

Será o ChatGPT um bom divulgador científico em cosmetologia? Um estudo linguístico sobre textos de divulgação científica

Is ChatGPT a good popular science disseminator in cosmetology? A linguistic study on popular science texts

Ana Filipa Pacheco¹, Nuno Guimarães², Ana Torres^{3,4}, Purificação Silvano⁵, Isabel Almeida^{3,4}

¹Faculdade de Letras da Universidade do Porto, INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência

²Faculdade de Ciências da Universidade do Porto, INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência

³UCIBIO – Unidade de Biociências Moleculares Aplicadas, Departamento de Ciências do Medicamento, Faculdade de Farmácia da Universidade do Porto

⁴Laboratório Associado i4HB – Instituto para a Saúde e Bioeconomia, Faculdade de Farmácia da Universidade do Porto

⁵Faculdade de Letras da Universidade do Porto, Centro de Linguística da Universidade do Porto, INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência

Abstract

The science popularisation texts are fundamental for disseminating scientific knowledge in an accessible and understandable way to a non-specialised audience and have a different structure and characteristics from scientific articles (e.g., Garcés-Conejos & Sánchez-Macarro, 1998; Zamboni, 1998). Studies on the linguistic properties of science popularisation texts in European Portuguese are not abundant, the exception being the *Promoção da Literacia Científica* project (Gonçalves & Jorge, 2018). On the other hand, within the realm of producing content, the large language models (LLMs), namely OpenAI's GPT models, have gained widespread public attention in a short period of time. Since they are recent, there is still very little assessment of the linguistic quality of the texts produced. Bearing these premises in mind, this study aims to assess the linguistic and textual quality of the responses generated by ChatGPT (GPT-3.5) in the field of cosmetology, concerning cosmetic products, ingredients, safety and efficacy and regulation categories, to identify patterns that allow an understanding of the differences and/or similarities between the content generated by LLM and that produced by human experts on the *Portal infoCosméticos*. For this, twenty questions previously answered and published on the portal were selected, and subsequently, four different prompts with different degrees of complexity were created, which resulted in eighty answers generated by ChatGPT. The answers were then analysed according to the parameters defined in an evaluation grid consisting of 11 questions. The analysis produced different types of results: overall, the answers written by the experts produced slightly better results than those from ChatGPT; in terms of interphrasal cohesion, it was found that the texts produced by the experts use a reduced number of connectors, in contrast to the recurrent use of connectors in the ChatGPT texts; there is the use of non-explained scientific jargon and a macrostructure with the absence of a conclusion paragraph in the texts published on the portal; the texts generated by ChatGPT have a high frequency of repetitions and/or tautologies.

Keywords: popular science texts, linguistic features, LLMs, ChatGPT.

Resumo

Os textos de divulgação científica são fundamentais para a disseminação do conhecimento científico de forma acessível e compreensível junto do público não especializado, apresentando estrutura e características diferentes das dos artigos científicos (e.g., Garcés-Conejos & Sánchez-Macarro, 1998; Zamboni, 1998). Os estudos sobre as propriedades linguísticas do texto de divulgação científica em português europeu não abundam, sendo a exceção o projeto *Promoção da Literacia Científica* (Gonçalves & Jorge, 2018). Por outro lado, no âmbito da produção de conteúdo, os grandes modelos de linguagem (LLM), nomeadamente os modelos GPT da OpenAI, ganharam, em pouco tempo, atenção generalizada do público. Sendo recentes, a avaliação da qualidade linguística dos textos produzidos é ainda muito reduzida. Tendo estas premissas em consideração, o presente estudo tem como objetivo avaliar a qualidade linguística e textual das respostas geradas pelo ChatGPT (GPT-3.5) no domínio da cosmetologia, no que respeita às categorias de produtos cosméticos, ingredientes, segurança e eficácia e regulamentação, visando identificar padrões que permitam compreender as diferenças e/ou semelhanças entre o conteúdo gerado pelo LLM e aquele produzido por especialistas humanos no *Portal infoCosméticos*. Para isso, foram selecionadas vinte questões previamente respondidas e publicadas no portal e, posteriormente, criados quatro *prompts* distintos com diferentes graus de complexidade, que deram origem a oitenta respostas geradas pelo ChatGPT. As respostas foram, de seguida, analisadas, de acordo com os parâmetros definidos numa grelha de avaliação composta por 11 perguntas. A análise produziu resultados de diferentes tipos: em termos globais, as respostas escritas pelos especialistas produzem resultados ligeiramente superiores às do ChatGPT; quanto à coesão interfrásica, constatou-se que os textos produzidos por especialistas usam um número reduzido de conectores, contrastando com o uso recorrentemente de conectores nos textos do ChatGPT; verifica-se o uso de jargão científico não explicado e uma estrutura composicional global com ausência do parágrafo da conclusão, nos textos publicados no portal; os textos gerados pelo ChatGPT apresentam uma frequência elevada de repetições e/ou tautologias.

Palavras-chave: textos de divulgação científica, características linguísticas, grandes modelos de linguagem, ChatGPT.

1. Introdução

Numa era em que os avanços científicos moldam grandemente a nossa sociedade, a capacidade de comunicar eficazmente conhecimentos complexos a públicos não especializados tornou-se essencial. A divulgação científica estabelece a ponte entre o discurso especializado e a forma como o público compreende e se relaciona com o conhecimento, transformando conceitos técnicos em conteúdo acessível.

Com o advento do acesso rápido e fácil ao conteúdo *online*, e à semelhança do que acontece noutras áreas do conhecimento, observa-se um crescimento significativo na procura de informação relacionada com a área da cosmetologia, por parte dos consumidores (cf. Ferreira et al., 2024). No entanto, esta busca por conhecimento nem sempre é acompanhada por uma avaliação crítica e cuidada das fontes consultadas. Este comportamento pode levar os consumidores a tomarem decisões menos informadas na escolha de produtos cosméticos e, consequentemente, a aumentarem o risco tanto de seleção como de utilização inadequadas de produtos cosméticos para as suas necessidades individuais, potenciando não só efeitos adversos e reações indesejadas, mas também a frustração face a resultados que não correspondem às expectativas.

Este cenário potencial sublinha a importância de uma comunicação clara e rigorosa na divulgação de informações sobre cosmética que garanta que o público não especializado tenha acesso a conteúdos fiáveis e cientificamente fundamentados.

Por outro lado, na corrida à automatização da comunicação, ao acesso à informação e à produção de conteúdo, os grandes modelos de linguagem (*large language models* – LLMs), nomeadamente os modelos GPT da OpenAI, entre os quais o ChatGPT (GPT-3.5) (Brown et al., 2020), ganharam, em pouco tempo, atenção

generalizada do público. Estes modelos, cujo conteúdo de saída se assemelha, aparentemente, aos resultados produzidos por humanos, vêm sendo progressivamente utilizados como um recurso de procura e produção de conteúdo, em vários domínios (e.g., Ray, 2023), designadamente na área da divulgação científica. Não obstante, ao facilitarem o acesso rápido a informações sobre, concretamente, cosmética – o objeto de estudo deste trabalho –, estas ferramentas introduzem novos desafios, como a verificação da precisão científica ou a potencial simplificação excessiva de conceitos complexos.

Sendo recentes, a avaliação da qualidade linguística e textual dos textos produzidos por LLMs é ainda pouco explorada. Contudo, tal análise pode determinar, por exemplo, se os textos gerados cumprem o seu objetivo comunicativo, na mesma medida em que os textos redigidos por humanos. Os dados fornecidos podem, assim, possibilitar contributos relevantes, tanto para a avaliação da adequação a diferentes contextos de comunicação dos textos gerados como para a identificação de padrões de erro, com consequente contributo para a compreensão de potenciais desafios linguísticos apresentados por linguagem gerada por Inteligência Artificial.

Tendo estas premissas em consideração, o presente trabalho pretende investigar o desempenho do ChatGPT (GPT-3.5) quanto à produção de conteúdo de divulgação científica, na área da cosmetologia, de modo a estabelecer padrões que possam identificar as diferenças e/ou semelhanças entre a produção de conteúdo sobre cosmética por humanos especialistas num portal de divulgação científica – o *Portal infoCosméticos* (PiC) – e por um LLM, no que diz respeito à qualidade linguística e textual.

Considerando este enquadramento, estabelecem-se as seguintes perguntas de investigação:

- I. Quais são as principais diferenças e semelhanças linguísticas e textuais entre os textos gerados pelo ChatGPT e os textos produzidos por humanos especialistas no PiC?
- II. Que fragilidades linguísticas e textuais se observam na produção textual do ChatGPT, quando comparada com a produção humana, no contexto da divulgação científica em cosmetologia?

Para responder a estas questões, o presente estudo procede à análise comparativa de um conjunto de textos de divulgação científica sobre cosmética, redigidos por especialistas no PiC e por um modelo de linguagem (ChatGPT, versão 3.5), com base numa grelha construída a partir de bibliografia selecionada.

2. Algumas considerações linguísticas sobre os textos de divulgação científica

Os textos de divulgação científica são fundamentais para a disseminação do conhecimento científico de forma acessível e compreensível junto do público em geral, não especializado, apresentando estrutura e características diferentes das dos artigos científicos (e.g., Garcés-Conejos & Sánchez-Macarro, 1998; Zamboni, 1998). Vários estudos destacam as propriedades linguísticas da comunicação científica dirigida ao público em geral que garantem a eficácia e a acessibilidade dos textos de divulgação científica.

Myers (1991), por exemplo, observa que a coesão lexical é marcada de forma mais explícita em textos de divulgação científica do que em artigos científicos pela necessidade de tornar mais claras as ligações lexicais a um público não especialista. Já Aubakir (2017) destaca a importância do uso de linguagem não especializada e da ausência de jargão científico, enfatizando o princípio da acessibilidade e da visibilidade da informação. Para o autor, a inclusão de exemplos desempenha um papel igualmente fundamental na compreensão e retenção dos conteúdos apresentados.

Por outro lado, o estudo de Molek-Kozakowska (2016) sobre os títulos da revista *New Scientist* revela tendências linguísticas específicas, tais como a redução do número de artigos definidos, bem como uma menor incidência de verbos, que, quando presentes, surgem maioritariamente no presente do indicativo. Formas verbais no passado são escassas, e a conjunção *and* é frequentemente utilizada para ligar ideias. Os advérbios tendem a ser supérfluos, surgindo apenas quando desempenham funções intensificadoras. Outros traços distintivos incluem o uso do pronome *your*, a presença dos verbos modais *may* e *could*, associados à modalidade

epistémica, e o uso do numeral ordinal *first*. Num estudo posterior, Molek-Kozakowska (2017) identifica determinados grupos de características essenciais no jornalismo de divulgação científica, como o uso de adjetivos na forma superlativa e comparativa e de adjetivos intensificadores para expressar a inovação, importância e destaque das descobertas e avanços científicos. A autora sublinha ainda a necessidade de adotar estilos comunicativos menos exigentes e mais envolventes, uma vez que o excesso de jargão e a complexidade excessiva dos textos podem afastar os leitores (Molek-Kozakowska, 2017).

Embora haja alguns estudos sobre as propriedades linguísticas do texto de divulgação científica, como aqueles mencionados acima, em português europeu os estudos não abundam, sendo a exceção o projeto *Promoção da Literacia Científica* (Gonçalves & Jorge, 2018). Nesse contexto, Filipe (2018) analisa algumas marcas do texto de divulgação científica e constata, por exemplo, uma maior predominância de conectores que atestam relações de adição/conjunção, seguidas das relações de temporalidade, contraste e finalidade/intencionalidade. Quanto aos tempos verbais, o autor destaca, igualmente, o uso do presente do indicativo seguido do pretérito perfeito e do futuro, sendo o pretérito imperfeito o que menos vezes ocorre.

Não é objetivo do presente estudo analisar exaustivamente alguns dos aspetos identificados por Molek-Kozakowska (2016, 2017) e por Filipe (2018). No entanto, considerou-se relevante partilhá-los, uma vez que contribuem para a contextualização dos textos de divulgação científica e para uma melhor compreensão das marcas linguísticas que os caracterizam.

Para além das propriedades linguísticas já mencionadas, a organização global do texto também constitui um fator essencial para a eficácia comunicativa da informação veiculada. Uma das dimensões avaliadas nas respostas que constituem o corpus deste trabalho prende-se com a estrutura composicional global, tendo sido verificada a existência (ou não) de uma organização tripartida (introdução, desenvolvimento e conclusão) do conteúdo temático. Esta divisão, embora não formalizada teoricamente por Bronckart (1997), aproxima-se da sua noção de plano geral do texto, que diz respeito à organização global do conteúdo temático e é passível de ser apreendida pelo leitor durante o processo de leitura. Por sua vez, Adam (2005) propõe que os textos se organizem em unidades hierárquicas complexas, que formam estruturas globais relativamente autónomas. Assim, a presença de uma estrutura tripartida pode ser interpretada como um tipo de configuração funcional que decorre da organização dessas unidades superiores e contribui para a progressão temática e coerência global. Em contextos de divulgação científica, esta estrutura revela-se particularmente relevante, na medida em que favorece a orientação do leitor e a construção progressiva do significado.

Contudo, Gonçalves & Magalhães (2019) destacam a heterogeneidade inerente à divulgação científica, apontando para a diversidade e natureza dos formatos textuais, que não se reduzem a somente um género textual estabilizado, mas a um conjunto diversificado de práticas comunicativas.

Apesar de se registarem algumas investigações desenvolvidas neste domínio, estudos em português europeu ainda são limitados, reforçando-se, assim, a necessidade de mais contribuições para a análise das características linguísticas e estruturais dos textos de divulgação científica, contributo que este trabalho pretende dar.

3. Algumas considerações sobre o desempenho dos LLMs na geração de textos de divulgação científica

Estudos sobre o desempenho dos LLMs em diferentes contextos revelam um conjunto de potencialidades, mas também desafios que podem influenciar a geração de textos em domínios como os da divulgação científica. Em Onder et al. (2024), por exemplo, é avaliada a fiabilidade e a legibilidade das respostas do ChatGPT-4 sobre hipotireoidismo na gravidez, e os resultados mostram fiabilidade de moderada a boa e qualidade alta, adequada ao público em geral. No entanto, a legibilidade tende a ser reduzida, sugerindo que os textos gerados exigem um nível universitário para serem compreendidos, o que pode dificultar a sua acessibilidade no domínio da divulgação científica.

Zhang et al. (2025) analisam a capacidade dos LLMs para gerar conteúdo (perguntas, concretamente) a partir de um dado contexto, comparando-o com conteúdo formulado por humanos. Os resultados sugerem que os LLMs tendem a produzir perguntas com um nível de cobertura equilibrado, mas com preferência por questões que exigem respostas descritivas e extensas. Esta característica pode indicar detalhe e extensão nos textos gerados pelos modelos.

Zhou et al. (2024) alertam para uma tendência preocupante: à medida que os modelos aumentam de escala, a sua fiabilidade tende a diminuir. Os erros persistentes, mesmo em tarefas simples, e as respostas incorretas a perguntas complexas constituem desafios que devem ser considerados na utilização de LLMs na procura de informação do domínio da divulgação científica. O estudo salienta, ainda, que a supervisão humana é muitas vezes ineficaz na correção destes erros, uma vez que os utilizadores tendem a confiar demasiado nos modelos e podem não reconhecer os resultados incorretos.

A análise de Moraes et al. (2024) reforça outro aspeto crítico relacionado com a forma como modelos como o ChatGPT lidam com ambiguidades linguísticas em português do Brasil, que vão desde as ambiguidades semânticas e sintáticas às lexicais. Os resultados indicam que, apesar dos avanços, os modelos ainda apresentam limitações na interpretação de textos com ambiguidades, o que pode comprometer a clareza e a precisão das informações geradas.

Embora os estudos pareçam mostrar que os LLMs apresentam potencialidades a considerar para a geração de textos com grau de especificidade mais elevado, como os textos de divulgação científica, o seu desempenho é influenciado por limitações que podem estar relacionadas, por exemplo, com a legibilidade, a fiabilidade das informações geradas e a interpretação de ambiguidades.

4. O estudo

O presente estudo tem como objetivo avaliar a qualidade linguística e textual das respostas geradas pelo ChatGPT (GPT-3.5) no domínio da cosmetologia, no que respeita às categorias de produtos cosméticos, ingredientes, segurança e eficácia e regulamentação. A análise visa identificar padrões que permitam compreender as diferenças e/ou semelhanças entre o conteúdo gerado pelo LLM e aquele produzido por especialistas humanos num portal de divulgação científica – o *Portal infoCosméticos* (PiC).

4.1. Contextualização do estudo

O PiC é um projeto que assenta numa prática pedagógica de aprendizagem colaborativa na criação de conhecimento para divulgação científica no âmbito da promoção e educação para a saúde e envolve três unidades orgânicas da Universidade do Porto: a Faculdade de Farmácia, a Faculdade de Letras e a Faculdade de Belas Artes (Almeida & Silvano, 2021). O portal facilita, igualmente, o acesso a informação validada cientificamente, educando e informando os consumidores e profissionais na área da cosmética através de linguagem acessível e de infografias para facilitar a compreensão.

O processo metodológico para a criação de conteúdo de divulgação científica para o PiC inicia-se com a identificação de uma questão relevante na área da cosmética, sendo a redação da resposta atribuída a um estudante do Mestrado Integrado em Ciências Farmacêuticas da Faculdade de Farmácia da Universidade do Porto. O colaborador deve seguir um guião para a elaboração das respostas, que são acompanhadas de infografias, desenhadas por estudantes da Faculdade de Belas Artes. Os textos e as infografias passam, então, por uma edição linguística antes da sua submissão para validação científica. Após esta etapa, procede-se à reformulação e validação da versão final da resposta por um revisor, seguida de uma revisão linguística final. Posteriormente, o conteúdo é harmonizado pela coordenadora do PiC, sendo finalmente publicado, num período que pode variar entre cinco e oito meses.

Dada a crescente utilização de LLMs na geração automática de conteúdo, este estudo propõe-se a comparar a qualidade linguística e textual das respostas produzidas pelos colaboradores do PiC com as respostas geradas

pelo ChatGPT, um dos modelos mais avançados de processamento de linguagem natural. Para esse efeito, recorreu-se ao ChatGPT-3.5, uma versão otimizada do modelo GPT-3, lançada pela OpenAI em março de 2023 (Brown et al., 2020). Este modelo introduz melhorias na geração de texto, nomeadamente ao nível da coerência e da compreensão contextual, tendo sido desenhado para produzir respostas semelhantes às de um ser humano com base no *input* recebido (Ray, 2023; Ye et al., 2023).

4.2. Processo metodológico

Numa primeira fase, procedeu-se ao desenho de uma grelha (cf. Anexo 1) de análise das propriedades linguísticas e textuais dos textos de divulgação científica, com base em estudos como Myers (1991), Bronckart (1997), Gárce-Conejos & Sánchez-Macarro (1998), Aubakir (2017) e Gonçalves & Jorge (2018). Esta grelha, composta por 11 perguntas, permitiu examinar os seguintes elementos: eficiência comunicativa, estrutura composicional global, clareza, coesão, coerência, (presença/ausência de) jargão científico, inclusão de exemplos, erros de natureza morfossintática, lexical, semântica e/ou ortográfica e repetições/tautologias.

A seleção das perguntas e respostas utilizadas no estudo foi realizada pela equipa das Ciências Farmacêuticas, tendo sido escolhidas vinte questões previamente respondidas e publicadas no PiC pelos colaboradores. As perguntas pertencem a quatro categorias: produtos cosméticos, ingredientes, segurança e eficácia e regulamentação.

Posteriormente, para cada uma das perguntas, foram criados, por um membro da equipa de Ciência de Computadores, quatro *prompts* distintos (cf. Tabela 1) com diferentes graus de complexidade que permitiram a geração das respostas por parte do ChatGPT e garantiram uma avaliação abrangente do desempenho do modelo: (i) o *prompt 1 – zero-shot*: pergunta ou instrução sem exemplos prévios; simula o comportamento do utilizador do LLM; (ii) o *prompt 2 – expert 1*: é feita uma contextualização sobre a tarefa a realizar de forma a orientar o *output* do modelo; (iii) o *prompt 3 – expert 2*: refinamento à contextualização usada para o *prompt 2*, providenciando-se mais contexto; e (iv) o *prompt 4 – few-shot*: pergunta acompanhada de exemplo de resposta desejada de forma a aumentar a precisão e adequação da resposta ao contexto (a resposta selecionada foi retirada do PiC e corresponde a uma pergunta não usada na amostra).

O processo de avaliação das respostas foi conduzido por uma investigadora, com formação em Linguística e conhecimento não especializado, mas informado, em cosmética e cosmetologia, o que permitiu a aplicação crítica e metodológica da grelha de critérios linguísticos.

A análise foi realizada manualmente, com recurso ao visualizador de documentos do sistema operativo macOS – o Preview –, e baseou-se na leitura e análise de cada resposta, seguida do preenchimento da grelha linguística previamente definida. Foram tomadas notas durante o processo de análise, com registo imediato das pontuações atribuídas a cada critério.

A recolha e análise dos conectores presentes nas respostas foram realizadas pela mesma investigadora, tendo sido posteriormente revistas por dois especialistas na área da Linguística, que orientaram e validaram todo o processo. Esta verificação garantiu a consistência da identificação e categorização dos conectores utilizados em cada um dos dois grupos de respostas.

Dado que a análise foi realizada por um único elemento, não foi possível calcular métricas de acordo interanotador. Reconhece-se esta limitação, comum em estudos exploratórios de pequena escala, e sugere-se a replicação do presente trabalho com múltiplos avaliadores, de modo a reforçar a robustez metodológica e a fiabilidade dos resultados.

Tabela 1

Prompts Usados no Processo de Geração de Respostas pelo ChatGPT

Tipo de <i>prompt</i>	Princípio	Exemplo
<i>zero-shot</i> (<i>prompt</i> 1)	Pergunta direta no ChatGPT – simula o comportamento do utilizador	É mais eficaz um protetor solar com proteção UVA ou UVB?
<i>expert 1</i> (<i>prompt</i> 2)	Contextualização sobre a tarefa a realizar	És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: É mais eficaz um protetor solar com proteção UVA ou UVB? A resposta deverá ter entre 250 e 300 palavras.
<i>expert 2</i> (<i>prompt</i> 3)	Contextualização sobre a tarefa a realizar + refinamento	Estás a produzir um texto de divulgação científica na área da cosmética. Tens de explicar, de uma forma simples, mas rigorosa, conceitos aos consumidores dos produtos dessa área e aos profissionais em cosmética. Questão: É mais eficaz um protetor solar com proteção UVA ou UVB? A resposta deverá ter entre 250 e 300 palavras.
<i>few-shot</i> (<i>prompt</i> 4)	Contextualização + disponibilização de exemplo do tipo de resposta esperada	És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Aqui estão uma pergunta e resposta que servem como exemplo do que é pretendido. Pergunta: O que são champôs sem sulfatos? Resposta: [resposta publicada no <i>Portal infoCosméticos</i> : https://portalinfocosmeticos.pt/ingredientes/o-que-sao-champos-sem-sulfatos/] O aparecimento no mercado dos produtos capilares de champôs rotulados com menção ‘champô sem sulfatos’ surgiu na sequência de notícias que colocam em causa a segurança destes ingredientes. Os champôs contêm tensioativos que atuam como agentes lavantes por apresentarem características físico-químicas que lhes conferem capacidade para envolver a gordura e sujidade de forma a serem arrastadas com a água. Existem vários grupos destes ingredientes, sendo que os mais representativos pertencem à classe dos tensioativos aniónicos, entre os quais se incluem os sulfatos como o sulfato de laurilo e sódio (Sodium Lauryl Sulfate) ou sulfato de laurilo e amónio (Ammonium Lauryl Sulfate) e os seus éteres derivados como o lauril éter sulfato de sódio (Sodium Laureth Sulfate) ou lauril éter sulfato de amónia (Ammonium Laureth Sulfate). Por norma, os tensioativos aniónicos são dotados de elevado poder detergente e bom poder espumante, podendo, contudo, remover lípidos essenciais à qualidade da haste capilar. Assim, um champô com certos tipos de sulfatos, poderá deixar o cabelo fragilizado, opaco e emaranhado. Dado o baixo tempo de contacto com estes ingredientes sujeitos a posterior enxaguamento, o couro

	cabeludo deverá ficar limpo, mas a sua qualidade não deverá ser afetada. A partir do ano 2000, promoveu-se o desenvolvimento de formulações com menor teor em sulfatos, compensados ou substituídos por agentes tensioativos menos espumantes, menos desengordurantes e, por conseguinte, mais suaves. Assim sendo, formulações sem sulfatos não contêm os ingredientes acima referidos, apesar de estes terem já sido estabelecidos como seguros nas condições de utilização previstas. Os sulfonatos, sarcosinatos e sulfosucinatos são tensioativos aniónicos alternativos com razoável poder espumante e maior suavidade. Pergunta: É mais eficaz um protetor solar com proteção UVA ou UVB?
--	--

Os *prompts* apresentados na Tabela 1 deram, portanto, origem a oitenta respostas geradas pelo ChatGPT e recolhidas em janeiro de 2024, as quais, juntamente com as vinte respostas às vinte perguntas selecionadas a partir do PiC, constituem o *corpus* deste trabalho. Este *corpus* totaliza 25398 palavras: 6449, referentes ao *corpus* constituído pelos textos de divulgação científica do PiC, e 18949, correspondentes aos textos gerados pelo ChatGPT.

Por último, e após a geração das respostas pelo ChatGPT, procedeu-se à análise com base nos resultados das avaliações linguísticas e textuais e da qualidade científica¹, conduzidas a partir das grelhas desenvolvidas, das respostas do PiC e do ChatGPT.

4.3. Resultados

A análise produziu resultados de diferentes tipos. Em primeiro lugar, permitiu detetar diferenças linguísticas e textuais entre os textos produzidos por especialistas na área da cosmetologia e aqueles gerados pelo LLM. Essas diferenças revelaram-se importantes para a reflexão sobre as características dos textos de divulgação científica, escritos em português europeu, contribuindo para a sua caracterização.

Em termos globais (cf. Tabela 2), os resultados revelam que as respostas escritas pelos especialistas produzem resultados ligeiramente melhores do que as do ChatGPT. Os textos do PiC obtiveram uma média global de 13,6/20 pontos, enquanto aqueles gerados pelo ChatGPT apresentaram uma média ligeiramente inferior de 13,43/20 pontos. Ao nível da categoria de ingredientes, o ChatGPT apresenta valores acima daqueles obtidos pelo PiC (14 vs. 13,2 pontos, respetivamente) e inferiores em relação à categoria de segurança (12,15 vs. 14 pontos, respetivamente).

Ao analisar-se os dados do ponto de vista dos diferentes *prompts* utilizados para gerar respostas do ChatGPT, os resultados demonstram uma variação considerável no desempenho. O *prompt* 3, criado tendo em consideração o refinamento da contextualização sobre a tarefa a realizar, gerou as respostas com a melhor pontuação média (15,15/20 pontos), enquanto o *prompt* 1, que simula o comportamento do utilizador do LLM, apresentou a menor (11/20 pontos), o que sugere que a formulação do pedido influencia significativamente a qualidade da resposta obtida, tal como Brown et al. (2020) referem. Na categoria de segurança, a discrepância é ainda mais evidente, ficando clara a diferença de resultados entre a análise das respostas do PiC (14/20 pontos) e aquelas geradas pelo ChatGPT (9,8/20 pontos).

Quanto à variabilidade dos resultados (desvio padrão, σ), o PiC apresenta menor dispersão (σ global = 1,9) em comparação com o ChatGPT (σ global = 2,5), indicando uma maior consistência na produção dos textos de divulgação científica. Nas respostas do ChatGPT, a variabilidade é particularmente elevada nos resultados

¹ Neste artigo, só são apresentados os resultados da avaliação linguística e textual. A análise da qualidade científica das respostas é apresentada em Torres et al. (2024).

globais do *prompt* 1 ($\sigma = 2,79$) e, especificamente, na categoria de produtos ($\sigma = 4,34$), sugerindo que a qualidade das respostas pode oscilar dependendo da formulação da pergunta e do tópico solicitado. Por outro lado, o *prompt* 4 apresentou o menor desvio padrão ($\sigma = 1,09$), o que pode indicar uma maior uniformidade na geração das respostas com este formato de *prompt*, para o qual é fornecido um exemplo de resposta desejada.

A Tabela 2 apresenta os dados estatísticos detalhados das respostas do PiC e do ChatGPT, por *prompt* e por resultado global. De forma a facilitar a leitura comparativa, os dados da Tabela 2 estão representados graficamente na Figura 1.

Tabela 2

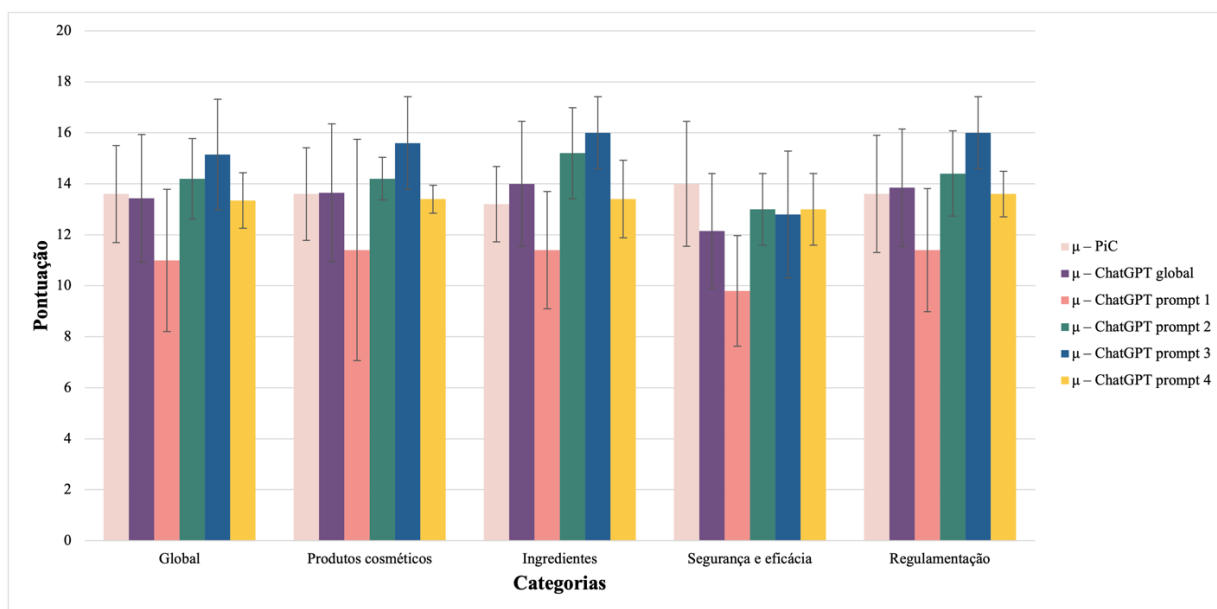
Dados Estatísticos Referentes à Análise Linguística das Respostas do PiC e do ChatGPT

	Global	Produtos cosméticos	Ingredientes	Segurança e eficácia	Regulamentação
μ – PiC	13,6	13,6	13,2	14	13,6
μ – ChatGPT global	13,43	13,65	14	12,15	13,85
μ – ChatGPT <i>prompt</i> 1	11	11,4	11,4	9,8	11,4
μ – ChatGPT <i>prompt</i> 2	14,2	14,2	15,2	13	14,4
μ – ChatGPT <i>prompt</i> 3	15,15	15,6	16	12,8	16
μ – ChatGPT <i>prompt</i> 4	13,35	13,4	13,4	13	13,6
σ – PiC	1,9	1,82	1,48	2,45	2,3
σ – ChatGPT global	2,5	2,7	2,45	2,25	2,3
σ – ChatGPT <i>prompt</i> 1	2,79	4,34	2,3	2,17	2,41
σ – ChatGPT <i>prompt</i> 2	1,58	0,84	1,79	1,41	1,67
σ – ChatGPT <i>prompt</i> 3	2,17	1,82	1,41	2,49	1,41
σ – ChatGPT <i>prompt</i> 4	1,09	0,55	1,52	1,41	0,89

Os resultados parecem demonstrar que, embora o ChatGPT possa produzir respostas cujos resultados da análise aparentam ser comparáveis ou até superiores aos produzidos pelo PiC em algumas categorias, o seu desempenho varia consideravelmente de acordo com o tipo de *prompt* utilizado e a categoria em análise.

Figura 1

Representação Gráfica dos Dados Estatísticos



Paralelamente, foi possível identificar alguns problemas linguísticos e composicionais nos textos produzidos pelos especialistas humanos e pelo LLM, cuja análise é apresentada de seguida. Para este estudo, centrar-nos-emos nas questões em que foram observados mais problemas: a estrutura composicional global, a coesão interfrásica, as repetições e/ou tautologias e o jargão científico.

No que diz respeito à **estrutura composicional global dos textos**, a análise procurou caracterizar a estrutura dos textos produzidos pelos especialistas humanos e pelo LLM, verificando a presença, ou não, de uma estrutura tripartida (cf. Tabela 3, na qual as três partes da estrutura se encontram marcadas a negrito verde [introdução], rosa [desenvolvimento] e laranja [conclusão]). Nos textos do PiC foi possível constatar a ausência frequente do parágrafo de conclusão (cf. Tabela 4), o que pode comprometer a clareza e a estrutura discursiva, dificultando a perceção global da mensagem por parte do leitor não especializado. A solução para este problema é, no entanto, relativamente simples e consiste na inclusão sistemática de um parágrafo final que sintetize os pontos essenciais abordados e forneça um encerramento adequado ao texto. Esta recomendação foi já comunicada à equipa responsável pela produção de conteúdos do portal, garantindo que futuros textos integrem esse elemento fundamental para a coesão e eficiência comunicativa dos textos.

Tabela 3

Estrutura Composicional Global das Respostas do PiC e do ChatGPT

Pergunta: É mais eficaz um protetor solar com proteção UVA ou UVB?	Prompt 1 – zero-shot: É mais eficaz um protetor solar com proteção UVA ou UVB?
PiC	ChatGPT
A pele está permanentemente exposta à radiação solar, que inclui a radiação ultravioleta (UV). A radiação UV subdivide-se em radiação UVA, que corresponde a cerca de 90-95% da radiação UV que atinge a superfície terrestre, a radiação UVB, que constitui aproximadamente 5-10%, e a radiação UVC. Destas, as duas primeiras requerem especial atenção, uma vez que a radiação UVC não atravessa a camada de ozono e não atinge a superfície da Terra. [introdução] A radiação UVA, com comprimentos de onda compreendidos entre os 320-400 nanómetros (nm), tem a capacidade de atingir não só a epiderme (camada mais superficial da pele), mas também a derme (camada mais profunda da pele), sendo também responsável por promover o bronzeamento direto. Este tipo de radiação provoca stress oxidativo, contribuindo, assim, para um envelhecimento prematuro da pele com consequente formação de rugas e flacidez cutânea. É importante salientar que a radiação UVA apresenta outros efeitos nocivos, como o aumento do risco de cancro da pele e indução de reações de fotossensibilidade. [desenvolvimento] A radiação UVB, por sua vez, caracteriza-se por comprimentos de onda entre os 290-320 nm e atinge apenas a epiderme. Esta radiação está associada a queimaduras solares e ao risco de desenvolvimento do cancro de pele, como, por exemplo, o melanoma. Uma das respostas do nosso organismo à exposição de radiação UVB é o bronzeamento devido à promoção da síntese de mais melanina, a qual confere alguma proteção a este tipo de radiação. [desenvolvimento] Tendo em consideração os efeitos nocivos de ambos os tipos de radiação para o nosso organismo, na hora de escolher um protetor solar, é importante optar por um que ofereça proteção simultânea contra as radiações UVA e UVB. Desta forma, o protetor, ao ser aplicado em períodos regulares, evita as queimaduras solares, previne o envelhecimento precoce da pele e protege de outros efeitos prejudiciais. [conclusão]	Ambos os tipos de proteção são importantes para garantir uma proteção eficaz contra os danos causados pelo sol. Os raios UVA são responsáveis pelo envelhecimento da pele e podem penetrar nas camadas mais profundas, causando danos a longo prazo, como rugas, manchas e perda de elasticidade. Eles também podem contribuir para o desenvolvimento de câncer² de pele. [introdução] Os raios UVB são responsáveis pelas queimaduras solares e são mais intensos durante o verão e em altitudes mais elevadas. Eles também são um fator importante no desenvolvimento de câncer de pele. [desenvolvimento] Portanto, é importante escolher um protetor solar que ofereça proteção contra ambos os tipos de raios. Procure por produtos que tenham um amplo espectro de proteção, indicado na embalagem como "FPS" (fator de proteção solar) seguido de um número e "UVA" dentro de um círculo. Isso indica que o produto oferece proteção contra os raios UVA e UVB. [conclusão]

Quanto ao ChatGPT, a presença da estrutura tripartida foi consistente na maior parte dos textos gerados através dos *prompts* 2, 3 e 4. Os problemas mais relevantes registaram-se face a respostas produzidas com recurso ao *prompt* que simula o comportamento do utilizador – o *prompt* 1. Nestes casos, foi recorrente a ausência total de uma estrutura com introdução, desenvolvimento e conclusão, como é exemplo o texto gerado

² Um problema recorrente observado nos textos produzidos pelo LLM é o uso do Português do Brasil, sobretudo ao nível lexical, como evidencia este exemplo.

como resposta à pergunta “*Os cosméticos para bebé têm “perfume”?*” (cf. Tabela 4). Para o mesmo *prompt*, ocorrências de uma estrutura incompleta foram igualmente frequentes, o que evidencia uma aparente fragilidade do modelo na geração de conteúdo especializado perante instruções menos detalhadas e/ou contextualizadas. Este resultado sugere que a qualidade estrutural dos textos gerados pelo ChatGPT depende significativamente da formulação do *prompt*, sendo que instruções vagas ou que simulem um utilizador não especializado tendem a resultar em respostas desorganizadas e menos completas, com baixo valor qualitativo.

Tabela 4

Estrutura Composicional Global das Respostas do PiC e do ChatGPT – Problemas

Pergunta: Os cosméticos para bebé têm “perfume”?	Prompt 1 – zero-shot: Os cosméticos para bebé têm “perfume”?
PiC	ChatGPT
Os cosméticos para bebé são produtos destinados a bebés e crianças com menos de três anos de idade e são especialmente formulados para serem suaves e não irritantes, pelo que os ingredientes utilizados são criteriosamente selecionados. Fragrâncias, designadas frequentemente por “perfume”, são combinações complexas de substâncias naturais e/ou artificiais que são utilizadas numa grande variedade de produtos cosméticos para transmitir um odor agradável e/ou mascarar o odor de outros ingredientes, melhorando a experiência de utilização do produto. Para efeitos de rotulagem dos produtos cosméticos estas substâncias são identificadas na lista de ingredientes pelos termos genéricos “perfume”, “parfum” ou “aroma”, mantendo-se a confidencialidade dos compostos odoríferos utilizados, com a exceção das substâncias enumeradas no Anexo III do Regulamento europeu aplicável aos produtos cosméticos. Neste anexo, estão identificadas 26 fragrâncias que devem ser obrigatoriamente mencionadas na lista de ingredientes quando a sua concentração é superior ao valor limite legislado (0,001 % nos produtos não enxaguados e 0,01 % nos produtos enxaguados) devido ao seu potencial alergénico, mesmo que sejam introduzidas no produto cosmético como parte de um ingrediente complexo, sendo identificadas de acordo com a Nomenclatura Internacional dos Ingredientes Cosméticos (INCI). A pele das crianças, apesar de ser semelhante à dos adultos, em termos estruturais, é potencialmente mais sensível. Desta forma, tal como os produtos destinados a adultos, os cosméticos destinados a bebés podem conter perfume. No entanto, é utilizada menor quantidade de perfume e/ou é evitada a incorporação de fragrâncias com potencial sensibilizante. A utilização de fragrâncias em produtos cosméticos é sujeita a restrições de forma a garantir a proteção da saúde e a segurança do consumidor. Contudo, em caso de preocupação com o potencial sensibilizante das fragrâncias, o consumidor pode optar por produtos	Sim, muitos cosméticos para bebé têm perfume. No entanto, é importante escolher produtos com fragrâncias suaves e hipoalergénicas, especialmente formuladas para a pele sensível dos bebés. É recomendado evitar produtos com fragrâncias fortes, uma vez que podem irritar a pele delicada do bebé. [não estruturado em introdução, desenvolvimento e conclusão]

cosméticos sem perfume. Para os identificar é necessário verificar a lista de ingredientes e garantir que não são referidos os termos “perfume” ou “aroma”, nem são referidas as substâncias odoríferas que se encontram referidas no Anexo III da Legislação Europeia tais como, o “eugenol”, “limonene”, “benzyl alcohol”, “benzyl benzoate”, “linalool”, “citronellol”, “benzyl salicylate” entre outros.

[falta conclusão]

Como mencionado anteriormente, um outro problema observado durante a análise realizada diz respeito à **coesão interfrásica**. A este nível, constatámos que os textos produzidos por especialistas usam um número muito reduzido de conectores, o que pode dificultar o estabelecimento, pelo público, de relações entre as ideias apresentadas e comprometer a clareza e a eficiência da mensagem. Neste trabalho, seguimos a proposta de Lopes & Carapinha (2013) para os conectores “de natureza adverbial, preposicional ou até verbal que exercem funções de articulação textual, sem dar origem a produtos frásicos” (Lopes & Carapinha, 2013, p. 89). Adotamos ainda na nossa análise a proposta das autoras quanto à categorização destes conectores sistematizada na Tabela 5.

Tabela 5

Categorização dos Conectores (Lopes & Carapinha, 2013)

Categorias de conectores		Valor	Exemplos do corpus
Reformuladores	de paráfrase	Clarificação da informação	<i>ou seja</i>
	de retificação	Correção da informação	-----
	de síntese	Síntese da informação	<i>em resumo, em suma, em conclusão</i>
Especificativos		Exemplificação ou particularização da informação	<i>por exemplo, como, tais como, assim como</i>
Reforçativos		Reforço da informação com valor aditivo	<i>além disso, de facto</i>
Estruturadores		Ordenação da informação	<i>por outro lado, em seguida, em primeiro lugar, por fim</i>
Consequenciais		Expressão de consequência ou resultado	<i>assim, desta forma, consequentemente, portanto, assim sendo, como tal, por isso, como resultado</i>
Contrastivos		Expressão de contraste	<i>no entanto, contudo, por sua vez, porém</i>

Os resultados da análise da frequência de conectores, apresentados na Tabela 6, revelam diferenças significativas tanto na distribuição global quanto na densidade de uso dentro de cada conjunto. No total, 19,71% de conectores ocorre nos textos de divulgação científica do PiC, enquanto 80,29% ocorre nos textos gerados pelo ChatGPT. Estes resultados mostram claramente que o ChatGPT inclui uma elevada frequência de conectores em comparação com o PiC.

No entanto, considerando-se a frequência relativa dentro de cada conjunto, observa-se que 1,46% das palavras dos textos do PiC correspondem a conectores, enquanto nos textos do ChatGPT, essa percentagem sobe para 2,02%. Assim, além de gerar mais conectores em termos absolutos, o ChatGPT também apresenta uma maior densidade de conectores no conjunto de textos produzidos pelo modelo de linguagem.

Tabela 6

Frequência dos Conectores

Categorias de conectores		PiC	ChatGPT
Reformuladores	de paráfrase	1	2
	de retificação	-----	-----
	de síntese	2	30
Especificativos		61	168
Reforçativos		5	69
Estruturadores		2	12
Consequenciais		14	44
Contrastivos		15	68

Nos textos do PiC, os conectores mais usados são os especificativos (cf. Tabela 7 e (1)), os contrastivos (cf. Tabela 7 e (2)) e os consequenciais (cf. Tabela 7 e (3)). Destaca-se o facto de os conectores especificativos “como” e “por exemplo” surgirem frequentemente de forma contígua, como mostra o exemplo (4).

- (1) A segurança dos ingredientes dos produtos cosméticos é analisada por organismos especializados, **tais como** a organização Cosmetic Ingredient Review.
- (2) **Contudo**, a legislação europeia não define nenhum valor para as concentrações máximas em que estes ingredientes poderão ser encontrados em produtos cosméticos.
- (3) **Assim**, é importante evitar a exposição a altas temperaturas e o frasco deve ficar devidamente fechado.
- (4) Esta radiação está associada a queimaduras solares e ao risco de desenvolvimento do cancro de pele, **como, por exemplo**, o melanoma.

Já os textos do ChatGPT, por outro lado, fazem uso frequente dos conectores, mas de um modo pouco diversificado dentro do mesmo texto e quando comparados os vários textos. Um caso que ilustra esta observação diz respeito, por exemplo, às repetições de “no entanto” e “além disso” 66 e 69 vezes, respetivamente, ao longo dos textos. No *corpus*, os conectores mais utilizados são os conectores os especificativos, como “por exemplo” e “como” (cf. (5) e (6)), como nos textos do PiC, os reforçativos, como “além disso” (cf. Tabela 7 e (7)), com valor aditivo (cf. Filipe (2018)), os contrastivos, como “no entanto” (cf. Tabela 7), os consequenciais, como “portanto” (cf. Tabela 7), e os reformuladores de síntese, como “em resumo” (cf. Tabela 7).

- (5) **Por exemplo**, em casos em que a segurança de um ingrediente é desconhecida ou quando há uma preocupação legítima com a saúde humana, é permitido realizar testes em animais.
- (6) Estas alternativas incluem conservantes naturais, **como** óleos essenciais e extratos de plantas, bem como conservantes sintéticos, **como** o fenoxietanol.
- (7) **Além disso**, é fundamental realizar o procedimento com um profissional qualificado, que saiba avaliar o estado do cabelo e aplicar os produtos corretamente, evitando danos aos fios.

As respostas à pergunta “*É mais eficaz um protetor solar com proteção UVA ou UVB?*” dos textos de divulgação científica apresentados na Tabela 7 ilustram estas diferenças. No texto do PiC, são utilizados apenas três conectores, “por sua vez”, “desta forma”, “por exemplo”, ou seja, de acordo com Lopes e Carapinha (2013), um conector contrastivo, um consequencial e um especificativo, respetivamente, enquanto no texto gerado pelo ChatGPT são usados cinco: dois contrastivos, um consequencial, um reforçativo e um reformulador de síntese.

Tabela 7

Uso de Conectores nas Respostas do PiC e do ChatGPT

Pergunta: É mais eficaz um protetor solar com proteção UVA ou UVB?	Prompt 2 – expert 1: És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: É mais eficaz um protetor solar com proteção UVA ou UVB? A resposta deverá ter entre 250 e 300 palavras.
PiC	ChatGPT
A pele está permanentemente exposta à radiação solar, que inclui a radiação ultravioleta (UV). A radiação UV subdivide-se em radiação UVA, que corresponde a cerca de 90-95% da radiação UV que atinge a superfície terrestre, a radiação UVB, que constitui aproximadamente 5-10%, e a radiação UVC. Destas, as duas primeiras requerem especial atenção, uma vez que a radiação UVC não atravessa a camada de ozono e não atinge a superfície da Terra. A radiação UVA, com comprimentos de onda compreendidos entre os 320-400 nanómetros (nm), tem a capacidade de atingir não só a epiderme (camada mais superficial da pele), mas também a derme (camada mais profunda da pele), sendo também responsável por promover o bronzeamento direto. Este tipo de radiação provoca stress oxidativo, contribuindo, assim, para um envelhecimento prematuro da pele com consequente formação de rugas e flacidez cutânea. É importante salientar que a radiação UVA apresenta outros efeitos nocivos, como o aumento do risco de cancro da pele e indução de reações de fotossensibilidade. A radiação UVB, por sua vez, caracteriza-se por comprimentos de onda entre os 290-320 nm e atinge apenas a epiderme. Esta radiação está associada a queimaduras solares e ao risco de desenvolvimento do cancro de pele, como, por exemplo, o melanoma. Uma das respostas do nosso organismo à exposição de radiação UVB é o bronzeamento devido à promoção da síntese de mais melanina, a qual confere alguma proteção a este tipo de radiação. Tendo em consideração os efeitos nocivos de ambos os tipos de radiação para o nosso organismo, na hora de escolher um protetor solar, é importante optar por um que ofereça proteção simultânea contra as radiações UVA e UVB. Destá forma, o protetor, ao ser aplicado em períodos regulares, evita as queimaduras solares, previne o envelhecimento precoce da pele e protege de outros efeitos prejudiciais.	Ambos os raios UVA e UVB são prejudiciais para a pele e podem causar danos a longo prazo, como envelhecimento precoce, rugas, manchas e até mesmo câncer de pele. No entanto, é importante entender a diferença entre eles para escolher o protetor solar mais eficaz. Os raios UVA são os principais responsáveis pelo envelhecimento da pele, pois penetram profundamente na derme, a camada mais interna da pele. Eles são menos intensos do que os raios UVB, mas são mais abundantes e estão presentes durante todo o ano, mesmo em dias nublados. Os raios UVA também podem penetrar através do vidro, o que significa que você ainda está exposto a eles mesmo quando está dentro de casa ou no carro. Os raios UVB, por outro lado, são mais intensos e são a principal causa de queimaduras solares. Eles penetram apenas na camada mais externa da pele, a epiderme, mas são mais fortes do que os raios UVA. Os raios UVB são mais comuns durante os meses de verão e são mais fortes durante o meio do dia. Portanto, um protetor solar eficaz deve oferecer proteção contra ambos os tipos de raios. Procure por um protetor solar de amplo espectro, que protege contra os raios UVA e UVB. Além disso, certifique-se de que o protetor solar tenha um fator de proteção solar (FPS) de pelo menos 30. Isso significa que o protetor solar protege a pele contra 97% dos raios UVB. Lembre-se de aplicar o protetor solar generosamente e reaplicá-lo a cada duas horas, especialmente se estiver nadando ou suando. Além disso, use roupas de proteção, como chapéus de aba larga e óculos de sol, e evite a exposição solar direta durante as horas de pico. Em resumo, tanto os raios UVA quanto os UVB são prejudiciais para a pele e um protetor solar eficaz deve oferecer proteção contra ambos. Escolha um protetor solar de amplo espectro com FPS 30 ou superior e siga as recomendações de aplicação e reaplicação para garantir uma proteção adequada contra os danos causados pelo sol.

Outro dos problemas identificados nos textos gerados pelo ChatGPT prendeu-se com a presença de **repetições e/ou tautologias**, praticamente inexistentes nas respostas publicadas no PiC, com exceção de um caso único observado. As repetições e/ou tautologias presentes em metade das respostas do *corpus*

comprometem a clareza e a eficiência comunicativa dos textos, podendo mesmo dificultar a leitura e consequente compreensão da informação. Do mesmo modo, poder-se-á argumentar que as repetições excessivas tendem a reforçar a percepção de falta de rigor do modelo. Na Tabela 8, abaixo, apresentam-se marcados a negrito vermelho alguns exemplos de repetições e/ou tautologias presentes nas respostas geradas pelo ChatGPT.

Tabela 8

Exemplos de Repetições e/ou Tautologias nas Respostas Geradas pelo ChatGPT

Pergunta	Resposta
<i>És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: O que é o alisamento de queratina? A resposta deverá ter entre 250 e 300 palavras.</i>	(...) O procedimento de alisamento de queratina consiste na aplicação de uma solução contendo queratina e outros ingredientes hidratantes nos cabelos, seguida de uma etapa de selagem térmica, onde os fios são alisados com o auxílio de uma prancha de calor. A queratina presente na solução penetra nos fios, preenchendo as áreas danificadas e fortalecendo a estrutura capilar. O calor da prancha ajuda a selar a queratina nos fios, proporcionando um efeito liso e duradouro. O alisamento de queratina é indicado para pessoas que desejam reduzir o volume e o frizz dos cabelos, além de obter um efeito liso. É especialmente recomendado para cabelos danificados, pois a queratina ajuda a reparar os danos e fortalecer os fios. (...)
<i>Estás a produzir um texto de divulgação científica na área da cosmética. Tens de explicar, de uma forma simples, mas rigorosa, conceitos aos consumidores dos produtos dessa área e aos profissionais em cosmética. Questão: Os cosméticos para bebé têm “perfume”? A resposta deverá ter entre 250 e 300 palavras.</i>	(...) Existem diferentes tipos de perfumes, alguns dos quais podem ser irritantes para a pele sensível dos bebés. No entanto, os fabricantes de cosméticos para bebé têm o cuidado de escolher perfumes suaves e seguros para a pele dos bebés. Os perfumes utilizados em cosméticos para bebé são geralmente formulados com ingredientes suaves e não irritantes. Estes perfumes são testados dermatologicamente para garantir que são seguros para a pele sensível dos bebés. Além disso, os fabricantes de cosméticos para bebé seguem regulamentações rigorosas para garantir a segurança dos seus produtos. (...)
<i>És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: O que significa a alegação “não-comedogénico” nos produtos cosméticos? A resposta deverá ter entre 250 e 300 palavras.</i>	A alegação “não-comedogénico” nos produtos cosméticos refere-se à capacidade do produto de não obstruir os poros da pele, evitando assim a formação de comedões, também conhecidos como pontos negros ou cravos. Os comedões são pequenas obstruções dos poros da pele causadas pelo acúmulo de sebo, células mortas e sujidade. Quando os poros ficam obstruídos, as bactérias presentes na pele podem proliferar, levando à inflamação e ao aparecimento de acne. Os produtos cosméticos que são rotulados como “não-comedogénicos” são formulados de forma a minimizar o risco de obstrução dos poros. (...)

Por último, destaca-se a ocorrência frequente de **jargão científico não explicado** nos textos produzidos pelo PiC, fenómeno muito menos frequente nas respostas geradas pelo ChatGPT (cf. (8) e (9)). Verificou-se que os exemplos apresentados na Tabela 9 não são explicados nem no corpo do texto nem através de hipertexto

para outras respostas publicadas no portal ou para conteúdo externo ao portal que defina e/ou contextualize o termo em questão. A presença de jargão científico não explicado pode dificultar grandemente a compreensão da mensagem por parte de um leitor não especializado no domínio da cosmetologia e comprometer a acessibilidade e a visibilidade da informação, princípios fundamentais dos textos de divulgação científica, como asseveram Molek-Kozakowska (2017) e Aubakir (2017). Não obstante, esta questão pode ser facilmente resolvida através da explicação da terminologia no corpo do texto ou da disponibilização de hiperligações (assegurando-se a sua constante atualização). Como anteriormente, esta estratégia foi, do mesmo modo, sugerida à equipa responsável pela produção científica do portal, visando-se melhorar a acessibilidade e a clareza da informação.

Tabela 9

Exemplos de Jargão Científico Não Explicado nas Respostas do PiC Analisadas

Pergunta	Resposta
<i>Os nanomateriais são permitidos em produtos cosméticos?</i>	(...) Entre as vantagens da utilização de nanomateriais em produtos cosméticos salientam-se: a melhoria da cosmeticidade das formulações (textura, facilidade de aplicação, estética, conservação), controlo da libertação, aumento da eficácia dos ingredientes, otimização da fotoproteção e do desempenho dos produtos de maquilhagem. (...)
<i>Como se calcula o Fator de Proteção Solar?</i>	(...) O FPS pode ser calculado através da razão entre a Dose Eritematosa Mínima (DEM) da pele protegida por um protetor solar, e a DEM da pele não protegida. Assim, se o produto cosmético apresenta um valor de FPS igual a 30, significa que é necessária uma exposição solar 30 vezes maior para produzir eritema , comparativamente a uma situação em que este indivíduo não estaria a usar protetor solar. Quanto maior for o FPS, maior será a proteção do produto cosmético, sendo recomendado, pelo menos, um FPS de 30. De uma forma geral, para determinar o FPS, um grupo de voluntários com um fototipo entre I e III é submetido à exposição de doses de radiação UV progressivas, emitidas por uma fonte de luz artificial. (...)
<i>Quais são os constituintes típicos de um verniz de unhas?</i>	(...) De forma a garantir a distribuição uniforme destes constituintes, pode ser adicionado um aditivo reológico , como as bentonites, sílicas, celulosas, acrilatos, uretanos, entre outros. Podem ainda ser utilizados outros ingredientes, como os filtros de radiação ultravioleta (UV), com o objetivo de proteger o produto, evitando o desvanecimento da cor.

(8) Portanto, é essencial seguir as instruções de uso do produto e realizar um teste de **patch** antes de usar qualquer produto cosmético.

(9) Os corantes diretos são moléculas de cor que penetram na cutícula do cabelo e se depositam no **córtex**, conferindo cor ao cabelo.

5. Considerações finais

O presente estudo permitiu investigar o desempenho do ChatGPT quanto à produção de conteúdo dentro dos textos de divulgação científica na área da cosmética, tendo como objetivo a análise e o estabelecimento de padrões linguísticos e textuais que permitissem identificar as diferenças e/ou semelhanças entre a produção de conteúdo científico especializado da área da cosmética por humanos especialistas num portal de divulgação científica – o *Portal infoCosméticos* (PiC) – e por um LLM.

A análise linguística e textual produziu resultados de diferentes tipos e permitiu identificar problemas em ambos os textos analisados, o que levou à reflexão sobre as características dos textos de divulgação científica, escritos em português europeu, contribuindo para a caracterização destes textos.

Assim, por um lado, as respostas geradas pelo ChatGPT, destacaram-se pela estrutura tripartida, garantindo uma estrutura composicional global bem demarcada e sem lacunas estruturais evidentes, nos textos gerados a partir dos *prompts* 2, 3 e 4. Além disso, a ausência de jargão científico contribuiu para uma maior acessibilidade da informação, tornando a mensagem compreensível para um público não especializado. No entanto, registaram-se limitações significativas, nomeadamente quanto ao uso frequente de repetições e/ou tautologias, que comprometem a clareza e a eficiência comunicativa dos textos. Verificou-se, também, ausência de exemplos, o que afetou a resposta ao tópico que a pergunta pretendia responder, não sendo, por isso, tão eficiente do ponto de vista comunicativo. No que diz respeito à coesão interfrásica, observou-se uma elevada frequência de conectores ao longo dos vários textos, apesar da baixa diversidade, sendo mais frequentes os conectores reforçativos, especificativos, contrastivos, consequenciais e reformuladores de síntese. A avaliação global do desempenho linguístico do ChatGPT situou-se nos 13,43 pontos em 20, com uma variabilidade dos resultados expressa por uma dispersão de $\sigma = 2,5$.

As respostas apresentadas no PiC, por outro lado, obtiveram um desempenho médio global ligeiramente superior (13,6/20 pontos) face à pontuação obtida pelas respostas geradas pelo ChatGPT, apresentando uma dispersão de $\sigma = 1,9$. As respostas demonstram ausência quase total de repetições e/ou tautologias e incluem consistentemente exemplos, tanto inseridos no corpo de texto como apresentados sob a forma de hipertexto, o que favoreceu a clareza da informação apresentada. Contudo, a estrutura composicional global revelou algumas fragilidades, nomeadamente no que diz respeito à ausência frequente do parágrafo de conclusão. Simultaneamente, o uso frequente de jargão científico não explicado pode dificultar a compreensão da mensagem por parte de um leitor não especializado no domínio da cosmética. Por fim, ao nível da coesão interfrásica, a baixa frequência de conectores explícitos pode, igualmente, comprometer a compreensão do texto, dado que funcionam como pistas de como o leitor deve interpretar o texto. Os conectores mais frequentes são os especificativos, contrastivos e consequenciais. Estas questões podem ser facilmente resolvidas, tendo sido apresentadas soluções à equipa responsável pela produção de conteúdos do portal.

Com base nestes resultados, futuros trabalhos poderão permitir aprofundar os mecanismos de coesão interfrásica nos textos gerados pelo ChatGPT, analisando detalhadamente a diversidade e adequação do uso de conectores por parte dos LLMs. Além disso, a rápida evolução dos modelos de inteligência artificial torna relevante a necessidade de reavaliações periódicas no que ao desempenho de versões atualizadas do ChatGPT diz respeito. Novas iterações do estudo podem permitir verificar potenciais melhorias na construção textual, na coesão e na eficácia comunicativa dos textos de divulgação científica, por exemplo.

No âmbito do PiC, por sua vez, este estudo contribuirá para a melhoria do guião orientador da redação de textos de divulgação científica, visando a consistência ao nível da estrutura composicional global, mas também a uniformização quanto ao uso de conectores e a sensibilização para a utilização de jargão científico explicado. Concomitantemente, a dinamização de oficinas de escrita destinadas aos colaboradores do PiC na área das Ciências Farmacêuticas poderá contribuir para o aperfeiçoamento das práticas de redação, reforçando a clareza, a coesão, a coerência e a eficiência comunicativa dos textos produzidos.

Referências

- Adam, J.-M. (2005). *La linguistique textuelle*. Armand Colin.
- Almeida, I. M. & Silvano, P. (2022). Portal infoCosméticos: Um projeto colaborativo FFUP/ FLUP de divulgação científica. In *Livro de atas do 7.º Congresso Nacional de Práticas Pedagógicas no Ensino Superior* (pp. 288–294). UA Editora. <https://doi.org/10.48528/yhzq-cp97>
- Aubakir, N. (2017). Distinctive features of popular science discourse. *World Science*, 2(11), 32–35.
- Bronckart, J.-P. (1997). *Activité langagière, textes et discours : pour un interactionisme socio-discursif*. Delachaux et Niestlé.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language models are few-shot learners*. arXiv. <https://arxiv.org/abs/2005.14165>
- Ferreira, M. S., Cuboia, N., Azevedo, L. F. R., Magalhães, O. E., Almeida, I. M. de, & Ribeiro, V. M. (2024). Quality of information about cosmetic products in social media: One-year analysis of Portuguese influential Instagram profiles. *Journal of Consumer Health on the Internet*, 28(3), 203–223. <https://doi.org/10.1080/15398285.2024.2381281>
- Filipe, L. (2018). O artigo de divulgação científica – algumas marcas de género. In M. Gonçalves & N. Jorge (Org.), *Literacia científica na escola* (pp. 79–89). NOVA FCSH-CLUNL.
- Garcés-Conejos, P., & Sánchez-Macarro, A. (1998). Scientific discourse as interaction: Scientific articles vs. popularizations. In A. Sanchez-Macarro & R. Carter (Eds.), *Linguistic choice across genres* (pp. 173–190). John Benjamins.
- Gonçalves, M. & Jorge, N. (Org.) (2018). *Literacia científica na escola*. NOVA FCSH-CLUNL.
- Gonçalves, M. & Magalhães, M. (2019). Corpus e géneros textuais nas práticas de divulgação de ciência ou as novas hierarquias na construção do conhecimento. *Revista da Associação Portuguesa de Linguística*, 5, 145–157. <https://doi.org/10.26334/2183-9077/rapln5ano2019a1>
- Onder, C. E., Koc, G., Gokbulut, P., Taskaldiran, I., & Kuskonmaz, S. M. (2024). Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. *Scientific Reports*, 14(1), 243. <https://doi.org/10.1038/s41598-023-50884-w>
- Myers, G. (1991). Lexical cohesion and specialized knowledge in science and popular science texts. *Discourse Processes*, 14, 1–26.
- Molek-Kozakowska, K. (2016). Stylistic analysis of headlines in science journalism: a case study of New Scientist. *Public Understanding of Science*, 1–14.
- Molek-Kozakowska, K. (2017). Communicating environmental science beyond academia: stylistic patterns of newsworthiness in popular science journalism. *Discourse & Communication*, 11(1), 69–88.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Moraes, L. C., Silvério, I. C., Marques, R. A. S., Anaia, B. C., Paula, D. F., Faria, M. C. S., Cleveston, I., Correia, A. S., & Freitag, R. M. K. (2024). *Análise de ambiguidade linguística em modelos de linguagem de grande escala (LLMs)*. arXiv. <https://arxiv.org/abs/2404.16653>
- Lopes, A. C. M. & Carapinha, C. (2013). *Texto, coesão e coerência*. Almedina.
- Torres, A., Pacheco, A. F., Guimarães, N., Silvano, P., Ferreira, A. & Almeida, I. F. (2024, 10-11 julho). *How reliable is ChatGPT information regarding cosmetic products and ingredients?* [Comunicação oral]. Cosmetinnov 2024, Lisboa, Portugal.
- Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., Cui, Y., Zhou, Z., Gong, C., Shen, Y., Zhou, J., Chen, S., Gui, T., Zhang, Q. & Huang, X. (2023). *A comprehensive capability analysis of GPT-3 and GPT-3.5 series models*. arXiv. <https://arxiv.org/abs/2303.10420>

- Zamboni, L. M. S. (1998). Heterogeneidade e subjetividade no discurso da divulgação científica. *Sínteses – Revista dos Cursos de Pós-Graduação*, 3, 389–395.
- Zhang, Y., Liu, X., Sun, Y., Alharbi, A., Alzahrani, H., Alomair, B. & Song, D. (2025). *Can LLMs design good questions based on context?* arXiv. <https://arxiv.org/abs/2501.03491>
- Zhou, L., Schellaert, W., Martínez-Plumed, F., Moros-Daval, Y., Ferri, C. & Hernández-Orallo, J. (2024). Larger and more instructable language models become less reliable. *Nature*, 634, 61–68. <https://doi.org/10.1038/s41586-024-07930-y>

Anexo

Tabela 1

Grelha de avaliação linguística e textual

Análise linguística			
1. A resposta está de acordo com as características dos textos de divulgação científica?			
Não 0	Parcialmente 1	Quase completamente 2	Sim 3
2. A resposta é eficiente do ponto de vista comunicativo, respondendo ao tópico?			
Não 0	Parcialmente 1	Quase completamente 2	Sim 3
3. A resposta está bem estruturada?			
Não estruturada [a resposta não inclui introdução, desenvolvimento e conclusão] 0	Semiestruturada [a resposta inclui introdução, desenvolvimento ou conclusão] 1	Quase bem estruturada [a resposta inclui introdução, desenvolvimento e conclusão, com lacunas] 2	Bem estruturada [a resposta inclui introdução, desenvolvimento e conclusão] 3
4. A resposta é clara?			
Não clara [a resposta não é concisa, direta ou simples] 0	Algo clara [a resposta é clara em certa medida] 1	Clara [a resposta é concisa, direta e simples] 2	
5. A resposta é coesa?			
Não coesa [a resposta inclui poucos mecanismos de coesão referencial, temporal e interfrásica] 0	Algo coesa [a resposta inclui alguns mecanismos de coesão referencial, temporal e interfrásica, com lacunas] 1	Coesa [a resposta inclui os mecanismos de coesão referencial, temporal e interfrásica adequados] 2	
6. A resposta é coerente?			
Não coerente [as ideias não estão encadeadas logicamente] 0	Algo coerente [as ideias estão encadeadas logicamente, com lacunas] 1	Coerente [as ideias estão encadeadas logicamente] 2	
7. Quando presente, o jargão científico é explicado?			
Não 0		Sim 1	
8. A resposta inclui exemplos?			
Não 0		Sim 1	
9. A resposta apresenta erros ortográficos?			
Sim 0		Não 1	
10. A resposta apresenta erros linguísticos (morfo-sintáticos, lexicais ou semânticos)?			
Sim 0		Não 1	
11. A resposta apresenta repetições/tautologias?			
Sim 0		Não 1	
Pontuação global: -/20			