



21 A 23 DE NOVEMBRO DE 2025  
XXX ENAPET

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS:  
DESAFIOS ÉTICOS PARA O SÉCULO XXI

## **<sup>1</sup>PRODUÇÃO DINÂMICA E AVALIAÇÃO DE CRIATIVIDADE: ANÁLISE COMPARATIVA DA ESCRITA HUMANA E IA**

TSENG, H. H. <sup>1</sup>; FILHO, R. S. <sup>2</sup>;

<sup>1</sup>Grupo PET-Fronteira, Campus CPPP; <sup>2</sup>Tutor(a) do Grupo PET-Fronteira; UFMS, Campus CPPP

E-mail: tseng.hsin@ufms.br; reginaldo.uspoli@gmail.com; divulgapetfronteira@gmail.com

Grupo PET Sistemas de Informação, Matemática e Ciência da Computação (PET Fronteira);

UFMS; Universidade Federal do Mato Grosso do Sul

**RESUMO:** Este trabalho investiga a capacidade da inteligência artificial (IA) generativa na produção e avaliação de conteúdo literário, focando na divergência entre a percepção de qualidade humana e a computacional. O estudo realiza uma análise comparativa entre um conto de horror de autoria humana e três narrativas geradas pelos modelos Claude 4 Sonnet, Grok 4 e GPT-5, a partir de um mesmo prompt detalhado. A metodologia envolveu a avaliação cega das quatro obras por um painel de juízes (humanos e de IA), com base em critérios de novidade, surpresa, diversidade e complexidade linguística. Os resultados indicam que, embora as IAs reproduzam com fidelidade diretrizes estruturais e estilísticas, a produção humana se sobressai nos quesitos de novidade, surpresa e complexidade linguística. Entre os modelos de IA, o Claude 4 Sonnet apresentou o desempenho mais robusto, destacando-se no critério de diversidade. Conclui-se que, embora as IAs consigam emular com eficácia a diversidade estilística, a capacidade de gerar novidade e surpresa genuínas permanece um diferencial da criatividade humana. O estudo posiciona a IA generativa como uma sofisticada ferramenta de simulação estilística, sugerindo um futuro de colaboração simbiótica entre a eficiência da máquina e a engenhosidade humana na produção literária.

**Palavras-chave:** Inteligência Artificial Generativa; Produção Literária; Avaliação de Criatividade; Estrutura Narrativa.

### **DYNAMIC PRODUCTION AND CREATIVITY ASSESSMENT: A COMPARATIVE ANALYSIS OF HUMAN AND AI WRITING**

---

<sup>1</sup>Área do conhecimento: Linguística, Letras e Artes; Ecossistema de inovação CCT/PR: Educação, Sociedade & Economia; Transformação Digital.; Objetivos do Desenvolvimento Sustentável (ODS): ODS 4 (Educação de Qualidade); ODS 9 (Indústria, Inovação e Infraestrutura).



XXX ENCONTRO NACIONAL DOS GRUPOS PET  
UNIVERSIDADE DE BRASÍLIA (UnB)  
Campus Darcy Ribeiro, Asa Norte  
70910-900, Brasília - DF





**ABSTRACT :** This paper investigates the capacity of generative artificial intelligence (AI) in the production and evaluation of literary content, focusing on the divergence between human and computational perceptions of quality. The study conducts a comparative analysis between a human-authored horror short story and three narratives generated by the Claude 4 Sonnet, Grok 4, and GPT-5 models, based on a single detailed prompt. The methodology involved a blind evaluation of the four works by a panel of judges (both human and AI), based on the criteria of novelty, surprise, diversity, and linguistic complexity. The results indicate that while AIs faithfully reproduce structural and stylistic guidelines, human-produced work excels in the criteria of novelty, surprise, and linguistic complexity. Among the AI models, Claude 4 Sonnet showed the most robust performance, excelling in the criterion of diversity. It is concluded that while AIs can effectively emulate stylistic diversity, the ability to generate genuine novelty and surprise remains a key differentiator of human creativity. The study positions generative AI as a sophisticated tool for stylistic simulation, suggesting a future of symbiotic collaboration between machine efficiency and human ingenuity in literary production.

**Keywords:** Generative Artificial Intelligence; Literary Production; Creativity Assessment; Narrative Structure.

## Introdução

A ascensão da inteligência artificial (IA) generativa está redefinindo os conceitos de autoria e criatividade, posicionando modelos de linguagem de grande escala (LLMs) como potenciais parceiros no processo criativo. Este avanço tecnológico levanta questões fundamentais sobre originalidade e a natureza da criatividade, com um impacto na produção cultural comparável ao das revoluções industrial e científica. Embora os LLMs demonstrem crescente sofisticação estilística, persistem dúvidas sobre sua capacidade de gerar histórias com o mesmo nível de novidade e surpresa encontrado na produção humana. ISMAYILZADA et al. (2025), por exemplo, apontam uma lacuna entre a complexidade sintática (a forma) e a coerência semântica e criativa (o conteúdo) das obras geradas por IA.

Adicionalmente, a avaliação dessas obras constitui um desafio, pois os próprios LLMs, quando atuam como juízes, podem apresentar vieses que divergem significativamente da percepção humana. Este cenário demanda novas categorias analíticas para compreender a produção discursiva contemporânea. A pesquisa de Ismayilzada et al. (2025) evidencia que humanos e LLMs possuem entendimentos distintos do que significa "ser criativo": enquanto



humanos valorizam narrativas surpreendentes e originais, os LLMs tendem a associar criatividade a uma sobrecarga lexical e sintática.

A presente pesquisa busca, portanto, responder à seguinte questão: ao partir de um mesmo conjunto de diretrizes criativas e estruturais, a IA consegue gerar uma obra literária comparável à produção humana em termos de novidade, surpresa, diversidade e complexidade linguística? O objetivo geral é realizar uma análise comparativa da produção literária entre um autor humano e três modelos de IA de ponta, avaliando a fidelidade de emulação do conceito de assinatura autoral humana e investigando as divergências na percepção de qualidade entre avaliadores humanos e artificiais.

## Método

Este estudo utiliza uma abordagem de métodos mistos, combinando a análise de conteúdo qualitativa com a avaliação quantitativa por múltiplos juízes. O desenho metodológico foi organizado em três fases sequenciais: produção de um texto-base humano, geração de narrativas por LLMs e avaliação cega das narrativas.

Para garantir o controle das variáveis criativas, a produção de todos os contos foi guiada por um conjunto estrito de diretrizes: (1) adesão ao modelo de "Enredo de Descobrimento" de Noël Carroll (1999), estruturado em Irrupção, Descobrimento e Confronto; (2) uso de uma voz poética em primeira pessoa; e (3) conclusão com um epílogo em formato de documento oficial, funcionando como reviravolta temática e marcador estilístico.

Seguindo essas diretrizes, a primeira fase consistiu no desenvolvimento de um conto de horror cósmico de autoria humana, que serviu como padrão de referência. Na segunda fase, as mesmas diretrizes foram inseridas em um único e detalhado prompt, desenvolvido via engenharia de prompt (ZHAO et al., 2025), e submetido uma única vez a três modelos de IA de ponta: ChatGPT-5 (OpenAI), Claude 4 Sonnet (Anthropic) e Grok 4 (X).

Na terceira fase, o corpus textual com os quatro contos foi submetido a um processo de avaliação cega. Os textos foram anonimizados e apresentados em ordem aleatória a um painel de juízes composto por: (1) Juízes Humanos - leitores não especializados; e (2) Juízes de IA,



incluindo os modelos Deep Think R1, Gemini 2.5 Pro e Kimi-K2. A cada juiz foi solicitado que atribuísse notas de 1 (muito baixo) a 5 (muito alto) para quatro critérios de criatividade (Novidade, Surpresa, Diversidade e Complexidade Linguística), adaptados de Ismayilzada et al. (2025), e que classificasse a autoria de cada conto como "Humano" ou "IA".

### Resultados e Discussão

A análise dos dados consolidados na Tabela 1 revela que o conto de autoria humana foi o mais bem avaliado, obtendo a maior Média Geral (4.18). A obra humana dominou os critérios associados à criatividade, alcançando as pontuações mais altas em Novidade (4.57) e Surpresa (3.86). De forma notável, o texto humano também superou todas as IAs em Complexidade Linguística (4.43), desafiando a expectativa de superioridade dos LLMs na elaboração estilística. Em contrapartida, entre os modelos de IA, o Claude 4 Sonnet demonstrou o desempenho mais robusto (Média 3.57), destacando-se por alcançar a maior pontuação no critério Diversidade (4.00).

Além dos dados quantitativos, observações qualitativas dos juizes humanos apontaram para fraquezas recorrentes nos textos gerados por IA, como a tendência à repetição de estruturas frasais e vocabulário, bem como a ocorrência de lapsos de coerência interna, onde certas passagens narrativas careciam de uma conexão lógica. Esses elementos podem ter contribuído para as notas mais baixas em critérios como Surpresa e Novidade.

No entanto, em um contraste notável com essas fraquezas, um dos achados mais surpreendentes surgiu na tarefa de identificação da autoria. O conto gerado por Claude 4 Sonnet foi consistentemente confundido com uma obra humana, tanto pelos juizes de IA quanto por um dos avaliadores humanos. Este resultado indica um potencial significativo de mimetização estilística por parte deste modelo, que conseguiu, sob as condições do experimento, produzir uma obra indistinguível da criação humana.

Contudo, é fundamental ressaltar que estes resultados são baseados em uma amostra limitada de oito juizes. Portanto, devem ser interpretados como indicativos de uma tendência,

e não como conclusões generalizáveis, uma vez que a ampliação do painel de avaliadores poderia, potencialmente, redefinir os padrões aqui observados.

Tabela 1 – Avaliações Consolidadas (Médias)

CRITÉRIO	Humano	GPT-5	Claude 4 Sonnet	Grok 4
Novidade	4.57	2.86	3.29	1.71
Surpresa	3.86	3.00	3.14	2.14
Diversidade	3.86	3.71	4.00	3.43
Complexidade Linguística	4.43	3.43	3.86	3.14
Média Geral	4.18	3.25	3.57	2.61

Fonte: Elaborado pelo autor (2025). Média calculada a partir das avaliações dos leitores não-especialista e dos três juízes de IA.

## Conclusões

Os resultados deste estudo, embora preliminares, apontam para uma distinção clara entre a criação literária humana e a geração por IA. A superioridade do texto humano nos critérios de novidade e surpresa, aliada à sua vitória inesperada em complexidade linguística, sugere que a IA generativa opera de forma mais eficaz como uma sofisticada simuladora de estilo do que como uma criadora autônoma de conteúdo inovador. O modelo Claude 4 Sonnet é a maior evidência disso: ao mesmo tempo em que se destacou em diversidade, demonstrou um potencial de mimetização estilística tão eficaz que foi consistentemente confundido pelos juízes de IA com uma obra humana. Isso reforça a noção de uma "emulação competente" que, no entanto, ainda depende de uma faísca de originalidade distintamente humana.

Reconhecendo que um painel de juízes mais amplo poderia refinar estas observações, a principal implicação deste trabalho se mantém: o futuro da escrita criativa não aponta para a substituição, mas para uma simbiose. Nesta visão, a IA serve como uma poderosa ferramenta para expandir, ao em vez de replicar a criatividade do autor.

A discussão desses resultados sugere que, embora os modelos de IA sejam capazes de seguir diretrizes estruturais e temáticas com alta fidelidade, a emulação de aspectos mais sutis



21 A 23 DE NOVEMBRO DE 2025  
XXX ENAPET

## INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XXI

da criatividade humana permanece um desafio. A preferência humana por novidade e surpresa contrapõe-se à preferência da IA por complexidade formal. Essa discrepância evidencia que o conceito de qualidade literária não é monolítico e que a avaliação humana ainda opera com uma compreensão mais holística e profunda da criatividade narrativa.

### Agradecimentos

Agradecemos ao Programa de Educação Tutorial (PET) e ao Ministério da Educação (MEC) pelo suporte institucional e fomento que possibilitaram o desenvolvimento deste projeto.

Ao orientador, Prof. Dr. Reginaldo Inojosa da Silva Filho, pela orientação e apoio fundamentais.

Ao grupo PET Fronteiras e à Universidade Federal de Mato Grosso do Sul, Câmpus de Ponta Porã (UFMS/CPMP), pelo ambiente de aprendizado.

À minha família e amigos, pelo incentivo e paciência ao longo desta jornada.

### Referências

CARROLL, N. A Filosofia do Horror ou Paradoxos do Coração. Campinas: Papirus, 1999.

ISMAYILZADA, M.; STEVENSON, C.; VAN DER PLAS, L. Evaluating Creative Short Story Generation in Humans and Large Language Models. arXiv, 2024. Disponível em: <https://arxiv.org/abs/2411.02316>. Acesso em: 9 out. 2025.

ZHAO, W. X. et al. A Survey of Large Language Models. arXiv, 2025. Disponível em: <https://arxiv.org/abs/2303.18223>. Acesso em: 8 out. 2025.



XXX ENCONTRO NACIONAL DOS GRUPOS PET  
UNIVERSIDADE DE BRASÍLIA (UnB)  
Campus Darcy Ribeiro, Asa Norte  
70910-900, Brasília - DF

