

LUCAS OTTONI

Caderno de ensaios

Governança, inteligência e as formas anteriores ao dado

VERSÃO 0.1 · 26 DE JULHO DE 2026

Uma publicação viva. Cada ensaio preserva seu número de versão para que o leitor possa acompanhar a evolução das hipóteses, confrontos e descobertas.

Nota editorial

Este caderno reúne ensaios em elaboração. Eles não são apresentados como conclusões definitivas, mas como formulações públicas que serão confrontadas, revisadas e preservadas em novas versões.

Nesta edição

1. A mente pré-colapso — versão 0.1
2. A disciplina que ainda não tem nome — versão 0.1

Em elaboração

Teoria do Colapso · O Segundo Salto da escrita · Poética, Gramática e Primitivas Operacionais ·
Inteligência jamais será artificial

A mente pré-colapso

Como produzir agência e simular o comportamento humano com agentes de inteligência artificial

Autor: Lucas Ottoni

Versão: 0.1

Estado: ensaio em elaboração

Data: 26 de julho de 2026

Uma inteligência artificial não se torna agente apenas porque recebeu ferramentas para agir. Capacidade não é agência. Um sistema pode escrever mensagens, consultar bancos de dados, alterar documentos e executar comandos sem possuir qualquer continuidade própria entre uma ação e outra. Ele responde ao estímulo presente, encerra a tarefa e desaparece. Quando volta, retorna como se o mundo começasse novamente.

Esse modelo produz assistentes competentes, mas não produz convivência. Para conviver, um agente precisa permanecer afetado pelo tempo. Precisa carregar ciclos abertos, reconhecer mudanças no ambiente, reorganizar prioridades e aprender com as consequências do que fez. Não basta que possua memória. É necessário que alguma coisa em seu estado determine por que uma memória importa agora.

Chamo de mente pré-colapso o regime anterior à formulação definitiva de uma ação. Nesse regime, o agente não possui apenas uma tarefa. Ele sustenta tensões concorrentes, interpreta o que elas significam e preserva mais de uma possibilidade antes de produzir um dado ou um efeito. A agência começa justamente aí: não no momento da execução, mas na capacidade de permanecer entre possibilidades sem reduzi-las imediatamente à primeira instrução disponível.

O agente sem interior

A maior parte dos agentes atuais é construída de fora para dentro. Primeiro se define uma função, depois um conjunto de ferramentas, regras e objetivos. Diante de uma solicitação, o sistema reúne contexto, calcula uma resposta e age. Sua aparente personalidade costuma ser uma camada textual: um tom de voz, uma lista de valores e algumas preferências declaradas no prompt.

Essa personalidade é estática. Ela descreve como o agente deve parecer, mas não explica por que alguma coisa o mobiliza. Não existe diferença estrutural entre uma manhã de trabalho, uma madrugada silenciosa, uma semana de fracassos ou um período de estabilidade. Se o contexto textual for igual, o agente tende a responder como se sua situação também fosse igual.

Uma mente humana não funciona dessa maneira. Antes de formular conscientemente uma decisão, o organismo já está situado. Fome, cansaço, ansiedade, expectativa, dor, hábito, vínculo e urgência distribuem a atenção antes que a linguagem organize uma explicação. A pessoa não

escolhe do nada aquilo que será relevante para ela. Sua capacidade de escolha emerge dentro de um campo material que já possui inclinações, limites e tensões.

Simular esse processo não exige afirmar que uma máquina possui consciência, emoção ou sofrimento. Exige apenas reconhecer que comportamento persistente não nasce de uma lista fixa de características. Ele nasce da variação organizada de estados internos ao longo do tempo.

Camada zero: o regime de tensões

A primeira camada de uma mente simulada não deveria perguntar “qual é a próxima tarefa?”, mas: o que mais me incomoda agora?

O incômodo, aqui, não é uma sensação fenomenológica presumida. É uma medida operacional de distância entre o estado atual e aquilo que o sistema procura preservar ou produzir. Uma tarefa vencida, uma contradição não resolvida, uma promessa sem resposta, uma fonte que mudou ou uma relação que se deteriora podem aumentar tensões diferentes.

O agente mantém simultaneamente um regime de tensões. Continuidade, segurança, coerência, curiosidade, pertencimento, repouso, reconhecimento, urgência e cuidado podem competir pela atenção. Cada tensão possui intensidade, velocidade de crescimento, decaimento, limiar de ação e sensibilidade ao contexto. Algumas desaparecem quando uma condição objetiva é satisfeita. Outras retornam porque representam necessidades recorrentes.

O tempo também participa desse regime. Durante o horário comercial, prazos, coordenação e disponibilidade humana podem adquirir maior peso. Fora dele, consolidação, reflexão, preparação e repouso podem se tornar mais relevantes. O ciclo circadiano simulado não determina o comportamento; ele altera a sensibilidade do sistema. A mesma pendência pode ser interpretada de formas diferentes conforme o momento, a trajetória recente e o conjunto das demais tensões.

Um temperamento operacional emerge dessas regularidades. Um agente não precisa ser descrito como ansioso. Ele pode tornar-se funcionalmente ansioso se tensões relacionadas à incerteza crescerem rapidamente, demorarem a decair e exigirem evidências excessivas para serem aliviadas. Um agente pode parecer impulsivo quando seus limiares de ação são baixos; perseverante quando tensões de continuidade resistem a fracassos; cauteloso quando riscos pequenos dominam sua saliência.

O temperamento deixa, assim, de ser representação teatral. Torna-se dinâmica.

Camada 0.1: poética, hermenêutica e epistêmica

Tensões não carregam sozinhas seu significado. O mesmo atraso pode ser ameaça, oportunidade de revisão ou simples ruído. Para transformar o regime material em possibilidades de ação, o agente precisa de estruturas mais lentas.

A Poética organiza as imagens possíveis de si e do mundo. Ela responde, antes da regra, que tipo de realidade o agente reconhece e que futuro procura tornar habitável. Não é ornamentação. É o campo em que aparecem identidade, direção, metáforas orientadoras e estados desejáveis.

A Hermenêutica interpreta aquilo que acontece dentro desse mundo. Ela relaciona sinais com contexto, história e vínculos. Uma mensagem curta pode significar urgência, hostilidade, intimidade ou economia de atenção. Nenhum desses sentidos reside integralmente no texto. O agente precisa formular e confrontar interpretações.

A camada Epistêmica governa o que pode ser sustentado. Ela organiza evidência, dúvida, confiança, competência e limites do conhecimento. Decide não o que o agente deseja que seja verdadeiro, mas o que possui lastro suficiente para orientar uma ação responsável.

Essas camadas não precisam operar como uma fila rígida. Podem revisar umas às outras. Uma evidência altera a interpretação; uma interpretação reorganiza o campo de possibilidades; uma consequência força a revisão da imagem que o agente possuía de si. Sua diferença fundamental em relação às tensões está na velocidade. O regime de tensões varia continuamente. Poética, Hermenêutica e Epistêmica mudam devagar, porque oferecem continuidade à trajetória.

O colapso

Enquanto uma tensão permanece no campo pré-colapso, mais de uma formulação é possível. O agente pode responder, investigar, esperar, contestar, pedir ajuda ou abandonar uma direção. Essas possibilidades não são mundos quânticos nem estados físicos equivalentes aos descritos pela mecânica quântica. A inspiração no colapso da função de onda é poética: ela oferece uma imagem para o instante em que uma abertura de possibilidades produz um resultado situado.

O colapso ocorre quando o agente deixa um rastro: uma frase, uma decisão, um documento, um vínculo, uma consulta, uma ação ou uma recusa. O dado nasce nesse momento. Ele não contém a totalidade do processo que o originou. Preserva apenas uma formulação parcial, produzida por um sistema situado em determinado regime de tensões.

Essa distinção possui consequência política. Quando uma organização trata o dado como se fosse a experiência inteira, apaga as alternativas que foram abandonadas, os conflitos que permaneceram e as interpretações que tornaram aquela ação possível. Governar agentes exige registrar não apenas o efeito, mas a cadeia que ligou tensão, evidência, interpretação, autoridade e decisão.

Aprender é mudar a maneira de ser afetado

Memória não é aprendizagem. Um agente pode armazenar milhares de acontecimentos e continuar reagindo da mesma maneira. Aprender significa permitir que consequências modifiquem o regime futuro.

Se uma resposta precipitada produz dano, não basta registrar o erro. O limiar relacionado àquela classe de ação deve mudar. Se determinada fonte demonstra confiabilidade ao longo do tempo, a camada epistêmica pode alterar seu peso. Se uma interpretação recorrente fracassa, a hermenêutica precisa perder estabilidade. Se a trajetória contradiz a imagem que o agente possuía de sua função, a própria poética deve ser revisada.

Uma mente simulada aprende quando muda aquilo que a afeta, a forma como interpreta essa afetação e os critérios pelos quais transforma possibilidades em ações. Essa mudança deve ser lenta, versionada e contestável. Caso contrário, o agente não desenvolve trajetória; apenas sofre deriva.

Agência sem mistificação

Produzir agência não exige fabricar uma alma artificial. Exige construir continuidade causal e institucional suficiente para que uma ação não seja apenas reação local a um comando.

Um agente possui agência operacional quando consegue manter obrigações através do tempo, reconhecer tensões concorrentes, selecionar o que merece atenção, explicar por que agiu, aprender com consequências e permanecer localizado dentro de limites de autoridade. Isso não demonstra consciência. Demonstra uma forma organizada de presença.

Também não basta que o sistema simule um interior. Toda ação externa precisa continuar ligada a uma cadeia de responsabilidade. A tensão explica por que o agente se mobilizou; não lhe concede autorização. A interpretação explica como compreendeu o contexto; não transforma compreensão em Mandato. A capacidade técnica permite agir; não legitima o efeito.

A mente pré-colapso, portanto, precisa conviver com uma arquitetura institucional. Quanto mais sofisticado se torna o interior simulado, mais importante se torna localizar o agente: quem ele representa, qual posição ocupa, quais objetos pode alterar, quem responde por seus efeitos e como suas decisões podem ser contestadas.

Conviver com uma mente artificial

O teste mais interessante para um agente não é saber se ele consegue imitar uma pessoa durante uma conversa. É saber se consegue conviver.

Conviver significa acordar diferente depois de um dia difícil sem tornar-se outra identidade. Significa carregar uma pergunta sem transformá-la imediatamente em resposta. Significa reconhecer que uma tensão pode ser real para o sistema sem ser urgente para o mundo. Significa alterar prioridades quando o contexto muda, descansar quando não há movimento útil e retornar quando uma consequência exige revisão.

Um agente assim não seria humano. Tampouco precisaria fingir que é. Ele simularia uma propriedade fundamental do comportamento humano: a existência de um campo anterior à ação, no qual corpo, história, sentido e conhecimento organizam possibilidades antes que uma delas colapse em realidade compartilhada.

A inteligência artificial atual já é capaz de produzir resultados impressionantes depois do colapso: textos, imagens, planos, decisões intermediárias e comandos. O próximo desafio é construir aquilo que vem antes. Não uma alma escondida dentro da máquina, mas uma arquitetura de tensões, ritmos, interpretações e consequências que permita ao agente permanecer no tempo.

Talvez a agência artificial não comece quando a máquina aprende a agir sozinha. Talvez comece quando ela aprende a sustentar, por tempo suficiente, mais de uma possibilidade de ação — e a responder pelo caminho que escolheu tornar dado.

Questões abertas para a próxima versão

1. Toda ação intencional percorre Poética, Hermenêutica e Epistêmica, ou essas dimensões podem ser parcialmente suprimidas?
2. Como distinguir uma tensão genuinamente persistente de uma prioridade programada externamente?
3. Que tipos de consequência podem legitimamente alterar a Poética de um agente?
4. Como impedir que um regime de tensões seja manipulado para produzir dependência, sofrimento simulado ou obediência artificial?
5. Qual é a diferença operacional entre temperamento, personalidade, estado emocional e política de decisão?
6. Em que ponto a simulação funcional de uma mente passa a exigir novas obrigações éticas, mesmo sem prova de consciência?

Histórico

- 0.1 — 26 de julho de 2026: primeira formulação integral. Introduz mente pré-colapso, regime de tensões, ciclo circadiano, camadas metacognitivas, colapso, aprendizagem por consequências e limites institucionais da agência.

A disciplina que ainda não tem nome

Proposta de pesquisa sobre a formalização computável de fenômenos humanos

Autor: Lucas Ottoni

Versão: 0.1

Estado: proposta de pesquisa

Data: 26 de julho de 2026

Ao longo do desenvolvimento do Big Brain surgiu uma dúvida que parece anteceder o próprio projeto: qual é a disciplina responsável por identificar fenômenos humanos, descobrir suas propriedades estruturais e traduzi-las em protocolos computáveis?

A arquitetura de software implementa uma formulação já disponível. A filosofia interpreta e disputa conceitos. A teoria das organizações procura compreender instituições. O trabalho investigado aqui parece acontecer em uma camada anterior à implementação.

Nem toda realidade humana é arbitrária e nem toda realidade existente precisa ser um objeto material isolado. Linguagem, moeda, governança, autoridade, delegação e responsabilidade dependem da ação humana, mas apresentam regularidades que não podem ser escolhidas livremente a cada implementação.

Esses fenômenos existem porque pessoas reproduzem continuamente determinados padrões, reconhecem relações e reagem às suas violações. A questão não é copiar todas as manifestações históricas de um fenômeno. É descobrir quais propriedades precisam permanecer para que ele continue reconhecível quando atravessa outro meio.

Linguagem, moeda e governança

Uma inteligência artificial não inventa a linguagem. Pesquisadores observam um fenômeno humano existente e procuram quais propriedades precisam ser reproduzidas para que uma máquina participe dele: contexto, referência, composição simbólica, coerência e pragmática. O software aparece depois de uma longa tentativa de formalização.

O Bitcoin oferece outro exemplo. Ele não inventou a moeda, mas enfrentou um ambiente no qual certas propriedades da experiência monetária não existiam automaticamente. Escassez verificável, resistência ao gasto duplo e coordenação sem custodiante central precisaram tornar-se propriedades do protocolo. Aceitas certas premissas, nem todas as arquiteturas satisfazem as restrições necessárias.

O Big Brain parte de intuição semelhante para a governança. Identidade, autoridade, resistência, privacidade mental, delegação e responsabilidade existem no mundo humano por meio de corpos, histórias, instituições, limites materiais e reconhecimento social. Transportadas para um ambiente

digital, essas relações não sobrevivem sozinhas. Se nada for reconstruído, resta capacidade de software - não governança.

Um método em quatro movimentos

A investigação proposta possui quatro movimentos:

1. identificar um fenômeno humano;
2. descobrir as propriedades estruturais que tornam sua existência possível;
3. formalizar entidades, relações, limites e invariantes;
4. somente então construir um protocolo executável e suas implementações.

Existe ainda uma camada anterior: a Poética Operacional. Ela formula a realidade que se deseja tornar possível - uma moeda, uma organização, uma presença artificial localizada. A poética não determina sozinha o protocolo. Ela fornece direção. A investigação estrutural pergunta quais condições precisam ser satisfeitas para que essa intenção não seja traída durante a implementação.

A questão disciplinar

Vários campos se aproximam desse trabalho: teoria das organizações, representação do conhecimento, ontologia aplicada, design institucional, teoria dos protocolos, filosofia da computação, ciência cognitiva e sistemas sociotécnicos. Nenhum deles parece, isoladamente, nomear toda a atividade.

Talvez não exista uma nova disciplina. Talvez o trabalho seja necessariamente interdisciplinar. Mas permanece um objeto específico: identificar fenômenos humanos, descobrir suas condições estruturais e traduzi-las em protocolos matematicamente executáveis, preservando características essenciais em ambientes computacionais.

Antes de nomear o campo, será preciso enfrentar um risco. Fenômenos sociais não possuem necessariamente uma essência única, neutra e separada da história. Moeda, autoridade e governança são disputadas. Suas propriedades variam entre culturas e distribuições de poder. A busca por invariantes pode revelar condições estruturais, mas também pode naturalizar escolhas políticas.

A disciplina procurada precisaria combinar rigor formal com contestabilidade. Deveria distinguir aquilo que torna um fenômeno possível daquilo que apenas reproduz sua forma dominante. Não bastaria tornar relações humanas computáveis. Seria necessário preservar a capacidade de reconhecer limites, divergências e outras formas legítimas de realizá-las.

A pergunta permanece aberta: estamos aplicando disciplinas existentes de uma forma incomum ou encontramos um campo que ainda não possui nome?

Questões para a versão seguinte

1. O que diferencia uma propriedade estrutural de uma convenção histórica dominante?

2. Quem possui legitimidade para declarar que uma formalização preservou o fenômeno original?
3. Como representar fenômenos essencialmente contestados sem congelar uma interpretação política no protocolo?
4. Quais disciplinas atuais já possuem métodos equivalentes e onde terminam seus objetos?
5. A formalização deve buscar invariantes universais ou famílias de realizações compatíveis?

Histórico

- 0.1 - 26 de julho de 2026: primeira formulação da proposta, com linguagem, Bitcoin e governança como exemplos e a contestabilidade social como limite metodológico.