta. A intervenção de Duchamp com os seus *readymades* desafiou a ideia de que a arte reside na habilidade manual do artista, deslocando o foco para a ideia, a escolha e o gesto institucional. Essa mudança é central para compreender como a autoria na escrita distante pode residir mais no *design* conceitual e na curadoria do que na produção linguística. Para uma análise filosófica aprofundada, ver Goldie e Schellekens (2007).

argumento maior. No entanto a analogia só se sustenta sob uma condição crucial: a transparência. Duchamp nunca fingiu ter esculpido o urinol. A sua honestidade sobre a natureza do “objeto encontrado” era parte integrante da sua intervenção artística. O filósofo que oculta o uso de um LLM está a cometer o equivalente a Duchamp a afirmar que moldou a porcelana com as suas próprias mãos. O erro não está no método não tradicional de criação, mas na desonestade sobre o uso desse método.

Isso nos leva a repensar a própria noção de “*originalidade*” na escrita filosófica. A era da escrita distante pode acelerar o afastamento de um ideal romântico de *creatio ex nihilo* (criação a partir do nada) em direção a uma apreciação mais sofisticada da originalidade como um ato de síntese, curadoria e conexão. A “voz” de um autor pode manifestar-se menos no seu estilo de prosa idiossincrático e mais nas suas escolhas de ordem superior: na originalidade das questões que coloca à máquina (os *prompts*), na profundidade do seu julgamento crítico ao avaliar os resultados e na arquitetura geral do argumento que constrói a partir dos materiais gerados. A contribuição única do autor se torna, assim, a sua capacidade de orquestrar uma colaboração produtiva com a tecnologia e de assumir plena responsabilidade pelo produto final.

Em suma, a inadequação dos modelos tradicionais de plágio exige a adoção de um novo quadro normativo para a autoria na era da IA. Esse quadro deve abandonar a lógica do “*roubo*” de um não-agente e focar-se no imperativo ético da transparência. A utilização de uma ferramenta tão poderosa como um LLM é um fato metodologicamente relevante, que afeta as condições de produção do conhecimento. Como tal, a sua declaração não deve ser vista como uma confissão de culpa, mas como uma prática de rigor e honestidade intelectual, semelhante à forma como os cientistas declaram os seus métodos, o *software* que utilizam e as suas fontes de financiamento (Nature, 2023). Apenas através de uma cultura de transparência radical podemos preservar a confiança, que é a base de todo o empreendimento filosófico e acadêmico.

4.2 Uma Taxonomia para a Divulgação Ética

Tendo estabelecido a transparência como o imperativo ético central para a escrita distante, a questão pragmática que se segue é: como deve ser essa transparência na prática? Uma declaração genérica e binária, como “Foi utilizada IA na preparação deste manuscrito”, é manifestamente insuficiente. Tal afirmação é demasiadamente vaga para ser informativa, obscurecendo a diferença crucial entre usar uma IA para corrigir uma vírgula e usá-la para gerar um argumento central. Para que a divulgação seja eticamente significativa, ela deve ser granular e proporcional à natureza e à profundidade do envolvimento da máquina no trabalho intelectual. O que se propõe aqui é uma taxonomia do envolvimento da IA, um quadro conceptual que distingue diferentes níveis de colaboração e que pode guiar os filósofos na sua responsabilidade de divulgação¹⁸. Essa taxonomia funciona de forma análoga aos créditos de um filme: em vez de um único crédito de “*autoria*”, reconhece-se que a criação é um processo com diferentes funções (realizador, editor, argumentista), cada uma merecendo um tipo específico de reconhecimento.

No nível mais básico da taxonomia, encontramos a assistência lexical e estilística menor. Isso abrange a utilização de LLMs para tarefas que são funcionalmente equivalentes às de ferramentas de escrita já estabelecidas, ainda que mais poderosas. Inclui a correção gramatical e ortográfica, a sugestão de sinônimos, ou o parafraseamento de frases individuais para melhorar a clareza ou a fluidez, sem alterar a sua substância semântica. Por exemplo, um filósofo que tenha escrito uma frase densa e complexa sobre a “unidade transcendental da apercepção” de Kant poderia solicitar ao LLM: “Reescreva essa frase de forma mais simples, mantendo o significado original”. Nesse caso, o LLM atua como um editor de texto ou um dicionário de

¹⁸ Muitas instituições acadêmicas e editoras estão começando a desenvolver suas próprias políticas sobre o uso de IA, embora estas variem significativamente. Enquanto algumas proíbem o uso de IA para gerar texto, outras permitem-no com a condição de divulgação. A taxonomia aqui proposta visa oferecer um quadro conceptual mais robusto do que as atuais abordagens, muitas vezes binárias, permitindo uma política de divulgação mais informada e consistente. Para um exemplo de uma política editorial influente, ver Nature (2023).

sinônimos avançado. A estrutura do argumento, as ideias e a intenção autoral permanecem inteiramente humanas. A contribuição da máquina é superficial e não afeta o núcleo intelectual do trabalho. Consequentemente, a necessidade de divulgação nesse nível é mínima ou nula, tal como não se costuma creditar um processador de texto pelas suas sugestões gramaticais.

Um degrau acima, encontramos o aumento estilístico. Nesse nível, o autor utiliza o LLM não apenas para pequenas correções, mas para moldar significativamente a “voz”, o tom ou o efeito retórico de passagens inteiras, enquanto o argumento subjacente permanece do autor. Imagine um filósofo que escreveu uma análise acadêmica sobre a ética da virtude de Aristóteles e pretende adaptá-la para um público mais vasto. Ele poderia instruir o LLM: “Reescreva esse parágrafo sobre a *phronesis* num estilo mais envolvente e narrativo, usando uma analogia contemporânea”. Aqui, a ideia central sobre a sabedoria prática é do autor, mas a sua execução estilística e o seu invólucro retórico são o produto de uma colaboração híbrida. A voz do texto já não é puramente a do autor. Como isso pode afetar a percepção que o leitor tem do estilo e da capacidade de comunicação do autor, uma divulgação discreta torna-se uma prática de boa honestidade intelectual. Uma nota de rodapé ou uma menção na seção de agradecimentos, como “Foram utilizadas ferramentas de IA para auxiliar no refinamento estilístico de algumas secções”, seria apropriada e suficiente.

O terceiro nível, que marca a entrada num território de colaboração profunda, é a geração argumentativa. Aqui, o LLM deixa de ser um mero estilista para se tornar um participante ativo na construção do conteúdo filosófico. O autor pode utilizar a ferramenta como um parceiro de *brainstorming* ou um gerador de hipóteses. Por exemplo, um filósofo a trabalhar numa crítica ao dualismo cartesiano poderia solicitar: “Gere um argumento original contra a divisibilidade do corpo como prova da sua distinção da mente, utilizando conceitos da física moderna”. O LLM poderia produzir um esboço de argumento que o autor, por si só, não tinha considerado. O filósofo pode,

então, adotar essa linha de raciocínio, refiná-la, reforçá-la e integrá-la no seu trabalho. Nesse cenário, a gênese de uma parte da substância intelectual do artigo é não-humana. Apresentar esse argumento como sendo exclusivamente fruto da sua própria deliberação seria uma clara deturpação do processo criativo. Consequentemente, esse nível exige uma divulgação explícita e proeminente, preferencialmente numa seção de metodologia ou numa nota introdutória, detalhando o papel da IA no desenvolvimento de seções específicas do argumento.

No topo da taxonomia reside o nível mais profundo de colaboração: a inovação temática. Nesse modo, o LLM funciona como um catalisador para a descoberta de novas conexões, analogias ou mesmo teses. Um filósofo interessado em filosofia comparada poderia usar um *prompt* exploratório e aberto: “Identifique e descreva possíveis, mas pouco exploradas, ressonâncias estruturais entre o conceito de ‘*devir*’ em Deleuze e a noção de ‘*impermanência*’ no pensamento budista”. Ao processar e sintetizar padrões de vastos *corpora* de texto, o LLM pode destacar uma analogia ou uma tensão conceitual que o autor não tinha vislumbrado, abrindo, assim, uma linha de investigação inteiramente nova e original. Nesse caso, a IA desempenhou um papel que se assemelha ao de um colaborador de investigação ou de uma musa computacional. A sua contribuição para a originalidade da tese central é inegável. Ocultar essa contribuição seria a forma mais grave de deturpação epistêmica. A divulgação deve ser, portanto, total e explícita, reconhecendo abertamente o papel da ferramenta como parceira no processo de descoberta conceptual, um ato que preserva a honestidade do autor e enriquece a compreensão da comunidade sobre as novas formas como o conhecimento pode ser gerado.

A adoção de uma taxonomia como essa não é apenas um exercício de rigor processual; é a manifestação prática de uma ética da autoria adaptada ao século XXI. Ao aplicá-la, o filósofo exercitaativamente as virtudes relacionadas à honestidade intelectual, assumindo a responsabilidade tanto do conteúdo do seu trabalho quanto da transparência do processo empregado, garantindo que a sua relação com a

comunidade de leitores se mantenha fundamentada na confiança. Dessa forma, a integridade da autoria filosófica não é diminuída pela tecnologia, mas, pelo contrário, é reforçada através de um compromisso mais profundo com a clareza e a verdade em todas as fases da produção de conhecimento.

5 Conclusão

O advento da escrita distante, conforme teorizado por Luciano Floridi (2025a), representa mais que uma mera inovação tecnológica na produção textual; ele impõe uma reavaliação fundamental das práticas e dos pressupostos que sustentam a escrita filosófica. Neste artigo, defendemos a tese de que, perante a capacidade geradora dos Modelos de Linguagem Ampla (LLMs), o papel do filósofo não é diminuído nem substituído, mas sim reconfigurado para o de um agente epistêmico curador. Longe de ser um mero operador de uma ferramenta, o autor na era da IA é chamado a exercer a sua agência a um nível mais elevado de abstração, orquestrando, avaliando e assumindo a responsabilidade por um processo de criação de significado que é intrinsecamente híbrido.

Para sustentar essa tese, a nossa análise começou por estabelecer uma distinção crucial na arquitetura epistêmica da escrita distante: a diferença entre a agência epistêmica do autor humano e a instrumentalidade epistêmica do LLM. Argumentamos que, embora o LLM seja um instrumento de uma potência sem precedentes, ele permanece desprovido de compreensão, intencionalidade ou responsabilidade, funcionando como um sofisticado motor sintático. O trabalho genuinamente filosófico, de justificar crenças, construir argumentos válidos e gerar conhecimento, permanece como domínio exclusivo do filósofo-curador, que mobiliza o seu conhecimento processual, avaliativo e semântico para guiar e validar o processo.

No entanto essa nova forma de cognição distribuída não é isenta de perigos. Analisamos como a fluência retórica dos LLMs pode criar simulacros epistêmicos, textos que imitam a forma do conhecimento filosófico sem a sua substância,

introduzindo vulnerabilidades como a desinformação, os argumentos falaciosos e os preconceitos ocultos. A resposta a esses riscos, argumentamos, não é puramente técnica, mas normativa, e reside no cultivo de virtudes intelectuais. A vigilância, a humildade e a autorreflexão crítica foram apresentadas como as disposições essenciais que permitem ao filósofo navegar nesse ecossistema de forma segura e responsável, garantindo a integridade do seu trabalho.

Finalmente, exploramos as implicações dessa nova realidade para os conceitos de plágio e autoria. Concluímos que os modelos tradicionais de plágio, baseados na ideia de roubo, são inadequados. O erro ético da utilização não declarada de IA reside no ardil e na quebra da confiança com a comunidade de leitores. Para resolver essa questão, propusemos uma taxonomia para a divulgação ética, um quadro que permite uma transparência granular e honesta sobre o nível de envolvimento da IA, desde a simples assistência estilística até à inovação temática. A aplicação dessa taxonomia não é apenas uma formalidade, mas um exercício de responsabilidade e consciência epistêmica.

Em última análise, a escrita distante não assinala o fim da autoria filosófica, mas sim a sua evolução. Ela desafia-nos a sermos mais rigorosos, não apenas no conteúdo dos nossos argumentos, mas também na reflexão sobre os métodos que utilizamos para os produzir. A responsabilidade pela coerência, pela veracidade e pela originalidade de um trabalho filosófico permanece, de forma indivisível, sobre os ombros do autor humano. A era da IA na filosofia não é um convite à abdicação do esforço intelectual, mas uma chamada para uma forma mais profunda de mestria: a do pensador que não só domina as suas ideias, mas que também domina, de forma crítica e virtuosa, as poderosas ferramentas que ajudam a dar-lhes forma.

Referências

ARISTOTLE. *Nicomachean ethics*. Translated by W. D. Ross. c. 350 BCE.

ATLAS Collaboration; CMS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson at the LHC. **Physics Letters B**, v. 716, n. 1, p. 1–29, 2012.

BAEHR, J. The epistemic virtues and the nature of knowledge. In: PRITCHARD, D. (ed.). *Epistemology: The key thinkers*. London: Routledge, 2011. p. 148-164.

BATTAGLIA, F. Intellectual virtues, extended cognition, and the epistemology of AI. **Philosophies**, v. 7, n. 4, p. 84, 2022. DOI

BAUDRILLARD, J. **Simulacra and simulation**. Ann Arbor: University of Michigan Press, 1994. DOI <https://doi.org/10.3998/mpub.9904>

BENDER, E. M. *et al.* On the dangers of stochastic parrots: Can language models be too big? In: **Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency**. 2021. p. 610–623. DOI <https://doi.org/10.1145/3442188.3445922>

BODEN, M. A. **The creative mind: Myths and mechanisms**. 2. ed. London: Routledge, 2004. DOI <https://doi.org/10.4324/9780203508527>

BUCKNER, C. Deep learning: A philosophical introduction. **Philosophy Compass**, v. 14, n. 10, e12625, 2019. DOI <https://doi.org/10.1111/phc3.12625>

CLARK, A.; CHALMERS, D. The extended mind. **Analysis**, v. 58, n. 1, p. 7–19, 1998. DOI <https://doi.org/10.1093/analys/58.1.7>

CLOWES, R. W. AI tools, epistemic dependence and the demotion of human agency. **AI and Ethics**, p. 1–13, 2023. DOI

COLE, D. The Chinese Room Argument. In: ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy** (Winter 2020 Edition). 2020. Disponível em: <https://plato.stanford.edu/archives/win2020/entries/chinese-room/>. Acesso em: 11 out. 2025.

EATON, S. E.; GÁLIK, M. A view on generative artificial intelligence and the future of academic integrity. **Journal of Education, Culture and Society**, v. 15, n. 1, p. 11–26, 2024. DOI

ELGIN, C. Z. **True enough**. Cambridge: MIT Press, 2017. DOI <https://doi.org/10.7551/mitpress/9780262036535.001.0001>

FLORIDI, L. Distant writing: Literary production in the age of artificial intelligence. *SSRN*, 2025a. DOI <https://doi.org/10.2139/ssrn.5232088>

FLORIDI, L. AI as agency without intelligence: On artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. *Philosophy & Technology*, v. 38, n. 30, 2025b. DOI <https://doi.org/10.1007/s13347-025-00858-9>

FRANKFURT, H. G. **On bullshit**. Princeton: Princeton University Press, 2005.

FRIEDMAN, B.; NISSENBAUM, H. Bias in computer systems. *ACM Transactions on Information Systems*, v. 14, n. 3, p. 330-347, 1996. DOI <https://doi.org/10.1145/230538.230561>

GALISON, P. **Image and logic**: A material culture of microphysics. Chicago: University of Chicago Press, 1997. DOI <https://doi.org/10.29173/istl1396>

GOLDIE, P.; SCHELLEKENS, E. (ed.). **Philosophy and conceptual art**. Oxford: Oxford University Press, 2007. DOI <https://doi.org/10.1093/oso/9780199285556.001.0001>

GOLDMAN, A. I. **Knowledge in a social world**. Oxford: Oxford University Press, 1999. DOI <https://doi.org/10.1093/0198238207.001.0001>

GRASSO, F. On the epistemic risks of large language models. *Minds and Machines*, v. 33, p. 685–690, 2023. DOI

GRECO, J. **Achieving knowledge**: A virtue-theoretic account of epistemic normativity. Cambridge: Cambridge University Press, 2010. DOI <https://doi.org/10.1017/CBO9780511844645>

JAKESCH, M.; LEDERER, M.; MEYER, M. Epistemic risks of AI-generated texts. *Digital Society*, v. 2, n. 1, p. 10, 2023. DOI

KAHNEMAN, D. **Thinking, fast and slow**. New York: Farrar, Straus and Giroux, 2011.

KIDD, I. J.; BATTAGLIA, F. AI and intellectual virtue: An introduction. *Philosophies*, v. 8, n. 6, p. 110, 2023. DOI

MORETTI, F. Conjectures on world literature. *New Left Review*, n. 1, p. 54–68, 2000. DOI <https://doi.org/10.64590/hxj>

MORETTI, F. **Distant reading**. London: Verso Books, 2013.

NATURE. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. **Nature**, v. 613, p. 612, 2023. DOI <https://doi.org/10.1038/d41586-023-00191-1>

NYHOLM, S. The ethics of AI ethics: An introduction to the special issue. **Ethics and Information Technology**, v. 25, n. 1, p. 1, 2023. DOI

PRITCHARD, D. Epistemic risk. **Journal of Philosophy**, v. 113, n. 11, p. 550-571, 2016. DOI <https://doi.org/10.5840/jphil20161131137>

ROIG, M. **Avoiding plagiarism, self-plagiarism, and other questionable writing practices**: A guide to ethical writing. U.S. Department of Health & Human Services, Office of Research Integrity, 2006. Disponível em: <https://ori.hhs.gov/plagiarism-0>. Acesso em: 11 out. 2025.

ROSS, J. S. et al. Guest authorship and ghostwriting in publications related to rofecoxib: A case study of industry documents from rofecoxib litigation. **JAMA**, v. 299, n. 15, p. 1800–1812, 2008. DOI <https://doi.org/10.1001/jama.299.15.1800>

RYLE, G. **The concept of mind**. Chicago: University of Chicago Press, 1949.

SCANLON, P. M. Student online plagiarism: How do we respond? **College Teaching**, v. 51, n. 4, p. 161–165, 2003. DOI <https://doi.org/10.1080/87567550309596432>

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences**, v. 3, n. 3, p. 417–457, 1980. DOI <https://doi.org/10.1017/S0140525X00005756>

STRZELECKI, A. To use or not to use ChatGPT in research? Researchers' acceptance of AI-based conversational agents. **Journal of Information Science**, v. 0, n. 0, 2023. DOI

U.S. COPYRIGHT OFFICE. Copyright registration guidance: Works containing material generated by artificial intelligence. **Federal Register**, v. 88, n. 59, p. 16190-16194, 2023. Disponível em: <https://www.federalregister.gov/documents/2023/03/16/2023-05321>. Acesso em: 11 out. 2025.

ZAGZEBSKI, L. **Virtues of the mind**: An inquiry into the nature of virtue and the ethical foundations of knowledge. Cambridge: Cambridge University Press, 1996. DOI <https://doi.org/10.1017/CBO9781139174763>

ZYLINSKA, J. **AI art**: Machine visions and warped dreams. London: Open Humanities Press, 2020.