

# Estimulando a Metacognição de Estudantes por meio da Tutoria Virtual com IA Generativa

Vitória C. S. Camelo<sup>1,2</sup>, Claurton A. Siebra<sup>1,2</sup>

<sup>1</sup>Centro de Informática – Universidade Federal da Paraíba (UFPB)  
58055-000 – João Pessoa – PB – Brasil

<sup>2</sup>SAIL – Social Artificial Intelligence Lab – Universidade Federal da Paraíba  
58055-000 – João Pessoa – PB – Brasil

vitoria.cristhyna@academico.ufpb.br, claurton@ci.ufpb.br

**Abstract.** *Generative AI has emerged as a potential solution to extend the abilities of tutoring systems. However, while beneficial at first glance, features such as high flexibility and ability to create ready-to-use answers also pose risks to the educational process, potentially fostering dependencies that reduce students' critical thinking, creativity, and problem-solving abilities. This paper proposes the use of generative AI based on a metacognitive teaching model, encouraging students to reflect on their learning process by planning and evaluating their actions, going beyond the traditional question-answer paradigm. The evaluation was conducted via a multi-layered analytical framework, which highlighted the potential benefits for learning.*

**Resumo.** *A IA generativa surgiu como uma solução promissora para os sistemas de tutoria. Porém, embora benéfica à primeira vista, características como alta flexibilidade e capacidade de gerar respostas prontas geram riscos para o processo educacional, podendo fomentar dependências que reduzem o pensamento crítico, a criatividade e a capacidade de resolução de problemas dos estudantes. Este artigo propõe o uso da IA generativa com base em um modelo de ensino metacognitivo, o qual incentiva estudantes a refletirem sobre seu próprio processo de aprendizagem por meio do planejamento e da avaliação de suas ações. Como validação, foi usado um modelo analítico em múltiplas camadas, o qual evidencia os potenciais benefícios da abordagem.*

## 1. Introdução

A Inteligência Artificial (IA) generativa está transformando a educação em diversos níveis e disciplinas, ao personalizar conteúdos, apoiar professores e aumentar o engajamento dos estudantes. De fato, essa tecnologia apresenta um grande potencial para democratizar o acesso ao conhecimento, sendo os estudantes alguns dos seus primeiros e mais frequentes usuários (e.g., ChatGPT) [Chan and Lee 2023]. No entanto, recentes discussões [De Cremer et al. 2023; Mittal and Chamola 2024] têm mostrado preocupações quanto ao uso indevido da IA generativa em contextos educacionais, especialmente porque a maioria desses modelos não foi projetada com fundamentos pedagógicos. Ou seja, em vez de promover a aprendizagem, esses modelos são otimizados para serem úteis, uma definição que frequentemente tende a entrar em conflito com uma educação eficaz. Por exemplo, os estudantes podem facilmente obter respostas diretas para as tarefas, em vez de se engajarem na resolução de problemas e condução de prática [De Cremer et al. 2023]. Além disso, o uso de IA para geração de respostas por parte do aluno pode criar

uma falsa ilusão de domínio do assunto, resultando em lacunas de aprendizagem ao longo do tempo [Bauschard e Quidwa 2024]. Enfrentar esses desafios exige uma integração cuidadosa da IA generativa na educação, garantindo que ela fortaleça, e não prejudique, o processo de aprendizagem.

Uma abordagem para tornar a IA generativa mais eficaz na educação é integrar sua flexibilidade com teorias da aprendizagem, que explicam como as pessoas adquirem conhecimento, desenvolvem habilidades e mudam comportamentos. Diversas teorias fundamentais influenciaram a educação, cada uma destacando diferentes aspectos do processo de aprendizagem. Por exemplo, o behaviorismo [Ng'andu et al. 2013] propõe que a aprendizagem ocorre por meio do condicionamento, no qual as respostas a estímulos são reforçadas ou enfraquecidas por recompensas e punições. Em contraste, o construtivismo [Steffe and Gale 1995] enfatiza a aprendizagem como um processo ativo, no qual os indivíduos constroem sua própria compreensão com base nas experiências. No entanto, apesar do crescente número de pesquisas sobre IA generativa na educação, revisões recentes não exploram como alinhar essa tecnologia às teorias de aprendizagem consolidadas [Mittal et al. 2024], deixando uma lacuna crítica na discussão.

Nesse contexto, o presente artigo explora soluções para essas limitações por meio do novo conceito de Tutoria Metacognitiva Generativa (TMG). Na educação, a metacognição [Mahdavi 2014] refere-se à consciência e à regulação do próprio pensamento e dos processos de aprendizagem. Ela permite que os estudantes planejem, monitorem e avaliem suas estratégias de aprendizagem de modo a melhorar o desempenho acadêmico. Educadores utilizam a metacognição para promover uma aprendizagem mais profunda, o pensamento crítico e habilidades de resolução de problemas [Mahdavi 2014]. Ao incorporar a TMG, buscamos:

- Incentivar a aprendizagem autorregulada, orientando os estudantes a definirem metas de aprendizagem, monitorar seu progresso e ajustar estratégias com base em sua compreensão;
- Apoiar estratégias de resolução de problemas, ajudando os estudantes a dividirem problemas complexos em etapas gerenciáveis;
- Promover a reflexão e a autoavaliação, incentivando os estudantes a avaliarem sua aprendizagem e identificarem áreas de melhoria.

Portanto, a combinação da IA generativa com os princípios da metacognição possibilita o desenvolvimento de tutores virtuais que não apenas transmitem conhecimento, mas também capacitam os estudantes a aprenderem a aprender, promovendo maior autonomia e pensamento crítico ao longo de seu desenvolvimento educacional.

## 2. Trabalhos Relacionados

Diversos artigos avaliam o uso da IA generativa aplicada, especialmente, à educação fundamental, considerando sua forma mais tradicional (chatbots de livre conversação) [Bai et al. 2024] ou geração de conteúdo sob demanda [Chen et al. 2023]. No entanto, outros trabalhos tentam avançar esse uso para formas não convencionais. Por exemplo, o estudo de Xie et al. (2025) discute técnicas de *fine-tuning* (refinamento) para aprimorar modelos de IA de forma a alinhá-los melhor com os níveis de linguagem e as necessidades dos estudantes. De forma semelhante, o estudo de Lee e Zhai (2024) menciona a possibilidade de desenvolver ferramentas baseadas em LLMs (Modelos de Linguagem de

Grande Escala) adaptadas a conteúdos científicos específicos e métodos instrucionais, bem como o papel de modelos de código aberto como Llama, Bloom e Falcon para promover um uso de IA mais equitativo e transparente.

O estudo de Lee (2024) também aborda a importância do *fine-tuning*, mas com foco no aspecto de eficiência computacional. Assim, a ideia é reduzir os requisitos computacionais mantendo o desempenho, com o desenvolvimento de modelos como Llama2 e Polyglot, otimizados para contextos educacionais. Os autores em [Li e Zhang 2024] abordam o potencial da IA generativa em automação, interatividade e aprendizagem personalizada por meio de ferramentas como sistemas de tutoria inteligentes, que podem integrar funções como correção automática e monitoramento do progresso em tempo real. Essas ideias estão alinhadas com os avanços no ensino personalizado impulsionado por IA, geração automática de materiais didáticos digitais e análise em tempo real dos dados de aprendizagem dos alunos para melhorar as intervenções em sala de aula [Huang and Yu 2024]. Contudo, o único uso da personalização não garante um processo de aprendizagem eficiente. Por isso, estudos semelhantes ao de Dai (2024) estão avaliando aspectos pedagógicos como forma de melhoria da alfabetização em IA, o que pode ser considerado uma inovação educacional positiva.

Esses e outros trabalhos focados na educação fundamental não exploram como alinhar a tecnologia de IA generativa com paradigmas educacionais consolidados. Mesmo ao extrapolarmos para além da educação fundamental, não encontramos iniciativas que de fato implementem esse alinhamento. Por exemplo, o artigo de Zhou e Schofield (2024) apenas discute a aplicação da IA generativa em ambientes de aprendizagem colaborativa, com base em teorias construtivistas sociais. O trabalho destaca o potencial da IA como participante no processo de aprendizagem, facilitando a cocriação de conhecimento e aprimorando experiências educacionais colaborativas. No entanto, apenas aspectos gerais são discutidos, sem direções pragmáticas para implementações.

O estudo de caso discutido por Santos (2023) analisa o uso de chatbots com IA generativa como tutores virtuais no ensino de física, dentro de um enquadramento construtivista. Ou seja, ele avalia como esses agentes podem estimular o pensamento crítico e as habilidades de resolução de problemas, atuando como "agentes para pensar junto" no processo de aprendizagem. A ideia de construtivismo, porém, não é implementada diretamente nos tutores, e os próprios usuários precisam controlar o tutor por meio de instruções para que ele atue de acordo com os princípios construtivistas. Essa ausência de alinhamento identificada em nossa pesquisa é corroborada pela análise de Griffiths et al. (2024), que evidenciam que os estudos revisados sobre a integração da IA generativa na educação, em sua grande maioria, não se fundamentam em teorias da comunicação e da aprendizagem como base estruturante de suas propostas.

### **3. Ferramentas e Métodos**

#### **4.1. Metacognição**

A metacognição é um conceito que já foi bem explorado no ambiente computacional, principalmente nos trabalhos de Cox, os quais abordaram tanto aspectos teóricos [Cox 2005] como práticos [Cox 2022] para a implementação de arquiteturas cognitivas. No contexto da educação, metacognição é o processo por meio do qual os aprendizes se envolvem ativamente no planejamento, monitoramento e avaliação da própria

aprendizagem, utilizando sua compreensão da tarefa, sua consciência sobre estratégias de aprendizagem eficazes e o conhecimento sobre si mesmos [Mahdavi 2014]. Esse processo de autorregulação permite que os estudantes estabeleçam metas, acompanhem seu progresso e avaliem os resultados, promovendo, assim, uma compreensão mais profunda e uma resolução de problemas mais eficaz. Por exemplo, antes de resolver um problema de matemática, os estudantes podem definir metas alcançáveis, como minimizar erros de cálculo ou aplicar a abordagem correta de resolução, com base em sua consciência sobre seus próprios pontos fortes e fracos na matemática (*conhecimento pessoal*), bem como sua compreensão sobre o tipo específico de problema e suas exigências (*conhecimento da tarefa*). Em seguida, os estudantes podem analisar o problema, dividi-lo em etapas menores e decidir por estratégias como desenhar um diagrama, usar fórmulas algébricas ou revisar sistematicamente sua solução (*conhecimento estratégico*) para melhorar seu processo de resolução de problemas. O potencial da metacognição na educação foi demonstrado em diversos estudos [Tsai et al. 2018; Li e Yuan 2022], os quais discutem estratégias práticas (Tabela 1) para aplicar os princípios metacognitivos em ambientes de aprendizagem.

**Tabela 1. Exemplos de estratégias utilizadas na aplicação de princípios da metacognição**

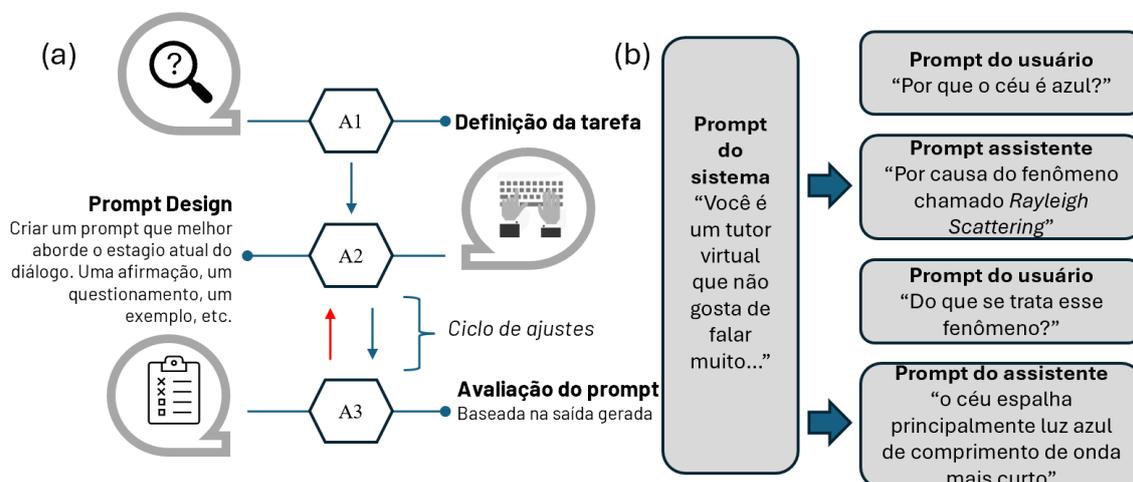
Princípios	Estratégia
Ensino da autorregulação	Usar organizadores gráficos, como mapas conceituais, ou ferramentas de suporte ao planejamento para ajudar os estudantes a estruturarem sua abordagem.
Promoção do automonitoramento	Fornecer recursos de verificação para ajudar os estudantes a acompanharem seu progresso.
Promoção do pensamento reflexivo	Facilitar atividades de autoavaliação ou avaliação por pares, nas quais os estudantes critiquem seu próprio trabalho com base em critérios específicos.
Desenvolvimento da capacidade de aprendizagem adaptativa	Utilizar estudos de caso ou problemas do mundo real para permitir que os estudantes apliquem e ajustem suas estratégias de aprendizagem.

## 4.2. Engenharia de Prompt

Os modelos de IA generativa não foram originalmente desenvolvidos para atuar como assistentes educacionais, portanto, seu comportamento padrão precisa ser ajustado. Existem duas abordagens principais para modificar esse comportamento: a engenharia de prompts e o *fine-tuning* (ajuste fino supervisionado e por reforço). De acordo com Jurenka et al. (2024), o *fine-tuning* oferece o potencial de explorar totalmente as capacidades do aprendizado profundo, possibilitando a integração de estratégias educacionais. No entanto, essa abordagem exige muitos recursos, requerendo contribuições adicionais em ambas as etapas de aprendizado supervisionado e por reforço. A fase de ajuste supervisionado demanda um conjunto de dados que reflita paradigmas pedagógicos, enquanto a etapa de aprendizado por reforço exige a criação de uma função de recompensa para avaliar as interações. Atualmente, a literatura carece de soluções consolidadas para esses desafios.

Por outro lado, a engenharia de prompt é o método mais simples e amplamente utilizado para ajustar o comportamento de IAs generativas. Assim como no desenvolvimento de

software, a engenharia de prompts envolve refinamento iterativo (Figura 1a) para melhorar as respostas do modelo. Portanto, após a definição da tarefa, o prompt, que pode ser uma pergunta, afirmação, exemplo, ou uma combinação desses elementos, é composto e avaliado em um ciclo contínuo de ajustes.



**Figura 1. (a) Engenharia de prompt como um processo de desenvolvimento iterativo, e (b) tipos de prompts**

Em sistemas tradicionais de conversação baseados em IA generativa, a engenharia de prompts é empregada para compor um *prompt* do sistema e múltiplos *prompts* do usuário (Figura 1b). O prompt do sistema serve como uma diretriz geral para a conversa e geralmente inclui regras abrangentes e personas a serem seguidas. Portanto, ele sempre controla o comportamento do assistente (Figura 1b). Em sistemas de conversação atuais, o fluxo de prompts do usuário é, em sua maioria, predefinido ou controlado por meio de menus com opções (e.g., chatbots de assistentes bancários em dispositivos móveis). Porém, esta abordagem é bastante inflexível, não sendo natural para uma sessão de aprendizagem.

Podemos observar que o uso da engenharia de prompt também apresenta limitações. Notadamente, ele requer definições explícitas sobre os comportamentos ideais de tutoria em linguagem natural, o que envolve detalhar quais ações devem ser tomadas e quando, o que deve ser evitado e quais são as exceções a essas regras. A Seção 4.3 detalha a abordagem que utilizamos para lidar com essa limitação. Uma última definição importante diz respeito aos métodos de engenharia de prompts. Em geral, esses métodos são agrupados em oito classes: *Zero-Shot Learning*, *One-Shot Learning*, *Few-Shot Learning*, *Chain-of-Thought Prompting*, *Iterative Prompting*, *Negative Prompting*, *Hybrid Prompting* e *Prompt Chaining*. Os detalhes de cada um desses métodos estão fora do escopo deste artigo, podendo-se consultar a referência em [Giray 2023]. No entanto, é importante compreender a ideia por trás do *Few-Shot Learning* e do *Chain-of-Thought Prompting*. O método *Few-Shot Learning* envolve fornecer alguns exemplos (geralmente de 2 a 5) para ajudar a IA a entender o padrão ou estilo da resposta desejada. Por exemplo:

**Prompt:** Escreva um poema rimado de quatro versos com base no tema indicado. Siga o padrão dos exemplos abaixo.

**Exemplo 1 (Tema: Noite)** A luz da lua dança no mar, um brilho prateado a flutuar. As estrelas sussurram lá no céu, uma canção de ninar sob o véu.

**Exemplo 2** (*Tema: Outono*) Folhas douradas caem devagar, uma brisa suave começa a soprar. O ar é fresco, o céu reluz, um mundo pintado com toque de luz.

**Agora, escreva um poema com o tema:** *Primavera*.

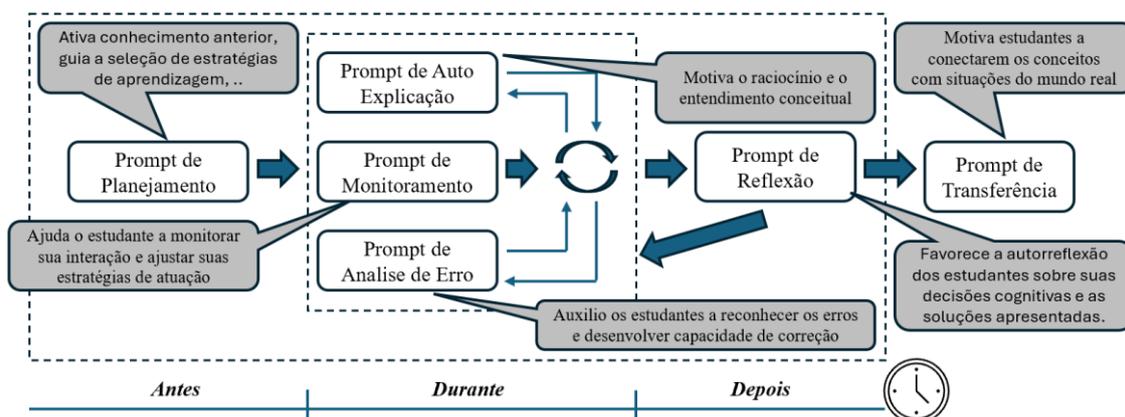
O modelo então irá gerar um poema de quatro versos seguindo a estrutura e o esquema de rimas demonstrados nos exemplos. Já a técnica de *Chain-of-Thought Prompting* consiste na indicação de uma estrutura de ação a ser seguida pelo modelo, manifestando-se como um passo a passo. Essas estratégias são muito úteis para a nossa abordagem, conforme discutido na Seção 4.3.

### 4.3. Definindo Prompts Metacognitivos

De acordo com a Figura 1a, a definição de cada prompt começa com a definição da tarefa que ele deve realizar. Para tal, criamos o termo *seed-prompt* (ou prompt-semente), que representa um prompt básico que descreve a tarefa e é posteriormente refinado durante o processo de interação em tempo de execução. Desta forma, os *prompts* orientados por abordagens metacognitivas compilam estratégias para promover comportamentos metacognitivos. No nosso estudo, em particular, estamos tratando dos seguintes *prompts*:

- **Prompts de planejamento (antes da aprendizagem):** ativam conhecimentos prévios, estabelecem metas e incentivam a seleção de estratégias. Exemplo: *Antes de iniciar esta tarefa, o que você já sabe sobre este tópico? Quais estratégias você usará para abordar o problema e por quê?*
- **Prompts de monitoramento (durante a aprendizagem):** ajudam os estudantes a acompanharem sua compreensão, reconhecerem dificuldades e ajustarem estratégias. Exemplo: *Enquanto você trabalha neste problema, está seguindo o plano inicial? Se não, que mudanças está fazendo e por quê?*
- **Prompts de reflexão (após a aprendizagem):** incentivam os estudantes a avaliarem seu processo e resultados de aprendizagem. Exemplo: *Agora que terminou esta tarefa, como você avalia o que aprendeu? O que funcionou bem e o que poderia melhorar na próxima vez?*
- **Prompts de auto explicação (aprofundam a compreensão):** incentivam o raciocínio e a compreensão conceitual. Exemplo: *Você pode explicar por que usou esse método em vez de outro? Haveria outra forma de resolver isso?*
- **Prompts de análise de erros (identificação de lacunas):** ajudam os estudantes a reconhecerem erros e desenvolverem habilidades de resolução de problemas. Por exemplo: *Você consegue encontrar e explicar algum erro na sua solução? O que faria de diferente da próxima vez?*
- **Prompts de transferência (aplicação na vida real):** incentivam os estudantes a conectarem conceitos com situações do mundo real. Por exemplo: *Como você poderia usar esse conceito em uma situação da vida real? Consegue pensar em um exemplo?*

A semântica desses prompts impõe uma ordem mínima de aplicação, conforme ilustrado na Figura 2, que relaciona os seis tipos anteriores de prompts metacognitivos ao processo de auxiliar os estudantes no planejamento, monitoramento e avaliação de sua aprendizagem. Ou seja, existem etapas (antes, durante e depois das interações de aprendizagem) nas quais cada um desses prompts é melhor aplicado.



**Figura 2. Esquema de uso e transição entre os prompts metacognitivos**

A implementação das transições utiliza o modelo de linguagem (LLM) para classificar as respostas dos estudantes em uma das categorias de atos de fala [Degen 2023]. No entanto, esse processo não se resume a um mapeamento simples, pois depende de ajustes dinâmicos da aprendizagem. Por exemplo, se for identificado um “diretivo” e o estudante frequentemente pedir esclarecimentos (diretivos), então a resposta será adaptada, como uma explicação passo a passo ou atividades adicionais de prática. Seguindo essa lógica, outras situações também podem ser definidas, como a progressão para tópicos avançados quando um estudante explica corretamente os conceitos (assertivos), ou a simplificação do conteúdo quando o estudante expressa frustração (expressivos). É interessante observar que, ao longo desse processo, existe um ciclo de monitoramento e retroalimentação, o qual analisa continuamente interações passadas para ajustar e aprimorar a experiência de aprendizagem. Neste artigo, em particular, foram explorados os *prompts* de planejamento, monitoramento, auto-explicação e reflexão.

#### 4.4. Framework de Avaliação

Na avaliação proposta, cada estudante passa um determinado tempo  $t$ , previamente configurado, interagindo com o tutor virtual na busca de solucionar dúvidas ou se aprofundar em algum tema previamente exposto. Desta forma, o objeto de validação é composto pelo diálogo gerado desta interação. Porém, a análise de um diálogo entre um estudante e um tutor virtual com o objetivo de avaliar se a aprendizagem está ocorrendo de forma eficaz continua sendo um desafio central na IA educacional. Desta forma, nossa proposta se baseia em uma abordagem analítica em múltiplas camadas que combine análise do discurso, marcadores comportamentais de aprendizagem e indicadores cognitivos ou metacognitivos. Os seguintes passos fazem parte desta proposta:

Identificação de padrões de discurso e movimentos no diálogo. O objetivo é a identificação de padrões de discurso e movimentos no diálogo, ou seja, a detecção de atos de fala ou de estruturas específicas que estão comumente associadas a uma aprendizagem eficaz. Nossa proposta baseou-se no estudo de Nguyen (2023), que utilizou o ICAP (*Interactive, Constructive, Active, Passive*) [Chi, 2009] como framework teórico de análise voltado ao estudo da influência de tutores virtuais com diferentes papéis no processo de aprendizagem. A principal vantagem do ICAP está na clareza de sua hierarquia e na forma como conecta diretamente tipos de atividade a níveis de aprendizagem esperados. Essa hierarquia já foi validada empiricamente em diversos

estudos e contextos educacionais, mostrando que níveis mais elevados tendem a produzir maiores ganhos de aprendizagem. Entretanto, o ICAP foi estendido com a Teoria da Argumentação [Van Eemeren et al., 2004], de modo a capturar uma maior variedade de padrões discursivos e a complexidade dos diálogos. A Tabela 2 detalha a codificação proposta, baseada na integração dessas teorias

**Tabela 2. Tipos de movimento contidos em um diálogo e suas descrições**

<b>Padrão/(código)</b>	<b>Definição</b>	<b>Exemplo</b>
Pedido de Esclarecimento (CRQ)	Um pedido para tornar uma afirmação mais clara ou específica.	<b>Estudante:</b> “O que você quer dizer com ‘inibidor?’”
Estímulo /Indagação (PRM)	Uma pergunta ou instrução com o objetivo de provocar uma resposta.	<b>Tutor:</b> “Você pode explicar por que pensa isso?”
Elaboração (ELB)	Adição de informações além do que foi solicitado ou declarado anteriormente.	<b>Estudante:</b> “Não é só rápido, é porque as moléculas estão se movendo mais.”
Justificativa (JST)	Apresentação de evidências ou razões para sustentar uma afirmação.	<b>Estudante:</b> “Escolhi a A porque a taxa aumenta com a temperatura.”
Refutação /Contra-argumento (REF)	Um desafio ou contradição a uma afirmação ou argumento anterior.	<b>Estudante:</b> “Mas isso só acontece se o pH cair, não é?”
Afirmativa (CLM)	Declaração que afirma um fato, opinião ou conclusão.	<b>Estudante:</b> “A resposta é B.” (sem justificativa)
Repetição (RPT)	Repetição de um conceito ou frase sem adicionar novas informações.	<b>Estudante:</b> “Como eu disse, é mais rápido.”
Metadiscorso (MET)	Enunciados sobre o processo do diálogo, a tarefa de aprendizagem ou a estrutura da interação.	<b>Tutor:</b> “Vamos revisar o que acabamos de discutir.”
Reconhecimento (ACK)	Confirmações breves ou sinais de compreensão/acordo.	<b>Estudante:</b> “Entendi.” Ou “Certo.”
Correção (COR)	Correção de um erro.	<b>Estudante:</b> “Ah, quis dizer difusão, não osmose.”
Síntese (KCS)	Resumir ou sintetizar múltiplas ideias em uma conclusão coerente.	<b>Tutor:</b> “Então podemos concluir que o pH afetou a osmose.”

Execução da análise de sequências. Realiza um processo de mineração de padrões sequenciais, tais como “Estímulo → Elaboração → Esclarecimento → Confirmação”, e sua interpretação de acordo com evidências de aprendizagem estabelecidos na literatura (Tabela 3). Tais evidências indicam, por exemplo, modificações conceituais (ocorre quando conceitos inicialmente equivocados são corrigidos), aprimoramento de habilidades (ocorre quando se percebe respostas mais precisas ou completas ao longo do tempo), crescimento metacognitivo (ocorre quando se observa um pensamento mais regulado em termos de planejamento e reflexão) e engajamento e persistência (o estudante continua explorando ou fazendo perguntas com uma evolução no nível de complexidade).

Observe que, enquanto a maioria desses padrões envolve sequências produtivas, algumas delas (e.g., ACK → RPT → CLM) representam um engajamento menos produtivo e mais superficial.

**Tabela 3. Padrões sequenciais e suas interpretações**

Padrão sequencial	Interpretação	Referência
PRM → ELB → JST	Produção construtiva do conhecimento. O estímulo leva à elaboração e à justificação.	[Chi 2009; Dillenbourg e Hong 2008]
REF → COR → JST → KCS	Resolução de conflitos e mudança conceitual. Uma refutação é seguida por correção, justificação e síntese.	[Suthers 2003; Andriessen et al. 2013]
CLM → JST → ELB	Afirmação justificada levando a uma explicação mais profunda. Da reivindicação à justificação e elaboração.	[Stahl 2013]
PRM → ACK → PRM → ELB	Reformulação do estímulo pelo tutor devido a um reconhecimento superficial, até que o aluno elabore.	[Walker et al. 2012]
ELB → ELB → JST → KCS	Construção colaborativa. Elaboraões repetidas culminando em justificação e síntese.	[Stahl 2013; Barron 2003]
JST → REF → COR → ELB	Refinamento do conhecimento. Uma justificação é contestada, levando à correção e a uma elaboração mais profunda.	[Suthers 2003; Andriessen et al. 2013]
ACK → RPT → CLM	Baixo engajamento cognitivo. Respostas mínimas e repetições sem justificação.	[D’mello et al. 2008]
MET → PRM → JST → KCS	Auxílio pelo tutor. Um estímulo metadiscursivo desencadeia justificação e síntese.	[Mercer e Littleton 2007]
CLM → REF → JST	Aprendizagem argumentativa. Uma afirmação é desafiada e precisa ser justificada, promovendo pensamento crítico.	[Von Aufschnaite et al. 2008]
CRQ → ELB → JST	Esclarecimento e elaboração iniciados pelo aluno. Motivados pela incerteza do próprio estudante.	[Kumar et al. 2007]

## 5. Resultados

O tutor foi baseado no modelo de linguagem *llama3-70b-8192*, sendo inicializado com os seguintes hiperparâmetros: *max\_tokens=502*, *stream=False*, *temperature=0.8* e *top\_p=0.9*. Tais valores foram obtidos de forma empírica por meio de ciclos de testes. A aplicação foi primeiramente configurada para agir de forma livre (interação direta entre o modelo e o estudante) e, então, por meio de um tutor que promove a metacognição. Ao executar a aplicação, o estudante recebe uma mensagem de boas-vindas e é questionado sobre o que deseja aprender naquela sessão. O design de interação foi baseado em modelos de salas de bate-papo virtuais existentes. Os resultados são discutidos nas subseções seguintes.

### 5.1. Análise de Diálogo: tutor generativo tradicional

Sessões de estudo de 15 minutos foram realizadas por 9 estudantes universitários com o modelo Lhama no modo livre. Então, os diálogos foram codificados de acordo com a Tabela 2. O resultado abaixo (Quadro 1) foi uma das saídas obtidas neste experimento, utilizando a Revolução Francesa como tema, o qual foi definido em tempo real. Os índices  $T$  e  $E$  indicam a fonte da classificação da fala como sendo o tutor ou o estudante respectivamente.

**Quadro 1. Codificação de um experimento com o modelo Lhama no modo livre**

[PRM<sup>T</sup>, CLM<sup>E</sup>, CLM<sup>T</sup>, CRQ<sup>E</sup>, ELB<sup>T</sup>, MET<sup>E</sup>, ACK<sup>T</sup>]

Por exemplo, o primeiro parágrafo “Você quer conversar sobre o quê?” foi classificado como um estímulo (PRM), enquanto o segundo parágrafo “quero aprender sobre Revolução Francesa” foi classificado como uma afirmativa (CLM). Nossa análise demonstrou que nenhum dos padrões de aprendizagem foi identificado nesta codificação. De fato, foram identificados diversos pedidos de esclarecimento (CRQ) seguidos por elaborações (ELB). Porém, este é exatamente o padrão que se deseja evitar, uma vez que a elaboração está sendo feita sempre pelo tutor e o pedido de esclarecimento pelo estudante. Ou seja, tal padrão configura o estilo “Educação Bancária” de aprendizagem, tão criticado por autores como Freire (1997).

### 5.2. Análise de Diálogo: tutor generativo baseado em metacognição

Da mesma forma, os 9 estudantes de nível superior participaram de sessões de estudo de 15 minutos com o modelo Lhama, mas agora configurado com alguns dos princípios discutidos na seção 4.3 (planejamento, monitoramento e reflexão). Outro tema foi aleatoriamente escolhido (para o exemplo a seguir, Revolução Industrial) para não enviesar os resultados. Novamente os diálogos foram codificados de acordo com a Tabela 2. O resultado abaixo (Quadro 2) foi uma das saídas obtidas neste experimento.

**Quadro 2. Codificação de um experimento com o modelo Lhama no modo metacognitivo**

PRM<sup>T</sup>, CLM<sup>E</sup>, CRQ<sup>T</sup>, CLM<sup>E</sup>, MET<sup>T</sup>, ACK<sup>E</sup>, CRQ<sup>T</sup>, CLM<sup>E</sup>, MET<sup>T</sup>, CLM<sup>E</sup>, (MET/PRM)<sup>T</sup>, CLM<sup>E</sup>, (ACK/CRQ)<sup>T</sup>, JST<sup>E</sup>, (ACK/KCS/PRM)<sup>T</sup>, CLM<sup>E</sup>, (ELB/CRQ)<sup>T</sup>, ELB<sup>E</sup>, (ACK/JST/KCS/CRQ)<sup>T</sup>, ELB<sup>E</sup>, (ACK/MET/PRM)<sup>T</sup>, (ELB/JST)<sup>E</sup>, CRQ<sup>T</sup>, (CLM/CRQ)<sup>E</sup>, (ELB/CRQ)<sup>T</sup>, (CLM/CRQ)<sup>E</sup>, (ELB/CRQ)<sup>T</sup>, ELB<sup>E</sup>, (ACK/KCS/PRM)<sup>T</sup>, KCS<sup>E</sup>, ELB<sup>T</sup>, (ACK/PRM)<sup>E</sup>, (KCS/PRM)<sup>T</sup>, CLM<sup>E</sup>, (ACK/ELB)<sup>T</sup>, ACK<sup>E</sup>, MET<sup>T</sup>

Diferentemente do caso anterior, esta sequência já demonstra indícios de pelo menos três padrões de aprendizagem definidos na Tabela 3. Vamos discutir os trechos de diálogo que caracterizam estes padrões:

Padrão: PRM → ELB → JST: nesse caso (Tabela 4) o estudante é instigado a elaborar suas respostas em relação aos meios de produção da época e justificar a sua resposta.

**Tabela 4. Trecho de diálogo com o padrão PRM → ELB → JST**

Parte do diálogo	Codificação
Tutor: Excelente ideia! Sim, investir na construção de máquinas mais eficientes e que poluíssem menos teria sido uma forma de reduzir o impacto ambiental negativo da Revolução Industrial. Isso teria ajudado a minimizar a poluição do ar e a proteger o meio ambiente. Agora, vamos voltar ao Objetivo 3: Compreender como a Revolução Industrial afetou a sociedade e a economia. Você acha que a Revolução Industrial beneficiou todas as classes sociais ou apenas algumas delas?	(ACK/MET/PRM) <sup>T</sup>
Estudante: Acho que beneficiou os donos dos meios de produção, sei que os trabalhadores tinham longas, perigosas e exaustivas jornadas de trabalho	(ELB/ JST) <sup>E</sup>

Padrão: ELB → ELB → JST → KCS: na Tabela 5, é interessante notar que tutor e estudante colaborativamente engajam na construção de um conhecimento. No caso, sobre uma das inovações da época que foi citada pelo próprio estudante.

**Tabela 5. Trecho de diálogo com o padrão ELB → ELB → JST → KCS**

Parte do diálogo	Codificação
Estudante: já ouvi falar sobre a locomotiva a vapor	CLM <sup>E</sup>
Tutor: A locomotiva a vapor é um exemplo icônico da Revolução Industrial. Ela permitiu o transporte de mercadorias e pessoas de forma mais rápida e eficiente. Agora, vamos pensar sobre como a locomotiva a vapor afetou a sociedade e a economia. Qual é sua opinião sobre como ela mudou a forma como as pessoas viviam e trabalhavam?	(ELB/CRQ) <sup>T</sup>
Estudante: acho que deve ter ajudado a escoar produtos e a transportar pessoas, mas também imagino que deva ter contribuído para o aumento da poluição do ar	ELB <sup>E</sup>
Tutor: Muito bem! Sim, a locomotiva a vapor ajudou a escoar produtos e a transportar pessoas de forma mais rápida e eficiente, o que contribuiu para o desenvolvimento econômico. No entanto, como você mencionou, também contribuiu para o aumento da poluição do ar, pois as locomotivas a vapor emitiam grandes quantidades de fumaça e fuligem. Essa é uma boa oportunidade para discutir sobre o impacto ambiental da Revolução Industrial. O que você acha que poderia ter sido feito para reduzir o impacto ambiental negativo da Revolução Industrial?)	(ACK/JST/KCS/CRQ) <sup>T</sup>

Padrão: MET → PRM → JST → KCS: na (Tabela 6), vemos o uso de elementos do metadiscursos, onde o tutor foca na estrutura que será utilizada no diálogo como preparação para o estímulo, o que continua até que o estudante apresente uma justificativa para sua resposta. Finalmente o tutor faz uma síntese do conhecimento gerado.

**Tabela 6. Trecho de diálogo com o padrão MET → PRM → JST → KCS**

Parte do diálogo	Codificação
Tutor: Vamos usar a estratégia de ensino recíproco**. Isso significa que eu vou explicar um conceito e, em seguida, você vai explicar para mim como se você fosse o professor. Isso vai ajudar a consolidar seu conhecimento e a identificar áreas em que você precisa praticar. Vamos começar com o Objetivo 1: Entender o que é a Revolução Industrial e sua importância na história. O que você já sabe sobre a Revolução Industrial? Qual é a sua definição inicial?	(MET/PRM) <sup>T</sup>

Estudante: foi uma mudança da forma de produção	CLM <sup>E</sup>
Tutor: Excelente começo! Sim, a Revolução Industrial foi uma mudança significativa na forma de produção. Você poderia expandir um pouco mais sobre o que você entende por "mudança na forma de produção"? O que mudou exatamente? (Lembre-se de que você é o professor agora)	(ACK/CRQ) <sup>T</sup>
Estudante: Antes da Revolução, a produção de bens, como tecidos, era feita por manufatura. Depois passou a ser com máquinas em indústrias, certo?	JST <sup>E</sup>
Tutor: Muito bem! Sim, você está correto. Antes da Revolução Industrial, a produção de bens, como tecidos, era feita por meio de manufatura, que era um processo lento e artesanal. Com a Revolução Industrial, a produção passou a ser feita com o uso de máquinas em indústrias, o que permitiu uma produção mais rápida e em maior escala. Agora, vamos falar sobre as inovações tecnológicas que permitiram essa mudança. Você conhece alguma dessas inovações, como a máquina de tecer de Jacquard ou a locomotiva a vapor?	(ACK/KCS/PRM) <sup>T</sup>

## 6. Conclusão

O principal efeito do uso dos princípios da metacognição é levar o tutor a sair do estilo “Educação Bancária”, onde o estudante apenas pergunta e recebe respostas prontas, para um diálogo onde a construção do conhecimento é feita de forma mais colaborativa e a autonomia dos alunos é estimulada. Observe também que os dois experimentos utilizaram o mesmo tempo. Porém, a quantidade de mensagens no segundo caso é bem maior (em média, 21 vs 37 mensagens), sendo quase o dobro, o que indica um diálogo mais dinâmico e sugere uma maior troca de ideias. Outro fato interessante é que a participação do tutor é dialogicamente mais rica. Ou seja, um único turno de fala apresentou elementos linguísticos que caracterizam múltiplas codificações.

Em termos de aplicação prática, a proposta é computacionalmente escalável uma vez que todos os sistemas de tutoramento executam como aplicações individuais e utilizam um modelo de linguagem (LLaMA) que foi projetada justamente para ser escalável. Um problema é o custo envolvido. Como cada sistema é individual, ele possui seu próprio token. Esses tokens funcionam como créditos que vão sendo consumidos na medida que o modelo é requisitado. Desta forma, para uma grande quantidade de estudantes, tal abordagem pode se tornar financeiramente inviável.

O projeto, porém, ainda apresenta diversas limitações. No momento, alguns princípios da metacognição ainda não foram implementados ou foram apenas parcialmente implementados. O monitoramento, por exemplo, precisa de recursos que extrapolam a engenharia de prompt, uma vez que é necessário manter uma referência para entender a evolução do estudante e se ele está mantendo o foco no assunto, como um mapa cognitivo sobre a temática. O trabalho de Nguyen (2023) discute algumas estratégias nesta direção. Também se tinha uma expectativa de que situações de refutação (REF) emergissem dos diálogos, mas essa codificação não foi utilizada, o que impossibilitou o surgimento de importantes padrões de aprendizagem. Por exemplo, a aprendizagem argumentativa, onde uma afirmação é desafiada e precisa ser justificada, promovendo pensamento crítico. A busca de soluções para tais limitações é uma oportunidade que está sendo considerada nos nossos trabalhos em andamento.

## Agradecimentos

Os autores gostariam de agradecer a Dra. Patrícia Fernanda da Silva, da Universidade Federal do Rio Grande do Sul, pela inspiração no uso dos conceitos de metacognição, assim como a todos os membros do *Transformative Learning Technologies Lab* (Columbia University) e Fundação Itaú pelo suporte a este projeto.

## Referências

- Andriessen, J., Baker, M., & Suthers, D. D. (Eds.). (2013). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (Vol. 1). Springer Science & Business Media.
- Bai, Y., Li, J., Shen, J., & Zhao, L. (2024). Investigating the efficacy of ChatGPT-3.5 for tutoring in Chinese elementary education settings. *IEEE Transactions on Learning Technologies*.
- Barron, B. (2003). When smart groups fail. *The journal of the learning sciences*, 12(3), 307-359.
- Bauschard, S., & Quidwai, S. (2024). From insight to implementation: How to create your AI school guidance. *Available at SSRN*.
- Chan, C. K. Y., & Lee, K. K. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers?. *Smart learning environments*, 10(1), 60.
- Chen, S., Liu, Q., & He, B. (2023, November). A Generative AI-based Teaching Material System Using a Human-In-The-Loop Model. In *2023 International Conference on Intelligent Education and Intelligent Research (IEIR)* (pp. 1-8). IEEE.
- Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, 1(1), 73-105.
- Cox, Michael T. Metacognition in computation: A selected research review. *Artificial intelligence*, v. 169, n. 2, p. 104-141, 2005.
- Cox, Michael et al. Computational metacognition. *arXiv preprint arXiv:2201.12885*, 2022.
- Dai, Y. (2024). Dual-contrast pedagogy for AI literacy in upper elementary schools. *Learning and Instruction*, 91, 101899.
- De Cremer, D., Bianzino, N. M., & Falk, B. (2023). How generative AI could disrupt creative work. *Harvard Business Review*, 13.
- Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, 9(1), 519-540.
- D'mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User modeling and user-adapted interaction*, 18, 45-80.

- Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, 3, 5-23.
- Freire, P. (1997). Educação “bancária” e educação libertadora. *Introdução à psicologia escolar*, 3, 61-78.
- Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering*, 51(12), 2629-2633.
- Griffiths, D., Frías-Martínez, E., Tlili, A., & Burgos, D. (2024). A cybernetic perspective on generative AI in education: From transmission to coordination.
- Huang, C., & Yu, H. (2024, June). Technological Dialogue and Intelligent Generation: An Empirical Study of Primary Chinese Teachers Using Generative AI for Lesson Preparation. In *Proceedings of the 2024 International Conference on Intelligent Education and Computer Technology* (pp. 606-611).
- Jurenka, I., Kunesch, M., McKee, K. R., Gillick, D., Zhu, S., Wiltberger, S., ... & Ibrahim, L. (2024). Towards responsible development of generative AI for education: An evaluation-driven approach. *arXiv preprint arXiv:2407.12687*.
- Kasonde Ng’andu, F. H., Haambokoma, N., & Tomaida, M. (2013). The Contribution of Behaviourism Theory to Education. *Journal: Zambia Journal of Education*, 4(1), 58-74.
- Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Frontiers in artificial intelligence and applications*, 158, 383.
- Lee, Y. (2024). Development of Open Large Language Models for Artificial Intelligence Digital Textbooks. *TEM Journal*, 13(4), 2773-2783.
- Lee, G. G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies*.
- Li, M., & Yuan, R. (2022). Enhancing students’ metacognitive development in higher education: A classroom-based inquiry. *International Journal of Educational Research*, 112, 101947.
- Li, M., & Zhang, H. (2024). A Study on the Behavioral Intention of Primary and Secondary School Teachers Using Generative Artificial Intelligence in Teaching. In *Proceedings of the 2024 9th International Conference on Distance Education and Learning* (pp. 35-41).
- Mahdavi, M. (2014). An overview: Metacognition in education. *International Journal of Multidisciplinary and current research*, 2(6), 529-535.
- Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Routledge.
- Mittal, U., Sai, S., & Chamola, V. (2024). A comprehensive review on generative ai for education. *IEEE Access*, 12, 142733-142759.
- Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education*, 192, 104661.

- Santos, R. P. D. (2023). Enhancing physics learning with ChatGPT, Bing Chat, and Bard as agents-to-think-with: a comparative case study. *arXiv preprint arXiv:2306.00724*.
- Stahl, G. (2013). A model of collaborative knowledge-building. In *International conference of the learning sciences* (pp. 70-77). Psychology Press.
- Steffe, L. P., & Gale, J. E. (Eds.). (1995). *Constructivism in education*. Psychology Press.
- Suthers, D. D. (2003). Representational guidance for collaborative inquiry. In *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 27-46). Dordrecht: Springer Netherlands.
- Tsai, Y. H., Lin, C. H., Hong, J. C., & Tai, K. H. (2018). The effects of metacognition on online learning interest and continuance to learn with MOOCs. *Computers & Education, 121*, 18-29.
- Van Eemeren, F. H., Grootendorst, R., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge University Press.
- Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 45*(1), 101-131.
- Walker, M., Anand, P., Abbott, R., & Grant, R. (2012, June). Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 592-596).
- Xie, W., Li, J., Mu, Y., Zhang, H., Zhao, S., Zheng, X., & Song, K. (2025). The Power of Personalized Datasets: Advancing Chinese Composition Writing for Elementary School through Targeted Model Fine-Tuning. *International Journal of Asian Language Processing, 2450017*.
- Zhou, X., & Schofield, L. (2024). Using social learning theories to explore the role of generative artificial intelligence (AI) in collaborative learning. *Journal of Learning Development in Higher Education, (30)*.