

O processo de alucinação em Large Language Models de Inteligência Artificial (LLMS): perspectivas da Computação e da Ciência da Informação

LEONARDO AGOSTINELLI PEREIRA PINHEIRO DE SOUZA
CÁSSIA REGINA BASSAN DE MORAES

Resumo

O uso de inteligência artificial generativa se tornou grandemente popular entre leigos e cientistas, por meio de ferramentas como o Chat GPT. Isto só foi possível graças ao advento das redes neurais *transformers* com mecanismo de autoatenção. Entretanto, a geração de erros e informações falsas pelas ferramentas, processo definido como alucinação, é inextricável à sua constituição, por motivos técnicos. Este trabalho objetiva analisar e discutir as características da produção científica sobre alucinação em inteligência artificial nas áreas de Computação e Ciência da Informação, verificando tendências, lacunas e oportunidades neste respeito. Como procedimentos metodológicos, artigos da base *Web of Science* foram analisados com linguagem R, de acordo com área do conhecimento, lapso temporal e país de origem. Como resultados verificou-se um aumento exponencial de produção a partir de 2023, sendo a participação da Ciência da Informação e do Brasil ainda tímida, mas com tendência de crescimento. Argumenta-se que a Ciência da Informação tem condições de contribuir para o aprofundamento do tema tanto conceitual como tecnicamente, em disciplinas como a organização do conhecimento, recuperação da informação e competência informacional.

Palavras-chave: Inteligência artificial; Large Language Models; Alucinação; Cientometria; Deep Learning.

***THE HALLUCINATION PROCESS IN LARGE LANGUAGE MODELS (LLMS) OF
ARTIFICIAL INTELLIGENCE: perspectives from Computing and Information Science***

Abstract

The use of generative artificial intelligence has become widely popular among laypeople and scientists alike, through tools such as GPT Chat. This was only possible thanks to the advent of transformer neural networks with a self-attention mechanism. However, the generation of errors and false information by these tools, a process defined as hallucination, is inextricably linked to its nature, for technical reasons. This work aims to analyze and discuss the characteristics of scientific production on hallucination in artificial intelligence in the fields of Computer Science and Information Science, identifying trends, gaps, and opportunities in this regard. As methodological procedures, articles from the Web of Science database were analyzed using the R language, according to field of knowledge, time period, and country of origin. The results showed an exponential increase in production starting in 2023, with the participation of Information Science and Brazil still modest, but with a growing trend. It is argued that Information Science can contribute to the deepening of the topic, both conceptually and technically, in disciplines such as knowledge organization, information retrieval, and information literacy.

Keywords: Artificial Intelligence; Large Language Models; Hallucination; Scientometrics; Deep Learning.

1 INTRODUÇÃO

Em anos recentes, desde 2022, com o advento de Inteligências artificiais (IA) generativas, como Chat GPT e Deepseek, o interesse por parte tanto da população em geral quanto da comunidade acadêmica foi despertado para explorar vários aspectos desse tema. Os modelos de IA generativos atuais se baseiam em uma tecnologia revolucionária de redes neurais do tipo *transformer* com mecanismo de autoatenção, que conseguem se ‘lembrar’ de

informações complexas e guardar dados de contexto, requisitos essenciais para o *Natural Language Processing* (NLP - Processamento de Linguagem Natural) por meio de *Large Language Models* (LLM - modelos de linguagem amplos) (Barcaui, 2025; Faceli *et al.*, 2025).

Contudo, o uso de IAs generativas deve ser feito com cautela e análise crítica devido à sua tendência de produzir informações inconsistentes e/ou falsas. Este fato tem sido destacado por pesquisadores da Ciência da Informação (CI), especialmente e, discussões sobre competência informacional e produção científica (Trindade; Oliveira, 2024). A geração de informação incorreta pela IA é consequência de um problema técnico em redes neurais artificiais definido como alucinação. Visto que essas redes emulam o funcionamento do cérebro humano com seus neurônios artificiais, o processo de alucinação é análogo ao um delírio para seres humanos, onde o resultado produzido é coerente, convincente, porém incorreto.

Isso significa especificamente o conteúdo produzido pelo modelo que se desvia de sua entrada, carece de evidências empíricas, é desprovido de coerência significativa ou contradiz fatos do mundo real (Lin *et al.*, 2024, p. 243, tradução nossa).

São vários os fatores que podem levar à ocorrência de alucinações: desde a má formulação das perguntas, a qualidade dos dados de treinamento da rede e até mesmo a própria ‘opacidade’ de sua estrutura interna inerente, que se torna complexa demais para que se possa auditar de modo exato como as entradas estão sendo processadas e os resultados estão sendo gerados. Por ora, não existe ainda solução técnica totalmente eficaz e factível para resolução do problema.

Em última análise, temos que aceitar que, por agora, as alucinações nos LLMs são inevitáveis. Compreender o comportamento desses modelos e adotar estratégias – como a regeneração de respostas e a validação cuidadosa dos dados gerados – podem ajudar a navegar por esses desafios (Barcaui, 2025, p. 157).

É neste respeito que se argumenta que a CI pode contribuir grandemente em problematizar, discutir e propor estratégias para lidar com o problema da alucinação, já que esta área do conhecimento, por sua essência interdisciplinar, teria condições de desenvolver abordagens tanto teóricas quanto técnicas para o tema.

Deste modo, o presente trabalho objetiva examinar a produção científica sobre o tema da alucinação em LLM pelas óticas combinadas da CI e da computação, identificando tendências e lacunas no *corpus* de publicações. De modo específico, os objetivos são: discutir de modo didático alguns pormenores técnicos da constituição das redes *transformers*, LLM e NLP; examinar as publicações da base de dados *Web of Science* (WoS) sobre o tema em relação ao lapso temporal, áreas do conhecimento e países que publicam; analisar os resultados por meio de aportes da estatística descritiva, elaborando tabelas, gráficos do tipo *boxplot* e histogramas; realizar projeções de níveis de publicação para uma janela de futuro próximo. Assim, é de suma relevância verificar como a CI está contribuindo para a compreensão e mitigação de um problema que afeta os contextos científico, educacional e a vida cotidiana dos usuários de IA.

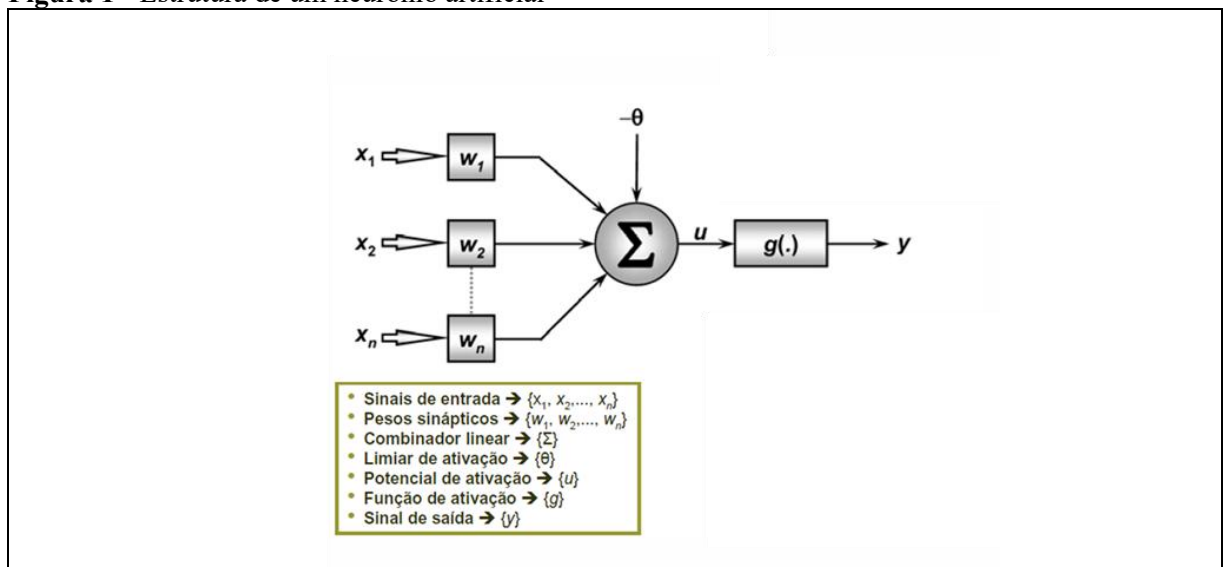
2 REFERENCIAL TEÓRICO

O alicerce de aplicações de IA generativa, como ChatGPT, DeepSeek e semelhantes é o chamado *deep learning*, ou aprendizado profundo. Enquanto pode-se utilizar modelos de IA mais simples para tarefas mais determinísticas, ‘mecanizadas’, como classificação de dados, clusterização (agrupamento) e regressão linear (inferir a fórmula matemática que possa ser usada para prever o valor, como preço de um imóvel, por exemplo), o *deep learning* utiliza estruturas mais complexas para executar tarefas que exigiriam o discernimento humano, tais como: reconhecimento de imagens, tradução, geração de textos e afins. O *deep learning*, por sua vez, é uma subárea do *machine learning* (aprendizado de máquina), baseado em estruturas que emulam o cérebro humano, as redes neurais artificiais.

Redes Neurais podem ser caracterizadas como modelos computacionais com capacidades de adaptar, aprender, generalizar, agrupar ou organizar dados, nos quais a estrutura operacional é baseada em processamento paralelo. A origem das pesquisas nesta área data da época em que foi apresentado o modelo computacional de um neurônio biológico [...] (Lima, 2014, p. 47).

Na Figura 1, apresenta-se um esquema visual da constituição e funcionamento de um neurônio artificial, unidade básica de processamento de informação em redes neurais. Este elemento é composto de: valores de entrada ($x_1, x_2 \dots$); pesos sinápticos ($w_1, w_2 \dots$), combinador linear (Σ), que realiza as somas ponderadas dos valores de entrada vezes seus pesos; limiar de ativação (θ), que determina um valor mínimo para que o neurônio seja ativado e uma saída possa ser gerada; o potencial de ativação (u) se refere à diferença entre o cálculo do combinador linear e o limiar de ativação, tendo poder de ativar a saída se for um valor positivo ou inibi-la, caso seja negativo; função de ativação (g), função matemática que transforma o valor rumo à saída do neurônio, de acordo com o que é requerido; saída do neurônio (y), resultado final gerado conforme o processamento do neurônio, que pode ser passado adiante numa estrutura de rede para ser mais profundamente processado por outro(s) neurônio(s) (Silva, 2010).

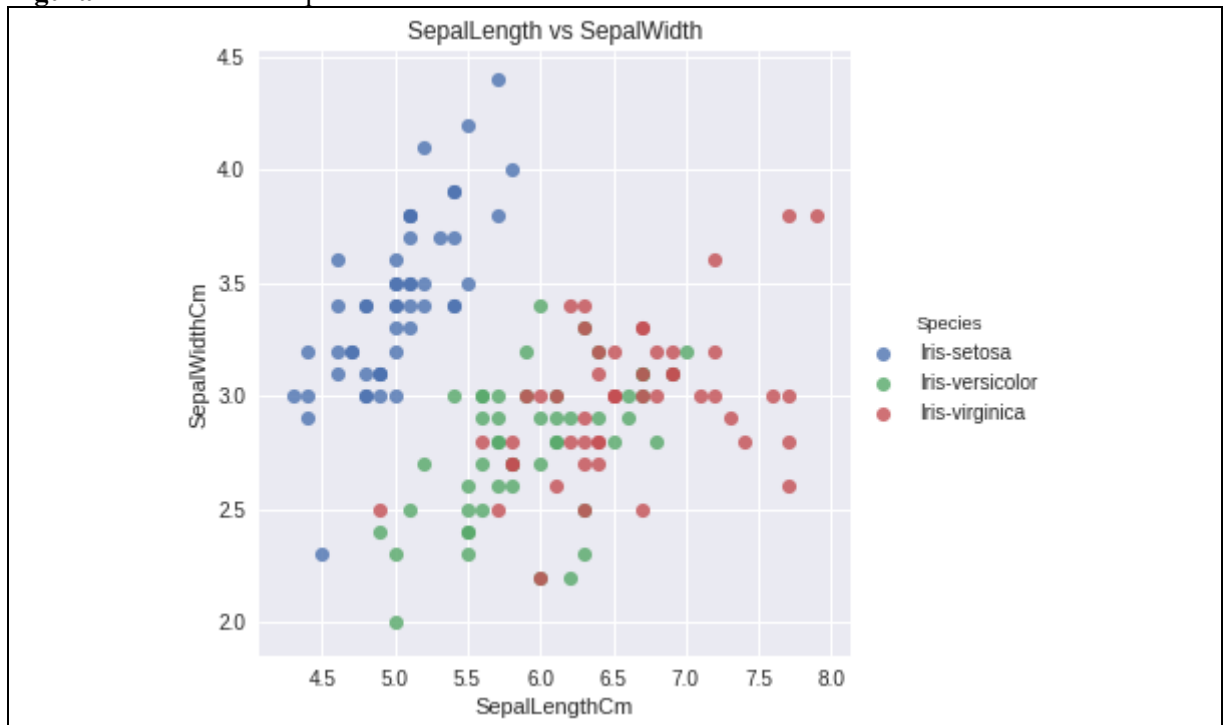
Figura 1 - Estrutura de um neurônio artificial



Fonte: Silva (2010, p. 34).

A estrutura mais simples de rede neural existente é a *perceptron* de uma camada, que utiliza apenas um neurônio artificial e serve para classificar dados binariamente, em duas categorias, quando estes apresentam uma distribuição linearmente separável, ou seja, quando se pode traçar uma reta divisória clara entre as classes 0 e 1, A e B, etc., num gráfico de dispersão. Um exemplo de dados classificáveis pela *perceptron* simples está o famoso *dataset* íris, contendo dados (dimensões das pétalas e sépalas) de flores de íris das espécies setosa, versicolor e *virginica*. Tais dados foram coletados nos Estados Unidos durante as décadas de 1920 e 1930, originalmente, pelo botânico Edgar Anderson, vindo pela primeira vez a público em discussões sobre a técnica estatística de análise discriminante linear (Unwin, Kleinman, 2021). Abaixo é exibido um gráfico de dispersão referente aos dados de dimensões das sépalas das flores, onde se pode notar uma separação entre classes.

Figura 2 - Gráfico de dispersão das medidas de flores de íris



Fonte: Ong (2016, *online*).

Como exemplo prático, uma rede *perceptron* simples poderia ser alimentada e treinada com dados de largura e altura das sépalas e pétalas das espécies setosa e versicolor, estando, ao final do treinamento, apta a classificar, com o mínimo de margem de erro, novos dados de entrada de flores de íris. Considerando-se que os rótulos, ou informações da espécie de flor, foram dados *a priori* (aprendizado supervisionado), o processo de treinamento consiste em: alimentar a rede com valores de entrada do *dataset* (x), ajustando iterativamente os valores de pesos (w) e limitar de ativação (θ) até que os valores de saída (y) gerados estejam iguais aos valores reais de rótulos dados (Silva, 2010). Desta maneira, a rede aprende a identificar padrões nas proporções de dados apresentados e classificar os resultados corretamente.

Ainda que a *perceptron* simples tenha sido pioneira, ela se mostra limitada à classificação binária, sendo necessário utilizar arquiteturas mais robustas para tarefas complexas, com grande número de camadas de neurônios interligados entre a entrada e a saída da rede (camadas ocultas), que processam a informação de modo cada vez mais

profundo. Tratando-se especificamente de tarefas de PNL, o estado da arte da computação determina o uso de redes do tipo transformador ou *transformer*. Tais arquiteturas aliam redes densas, com várias camadas de neurônios ocultas, com mecanismos de atenção, que mantêm registro de informações contextuais que ajudam a analisar a relevância dos termos (*tokens*) em relação uns aos outros, para compreensão de seu papel sintático/semântico na frase.

Considere, por exemplo, que a rede tenha que analisar um período longo, como, por exemplo, um trecho do poema Canção do exílio, de Dias Gomes: ‘Minha terra tem palmeiras, Onde canta o Sabiá; As aves, que aqui gorjeiam, Não gorjeiam como lá’. No momento em que uma rede tradicional, como uma rede neural recorrente (RNN) chegasse no ponto de analisar a palavra ‘lá’, ao final, possivelmente teria sofrido com a perda de gradiente (perda contínua de sinais de ajuste dos pesos sinápticos) e já não conseguiria identificar que ‘lá’ se refere a ‘minha terra’ no início da frase. A seguir, conceitua-se o que vem a ser uma rede recorrente, bem como uma rede *feed forward*, aspectos importantes de serem entendidos para sustentar as argumentações que se seguem:

Uma rede *feedforward* ou de alimentação progressiva, como o nome sugere, tem conexões apenas em uma direção – ou seja, ela forma um grafo acíclico direcionado com nós de entrada e saída designados. Cada nó calcula uma função de suas entradas e passa o resultado para seus sucessores na rede. As informações fluem pela rede dos nós de entrada para os nós de saída e não existem laços. Uma rede recorrente, por outro lado, alimenta suas saídas intermediárias ou finais de volta em suas próprias entradas. Isso significa que os valores do sinal dentro da rede formam um sistema dinâmico que inclui estado interno ou memória (Russel *et al.*, 2022, p. 680).

Para resolver a situação de perda de gradientes, surgiram as *transformers* com mecanismo de autoatenção, capaz de se lembrar da informação no início da frase, e dar a ela seu devido peso na interpretação geral de significado.

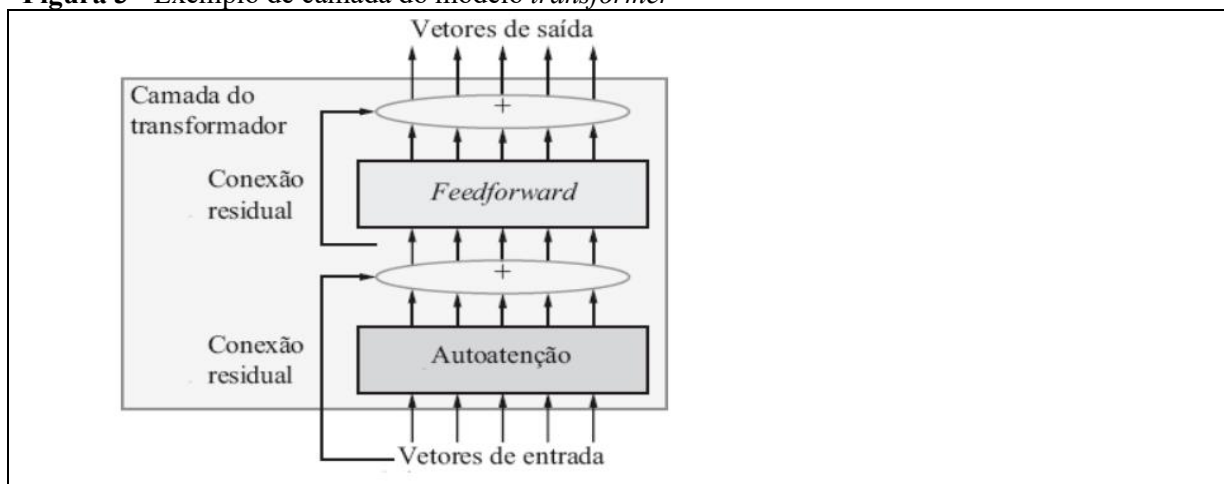
As redes neurais de transformadores, ou simplesmente transformadores (do inglês, *transformers*) revolucionaram o campo do processamento de linguagem natural e de outras tarefas envolvendo dados sequenciais, oferecendo uma abordagem altamente paralelizável e eficaz para lidar com dependências de longo alcance em sequências. Introduzidos por Vaswani [...], os transformadores se destacam por utilizar mecanismos de atenção em vez de conexões recorrentes (como nas RNNs) para processar dados sequenciais (Faceli *et al.*, 2025. p. 130).

Simplificadamente, a arquitetura de uma *transformer* é composta de um codificador e um decodificador, contendo várias camadas, cada qual com seu mecanismo de autoatenção e uma rede neural densa, do tipo *feed forward* (alimentação para adiante). A arquitetura detalhada da rede é explicada pelos autores supracitados (2025): o codificador recebe uma entrada (frase), a desmembra em unidades menores, *tokens*, e a converte esses *tokens* em uma representação processável (vetor numérico, ou *embedding*), passando essa sequência para o mecanismo de autoatenção, que possibilita que cada parte da *token* preste atenção em todas as outras partes, sendo cada *token* repassado para a rede *feed forward*, para que esta aprenda representações mais complexas; no decodificador, a representação gerada pelo codificador é recebida para geração do resultado de saída, passando também por camadas de *feed forward* e autoatenção, tendo como diferencial o mecanismo de atenção cruzada, que mantém o foco na sequência de entrada e também em todo resultado já gerado, garantindo uma saída final que leve em conta o contexto completo da sequência.

Ademais, o advento das *transformers* permitiu o processamento paralelo (concomitante) dos *tokens* de dados, permitindo um ganho de desempenho sem precedentes, se comparado a soluções anteriores, como as redes neurais recorrentes, que processavam os dados em ordem sequencial, o que permite a análise eficiente e rápida de grandes volumes de dados (Stryker, Bergmann, 2025).

A Figura 3 mostra o esquema de uma camada de rede *transformer* simplificada, constando os elementos acima descritos.

Figura 3 - Exemplo de camada do modelo *transformer*



Fonte: Russel *et al.* (2022, p. 789).

Regressando ao exemplo do trecho de Dias Gomes, mencionado anteriormente nesta seção, o mecanismo de autoatenção, analisando a frase como um todo, atribuiria ao pronome ‘lá’ um peso de atenção de valor 1 (100%) em relação à expressão ‘minha terra’, indicando forte correlação entre esses termos, ao mesmo tempo que em poderia atribuir peso 0 à relação entre ‘lá’ e ‘sabiá’, indicando que estas palavras não estão diretamente relacionadas. Mesmo que ‘lá’ e ‘minha terra’ estejam bastante distantes na frase, a *transformer* consegue se lembrar de relações contextuais, evitando a perda de gradiente. Desta forma, a rede consegue ‘entender’ o sentido geral da frase e a relação de seus componentes entre si.

Stryker e Bergmann (2025, *online*) asseveram que as *transformers* são empregadas em processamento de linguagem natural - NLP, executando funções como “[...] *chatbots*, geração de texto, sumarização, resposta a perguntas e análise de sentimento”, servindo ainda de base para “LLMs somente de decodificadores autorregressivos, como o modelo GPT-3 (abreviação de *Generative Pre-trained Transformer*) que impulsionou o lançamento do ChatGPT da OpenAI, catalisou a era moderna da IA generativa (IA gen)”.

Conforme a empresa IBM (2025) os LLMs são redes neurais treinadas com grandes quantidades de dados, e bilhões de parâmetros, de modo a reconhecer padrões complexos, que os permitam gerar textos como seres humanos, dentre outras tarefas, reconhecendo contextos para gerar respostas coerentes. A curadoria dos dados fornecidos ao modelo é de crucial importância para a corretude e até mesmo para aspectos éticos dos resultados obtidos:

Os dados agem como o manual de instruções que guia o processo de aprendizado do modelo. Eles ensinam o LLM a entender a gramática e a sintaxe e a capturar nuances contextuais e culturais da linguagem. Para que seu desempenho seja bem-sucedido, é vital que ele seja exposto a um conjunto de dados diversificado e abrangente. Isso ajuda a minimizar o viés e a garantir que o modelo possa se comunicar de forma eficaz e precisa em

uma variedade de situações. Assim, a seleção e a curadoria de dados de treinamento são fundamentais para desenvolver um LLM que seja tecnicamente competente e culturalmente sensível e adaptável a diferentes contextos de uso (Barcaui, 2025, p. 137).

Contudo, mesmo sistemas tão intrincados não estão a salvo de cometer erros crassos, definidos como alucinação de IA. As alucinações, mais especificamente, se referem a respostas incorretas e/ou tendenciosas, mas que são aparentemente verossímeis (BBC Brasil, 2023). Dois pontos relevantes para a ocorrência de tais problemas são o fato de os LLMs como ChatGPT e afins serem treinados com dados relativamente antigos, além do fato de se alimentarem de dados produzidos corriqueiramente na Internet, que podem também conter inexatidões e vieses dos mais diversos (*Op. cit.* 2025).

As alucinações geralmente ocorrem quando o modelo preenche lacunas com base em contextos semelhantes de seus dados de treinamento, ou quando é criado usando dados de treinamento tendenciosos ou incompletos. Isso leva à suposições incorretas [...] (Moraes, 2025, *on line*).

É relevante destacar que o processo de alucinação pode ser classificado em tipologias que ajudam a entender suas diversas origens: alucinação de conflito de entrada, quando os resultados apresentados divergem notavelmente da questão colocada pelo usuário, o que representa a dificuldade da IA de entender sua solicitação; conflito factual, quando os resultados divergem perceptivelmente de fatos já sabidos; alucinação contextual, quando a IA perde o fio da argumentação durante repetidas interações com o usuário, quando este busca o aprofundamento sucessivo dos resultados (Zang *et al.*, 2023 *apud* Lin, 2024).

A análise estrutural da tecnologia de rede neural utilizada para PLN, como o Chat GPT e afins, pode elucidar algumas causas possíveis do processo de alucinação. Conforme Barcaui (2025), são elencadas como razões da alucinação: a rede não entende efetivamente o significado do que está processando, apenas tenta prever a resposta com maior probabilidade de acerto mediante os padrões já aprendidos; às vezes, por razões aleatórias, a palavra menos provável pode ser escolhida como saída; os modelos trabalham com recursos computacionais limitados, como tempo de processamento e memória, o que interfere em sua janela de contexto, a quantidade de texto anterior que conseguem se lembrar para gerar respostas; os modelos de IA não aprendem ao longo do tempo, sendo treinados apenas uma vez e depois usados para inferência; se a entrada do usuário for vaga e/ou ambígua, a probabilidade de alucinação é maior, pois o modelo tentará preencher as lacunas de sentido com os padrões que já aprendeu.

Para Lin *et al.* (2023) as causas da alucinação podem estar no nível dos dados ou no nível do modelo de IA: alimentar a rede com informações falsas ou imprecisas aumenta a ocorrência de erros, ainda mais quando existe redundância de informações falsas, o que fará a rede dar mais peso a elas na geração dos resultados; a arquitetura e tamanho dos modelos pode influenciar, sendo que modelos de código aberto, construídos com um número reduzido de parâmetros têm maior propensão a alucinar, principalmente quando treinados com conjuntos de dados estáticos; podem gerar erros ao serem ao acompanhar um fluxo de interações com o usuário, onde se deve aprofundar as respostas novas com base nas anteriores, tendo desempenho medíocre também em cenários não vistos anteriormente pela rede.

A própria complexidade de modelos de redes neurais utilizados em atividades de *deep learning* também pode facilitar a ocorrência de erros, no sentido em que sua vertiginosa quantidade de neurônios interconectados em diversas camadas e os inúmeros parâmetros de configuração impedem a explicabilidade do funcionamento da rede, do modo exato como ela

converte as entradas em resultados finais, tornando-se uma espécie de caixa preta (Reis, 2024; Barbosa *et al.*, 2024).

Concernente à avaliação do nível de alucinação, pelo menos três estratégias podem ser utilizadas: avaliação humana, baseada em estatística e baseada em modelos. De acordo com Lin *et al.* (2025): a avaliação humana implica na análise de resultados por especialistas reais que geram um *score* de pontuação mediante a corretude dos fatos; na avaliação estatística é medida quantitativamente a correspondência entre as respostas geradas e os rótulos reais/textos alvo; na avaliação baseada em modelos, se utiliza uma rede neural adicional para validar as respostas geradas e ambas são comparadas, podendo utilizar bases de conhecimento consolidadas para este fim.

Moraes (2025) menciona que algumas empresas de tecnologia tentam mitigar as alucinações por meio do treinamento de suas redes com dados de alta qualidade, bem como limitar as respostas de acordo com diretrizes pré-determinadas, aconselhando, contudo, que respostas que exigem alta precisão sejam validadas por especialistas humanos. Barcaui (2025), por sua vez, recomenda que a eficácia do uso de LLMs está atrelada a estruturar perguntas claras e com o máximo de detalhes possível, podendo o usuário ir elaborando pergunta após pergunta, até chegar ao nível de refinamento necessário. Em campos de conhecimento específicos, utilizar os conceitos e termos técnicos pertinentes de modo adequado é de grande valia.

Outro aspecto bastante preocupante do uso de IAs generativas é a questão da privacidade de dados. Conforme Lee *et al.* (2025), dados pessoais de usuários que por ventura sejam inseridos em um *prompt* podem aparecer para outros usuários sem as devidas permissões, sendo este fato agravado pelo uso de dados pessoais constantes na Internet no treinamento das redes, algo que frequentemente ocorre. Assim, depreende-se que o aspecto da alucinação aliado às violações de privacidade nas LLMs pode resultar inclusive na geração de pseudo-informações detrimenais para os indivíduos, o que traz sérias implicações legais. Estes autores (2025) ainda destacam que, embora as empresas de tecnologia se esforcem para mitigar os problemas da IA, a alucinação é algo inextricável e inerente aos modelos de LLMs devido à sua estratégia autorregressiva.

Tratando-se diretamente da CI, Park (2025) vislumbra a possibilidade de integração da IA em sistemas de organização do conhecimento, o que poderia melhorar lacunas nos métodos atuais de recuperação e recomendação da informação, que se baseiam geralmente em navegação por uma estrutura rígida de assuntos ou busca por palavras-chave específicas, vocabulários controlados, tesouros e afins, o que pode não corresponder às necessidades específicas do usuário. Contudo, a praticidade de criar consultas flexíveis e humanizadas via *prompt* de IA pode ser afetada pela possibilidade da alucinação, que, evidentemente, interfere na idoneidade das respostas. Como solução parcial para o problema da alucinação, esta autora (2025) aponta que as seguintes estratégias têm impacto sensível na qualidade das respostas geradas: o uso de metadados de qualidade e em abundância; a engenharia de *prompt*, com refinamento progressivo dos comandos providos; intervenção de especialistas, determinado a estrutura dos metadados, criando critérios de recomendações e avaliando resultados intermediários e finais.

Além da organização do conhecimento, Cunha (2025) faz uma ampla revisão de literatura onde destaca ainda outros usos possíveis das IAs, entenda-se, especificamente, aplicações de LLM e NLP: *chatbots* para auxiliar em serviços de referência; classificação automática de documentos; IA generativa na recuperação de informação. Esta vertente da IA também demonstra ter uma relação estreita com a Gestão do Conhecimento (GC), no sentido em que seus resultados são gerados a partir da combinação de conteúdos de conhecimento explícito (codificado), utilizados no treinamento das redes, trazendo saídas em formato

humanizado, que facilitam a sua assimilação, processo descrito como internalização, conforme Takeuchi e Nonaka (2008).

É também possível que com o refinamento progressivo das tecnologias de NLP novas utilidades sejam encontradas que venham a beneficiar a CI. É evidente que, pela simples constatação lógica, todos os âmbitos de emprego de tecnologias de LLM/NLP na CI também estão sujeitos às inconveniências do processo de alucinação. Por conseguinte, é de interesse crucial da CI pesquisar, problematizar, discutir e propor soluções para este problema.

3 PROCEDIMENTOS METODOLÓGICOS

O presente trabalho se caracteriza como uma pesquisa cientométrica, na qual foram computados dados referentes à produção acadêmico-científica sobre o tema de alucinação em IA na base de dados *Web of Science (WoS) - core collection*. A busca de artigos se deu em outubro de 2025, utilizando-se a *field tag* ‘TS’, que representa o título, resumo e palavras-chave de autor. Os termos de busca, em inglês, foram: *hallucination* (alucinação), *artificial intelligence* (inteligência artificial) e LLM. Os parâmetros completos de busca e filtros são descritos no Quadro 1:

Quadro 1 – Parâmetros de busca nas bases de dados

Atributo	Valor
String de busca	<i>TS="hallucination" and TS= ("AI" OR "Artificial intelligence") OR "LLM"</i>
Campos considerados	Tema (título, palavras-chave de autor e resumo)
Lapso temporal	Não aplicável
Áreas do conhecimento	Computer Science Artificial Intelligence; Computer Science Interdisciplinary Applications; Computer Science Theory Methods; Computer Science Information Systems; Computer Science Software Engineering; Computer Science Cybernetics; Computer Science Hardware Architecture; Information Science Library Science.

Fonte: Elaborado pelos autores (2016).

Devido ao seu caráter exploratório, não se restringiu, *a priori*, nenhum lapso temporal para recuperação dos trabalhos. Aplicou-se, contudo, um filtro concernente às áreas do conhecimento, retornando trabalhos da Ciência da Informação e de áreas afins à computação, a saber: Ciência da Computação Inteligência Artificial; Aplicações Interdisciplinares da Ciência da Computação; Métodos da Teoria da Ciência da Computação; Sistemas de Informação da Ciência da Computação; Engenharia de Software da Ciência da Computação; Cibernética da Ciência da Computação; Arquitetura de Hardware da Ciência da Computação.

O *corpus* de publicações foi então classificado e contado por lapso temporal, país e área do conhecimento, sendo elaborados tabelas e gráficos para análise. Efetuou-se a comparação dos dados obtidos ao longo do tempo das áreas correlatas à computação e à CI, com todas as demais áreas do conhecimento disponíveis na WoS, para verificar se haveria diferenças nas tendências de comportamento dos dados neste respeito. Utilizou-se a

linguagem R para calcular as medidas principais de estatística descritiva sobre os dados e geração de gráficos.

4 RESULTADOS E DISCUSSÃO

Inicialmente, foi executada uma busca na WoS, utilizando os parâmetros descritos na seção 3, Quadro 1, recuperando artigos referentes aos termos (em inglês): alucinação e IA ou LLM. Considerou-se as áreas do conhecimento correlatas à computação e à CI, resultando em 399 trabalhos, ficando em primeiro lugar no ranking de publicações da área de Ciência da Computação - Inteligência Artificial, com 268 trabalhos, e em último lugar, Ciência da Computação - Arquitetura de Hardware, com apenas sete artigos. A CI fica em penúltimo, com 19 trabalhos. Os resultados desta análise por áreas aparecem na Tabela 1, abaixo. É preciso esclarecer que a simples soma dos artigos em cada categoria ultrapassa o total geral de 399. Isto se explica pelo fato de que um mesmo trabalho pode ser classificado em duas ou mais áreas ao mesmo tempo.

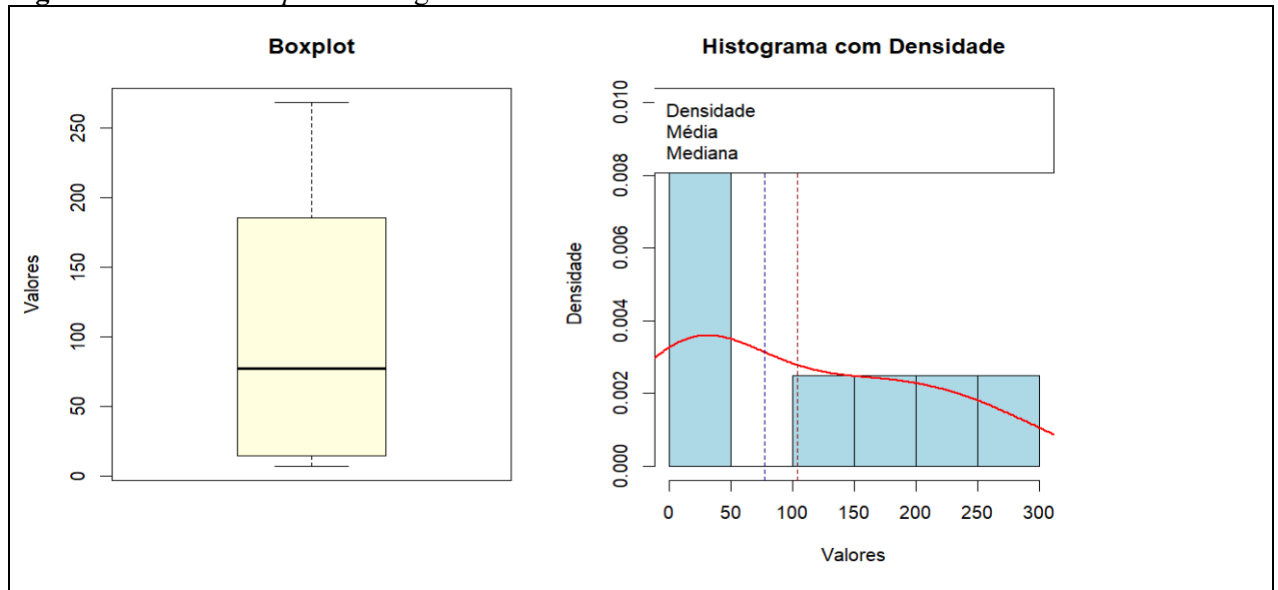
Tabela 1 - Distribuição de publicações por área

Área de Pesquisa	Quantidade
Ciência da Computação - Inteligência Artificial	268
Ciência da Computação - Aplicações Interdisciplinares	210
Ciência da Computação - Teoria e Métodos	161
Ciência da Computação - Sistemas de Informação	115
Ciência da Computação - Engenharia de Software	39
Ciência da Computação - Cibernética	10
Ciência da Computação - Arquitetura de Hardware	7
Ciência da Informação - Biblioteconomia	19
Total geral (sem sobreposição de áreas)	399
Total geral (com sobreposição de áreas)	829
Média trabalhos/área	103,62
Desvio padrão	100,85
Mediana	77

Fonte: Elaborado pelos autores (2025).

Observando os dados da Tabela 1 pode-se verificar que há uma disparidade grande de publicações por área, sendo que poucas áreas concentram a maior parte das publicações, e o restante publica próximo da média (103,62) ou bem abaixo dela. O desvio padrão, que mede a dispersão dos dados em relação à média, estando bastante elevado (100,85), salienta essa discrepância. Especialmente preocupante é a posição da CI, que está 81,66% abaixo do valor da média. Abaixo, a Figura 4 mostra a combinação dos gráficos de *boxplot* e histograma com densidade, referentes à distribuição de produção por área:

Figura 4- Gráficos *boxplot* e histograma com densidade

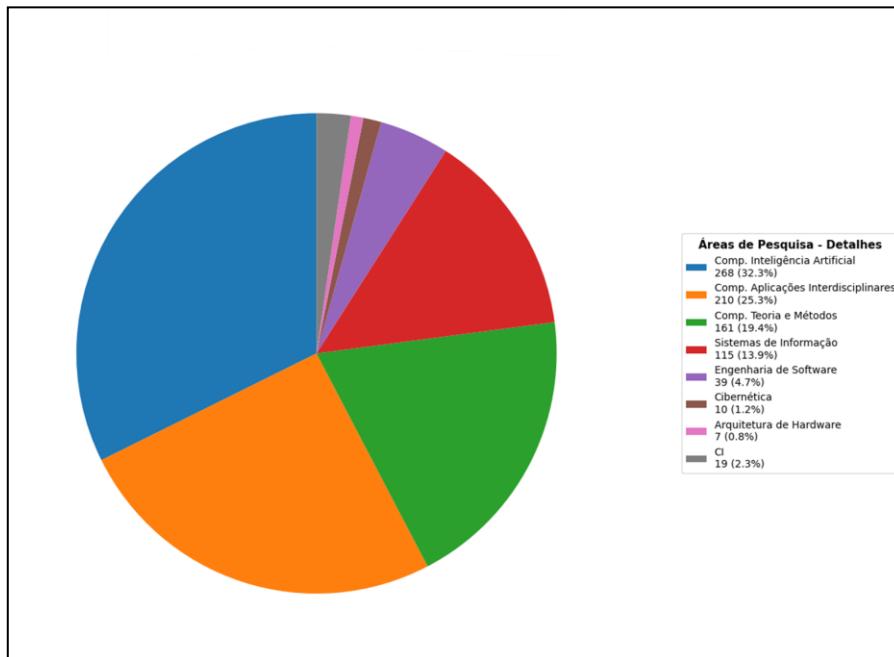


Fonte: Elaborado pelos autores (2026).

Na Figura 4, no gráfico à esquerda, a ‘caixa’ representa os quartis de Q_1 (valor que representa 25% da distribuição dos dados) a Q_3 (valor que representa 75% da distribuição dos dados), sendo que a linha interna representa a mediana, 77. A mediana se mostra baixa relativamente à amplitude total dos dados, tendo a maioria dos dados concentrados nos valores menores, representado pela caixa se situar em posição mais baixa. O “bigote” superior longo indica valores muito superiores ao restante dos dados.

Concernente ao gráfico à direita, verifica-se que a maioria dos valores está entre 0 e 100, com poucos valores muito altos. Observa-se uma curva de densidade assimétrica, com pico à esquerda e cauda alongada à direita. Observa-se ainda uma distância acentuada entre média (linha tracejada vermelha na vertical) e a mediana (linha tracejada azul). Percebe-se, assim, que a variabilidade dos dados é bastante alta, mesmo dentro de subáreas afins da computação, considerando-se também a CI, que tem uma interface de contato bastante relevante com essas áreas, tanto no aspecto prático quanto no teórico. O gráfico de setores da Figura 5, a seguir, traz uma representação mais simplificada e direta da contribuição de cada área para o tema de pesquisa, onde se pode verificar a grande fatia do gráfico ocupada pelas áreas de computação e inteligência artificial, computação e aplicações interdisciplinares e computação teoria e métodos. A CI ocupa seus tímidos 2% do gráfico. Esta situação, infere-se, é preocupante, pois o uso de LLMs/ NLP é de grande interesse para a área em aplicações como organização do conhecimento, sistemas de recuperação de informação e afins, sendo que o tema da alucinação impacta diretamente na qualidade e na corretude da informação gerada, atingindo um princípio que está no âmago da CI, que é a relevância pragmática e semântica da informação da perspectiva do ser humano.

Figura 5- Publicações sobre alucinação por área



Fonte: Elaborado pelos autores (2025).

A seguir são apresentados os dados de produção por país, onde se percebe Estados Unidos e China praticamente empatados no topo dos países mais produtivos, com uma diferença mínima de dois trabalhos entre eles. Este contexto mostra uma tendência chinesa de investimentos massivos em cooperações público-privadas para alavancar investimentos no campo da IA, sendo que a China tem até mesmo superado com larga vantagem os EUA em produção científica em certos nichos de IA, como mostra o trabalho de Souza e Mello (2025). No início do ano de 2025, empresas de tecnologia dos EUA sentiram forte queda no preço de suas ações com o lançamento de um novo modelo de IA de baixo custo, DeepSeek, que se equipara, ou mesmo supera o Chat GPT em desempenho (CNN, 2025). O Brasil, por sua vez, se encontra distante do *ranking* dos dez mais, na 27ª posição, com apenas dois artigos publicados.

O governo brasileiro, contudo, está concentrando esforços no desenvolvimento de ecossistemas de IA nacionais, que visam o atendimento de demandas socioeconômicas estratégicas para o país:

[...] durante a abertura do Painel Telebrasil 2025, a ministra da Gestão e da Inovação em Serviços Públicos, Esther Dweck, afirmou que o Governo Federal pretende investir R\$ 23 bilhões em uma política de Inteligência Artificial (IA) “que seja inclusiva, soberana, ética e centrada nas pessoas” (Agência Gov, 2025, *online*).

O investimento em tecnologias de IA, em realidade, é algo bastante complexo, que vai além do aspecto computacional. Visto a quantidade massiva de processamento que envolve o treinamento e operação de modelos do tipo GPT, já descritos anteriormente (Seção 2), Chang, Arcesati e Hmaidí (2025) descrevem alguns dos elementos necessários para um ecossistema de IA: pessoal capacitado; fontes de energia barata e abundante; fábricas para elaboração de chips especializados; *frameworks* de programação apropriados para o desenvolvimento de

soluções de IA. A Tabela 2 mostra o ranking dos dez de países que mais produzem sobre o tema alucinação, mais o Brasil, para fins de contraste.

Tabela 2- *Ranking* dos países que publicam sobre alucinação

Colocação	País	Quantidade
1º	Estados Unidos	134
2º	China	131
3º	Inglaterra	23
3º	Alemanha	23
5º	Singapura	21
6º	Canadá	20
7º	Austrália	19
7º	Japão	19
9º	Índia	18
10º	Coreia do Sul	16
⋮	⋮	⋮
27º	Brasil	2

Fonte: Elaborado pelos autores (2026).

Já a Tabela 3 exibe os dados gerais de produção em todos os países constantes na base WoS, referente às medidas básicas de estatística descritiva. Justifica-se que a base de dados faz a contagem dos países por meio da filiação dos autores. Sendo assim, um mesmo artigo pode contar para diferentes países.

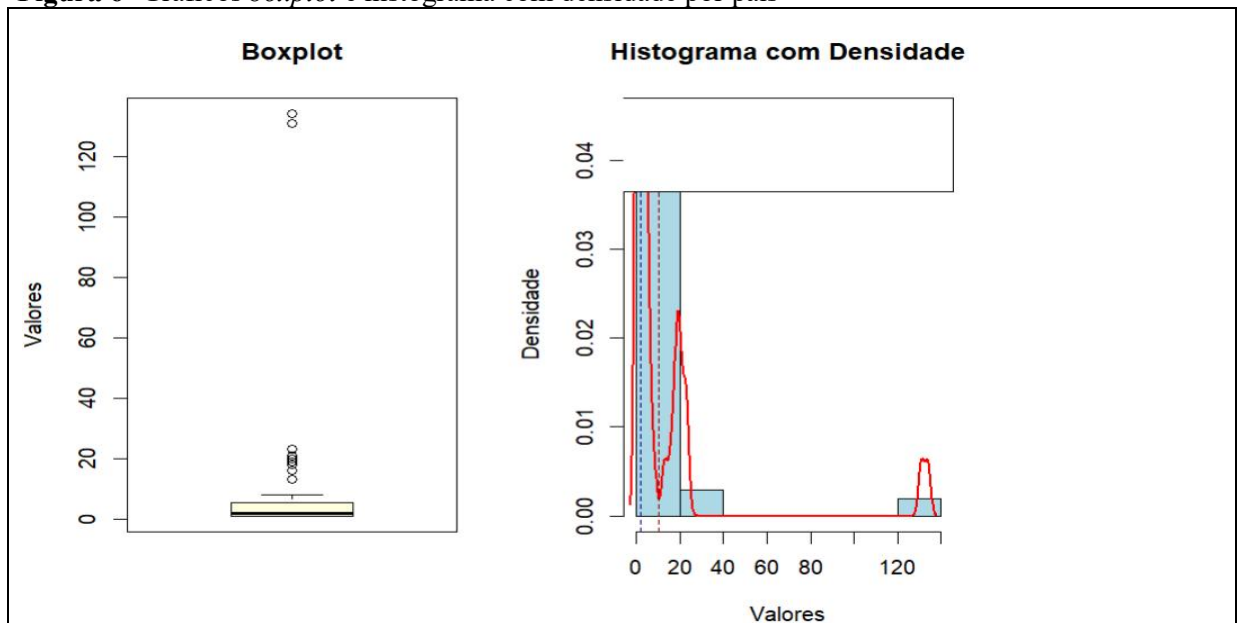
Tabela 3 - produção científica por país - dados totais

Parâmetro	Valor
Total de países	54
Total de produção por país	531
Média de artigos publicados	10,21
Mediana de artigos publicados	2
Desvio padrão dos artigos publicados	25,6

Fonte: Elaborado pelos autores (2026).

Observa-se na Tabela 3 que a mediana (2) está bastante baixa em relação à média (10,21), com 50% dos países com produção menor ou igual a 02. Existe uma concentração de dados extrema, com metade dos países com produção muito baixa, enquanto que 28 a 30% têm uma produção discrepantemente alta. Na Figura 6, são apresentados os gráficos que mostram a dispersão e densidade dos dados.

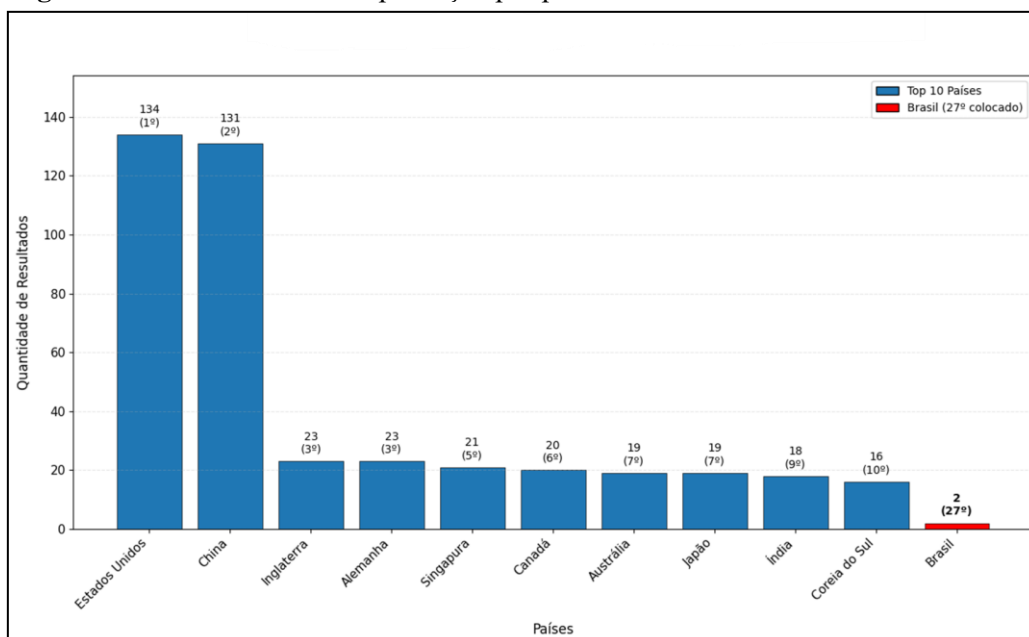
Figura 6- Gráficos *boxplot* e histograma com densidade por país



Fonte: Elaborado pelos autores (2025).

No gráfico *boxplot*, a caixa se mostra achatada e com valores concentrados em até 2, sendo que EUA e China têm valores tão distintos dos demais que são interpretados como *outliers*. Referente ao histograma com densidade, é observado um pico mais saliente à esquerda, com dados concentrados em valores mais baixos, com uma cauda alongada à direita e clara lacuna entre os países que mais publicam relativo aos que menos publicam. Para encerrar a análise desta parte dos dados, a Figura 7 traz um gráfico de barras com uma visualização mais simplificada, de fácil entendimento, da produção por país.

Figura 7: Gráfico de barras da produção por país



Fonte: Elaborado pelos autores (2026)

A Tabela 4 abaixo exhibe a quantidade de artigos publicados por ano. Pelo caráter exploratório da pesquisa, não foi definido nenhum lapso de tempo a priori. Os dados foram coletados até a primeira semana de outubro de 2025, quando foi realizada a busca na base WoS. Possivelmente não por coincidência, a série de dados se inicia em 2017, ano da publicação do artigo seminal “*Attention Is All You Need*”, de Vaswani *et al.* (2017), que lança as bases para a elaboração das redes *transformers* com mecanismo de autoatenção, revolucionando verdadeiramente o campo da NLP.

Na tabela, observa-se uma grande diferença entre a média e a mediana, significando uma curva assimétrica à direita. Sendo a mediana muito menor do que a média, isto representa que a maioria dos valores se concentra em números baixos. O desvio padrão é alto, indicando alta variabilidade nos dados. Como se pode observar, houve um grande salto de produção a partir de 2024.

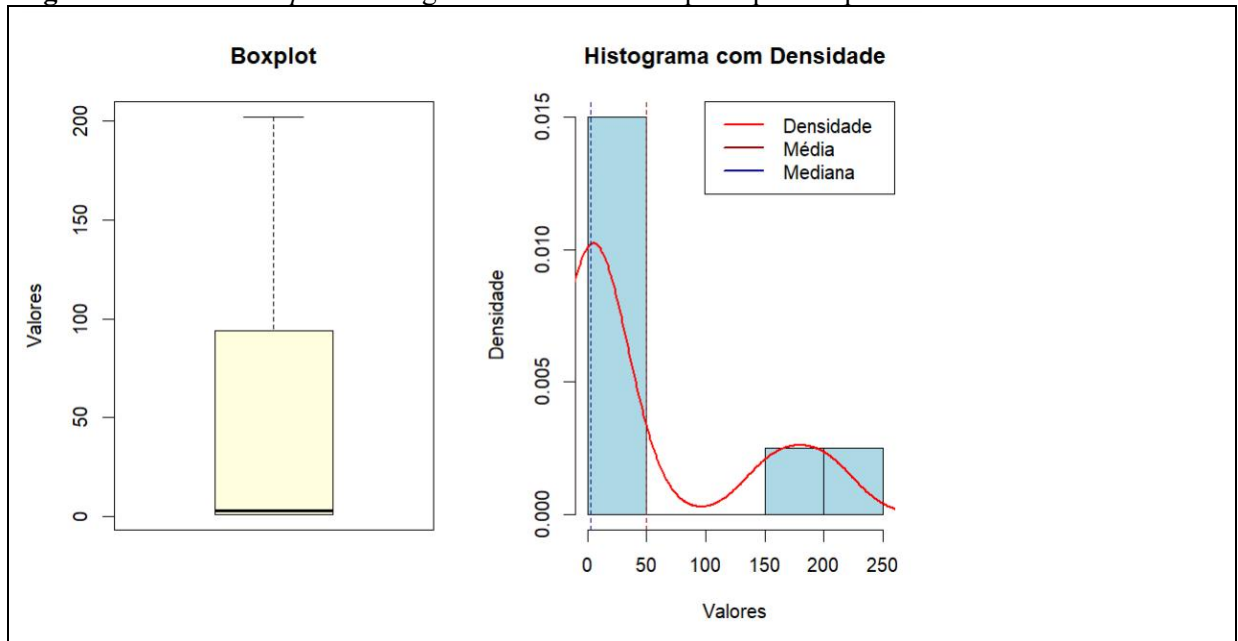
Tabela 4- Produção por ano nas áreas correlatas à computação e à CI

Ano	Quantidade de Publicações
2017	1
2019	1
2020	1
2021	2
2022	4
2023	31
2024	202
2025	157
Total	399
Média	49,88
Mediana	3
Desvio padrão	81,53

Fonte: Elaborado pelos autores (2026).

Na Figura 8 seguem os gráficos *boxplot* e histograma com densidade. A caixa do *boxplot*, que mostra dos 50% de dados centrais está muito próxima de zero e o “bigote” superior se mostra extremamente longo, representando a discrepância dos dados. No histograma se nota uma assimetria acentuada à direita, com uma cauda longa nessa direção, com um pico da curva de densidade à esquerda, indicando a concentração em números baixos.

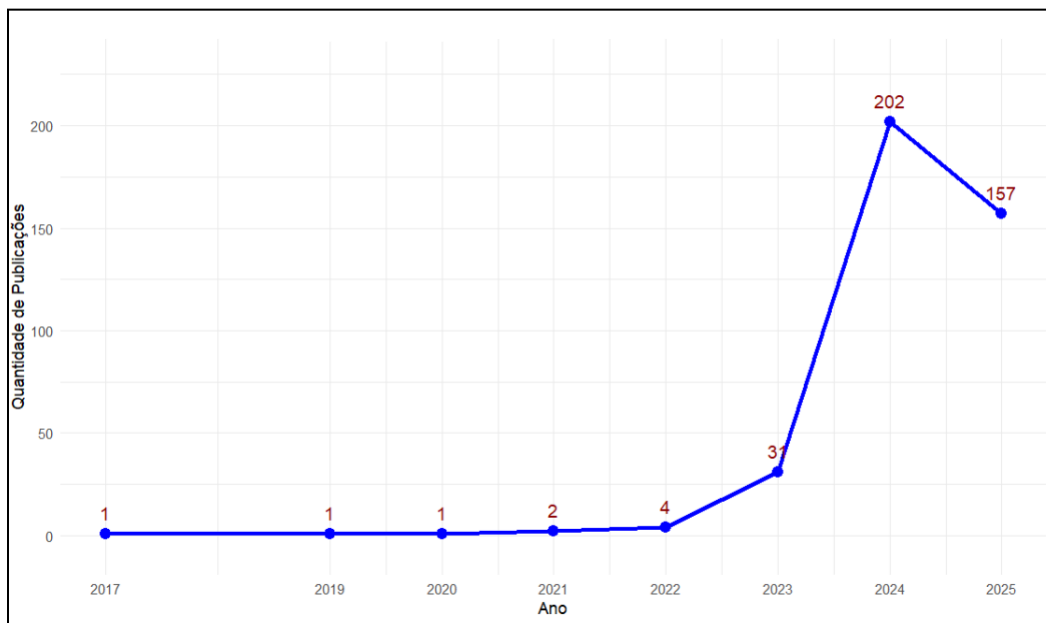
Figura 8- Gráficos *boxplot* e histograma com densidade por lapso temporal



Fonte: Elaborado pelos autores (2026).

A Figura 9 mostra um gráfico de linha, onde se vê um crescimento exponencial de publicações em 2023 e um gigantesco salto em 2024. É interessante notar que dois marcos históricos para a área de NLP/LLM se vêem representados no gráfico: o advento da tecnologia *transformer* em 2017, inaugurando a série de dados, e o crescimento acelerado um ano após o lançamento do Chat GPT, em 2022, que, possivelmente, despertou o interesse da comunidade acadêmica sobre o tema. Parece haver uma queda brusca de produtividade em 2025, mas ela é só aparente, pois, como já mencionado, os dados foram coletados na primeira semana de outubro de 2025, antes do término do ano, estando assim, ‘incompletos’.

Figura 9- Gráfico de linha de produção por lapso temporal



Fonte: Elaborado pelos autores (2025).

É necessário, para se ter uma visão mais realista do cenário que se apresenta, realizar uma projeção para o ano de 2025 ‘cheio’. A Tabela 5 mostra a projeção de publicação de 188 artigos, pelo menos, até o fim do ano, considerando a taxa de publicações nos dez meses (15,7%), o que mostra uma queda muito leve na produção, de -6,9%, o que também deve levar em conta a tendência de estabilização do tema após o pico de 2024 (+551%), demonstrando um possível amadurecimento da área de pesquisa, relativo às áreas da computação e CI combinadas. A tendência de produção no tema, portanto, ainda se mostra promissora.

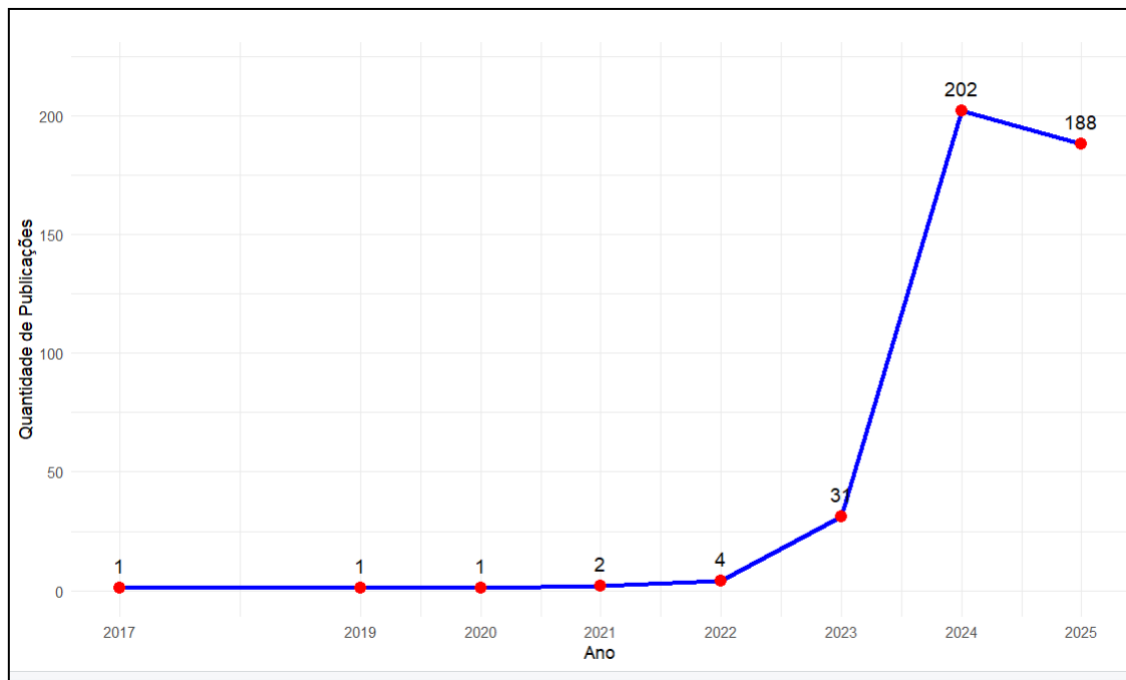
Tabela 5- Taxas anuais e projeções de publicação

Parâmetro	Valor
Meses 2024	12
Publicações 2024	202
Crescimento relativo ao ano anterior	+551%
Taxa mensal publ. 2024	16,8
Meses 2025	10
Publicações 2025	157
Taxa mensal 2025	15,7
Projeção total publ. 2025 (15,7 *12)	188
Crescimento relativo ao ano anterior	-6,9%

Fonte: Elaborado pelos autores (2025).

O gráfico da Figura 10 já incorpora as projeções realizadas e deixa claro o cenário de estabilização do tema de pesquisa.

Figura 10- Projeção de produção sobre o tema



Fonte: elaborado pelos autores (2026).

Enquanto que na análise combinada da computação com a CI a tendência é de estabilização, quando se compara a produção de todas as áreas da WoS no mesmo período, pode-se notar uma franca tendência de crescimento, como demonstrado na Tabela 6 que, inclusive, já traz alguns trabalhos já antecipadamente publicados nos números de 2026 de alguns periódicos.

Tabela 6: Publicações por lapso temporal em todas as áreas.

Ano	Quantidade de Publicações
2017	1
2019	1
2020	1
2021	2
2022	6
2023	57
2024	279
2025	271
2026	2

Fonte: Elaborado pelos autores (2026).

Na Tabela 7 pode-se notar uma projeção de crescimento de 16,49% para todas as áreas do conhecimento em termos de produção sobre o tema alucinação. Pode-se inferir que, enquanto a computação já tem o tema em vias de consolidação, possivelmente influenciada por não haver ainda caminhos técnicos para resolução definitiva do problema, as demais áreas do conhecimento parecem ter ainda curiosidade de explorar diferentes vertentes do assunto, problematizá-lo por suas distintas óticas.

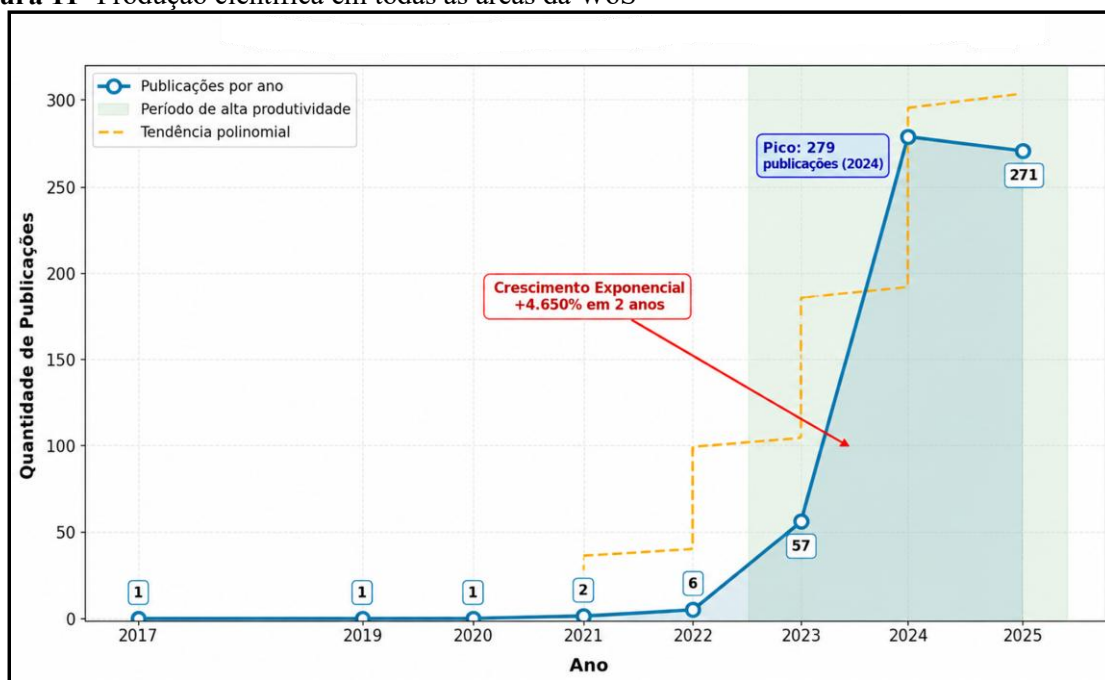
Tabela 7- Taxas anuais e projeções de publicação - todas as áreas

Parâmetro	Valor
Meses 2024	12
Publicações 2024	279
Crescimento relativo ao ano anterior	+389,47%
Taxa mensal publ. 2024	23,25
Meses 2025	10
Publicações 2025	271
Taxa mensal 2025	27,10
Projeção total publ. 2025 (27,10 *12)	325
Crescimento relativo ao ano anterior	+16,49%

Fonte: Elaborado pelos autores (2026).

O gráfico da Figura 11 traz uma dimensão visual da evolução da produção científica sobre o tema abordado ao longo do tempo, mostrando, na linha tracejada amarela, a projeção futura para o fim de 2025, com tendência de crescimento.

Figura 11- Produção científica em todas as áreas da WoS



Fonte: Elaborado pelos autores (2026).

Pode-se conjecturar que o crescimento exponencial mostrado no gráfico a partir de 2023 pode ter conexão com o lançamento do pioneiro chat GPT, em novembro de 2022, o que popularizou o uso amplo da IA generativa entre a população, ao mesmo tempo que trouxe à tona problemas apresentados pela alucinação.

Por fim, é relevante verificar, ainda, como a CI sozinha está lidando com o tema do presente trabalho ao longo do tempo. A Tabela 8 mostra o desenvolvimento do tema na CI, que passou a ser abordado timidamente apenas a partir de 2023 e, até o presente ano, tendo produzido apenas 13 trabalhos. Duas questões podem ser aqui ponderadas, que tocam a maneira como a relação da CI com a tecnologia é espelhada em uma das bases de dados científica mais prestigiadas do mundo. Primeiramente, percebe-se que a CI passou a tratar do assunto alucinação apenas seis anos após o advento da tecnologia *transformer* e seus derivados. Isto pode ser um tanto preocupante, no sentido em que toca temas caros à área, como sistemas de organização do conhecimento, recuperação de informação e até mesmo a literacia ou competência informacional, como se queira chamar. Trindade e Oliveira (2024, p. 12), ao tratarem da importância da competência em informação, destacam a análise crítica que o cientista deve exercitar ao utilizar IA generativa em seu trabalho, destacando que “[...] essas ferramentas podem sintetizar textos que têm dados/informações existentes e imprecisos [...]”. Claramente, estes efeitos comentados pelos autores são resultado do processo de alucinação, principalmente quando falam de alusão a informações inexistentes. O termo alucinação, contudo, não aparece em nenhum momento no trabalho dos autores supracitados.

Indaga-se se a CI já teria condições de problematizar o tema pioneiramente, contando com seu vasto arcabouço de saberes já construídos. O segundo ponto de crítica é constatar que até o momento presente seu nível de publicações continua ainda baixo em um tema de tanta relevância, em um repositório de publicações científicas de não menor relevância.

Tabela 8- Totais de produção por ano - CI

Ano	Quantidade de Publicações
2023	1
2024	4
2025	13

Fonte: Elaborado pelos autores (2026).

Na Tabela 9 pode-se observar que a taxa de crescimento de produção da CI está projetada para ser 225% maior em 2025 que no ano anterior. Isto em valores relativos. Tratando-se de números absolutos, os valores são ainda baixos, com total projetado de 15,6 publicações em 2025. Há, portanto, ainda muito que explorar no entendimento do tema pela ótica da CI, uma necessidade premente e urgente.

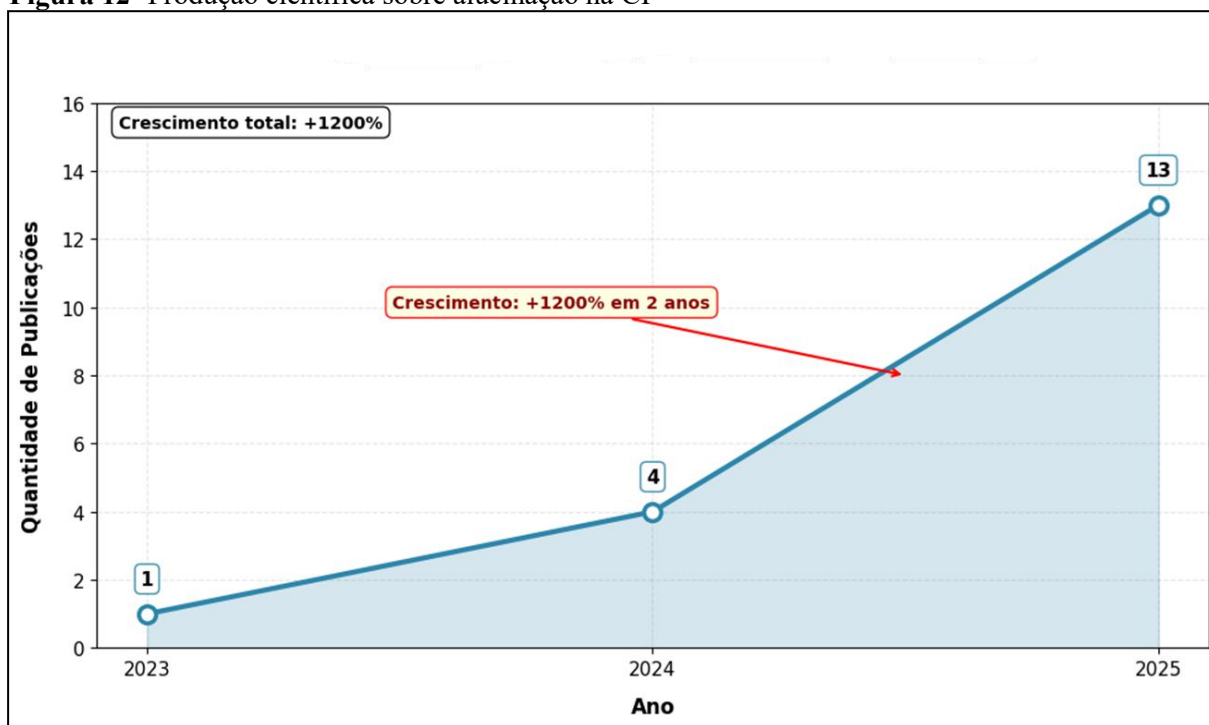
Tabela 9- Taxas anuais e projeções de publicação -CI

Parâmetro	Valor
Meses 2024	12
Publicações 2024	4
Crescimento relativo ao ano anterior	+300%
Taxa mensal publ. 2024	0,33
Meses 2025	10
Publicações 2025	13
Taxa mensal 2025	1,3
Projeção total publ. 2025 (1,3 *12)	15,6
Crescimento relativo ao ano anterior	+225%

Fonte: elaborado pelos autores (2026).

O Gráfico da Figura 12 mostra a evolução do interesse pela alucinação em IA especificamente pela perspectiva da CI, onde a área acompanha a tendência de um estirão de crescimento a partir de 2024, com os números projetados para continuarem subindo.

Figura 12- Produção científica sobre alucinação na CI



Fonte: Elaborado pelos autores (2026).

Diante dos dados apresentados, das análises e discussões realizadas, verificou-se a relevância crucial do tema alucinação para o momento contemporâneo. O corpus de publicações é ainda, de certo modo incipiente, o que traz oportunidades para a elaboração de mais pesquisas em diversas áreas, que aprofundem a problematização do tema e tragam diferentes perspectivas para sua mitigação ou mesmo resolução técnica. A CI, acredita-se tem condições de contribuir de modo mais efetivo nas pesquisas, tanto teórica quanto tecnicamente, construindo pontes interdisciplinares mais consistentes com a computação. Ademais, pode-se também conjecturar se o relativo atraso da CI em abordar o tema alucinação teria ocorrido pela possível dependência da utilização de resultados gerados primariamente pela computação. Entretanto, devido à longa tradição da CI em lidar com temas como a competência/literacia informacional, ética da informação e afins, poderiam dar condições à área de contribuir de modo mais intensivo para a pesquisa em alucinação.

O Brasil, particularmente, como nação, poderia direcionar os recursos governamentais recentemente aportados no esforço estratégico do desenvolvimento da IA para tratar desta questão emergente e premente.

5 CONSIDERAÇÕES FINAIS

Indubitavelmente, é urgente que se invista esforços em realizar pesquisas sobre o processo de alucinação em LLM/NLP, devido aos seus impactos científicos, educacionais e sociais. Esse tema toca no âmago de conceitos centrais para a CI, sendo que esta deve aliar-se de maneira mais efetiva à computação, podendo abordar o problema tanto do âmbito conceitual, humanístico como técnico. Por conseguinte, os objetivos do presente trabalho foram satisfatoriamente atingidos no que tange a: discutir aspectos técnicos de modo didático e compreensível; analisar o corpus de publicações por áreas de conhecimento, país e lapso

temporal; trazer à tona *insights* relevantes sobre os dados por meio de métodos estatísticos e representações visuais.

A despeito de a tecnologia de redes *transformer* ter sido lançada já em 2017, notou-se um início de crescimento expressivo no volume de publicações apenas após o lançamento do Chat GPT, que popularizou seu uso diante do público leigo. A CI, contudo, não acompanhou as discussões sobre alucinação desde o início, levando cerca de seis anos para fazê-lo. Sendo o processo de alucinação inextricável condição dos LLMs atuais, seria coerente esperar que este tema fosse tratado desde o início. No último ano da série de dados a CI tem demonstrado um crescimento relativo em interesse sobre alucinação, embora seus números absolutos sejam ainda muito baixos. Infere-se que o problema possa ser discutido no âmbito dos sistemas de organização do conhecimento, recuperação da informação, competência informacional e afins.

Enquanto que o assunto aparenta estar em vias de estabilização na computação, a CI e as demais áreas do conhecimento têm demonstrado potencial para avançar em novos picos de crescimento de volume de publicações. Espera-se que os atuais investimentos do governo brasileiro no aspecto estratégico da IA possam ser também direcionados para pesquisas sobre o processo de alucinação, para que o país possa diminuir sua grande disparidade com as demais nações, notadamente, Estados Unidos e China.

REFERÊNCIAS

AGÊNCIA GOV. Ministra Esther Dweck afirma que o Brasil investirá R\$ 23 bilhões em inteligência artificial. **Agência Gov**, Brasília, 2 set. 2025. Disponível em: <https://agenciagov.etc.com.br/noticias/202509/ministra-esther-dweck-afirma-que-o-brasil-investira-r-23-bilhoes-em-inteligencia-artificial>. Acesso em: 18 out. 2025.

BARBOSA, K. A. F. *et al.* Transformando Caixas-Pretas em Caixas de Vidro: Aumentando a Explicabilidade de Redes Neurais com Ferramentas de Visualização 3D. In: CONGRESSO LATINO-AMERICANO DE SOFTWARE LIVRE E TECNOLOGIAS ABERTAS, 21., 2024, **Anais [...]**. Foz do Iguaçu: Latinoware, 2024. p. 1-4. Disponível em: <https://latinoware.org/>. Acesso em: 18 out., 2025.

BARCAUI, André. **Guia da inteligência artificial**: do iniciante ao nerd. São Paulo: Almedina Brasil, 2025. E-book. ISBN 9786583400062. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9786583400062/>. Acesso em: 13 out. 2025.

BBC BRASIL. **O que é 'alucinação' de inteligência artificial**. 16 maio 2023. Disponível em: <https://www.bbc.com/portuguese/articles/cv24066nkpqo>. Acesso em: 07 out. 2025.

CNN BRASIL. Bolsas em NY abrem em queda com perdas provocadas por IA da China. São Paulo: **CNN Brasil**, 27 jan. 2025. Disponível em: https://www.cnnbrasil.com.br/economia/mercado/bolsas-em-ny-abrem-em-queda-com-perdas-provocadas-por-ia-da-china/#goog_rewarded. Acesso em: 16 out. 2025.

CHANG, W.; ARCESATI, R.; HMAIDI, A. **China's drive toward self-reliance in artificial intelligence**: from chips to large language models. Mercator Institute for China Studies, Berlin, 22/07/2025. Disponível em: <https://11nq.com/RH0TP>. Acesso em: 03 fev. 2025.

CUNHA, Murilo Bastos da. Bibliografia sobre o uso da Inteligência Artificial na biblioteca. **SciELO Preprints**, 2025. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/13055>. Acesso em: 17 out. 2025.

IBM. O que é LLM (large language models)? **IBM**, 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/large-language-models>. Acesso em: 07 out. 2025.

FACELI, Katti; LORENA, Ana C.; GAMA, João; ALMEIDA, Tiago Agostinho De; CARVA, André C. P. L. F De. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 3. ed. Rio de Janeiro: LTC, 2025. *E-book*. p.130. ISBN 9788521639213. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9788521639213/>. Acesso em: 12 out. 2025.

LEE, Chunsik; KIM, Junga; LIM, Joon Soo; SHIN, Donghee. Generative AI risks and resilience: How users adapt to hallucination and privacy challenges. **Telematics and Informatics Reports**, [S. l.], v. 19, p. 100221, 2025. Disponível em: <https://doi.org/10.1016/j.teler.2025.100221>. Acesso em: 08 out. 2025.

LIMA, Isaías. **Inteligência Artificial**. Rio de Janeiro: GEN LTC, 2014.

LIN, Zichao et al. Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. **Artificial Intelligence Review**, [S. l.], v. 57, n. 243, 2024. DOI: <https://doi.org/10.1007/s10462-024-10896-y>. Disponível em: <https://link.springer.com/article/10.1007/s10462-024-10896-y>. Acesso em: 09 out 2025.

MORAES, Ligia. O que acontece quando a inteligência artificial “inventa”. **Veja**, São Paulo, 29 mar. 2025. Seção Tecnologia. Disponível em: <https://veja.abril.com.br/ciencia/o-que-acontece-quando-a-inteligencia-artificial-inventa/>. Acesso em: [adicione a data de acesso].

ONG, Louis. Neural network approach to Iris dataset. **Kaggle**, [S. l.], 2016. Disponível em: <https://www.kaggle.com/code/louisong97/neural-network-approach-to-iris-dataset>. Acesso em: 11 out. 2025.

PARK, Ziyong. AI for KOS Discovery: Refining Search, Recommendation, and Hallucination Mitigation. **Knowledge Organization**, [S. l.], v. 52, n. 4, p. 458-489, 2025. Disponível em: <https://doi.org/10.31083/KO45889>. Acesso em: 08 out 2025.

REIS, Letícia. Evolução da Inteligência Artificial: o Deep Learning e a praticidade. **ICMC USP**, São Carlos, [s. d.]. Disponível em: <https://mba.iabigdata.icmc.usp.br/evolucao-da-inteligencia-artificial-o-deep-learning-e-a-praticidade/>. Acesso em: 18 out., 2025.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial: Uma Abordagem Moderna**. 4. ed. Rio de Janeiro: GEN LTC, 2022. *E-book*. p.680. ISBN 9788595159495. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9788595159495/>. Acesso em: 12 out. 2025.

Silva, Ivan Nunes da. **Redes neurais artificiais**: para engenharia e ciências aplicadas. São Paulo: Artliber, 2010.

SOUZA, Leonardo Agostinelli Pereira Pinheiro de; MELLO, Mariana Rodrigues Gomes de.

As tecnologias de inteligência artificial do paradigma simbólico e suas implicações interdisciplinares para a Ciência da Informação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 24, p. e026005, 2026. DOI: 10.20396/rdbci.v24i00.8680691. Disponível em: <https://doi.org/10.20396/rdbci.v24i00.8680691>. Acesso em: 16 out. 2025.

STRYKER, Cole; BERGMANN, Dave. O que é um modelo de transformer? **IBM**, [S. l.], 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/transformer-model>. Acesso em: 13 out. 2025.

TAKEUCHI, H.; NONAKA, I. **Gestão do conhecimento**. Porto Alegre: Bookman, 2008.

TRINDADE, Alessandra Stefane Cândido Elias da; OLIVEIRA, Henry Poncio Cruz de. Inteligência Artificial (IA) generativa e Competência em Informação: habilidades informacionais necessárias ao uso de ferramentas de IA generativa em demandas informacionais de natureza acadêmica-científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 29, e47485, 2024. DOI: <http://dx.doi.org/10.1590/1981-5344/47485>. Acesso em: 17 out. 2025.

UNWIN, Antony; KLEINMAN, Kim. The iris data set: In search of the source of virginica. **Significance**, [S. l.], 18 de dez. 2021. Disponível em: <https://academic.oup.com/jrssig/article/18/6/26/7038520>. Acesso em: 11 out. 2025.

VASWANI, Ashish et al. Attention is all you need. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (NIPS 2017), 2017, Long Beach, CA, USA. **Proceedings [...]**. Red Hook, NY: Curran Associates, Inc., 2017. p. 5998-6008. Disponível em: <https://arxiv.org/pdf/1706.03762>. Acesso em: 18 out. 2025.