XXVIII SEMINÁRIO DE FILOSOFIA · MINICURSO

Escrever com IA sem deixar de pensar

Texto de apoio do minicurso

Participante: _____

Faculdade Católica da Paraíba · Cajazeiras, 23 de outubro de 2026 · Ministrante: Gervásio Teixeira (mestre em Ciência da Computação, UFPE; graduado em Filosofia)

Pergunta-guia
Quando um texto é escrito com ajuda de uma máquina, o que nele ainda é meu?

Apresentação

Este texto acompanha o minicurso e foi escrito para quem estuda Filosofia e precisa, ou logo precisará, redigir resenhas, artigos e um trabalho de conclusão de curso numa época em que as máquinas também escrevem. Não é um manual de ferramentas, que envelhece em meses, mas uma leitura sobre três perguntas que não envelhecem: o que é um autor, o que uma inteligência artificial faz por dentro e como usá-la sem abrir mão do próprio juízo. Cada seção parte de uma ideia da tradição filosófica, explica o conceito de computação que lhe corresponde e termina, quando cabe, em uma prática no computador.

As três atividades do minicurso estão descritas ao fim das seções 3, 4 e 5. Para realizá-las você vai precisar de um computador com internet, uma conta em alguma IA de uso gratuito (ChatGPT, Gemini, Claude ou Copilot) e um tema seu, real ainda que provisório (um TCC, um artigo, um projeto). Como a leitura é curta, vale lê-la inteira antes e voltar a ela quando for escrever.

Cinco regras de ouro
1. Eu penso primeiro; depois peço ajuda. 2. Nada que a IA entrega vale como fonte antes de eu verificar. 3. Não colo dados pessoais nem trabalhos sigilosos em ferramentas públicas. 4. Registro o que fiz: é a minha declaração de uso. 5. Se a IA concordar demais comigo, peço que discorde.

Antes de tudo: de onde vem a IA e por que ela "estourou"

A inteligência artificial não nasceu com o ChatGPT. A ideia de mecanizar o raciocínio é bem mais antiga: no século XVII, Leibniz sonhava com um cálculo capaz de resolver disputas. Em 1943, McCulloch e Pitts propuseram um modelo matemático de neurônio; em 1950, Turing trocou a pergunta "as máquinas podem pensar?" por um jogo de imitação; e em 1956, na proposta do encontro de verão de Dartmouth (McCarthy et al., 1955), apareceu o nome "inteligência artificial", com a aposta de que aprender e raciocinar poderiam ser descritos com tanta precisão que uma máquina seria capaz de simulá-los.

Desde então, duas apostas disputaram o campo. A **IA simbólica** tentava escrever, à mão, regras e lógica que representassem o conhecimento. O **conexionismo** apostava em redes neurais que aprendem padrões a partir de exemplos, como o Perceptron, de 1958. Os resultados ficaram abaixo das promessas, e houve períodos de cortes de verba e descrença, chamados de "invernos da IA" (anos 1970 e do fim dos anos 1980 ao início dos 1990). Em 1986, a retropropagação, popularizada por Rumelhart, Hinton e Williams, tornou viável treinar redes de várias camadas, mas faltavam dados e poder de cálculo.

A virada veio em 2012, quando a rede AlexNet venceu o concurso de reconhecimento de imagens ImageNet, treinada em GPUs, as placas gráficas (Krizhevsky et al., 2012): começava a era do **aprendizado profundo**. Em 2016, o AlphaGo venceu um campeão mundial de Go. Em 2017, o artigo *Attention is all you need* (Vaswani et al., 2017) apresentou o **transformer**, arquitetura que permite treinar com enormes volumes de texto, em paralelo. Em 2020, o GPT-3 chegou a 175 bilhões de parâmetros (Brown et al., 2020), e as chamadas leis de escala mostraram que mais dados, mais parâmetros e mais cálculo melhoram o desempenho de modo previsível (Kaplan et al., 2020). Em 2022, o ajuste dos modelos com avaliações de pessoas (Ouyang et al., 2022) e o lançamento do ChatGPT, em 30 de novembro, levaram a tecnologia ao grande público: estimativas da época apontaram mais de 100 milhões de usuários em cerca de dois meses.

Quando	O que aconteceu
Séc. XVII	Leibniz sonha com um cálculo capaz de resolver disputas: a ideia de mecanizar o raciocínio é antiga.
1943	McCulloch e Pitts propõem um modelo matemático de neurônio.
1950	Turing publica *Computing machinery and intelligence* e propõe o jogo da imitação.

Quando	O que aconteceu
1956	Na proposta do encontro de Dartmouth (McCarthy et al., 1955) surge o termo “inteligência artificial”.
1958	Rosenblatt apresenta o Perceptron, rede neural que aprende com exemplos.
Anos 1970, 1980 e 1990	Os “invernos da IA”: promessas demais, resultados de menos, cortes de verba.
1986	A retropropagação, popularizada por Rumelhart, Hinton e Williams, permite treinar redes de várias camadas.
2012	A rede AlexNet vence o concurso ImageNet, treinada em GPUs: começa a era do aprendizado profundo (Krizhevsky et al., 2012).
2016	O AlphaGo, da DeepMind, vence um campeão mundial de Go.
2017	O artigo *Attention is all you need* apresenta o transformer (Vaswani et al., 2017).
2020	O GPT-3 chega a 175 bilhões de parâmetros (Brown et al., 2020); as leis de escala mostram ganhos previsíveis com mais dados e cálculo (Kaplan et al., 2020).
2022	O ajuste com avaliações de pessoas (Ouyang et al., 2022) e o ChatGPT (30 de novembro) levam a tecnologia ao grande público.

Por que a IA “estourou” há poucos anos?

Foi a combinação de cinco ingredientes:

- **Dados**: décadas de internet deixaram bilhões de textos, imagens e códigos disponíveis para treinar os modelos.
- **Cálculo**: GPUs e grandes centros de dados permitem treinar redes enormes em semanas.
- **Arquitetura**: o transformer aproveita esse hardware e capta relações entre palavras distantes.
- **Escala**: mais dados, mais parâmetros e mais cálculo trazem ganhos previsíveis de desempenho.
- **Conversa**: o ajuste com feedback humano e um chat aberto a todos levaram a IA do laboratório ao cotidiano.

Repare no que não está nessa lista: uma nova teoria do pensamento. O que mudou foi a quantidade de dados, de cálculo e a forma de conversar com a máquina. Por isso a pergunta que guia este minicurso continua de pé, e a Seção 2 vai abrir o que uma IA faz por dentro. *Para pensar:* se o que mudou foi a escala e a interface, e não a ideia de pensar, o que muda na pergunta “o que, num texto escrito com ajuda de uma máquina, ainda é meu?”

1 · Autoria na era da IA

Remédio e veneno

No *Fedro*, Platão (2011, 274c-275b) conta um mito egípcio. O deus Theuth, inventor dos números, do cálculo e das letras, apresenta ao rei Thamus a escrita como um *pharmakon*, palavra grega que significa ao mesmo tempo remédio e veneno. Seria, diz Theuth, um remédio para a memória e para a sabedoria. Thamus discorda: a escrita produzirá esquecimento, porque os homens passarão a confiar em marcas externas em vez de exercitar a memória, e dará apenas uma “aparência de sabedoria”, pois quem lê muito sem aprender julgar-se-á sábio sem o ser.

Mais adiante (275d-e), Sócrates acrescenta que o escrito é como uma pintura: parece falar, mas, interrogado, repete sempre a mesma coisa. A inteligência artificial que escreve é um novo capítulo da mesma história, com uma diferença notável: agora o texto *responde*, adapta-se, reformula. A impressão de estar diante de um interlocutor é muito mais forte, e por isso o risco do “veneno” também. Lévy (1993) e Ong (1998) lembram que cada tecnologia da palavra reorganiza o modo de pensar de quem a usa. A pergunta, como em Platão, não é se a técnica é boa ou má, mas o que ela faz de nós conforme o uso que lhe damos: remédio para quem já pensou e quer organizar e testar suas ideias, veneno para quem ainda não pensou e deixa de aprender.

Menoridade e esclarecimento

Kant (2005 [1784]) define o esclarecimento como a saída do ser humano de sua menoridade, isto é, da incapacidade de servir-se do próprio entendimento sem a direção de outrem. A menoridade é “por culpa própria”: sua causa não é falta de entendimento, mas de decisão e de coragem. Ele a ilustra com um exemplo célebre: como é cômodo ser menor! Se tenho um livro que entende por mim, um pastor que tem consciência por mim, um médico que decide minha dieta, não preciso esforçar-me eu mesmo. Seu lema, tomado de Horácio, é *Sapere aude!*: tem a coragem de servir-te do teu próprio entendimento.

A IA entra nessa lista? Kant não condenaria os guias; condena a comodidade de nunca deixar de precisar deles. Na área de fatores humanos, descreve-se algo parecido com o nome de *viés de automação*: a tendência a confiar excessivamente nas respostas de um sistema automático. A diferença entre a ferramenta que esclarece e a tutela que mantém na menoridade

está no que fazemos com o resultado: se o aceitamos sem exame, é tutela; se o examinamos, discordamos e refazemos, é ferramenta.

Quem é o autor?

Arendt (2007), no prólogo de *A condição humana*, propõe "pensar o que estamos fazendo": não aceitar que as técnicas façam por nós o que deveríamos examinar. Aplicada à escrita, a proposta se desdobra em três perguntas. **Quem pensou?** A ideia central, o recorte e a tese partiram de você ou do modelo? **Quem decidiu?** Cada escolha (o que entra, o que sai, qual fonte vale) passou pelo seu juízo? **Quem responde?** Se houver erro, plágio ou invenção, quem assume? A última é decisiva: um sistema de inteligência artificial não assina nem responde; a responsabilidade recai sobre pessoas e instituições (em computação, fala-se em *accountability*).

Daí uma definição de trabalho que usaremos até o fim: **autor é quem decide, responde e pode justificar o que escreveu**. Usar IA não retira a autoria, desde que as três perguntas tenham resposta clara e honesta. *Para pensar:* em um texto seu em que você já usou, ou pretende usar, ajuda de IA, quem pensou, quem decidiu e quem responde? Em que ponto a ajuda virou tutela?

2 · Como uma IA funciona

Duas perguntas clássicas

Em *Computing machinery and intelligence*, Turing (1950) julga a pergunta "as máquinas podem pensar?" vaga demais e a substitui por um jogo: um interrogador conversa por escrito com uma pessoa e com uma máquina, sem saber qual é qual; se não consegue distingui-las, a pergunta mudou de natureza. Os modelos de linguagem atuais conversam de modo convincente, mas o teste mede comportamento, não compreensão. É aí que entra Searle (1980) com o quarto chinês: uma pessoa que não sabe chinês, trancada num quarto, recebe símbolos, consulta um manual de regras e devolve outros símbolos; de fora, parece entender chinês, mas apenas manipula formas. A conclusão de Searle é que seguir regras sobre a forma dos símbolos (sintaxe) não basta para compreender o que significam (semântica).

Aplicar o argumento aos modelos atuais é uma *hipótese de leitura*, não um veredicto: Searle criticava programas de manipulação simbólica da sua época, os modelos de hoje são redes neurais estatísticas, e há objeções clássicas (a de que o sistema como um todo compreenderia, por exemplo). Use o quarto chinês como lente: diante de qualquer texto gerado, pergunte se há compreensão ou apenas forma convincente.

O percurso de uma frase

Um *modelo de linguagem de grande escala* (LLM, do inglês *large language model*) é uma rede neural treinada em enormes volumes de texto para uma tarefa aparentemente simples: **prever a próxima palavra**. Tudo o que ele faz vem daí. O percurso de uma frase dentro do modelo tem cinco etapas. Primeiro, o texto que você escreve (o *prompt*) é dividido em *tokens*, pedaços de palavras, e cada token vira um número. Depois, cada token é representado por um *vetor*, uma lista de números; é esse vetor que o modelo manipula. Em seguida, um mecanismo de *atenção*, ideia central da arquitetura *transformer* (Vaswani et al., 2017), pondera quais partes do texto anterior são mais relevantes para cada posição. Com isso, o modelo calcula a *probabilidade* de cada palavra possível vir a seguir e escolhe uma. Por fim, a palavra escolhida é acrescentada ao texto e todo o ciclo recomeça, palavra por palavra.

Repare no que *não* há nessas etapas: nenhuma delas consulta um banco de fatos. O que há são padrões numéricos aprendidos de muito texto. Quando o modelo escreve bem, é porque aprendeu muito bem como os textos costumam continuar; quando erra, é pela mesma razão.

Significado como uso, verossimilhança e temperatura

Por que tokens usados em contextos parecidos recebem vetores parecidos? Porque o treinamento ajusta os números de modo que palavras que aparecem em situações semelhantes fiquem próximas. Isso lembra Wittgenstein (1953, §43): para uma grande classe de casos, o significado de uma palavra é o seu uso na linguagem. É uma analogia produtiva, mas só até certo ponto: o modelo capta padrões de uso em textos; não participa de uma forma de vida em que as palavras servem para prometer, pedir, responsabilizar.

O modelo não escolhe sempre a palavra mais provável. A **temperatura** controla o quanto ele arrisca: com temperatura baixa, escolhe quase sempre as mais prováveis; com temperatura alta, sorteia entre várias, com mais variedade e mais risco. Por isso a mesma pergunta pode receber respostas diferentes, às vezes até contraditórias. Aristóteles (*Poética*, cap. 9) já distinguia o poeta do historiador: o primeiro diz o que poderia acontecer segundo o verossímil ou o necessário; o segundo, o que aconteceu. O modelo está mais perto do poeta: devolve o verossímil. **Provável não é o mesmo que verdadeiro.**

Resta saber de onde vem o que o modelo "sabe". Em um primeiro momento, o *pré-treinamento* faz o modelo aprender a prever a próxima palavra em volumes enormes de texto; é aqui que nascem o vocabulário, o estilo e também os vieses dos textos lidos. Em seguida, o *ajuste* refina o modelo com exemplos e com avaliações de pessoas, para que responda de modo útil, educado e mais seguro. Por fim, no *uso*, o seu prompt entra como contexto e a resposta é gerada. Três consequências práticas: o modelo tem uma **data de corte** do que conhece; não guarda as fontes como um catálogo, apenas padrões; e herda os vieses do que leu.

Alucinação

Chama-se **alucinação** o conteúdo gerado que não tem fundamento nos dados ou na realidade (Ji et al., 2023). Na escrita acadêmica, ela aparece como referências, autores, páginas e citações que não existem, entregues com total confiança. Há três maneiras complementares de compreendê-la. Do ponto de vista técnico, o modelo é treinado para produzir texto plausível, e não para verificar fatos. Do ponto de vista filosófico, em analogia com Frankfurt (2005), a mentira se opõe à verdade, ao passo que a conversa fiada (*bullshit*) é simplesmente indiferente a ela: o texto gerado pode soar convincente sem ter qualquer compromisso com a verdade (o modelo não tem intenção, e por isso a comparação é só uma analogia). Do ponto de vista crítico, Bender et al. (2021) chamam os modelos de "papagaios estocásticos": recombinam formas linguísticas sem acesso ao sentido.

Algumas ferramentas, como Perplexity, combinam a geração de texto com busca na internet e mostram links; isso ajuda, mas não elimina a necessidade de conferir a fonte. *Para pensar:* se a IA escreve bem e erra pela mesma razão, o que isso exige de quem a usa? *Experimento sugerido:* peça a uma IA cinco referências sobre um tema filosófico e tente encontrar cada uma em uma base confiável. Conte quantas são reais, quantas são reais com dados errados e quantas não existem.

3 · Pesquisa e referências

Do espanto ao problema

No *Teeteto* (155d), Platão (2001) diz que o espanto é o princípio da filosofia, e Aristóteles repete a ideia na *Metafísica* (A, 2). Em pesquisa, o espanto precisa ser lapidado até virar um problema que se possa investigar. Um caminho em quatro passos vai do **tema** ao **recorte**, deste à **pergunta** e dela à **hipótese**. Por exemplo: tema, "IA e autoria acadêmica"; recorte, "texto gerado por IA em trabalhos de graduação em Filosofia"; pergunta, "em que sentido o texto escrito com IA é de autoria de quem a aciona?"; hipótese, "autoria exige decisão e responsabilidade, e não apenas execução". A IA pode ajudar a afinar o recorte e a testar a pergunta; mas o espanto, isto é, a escolha do que merece ser investigado, é seu.

O prompt em cinco camadas

Como o modelo gera o texto a partir do contexto que recebe, quanto mais claro o contexto, mais útil a resposta. Um *prompt* bem construído tem cinco camadas: o **papel** (quem a IA deve ser), o **contexto** (seu tema e seu nível), a **tarefa** (em um verbo), o **formato** (tabela, lista) e as **restrições** (o que evitar). Em computação, o prompt é uma especificação, e especificar bem é metade do trabalho; em termos filosóficos, lembra que o sentido depende do jogo de linguagem em que se fala (Wittgenstein), e que dizer é também fazer.

Exemplo de prompt completo

```
Papel: Você é um professor de Filosofia Antiga.
Contexto: Sou graduando e escrevo um TCC sobre a noção de phronesis em Aristóteles.
Tarefa: Proponha 3 recortes de problema de pesquisa.
Formato: Tabela com recorte, pergunta e o que fica de fora.
Restrições: Não cite livros nem autores que você não possa indicar com certeza. Diga
quando não souber.
```

A restrição final não elimina a alucinação, mas reduz o risco de invenção. Nunca trate as referências sugeridas pela IA como verificadas.

Onde buscar

Separe sempre dois tipos de ferramenta. As **bases** são aquelas em que a fonte existe de fato: o Google Acadêmico, o SciELO, o Portal de Periódicos da CAPES, o PhilPapers (índice especializado em Filosofia) e a *Stanford Encyclopedia of Philosophy*. As **IAs de apoio** ajudam a localizar e resumir, mas seu resultado precisa ser conferido na base: além dos assistentes de conversa já citados, Perplexity, Consensus e Elicit buscam literatura e mostram links; o NotebookLM conversa apenas com os seus próprios documentos; o Zotero organiza referências. Os recursos gratuitos mudam sem aviso, mas o procedimento vale para ferramentas equivalentes. Uma dica técnica: as bases funcionam com lógica, e conhecer seus operadores melhora a busca. Use aspas para a expressão exata, AND para exigir os dois termos, OR para aceitar qualquer um e NOT (ou o sinal

de menos, no Google Acadêmico) para excluir um termo, como em *"inteligência artificial" AND (autoria OR plágio) AND "ensino superior"*.

A verificação em quatro passos

Descartes, nas *Meditações metafísicas*, propõe duvidar do que não é certo para chegar a um fundamento seguro. Usamos aqui o espírito do método: **nada que a IA entregue vale como fonte antes de passar pela verificação**. São quatro passos. (1) *Existe?* Procure o título exato em uma base ou no catálogo de uma biblioteca. (2) *Os dados conferem?* Autor, ano, veículo e, quando houver, o DOI, identificador permanente do documento. (3) *Diz isso?* Leia o trecho: o erro mais sutil é a obra existir, mas não dizer o que o modelo lhe atribui. (4) *Sustenta?* A fonte serve ao seu argumento ou apenas enfeita o texto? Os passos 1 e 2 são mecânicos; os passos 3 e 4 exigem leitura e juízo, e é a parte que não se delega.

Atividade 1 · Do espanto à referência verificada

Faça no computador, de preferência com um editor de texto aberto para registrar o que fizer. O produto final são duas coisas: **um problema de pesquisa em uma frase** e **um quadro de referências verificadas**.

1. Escreva em duas linhas o seu espanto: o tema ou o incômodo filosófico que realmente lhe chama a atenção.
2. Monte um prompt com as cinco camadas e peça à IA três recortes de problema de pesquisa. Escolha um e escreva seu problema em uma frase, em forma de pergunta.
3. Peça à IA três referências para o recorte escolhido.
4. Verifique cada uma pelos quatro passos e monte um quadro simples com cinco colunas: referência, existe?, dados conferem?, diz isso?, sustenta?; e classifique cada uma como **real**, **real com erro** (autor, ano, título ou livro trocados) ou **inexistente**.
5. Troque o quadro com um colega e confira uma referência dele pelos quatro passos.

Exemplo resolvido (didático). Tema: a prudência (*phronesis*) em Aristóteles. "ARISTÓTELES. *Ética a Nicômaco*, livro VI": existe, os dados conferem, trata da prudência e sustenta o argumento, logo **real**. "SILVA, M. *A prudência aristotélica na era digital*. Revista de Ética Aplicada, 2021" (referência inventada para este exemplo): não se encontra em nenhuma base, logo **inexistente**. "ARISTÓTELES. *Ética a Nicômaco*, livro X, sobre a prudência": a obra existe, mas a prudência é tratada no livro VI (o X trata do prazer e da felicidade), logo **real com erro**.

4 · A IA como interlocutora socrática

A maior utilidade filosófica de um modelo de linguagem não está em escrever por você, mas em **conversar com você**: obrigar a definir termos, apresentar objeções e apontar saltos de raciocínio. Aqui a máquina trabalha a favor do pensamento, e não no lugar dele. Duas tradições da nossa disciplina ensinam como.

Maiêutica

No *Teeteto* (148e-151d), Platão (2001) mostra Sócrates descrevendo-se como parteiro: filho de uma parteira, diz-se estéril de sabedoria; não gera o saber no outro, ajuda a dar à luz e examina se a criança é viável. O procedimento tem momentos reconhecíveis: a pergunta ("o que é X?"), uma definição provisória, a refutação por um caso que a desmente (*élenchos*), a aporia, o impasse produtivo em que se reconhece que não se sabe, e uma nova tentativa. Aplicado à IA, o recurso é simples e poderoso: peça que ela aja como Sócrates, só pergunte e nunca responda. Cada pergunta dela obriga você a definir melhor o conceito antes de escrever.

Disputatio

Nas escolas medievais, a *disputatio* organizava o saber como debate. Nos artigos da *Suma teológica*, Tomás de Aquino segue uma estrutura constante: a questão, as **objeções**, o *sed contra* (uma autoridade em sentido contrário), a **resposta** do mestre (*respondeo*) e a réplica a cada objeção. O valor para a escrita é grande: obriga a enfrentar as melhores objeções *antes* de afirmar. Há um paralelo na computação: antes de lançar um sistema, programadores aplicam **testes adversariais**, tentando deliberadamente quebrá-lo. A *disputatio* faz o mesmo com argumentos: procura as falhas antes que o leitor as encontre. Em termos práticos, peça à IA as objeções mais fortes à sua tese, responda a cada uma e só então redija.

Cuidado com a bajulação

Modelos de linguagem tendem a concordar com o usuário, fenômeno conhecido como bajulação (*sycophancy*). Se a IA elogiar demais a sua tese, escreva: "Discorde. Ataque o ponto mais frágil". No *Górgias* (463a-b), Platão chama de adulação (*kolakeia*) a fala que agrada sem compromisso com a verdade, e a retórica sem filosofia seria um de seus ramos. Uma IA que apenas concorda é adulação em forma de software.

Prompt socrático (copie e cole)

```
Aja como Sócrates. Vou apresentar uma tese. Você só pode fazer perguntas, uma por vez, para me levar a definir melhor meus conceitos e a perceber contradições. Nunca dê a resposta. Comece perguntando o que eu entendo por cada termo central. Minha tese: [cole aqui]
```

Advogado do diabo (copie e cole)

```
Aja como um professor rigoroso da tradição oposta à minha. Apresente as 3 objeções mais fortes à minha tese, cada uma com a premissa em que ataca. Não ofereça soluções. Depois, indique qual objeção eu teria mais dificuldade de responder e por quê. Minha tese: [cole aqui]
```

Caçando falácias

Nas *Refutações sofísticas*, Aristóteles estuda os argumentos que parecem válidos sem o ser; a petição de princípio é tratada nos *Analíticos anteriores* (II, 16). A IA pode ajudar a localizá-los no seu texto, mas também pode cometê-los: peça sempre o trecho exato e confira. Quatro frequentes: a **petição de princípio** (a conclusão já está dentro da premissa: "a IA é confiável porque não erra"); o **falso dilema** (só duas opções: "ou se proíbe a IA, ou se aceita tudo"); o **apelo à autoridade** ("disse o ChatGPT, então é verdade"); e o **ataque à pessoa**, ou *ad hominem* ("esse autor é leigo, logo está errado"). Um prompt útil: "Leia o parágrafo abaixo e indique, se houver, falácias. Para cada uma, cite o trecho exato, dê o nome e justifique. Se não houver, diga que não encontrou."

Atividade 2 · Seu argumento no tribunal da IA

O objetivo é testar uma tese sua contra perguntas e objeções fortes, reformulá-la e conseguir explicá-la em um minuto, sem ler.

1. Escreva sua tese em 3 a 5 linhas (pode ser a hipótese da Atividade 1).
2. Rode o prompt socrático por cinco ou seis rodadas e registre as perguntas da IA. Responda com as suas palavras; observe onde suas definições eram frágeis.
3. Rode o advogado do diabo e escolha a objeção mais forte, aquela que você teria mais dificuldade de responder.
4. Releia a tese procurando as quatro falácias acima.
5. Reescreva a tese respondendo à objeção escolhida.
6. Explique a nova tese à sua dupla em um minuto, sem ler, e ouça: ela entendeu? Qual ponto ficou obscuro?

5 · Estruturar, revisar e responder por si

A IA como leitora crítica

Depois de escrito o rascunho, a IA pode funcionar como leitora crítica: pergunta e aponta, sem escrever no seu lugar. Aristóteles (*Poética*, cap. 7) observa que um todo bem composto tem começo, meio e fim, e que cada parte se liga às outras; a estrutura mínima de um artigo ou TCC retoma a ideia: **introdução** (problema, tese, método), **desenvolvimento** (argumentos apoiados em fontes) e **conclusão** (que responde à pergunta da introdução). Para cada parte há uma boa pergunta à IA: o problema está claro? Qual tese você entendeu? Qual passo do argumento está sem justificativa? A conclusão responde à pergunta da introdução? Uma técnica útil é o *esboço reverso*: peça a ideia principal de cada parágrafo em uma frase e veja se a sequência é lógica. Mantenha o histórico de versões do seu editor de texto: ele ajuda a demonstrar a autoria.

Prompt do leitor crítico (copie e cole)

```
Leia a introdução abaixo como um orientador exigente. Responda: (1) qual é o problema? (2) qual é a tese? (3) qual é o método? (4) o que está vago? Não reescreva. Texto: [cole aqui o seu parágrafo]
```

O que delegar e o que não delegar

Algumas tarefas podem ser entregues à máquina com pouco risco: corrigir ortografia e gramática, formatar referências (e conferi-las), sugerir sinônimos, gerar perguntas para revisão. Outras exigem cuidado: resumir textos (leia sempre o original), sugerir estruturas, reformular parágrafos seus, traduzir trechos (confira os termos técnicos, que se distorcem com facilidade). E há o que não se delega: escolher o problema e a tese, interpretar os autores, decidir o que é verdadeiro, assinar e responder pelo texto. O critério é aristotélico. Na *Ética a Nicômaco* (livro VI), Aristóteles distingue a técnica (*poiesis*), que produz algo, da prudência (*phronesis*), a capacidade de deliberar bem sobre o que fazer em situações concretas. A técnica pode ser transferida a ferramentas; a prudência, não: ninguém delibera por você.

Ética, privacidade e lei

Floridi e Cowls (2019) propõem cinco princípios para uma IA socialmente benéfica: beneficência, não maleficência, autonomia, justiça e explicabilidade. Traduzidos para a escrita: use a IA para aprender mais, e não para aprender menos; não prejudique outros (nada de plágio nem de dados pessoais em ferramentas públicas); preserve sua capacidade de decidir e de pensar; reconheça os vieses e as fontes que a IA tende a omitir; e saiba dizer como chegou ao texto, declarando o uso. Jonas (2006) defende que a responsabilidade seja proporcional ao poder das técnicas: quanto mais poderosa a ferramenta, maior o dever de cuidado de quem a usa. A UNESCO (Miao e Holmes, 2023) propõe, para o uso de IA generativa na educação e na pesquisa, uma abordagem centrada no humano, que preserve a agência de estudantes e pesquisadores.

Na prática: não cole dados pessoais, nem trabalhos inéditos ou sigilosos de terceiros, em ferramentas públicas (a Lei nº 13.709/2018, a LGPD, protege dados pessoais); respeite os direitos autorais (Lei nº 9.610/1998), citando e referenciando as fontes; e confira as regras de sua instituição, de seu orientador e do periódico sobre o uso de IA, que variam e podem exigir declaração (veja a seção seguinte).

O que dizem as normas brasileiras hoje

Em 2026, o Brasil ainda não tem uma lei geral sobre inteligência artificial (o PL 2338/2023 segue em tramitação na Câmara dos Deputados), mas já existem orientações e normas setoriais que pesam sobre quem escreve na universidade. Em comum, elas não proíbem o uso: exigem **transparência**, **responsabilidade de quem assina** e **políticas próprias de cada instituição**.

O Conselho Nacional de Educação aprovou, em 1º de setembro de 2026, o **Parecer CNE/CP nº 15/2026**, com as Diretrizes para a Utilização da Inteligência Artificial na Educação Brasileira. Para a educação superior, o texto determina que cada instituição estabeleça políticas próprias para o uso de IA no ensino, na pesquisa e na extensão, com princípios de autoria, integridade acadêmica, transparência, responsabilidade intelectual e proteção de direitos, podendo prever formas de declaração de uso. Cabe ao estudante a responsabilidade intelectual pela seleção, verificação e veracidade das informações, e a transparência quanto ao uso da IA. Na pesquisa, a IA deve ser declarada sempre que sua contribuição for relevante para a elaboração de textos, a análise de dados, a revisão bibliográfica ou a programação, e ela não transfere a autoria nem a responsabilidade, que permanecem com as pessoas pesquisadoras. O texto veda também a sanção acadêmica fundada exclusivamente em ferramenta de detecção de conteúdo gerado por IA e o uso de conteúdos produzidos por estudantes para treinar modelos de terceiros sem autorização específica. **Atenção ao status:** a resolução que institui essas diretrizes ainda depende de homologação pelo MEC e de publicação; confira se isso já ocorreu.

O CNPq, na **Portaria nº 2.664/2026** (DOU de 11/03/2026), instituiu sua Política de Integridade na Atividade Científica. Na pesquisa apoiada pelo CNPq (por exemplo, com bolsas), é preciso declarar o uso de ferramentas de IA generativa em qualquer fase (concepção, redação, análise de dados, submissão), especificando a ferramenta e a finalidade; é vedado apresentar conteúdo gerado por IA como se fosse de autoria humana, e os autores respondem integralmente por plágios e imprecisões. O MEC divulgou em 2026 o *Referencial para o Uso e Desenvolvimento Responsáveis de IA na Educação*, que destaca a supervisão humana efetiva, a transparência e a explicabilidade, e a proteção de dados e da privacidade de estudantes e docentes.

Quanto às normas técnicas, valem a ABNT **NBR 6023:2025** (referências), a **NBR 10520:2023** (citações) e a **NBR 14724:2024** (trabalhos acadêmicos); até onde pudemos verificar, não há norma da ABNT específica para citar IA generativa. Por isso a prática mais segura é a que este minicurso adota: não tratar a IA como fonte de autoridade, citar apenas fontes verificadas e **declarar o uso** (ferramenta, finalidade e limites). Se sua instituição ainda adotar a edição de 2018 da NBR 6023, siga a orientação dela. Em todos os casos, a regra da sua instituição, do seu curso e do seu orientador vale mais do que qualquer guia genérico: pergunte e registre.

Declarar o uso

Declarar o uso de IA é o oposto do plágio: quem declara assume a responsabilidade pelo que escreveu. O modelo abaixo pode ser adaptado e anexado ao trabalho quando a instituição ou o orientador pedirem. Um teste prático, de inspiração kantiana (a publicidade como critério): se você teria vergonha de declarar o uso, reveja-o.

Declaração de uso de inteligência artificial (modelo)

Trabalho: [título]. Autor(a): [nome]. Data: __/__/____.

Ferramenta(s): [ex.: ChatGPT, Gemini ou Claude, versão gratuita].

Finalidade do uso: [ex.: sugestão de recortes de tema; leitura crítica do parágrafo 3; formatação de referências].

Não utilizei IA para: [ex.: formulação da tese, interpretação das obras, redação do texto final].

Verificação: todas as referências foram conferidas nas fontes originais.

Responsabilidade: assumo integralmente o conteúdo deste trabalho. Assinatura: _____

Atividade 3 · Do rascunho ao texto assinado

É a atividade de síntese: junta tese, escrita, revisão e declaração. **Regra de ouro: o primeiro parágrafo é escrito sem IA.** Ela só entra depois, como leitora crítica.

1. Escreva, sem IA, um parágrafo de seis a oito linhas com a sua tese (a da Atividade 2, de preferência).
2. Cole o parágrafo na IA com o prompt do leitor crítico e anote o que ela apontou.
3. Reescreva o parágrafo por sua conta, decidindo o que aceitar e o que recusar da leitura da IA, e justificando cada decisão para si mesmo(a).
4. Troque o texto com um colega. Cada um lê o do outro com cinco perguntas: o problema está claro nas duas primeiras frases? A tese aparece de forma explícita? Cada afirmação importante tem fonte ou justificativa? Há alguma objeção reconhecida e respondida? O texto soa como a voz de quem o assina?
5. Preencha a declaração de uso de IA do seu trabalho a partir do modelo acima.

6 · Fechamento: que é o homem?

Kant resumiu as perguntas da filosofia em três: *que posso saber? que devo fazer? que me é permitido esperar?* (*Crítica da razão pura*, A805/B833). Na *Lógica* acrescenta uma quarta, que as contém: **que é o homem?** A inteligência artificial devolve essa pergunta com urgência nova: o que fazemos quando uma máquina faz o que julgávamos próprio do humano, escrever? A resposta que este minicurso propõe é modesta e exigente: usar a ferramenta sem abrir mão do que só pode ser nosso, a pergunta, o juízo e a responsabilidade.

Podemos ler as quatro perguntas de Kant com as duas lentes deste minicurso, a da Filosofia e a da Ciência da Computação. A tabela abaixo mostra a correspondência e liga cada pergunta aos compromissos que vêm a seguir. É uma leitura proposta para fins didáticos, e não uma tese de Kant sobre a inteligência artificial.

Pergunta de Kant	Na Filosofia	Na Computação e no uso da IA	Compromisso
Que posso saber?	Os limites do conhecimento: só sabe quem examina (*Sapere aude*).	A IA prevê o verossímil e pode alucinar. Provável não é verdadeiro.	Eu verifico tudo
Que devo fazer?	Agir com autonomia e responsabilidade: autor é quem decide, responde e justifica.	Transparência, privacidade e integridade: declarar o uso e proteger dados.	Eu protejo dados; eu declaro e respondo
Que me é permitido esperar?	Esperança sem ingenuidade: a técnica a serviço do esclarecimento, e não da tutela.	A IA como interlocutora que testa nossos argumentos, em vez de pensar por nós.	Eu testo meus argumentos
Que é o homem?	O que a máquina não assume por nós: a pergunta, o juízo e a responsabilidade.	Escrever com a máquina sem terceirizar o pensar: um bom pedido exige ideia própria.	Eu penso primeiro; eu pergunto melhor

Para levar, seis compromissos. **Eu penso primeiro**: defino o problema e a tese antes de pedir ajuda. **Eu pergunto melhor**: uso prompts com papel, contexto, tarefa, formato e restrições. **Eu verifico tudo**: nenhuma referência entra no texto sem os quatro passos. **Eu testo meus argumentos**: uso a IA para encontrar objeções, e não para evitá-las. **Eu protejo dados**: não exponho dados pessoais nem trabalhos de terceiros. **Eu declaro e respondo**: informo o uso de IA e assino o que escrevo. Escolha três deles e comece esta semana.

Referências

ARENDT, Hannah. **A condição humana**. 10. ed. Rio de Janeiro: Forense Universitária, 2007.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023:2025**; **NBR 10520:2023**; **NBR 14724:2024**. Rio de Janeiro: ABNT.

BENDER, Emily M. et al. On the dangers of stochastic parrots: can language models be too big? In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021. **Proceedings** […]. New York: ACM, 2021. p. 610-623.

BRASIL. Ministério da Educação. Conselho Nacional de Educação. **Parecer CNE/CP nº 15/2026**: Diretrizes para a Utilização da Inteligência Artificial na Educação Brasileira. Aprovado em 1º set. 2026.

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico. **Portaria CNPq nº 2.664, de 6 de março de 2026**: Política de Integridade na Atividade Científica. Diário Oficial da União, 11 mar. 2026.

BRASIL. **Lei nº 9.610/1998** (direitos autorais); **Lei nº 13.709/2018** (Lei Geral de Proteção de Dados Pessoais); **Projeto de Lei nº 2.338/2023** (marco legal da IA, em tramitação).

BRASIL. Ministério da Educação. **Referencial para o Uso e Desenvolvimento Responsáveis de Inteligência Artificial na Educação**. 2026.

BROWN, Tom B. et al. Language models are few-shot learners. **Advances in Neural Information Processing Systems**, v. 33, 2020.

FLORIDI, Luciano; COWLS, Josh. A unified framework of five principles for AI in society. **Harvard Data Science Review**, v. 1, n. 1, 2019.

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

JI, Ziwei et al. Survey of hallucination in natural language generation. **ACM Computing Surveys**, v. 55, n. 12, art. 248, p. 1-38, 2023.

JONAS, Hans. **O princípio responsabilidade**. Rio de Janeiro: Contraponto; Ed. PUC-Rio, 2006.

KAPLAN, Jared et al. Scaling laws for neural language models. **arXiv**:2001.08361, 2020.

KANT, Immanuel. Resposta à pergunta: que é esclarecimento? In: **Textos seletos**. Petrópolis: Vozes, 2005.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. **Advances in Neural Information Processing Systems**, v. 25, 2012.

LÉVY, Pierre. **As tecnologias da inteligência**. Rio de Janeiro: Editora 34, 1993.

McCARTHY, John et al. **A proposal for the Dartmouth summer research project on artificial intelligence**. 1955.

MIAO, Fengchun; HOLMES, Wayne. **Guidance for generative AI in education and research**. Paris: UNESCO, 2023.

ONG, Walter J. **Oralidade e cultura escrita**. Campinas: Papirus, 1998.

OUYANG, Long et al. Training language models to follow instructions with human feedback. **Advances in Neural Information Processing Systems**, v. 35, 2022.

PLATÃO. **Fedro**. Trad. Carlos Alberto Nunes. Belém: Ed. UFPA, 2011. **Teeteto-Crátilo**. Belém: Ed. UFPA, 2001.

SEARLE, John R. Minds, brains, and programs. **Behavioral and Brain Sciences**, v. 3, n. 3, p. 417-424, 1980.

TURING, Alan M. Computing machinery and intelligence. **Mind**, v. 59, n. 236, p. 433-460, 1950.

VASWANI, Ashish et al. Attention is all you need. **Advances in Neural Information Processing Systems**, v. 30, 2017.

WITTGENSTEIN, Ludwig. **Philosophical investigations**. Oxford: Blackwell, 1953.

Clássicos citados por obra e passagem (use a edição de sua preferência): ARISTÓTELES, *Ética a Nicômaco* (VI), *Poética* (7 e 9), *Metafísica* (A, 2), *Refutações sofísticas*, *Analíticos anteriores* (II, 16); PLATÃO, *Górgias* (463a-b); TOMÁS DE AQUINO, *Suma teológica*; DESCARTES, *Meditações metafísicas*; KANT, *Crítica da razão pura* (A805/B833), *Lógica*.