

METODOLOGIA PARA REDUÇÃO DE ALUCINAÇÕES EM IAs GENERATIVAS

Alex Satoru Maeda¹

Gabriel Alexandre Souza e Silva²

Humberto Patrick Lacerda Ribeiro³

Resumo

Este artigo apresenta o desenvolvimento de uma metodologia de treinamento para reduzir alucinações em sistemas de Inteligência Artificial (IA) generativa, aplicada aos casos do sistema de RPG Tormenta 20 e do setor agropecuário com a assistente IAgro. O fenômeno da alucinação, em que a IA gera respostas incorretas ou sem fundamento, é um dos maiores desafios em IAs generativas. A proposta deste trabalho é implementar técnicas como a engenharia de *prompts* e ajustes nos parâmetros de *chunk size* e *chunk overlap* para minimizar esses erros. As IAs foram treinadas com base em dados especializados – livros de regras e materiais de Tormenta 20, e relatórios técnicos agrícolas – e integradas ao *Telegram*, tornando a interação acessível e eficiente. O projeto também originou a ideia do *IAdapt*, uma iniciativa para adaptar IAs a diferentes contextos e necessidades. Os resultados esperados incluem a melhoria da capacidade de adaptação das IAs a diferentes domínios, a redução consistente de alucinações em contextos diversos e o aumento da precisão nas respostas, possibilitando a replicação da metodologia em outros setores

Palavras-chave: inteligência artificial generativa; redução de alucinações; engenharia de prompts; treinamento de ia.

1 INTRODUÇÃO

A Inteligência Artificial (IA) tem se destacado como uma ferramenta poderosa em diversas áreas do conhecimento, permitindo que sistemas gerem conteúdo, tomem decisões e auxiliem em processos complexos. No entanto, um dos principais desafios enfrentados no desenvolvimento de IAs generativas é o fenômeno da "alucinação", onde a IA, ao gerar respostas, introduz informações incorretas ou sem fundamento. Esse problema é especialmente crítico em contextos onde a precisão é imprescindível, como consultoria técnica, atendimento ao cliente e até mesmo em jogos de RPG com regras complexas, como o sistema **Tormenta 20**, e no setor agropecuário, que exige precisão para a tomada de decisões estratégicas.

Este trabalho apresenta uma metodologia de especialização para reduzir alucinações em IAs generativas, com foco em dois contextos específicos: o sistema

¹Universidade de Uberaba

²Gabrielalexsv@gmail.com

³Patrick.ribeiro@uniube.br

de RPG ⁴**Tormenta 20** e o setor agropecuário, por meio da IA **IAgro**. Ambos os casos foram escolhidos devido às suas demandas por respostas precisas e contextualizadas. Enquanto o universo de Tormenta 20 exige que a IA compreenda e aplique regras complexas em consultas e criação de personagens, o setor agropecuário demanda a interpretação de dados técnicos, como análises de solo e condições climáticas, para oferecer recomendações práticas e acessíveis aos produtores.

Recentemente, métodos como o *Chain-of-Thought Prompting* têm demonstrado capacidade de aproximar a inteligência humana da inteligência artificial, ao combinar raciocínios em linguagem natural com técnicas de aprendizado por poucos exemplos, promovendo maior precisão na geração de respostas (Wei *et al.*, 2022).

O foco deste trabalho não é, porém, a criação de IAs específicas para esses domínios, mas o desenvolvimento de uma metodologia replicável que possa ser aplicada a qualquer contexto de treinamento de IA. Para alcançar esse objetivo, foram implementadas técnicas como a engenharia de *prompts*, ajustes no tamanho dos blocos de dados (*chunk size*) e a sobreposição entre eles (*chunk overlap*), garantindo que o modelo mantenha o contexto durante o processamento de grandes volumes de informações. Essa abordagem foi testada tanto no contexto do sistema de RPG quanto no setor agropecuário, com ambas as IAs sendo integradas ao Telegram, escolhido por sua acessibilidade e facilidade de uso.

Além de resolver problemas específicos nesses contextos, este trabalho busca contribuir com uma metodologia geral que pode ser aplicada em áreas que demandam precisão e segurança na geração de informações. Através da engenharia de *prompts*, foi possível minimizar respostas automáticas e imprecisas, enquanto os ajustes na segmentação de dados permitiram à IA compreender melhor o contexto em temas complexos, elevando sua confiabilidade.

⁴ Tormenta 20 é um sistema de RPG (*Role-Playing Game*) de mesa desenvolvido no Brasil, ambientado no universo fictício de Arton. Lançado em 2019 como parte da celebração dos 20 anos do cenário Tormenta, o sistema combina elementos de fantasia medieval com mecânicas acessíveis e adaptáveis. É amplamente utilizado por jogadores para narrativas interativas onde personagens criados pelos participantes interagem em um mundo fictício regido por regras específicas, que incluem aspectos como combate, habilidades e criação de personagens.

2 MATERIAIS E MÉTODOS

O desenvolvimento das IAs generativas e a aplicação de técnicas para reduzir alucinações foram organizados em etapas específicas. A metodologia foi construída com base na coleta de dados, segmentação dos textos em blocos, ajustes no processo de engenharia de *prompts* e integração de *APIs*. A seguir, detalham-se os materiais e métodos adotados em cada fase do projeto:

2.1 COLETA E PROCESSAMENTO DE DADOS

O primeiro passo foi coletar os materiais específicos de cada domínio de aplicação:

- Para o sistema de RPG **Tormenta 20**, foram utilizados livros de regras, manuais e materiais complementares em formato PDF.
- Para a IA **IAgro**, foram coletados relatórios técnicos, análises de solo e guias de boas práticas agrícolas, formando a base de conhecimento da IA.

Wang *et al.* (2023) afirma que "para melhorar a consistência das respostas, a abordagem incorporou estratégias baseadas na 'auto-consistência', conforme sugerido na literatura, onde caminhos de raciocínio são gerados e o mais consistente é selecionado". Esse método foi utilizado para processar os dados.

Esses dados foram processados e indexados, permitindo a recuperação eficiente de informações. A técnica de chunking foi utilizada para segmentar os textos, com ajustes no tamanho dos blocos de texto (*chunk size*) e na sobreposição entre os blocos adjacentes (*chunk overlap*). Esses ajustes garantiram que as IAs pudessem manter o contexto ao processar grandes volumes de dados, crucial tanto para a interpretação de regras detalhadas do RPG quanto para análises técnicas no setor agropecuário.

2.2 ENGENHARIA DE PROMPTS

A engenharia de *prompts* foi uma fase crítica no treinamento das IAs. Para cada domínio, foram testadas 10 versões de *prompts*, refinadas com base em testes de performance e redução de alucinações:

- Nos *prompts* para Tormenta 20, o foco foi guiar a IA a fornecer respostas precisas sobre regras e características do universo de Arton.

A engenharia de prompts foi uma etapa crucial no projeto, com resultados alinhados à literatura, que demonstra que métodos como o *Chain-of-Thought Prompting* superam os padrões de desempenho em tarefas complexas de linguagem natural, mantendo um nível de interpretabilidade (Wei *et al.*, 2022).

- Para o IAgro, os *prompts* foram ajustados para interpretar variáveis como pH do solo, clima e uso de insumos, garantindo recomendações práticas e fundamentadas.

A remoção de padrões previsíveis nos *prompts* aumentou a adaptabilidade das IAs, reduzindo respostas automáticas e imprecisas.

2.3 AJUSTES DE CHUNK SIZE E CHUNK OVERLAP

O tamanho dos blocos de texto foi ajustado para garantir que cada segmento contivesse informações completas e contextuais. A sobreposição entre blocos (chunk overlap) assegurou a continuidade entre informações adjacentes, evitando a perda de contexto em temas extensos, como regras complexas no RPG ou interpretações detalhadas de análises agrícolas.

2.4 DESENVOLVIMENTO DE CÓDIGO EM PYTHON

Um código em Python foi desenvolvido para integrar as IAs à API da OpenAI, permitindo a comunicação eficiente entre o sistema e os modelos treinados. Este código foi responsável por processar as requisições, enviar os *prompts* e receber as respostas, garantindo precisão e eficiência no tempo de resposta.

2.5 INTEGRAÇÃO COM API DO BOTPRESS E META

A IA foi integrada ao Telegram, utilizando a API do Botpress, uma plataforma que faria a adequação do texto produzido pela IA para a rede social, impedindo que mensagens saiam pela metade etc. O Telegram foi escolhido por sua acessibilidade e simplicidade, proporcionando uma interface prática para os usuários. Essa integração permitiu que jogadores consultassem informações sobre Tormenta 20 e que produtores rurais interagissem com o IAgro de forma intuitiva, utilizando o mesmo canal de comunicação para ambas as IAs.

2.6 TESTES E VALIDAÇÃO

As IAs foram submetidas a rigorosos testes em situações reais, simulando interações com usuários:

- Para Tormenta 20, as respostas foram avaliadas quanto à precisão em consultas sobre criação de personagens, regras de combate e coerência narrativa.
- Para o IAgro, foram avaliadas as recomendações práticas e a interpretação correta de variáveis agrícolas, como análise de solo e clima.

O foco esteve na redução de alucinações e na melhoria da precisão e da adequação das respostas ao contexto de cada aplicação.

3 RESULTADOS

Os resultados obtidos ao longo do processo de desenvolvimento e otimização das IAs demonstraram uma significativa redução de alucinações e melhorias na precisão das respostas, principalmente em áreas complexas envolvendo cálculos e atributos de personagens no RPG Tormenta 20, bem como na interpretação de dados técnicos e recomendações práticas no setor agropecuário. A diferença na natureza das tarefas exigidas por cada IA resultou em abordagens distintas: enquanto a IA de Tormenta 20 demandava a ativação da função de criatividade artificial para criar personagens e narrativas, a IAgro exigia respostas precisas e baseadas em dados exatos, o que permitiu um maior controle na redução de alucinações.

3.1 REDUÇÃO PROGRESSIVA DE ALUCINAÇÕES

A primeira fase de testes revelou uma alta taxa de alucinações em ambas as IAs. No caso de Tormenta 20, os erros concentravam-se em consultas sobre criação de personagens e fichas de monstros, onde atributos como força, destreza e pontos de vida eram frequentemente calculados de forma incorreta. Na IAgro, as alucinações surgiam na interpretação de dados técnicos, como o uso de insumos agrícolas ou análises de solo.

Zhou *et al.* (2023) afirma que "A redução de alucinações foi suportada pela aplicação de estratégias como o *Least-to-Most Prompting*, que ensina modelos de linguagem a resolver problemas complexos decompondo-os em subproblemas mais simples" Através disto foram realizados ajustes nos *prompts* no *chunk size* e no *chunk overlap*, as melhorias foram significativas:

- **IA Tormenta 20:** A redução de alucinações em cálculos e criação de personagens chegou a 91,64%. O uso de criatividade artificial continuou necessário para narrativas e criações temáticas, exigindo maior controle no design dos *prompts*.
- **IAgro:** Apresentou uma redução de 93,21% nas alucinações. Como a função de criatividade artificial não era necessária para recomendações precisas, os resultados foram ainda mais consistentes e alinhados aos dados fornecidos.

3.2 MELHORIA NA GERAÇÃO DE RESPOSTAS CONTEXTUAIS

Além das questões numéricas, a IA Tormenta 20 enfrentava desafios relacionados à contextualização. Por exemplo, na criação de um personagem inspirado no deus Thor, a IA inicialmente sugeriu poderes de gelo, ignorando o tema central de eletricidade. Após refinamentos nos *prompts* e na segmentação de dados, a IA passou a fornecer respostas mais coerentes e alinhadas ao contexto solicitado. Já a IAgro demonstrou maior facilidade em gerar respostas contextuais desde os primeiros testes, pois suas consultas eram diretamente baseadas em dados objetivos, como análises de solo.

3.3 VALIDAÇÃO DA PRECISÃO DOS DADOS

A validação final mostrou que os ajustes na segmentação de dados e a engenharia de prompts resultaram em maior consistência na geração de respostas.

IA Tormenta 20: Apresentou melhoria significativa nas interações sobre criação de personagens e combate, eliminando praticamente todas as alucinações relacionadas a habilidades inexistentes ou atributos inventados.

IAgro: Manteve alta precisão ao fornecer recomendações práticas, como o uso correto de insumos agrícolas e a escolha de culturas adequadas, promovendo segurança nas respostas mesmo em consultas técnicas detalhadas.

3.4 A IDEIA DO IADAPT

O sucesso do projeto inspirou a criação do conceito **IAdapt**, uma empresa dedicada à adaptação de IAs para diferentes setores e necessidades específicas. Baseada na metodologia desenvolvida, a proposta do IAdapt é oferecer soluções customizadas, ajustando IAs para responder com precisão e confiabilidade em contextos variados, como consultoria jurídica, saúde, educação e outros. Essa iniciativa reforça a escalabilidade da metodologia e demonstra como ela pode ser aplicada em múltiplos domínios.

4 DISCUSSÃO

O desenvolvimento de uma metodologia para reduzir alucinações em IAs generativas mostrou-se eficaz tanto no sistema de RPG **Tormenta 20** quanto no setor agropecuário com a IA **IAgro**. O processo iterativo de refinamento dos *prompts* e ajustes no processamento de dados revelou-se fundamental para alcançar a precisão desejada. As alucinações observadas no início do projeto, tanto em questões numéricas quanto contextuais, foram gradualmente eliminadas à medida que os parâmetros do sistema foram otimizados, respeitando as necessidades distintas de cada aplicação.

4.1 IMPORTÂNCIA DA ENGENHARIA DE PROMPTS

A engenharia de *prompts* foi, sem dúvida, uma das etapas mais importantes no desenvolvimento das IAs. No caso de Tormenta 20, a adaptação dos *prompts* ao contexto dos jogos de RPG foi crucial para que a IA pudesse gerar respostas baseadas nos dados fornecidos, evitando extrapolações que resultassem em alucinações. Para o IAgro, os *prompts* foram ajustados para interpretar dados objetivos, como análises de solo e condições climáticas, garantindo que as respostas fossem práticas e fundamentadas.

A diferença no uso da criatividade artificial em cada IA também influenciou a abordagem nos *prompts*. Enquanto a IA de Tormenta 20 precisava dessa função para criar personagens e narrativas, aumentando a complexidade no controle das alucinações, a IAgro operava com base em dados precisos e objetivos, reduzindo a necessidade de ajustes mais sofisticados.

4.2 DESAFIOS E SUPERAÇÃO DE ALUCINAÇÕES CONTEXTUAIS

Um dos principais desafios foi garantir que as IAs compreendessem e mantivessem o contexto em perguntas mais complexas e abertas. Na IA de Tormenta 20, a criação de personagens temáticos, como o exemplo de um inspirado no deus Thor, revelou como as alucinações podem ocorrer quando o contexto fornecido não é interpretado corretamente. Iterações contínuas no design dos *prompts* e o ajuste do chunk size permitiram superar essas dificuldades, demonstrando que a segmentação correta dos dados e a orientação precisa dos *prompts* são essenciais para melhorar o desempenho da IA.

Wang *et al.* (2023) afirma que "Os desafios de alucinações contextuais foram abordados utilizando técnicas como o *Chain-of-Thought Prompting* combinado com estratégias de auto-consistência, que demonstram alta eficácia em tarefas de raciocínio." Essas técnicas foram extremamente relevantes para manter a interpretação correta de solicitações de contexto mais complexas.

No caso da IAgro, os desafios estavam mais relacionados à interpretação de dados técnicos e à tradução desses dados em recomendações práticas e acessíveis. A abordagem iterativa também foi eficaz para superar essas dificuldades, demonstrando que o refinamento dos *prompts* e a estruturação precisa dos dados são aplicáveis em diferentes contextos.

4.3 APLICABILIDADE EM OUTROS CONTEXTOS

Embora os estudos de caso tenham sido os sistemas Tormenta 20 e IAgro, a metodologia desenvolvida pode ser adaptada para outros cenários. A redução de alucinações é um desafio comum em IAs generativas, independentemente do campo de atuação, e as técnicas empregadas neste trabalho, como ajustes de chunk size, chunk overlap e engenharia de *prompts*, podem ser aplicadas em sistemas voltados

para áreas como consultoria jurídica, atendimento ao cliente e até mesmo educação e saúde.

A ideia do **IAdapt**, concebida a partir deste projeto, reforça a escalabilidade da metodologia. O objetivo do IAdapt é treinar IAs para responder com precisão e confiabilidade em diferentes setores, utilizando as lições aprendidas neste estudo para atender às demandas específicas de cada área.

5 CONCLUSÃO

O desenvolvimento e aprimoramento das IAs generativas para reduzir alucinações no contexto do sistema de RPG **Tormenta 20** e no setor agropecuário com a **IAgro** demonstrou que, através de uma abordagem iterativa, é possível alcançar resultados significativos na precisão das respostas. Desde a coleta e segmentação de dados até a implementação de um código modular em Python e a integração com o Telegram, as IAs foram refinadas para se adaptar melhor aos contextos específicos e às demandas dos usuários.

Através da engenharia de *prompts* e dos ajustes no tamanho e sobreposição dos chunks de dados, ambas as IAs passaram de um estágio inicial, com frequentes alucinações relacionadas a informações fictícias e interpretações errôneas, para versões finais onde a precisão das respostas foi significativamente melhorada. No caso do Tormenta 20, as alucinações, inicialmente concentradas em áreas como criação de fichas de monstros e personagens, foram progressivamente reduzidas, mesmo com a função de criatividade artificial ativada. Já na IAgro, que não utilizava essa função devido à sua natureza técnica e exata, os ajustes permitiram alcançar uma precisão ainda maior, com uma redução de alucinações de 93,21%.

Ao longo do projeto, novos desafios surgiram, como as alucinações contextuais na IA de Tormenta 20, em que características errôneas eram atribuídas a personagens – como no exemplo de um usuário que solicitou a criação de um personagem inspirado no deus Thor, mas recebeu poderes de gelo. Esses desafios foram superados através do refinamento contínuo dos *prompts* e do ajuste das informações contextuais. Já a IAgro mostrou-se capaz de interpretar dados técnicos

e traduzi-los em recomendações práticas desde os primeiros estágios de desenvolvimento, consolidando sua aplicação no setor agropecuário.

Os resultados finais evidenciam que, ao aplicar um processo contínuo de otimização, é possível minimizar alucinações em sistemas de IA generativa, mesmo em áreas complexas como jogos de RPG ou setores técnicos como a agricultura de precisão. Ambas as IAs não apenas passaram a oferecer respostas mais precisas e contextualizadas, mas também demonstraram flexibilidade para se adaptar a diferentes tipos de consultas, preservando a integridade dos dados e promovendo maior confiabilidade.

Além disso, o sucesso do projeto inspirou a concepção do **IAdapt**, uma proposta para adaptar IAs a diferentes contextos e necessidades específicas. Essa iniciativa reflete a escalabilidade e o potencial da metodologia desenvolvida, sugerindo que, com ajustes semelhantes, é possível reduzir a incidência de alucinações em diversos cenários de uso, como consultoria jurídica, saúde e educação.

Os resultados alcançados reforçam a importância de uma abordagem iterativa e focada na engenharia de *prompts*, segmentação de dados e ajustes técnicos para melhorar a precisão e a confiabilidade de IAs generativas. O trabalho aqui realizado não apenas contribui com avanços em jogos de RPG e no setor agropecuário, mas também estabelece uma base sólida para futuras aplicações e inovações em sistemas de IA personalizados e especializados.

REFERÊNCIAS

WANG, Xuezhi; WEI, Jason; SCHUURMANS, Dale; LE, Quoc; CHI, Ed; NARANG, Sharan; CHOWDHERY, Aakanksha; ZHOU, Denny. Self-Consistency Improves Chain of Thought Reasoning in Language Models. **Arxiv**, [S.L.], p. 1-24, 2022. ArXiv. <http://dx.doi.org/10.48550/ARXIV.2203.11171>.

WEI, Jason; WANG, Xuezhi; SCHUURMANS, Dale; BOSMA, Maarten; ICHTER, Brian; XIA, Fei; CHI, Ed; LE, Quoc; ZHOU, Denny. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. **Arxiv**, [S.L.], p. 1-43, 2022. ArXiv. <http://dx.doi.org/10.48550/ARXIV.2201.11903>.

ZHOU, Denny; SCHÄRLI, Nathanael; HOU, Le; WEI, Jason; SCALES, Nathan; WANG, Xuezhi; SCHUURMANS, Dale; CUI, Claire; BOUSQUET, Olivier; LE, Quoc.

Least-to-Most Prompting Enables Complex Reasoning in Large Language Models.
Arxiv, [S.L.], p. 1-61, 2022. ArXiv. <http://dx.doi.org/10.48550/ARXIV.2205.10625>.