

Caracterização da dinâmica informacional sobre semaglutida no YouTube: análise de conteúdo, engajamento e sinais sociais

Samira R. Freitas¹, Michele A. Brandão², Leonardo R. Oliveira³,
Helen C. S. C. Lima¹

¹ Departamento de Computação e Sistemas
Universidade Federal de Ouro Preto (UFOP)
samira.resende@aluno.ufop.edu.br
helen@ufop.edu.br

² Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
michele.brandao@dcc.ufmg.br

³ Instituto de Ciências Biológicas
Universidade Federal de Minas Gerais (UFMG)
rossideoliv@gmail.com

Abstract. *The use of digital platforms as a source of health information has expanded access to knowledge, but also exposure to inaccurate content. This study characterizes the informational ecosystem surrounding semaglutide on YouTube, integrating content analysis, engagement metrics, and topic modeling of comments. Data were collected through the platform's official API, covering the period from 2019 to 2025. The results indicate a predominance of content classified as "Information with Limitations" (49%), while misinformation, although less prevalent (12%), shows high engagement, with over 4 million views and 200,000 comments. Topic modeling reveals that public discussion is largely driven by personal experiences, with limited presence of content related to clinical aspects and risks.*

Resumo. *O uso de plataformas digitais como fonte de informação em saúde tem ampliado o acesso ao conhecimento, mas também a exposição a conteúdos imprecisos. Este estudo caracteriza o ecossistema informacional sobre a semaglutida no YouTube, integrando análise de conteúdo, métricas de engajamento e modelagem de tópicos dos comentários. A base foi coletada por meio da API oficial da plataforma, considerando o período de 2019 a 2025. Os resultados indicam predominância de conteúdos classificados como Informação com Ressalvas (49%), enquanto conteúdos desinformativos, embora minoritários (12%), apresentam alto engajamento, com mais de 4 milhões de visualizações e 200 mil comentários. A modelagem de tópicos revela que a discussão pública é fortemente orientada por relatos de experiências pessoais, com menor presença de conteúdos relacionados a aspectos clínicos e riscos.*

1. Introdução

O ambiente informacional contemporâneo, impulsionado pela expansão da internet e das plataformas digitais, ampliou o acesso à informação, mas também a exposição a conteúdos imprecisos, especialmente em saúde pública [International Telecommunication Union 2023]. Nesse cenário, a Organização Mundial da Saúde define *infodemia* como o excesso de informações que dificulta a identificação de fontes confiáveis e pode induzir comportamentos de risco [World Health Organization 2020]. A discussão sobre medicamentos para diabetes tipo 2 e obesidade insere-se nesse contexto. A semaglutida apresenta eficácia clínica comprovada [Souza et al. 2025], mas sua popularização em plataformas digitais ampliou o uso para além das indicações aprovadas, incluindo práticas *off-label*¹ e a circulação de informações nem sempre alinhadas às evidências científicas.

Diante disso, este trabalho tem como objetivo caracterizar o ecossistema informacional sobre a semaglutida no YouTube, integrando análise de conteúdo, métricas de engajamento e modelagem de tópicos. A base foi coletada por meio da API da plataforma² (2019–2025), totalizando 206 vídeos, 90 canais, 28.558 comentários e 20.121 usuários. Como contribuições, o estudo: (i) analisa interações em larga escala; (ii) investiga a relação entre padrões informacionais e engajamento; (iii) caracteriza o ecossistema informacional da plataforma.

2. Trabalhos Relacionados

O aumento na discussão em torno de medicamentos à base de semaglutida tem atraído atenção de pesquisadores. *Fantaus* [Fantaus 2023] e *Carvalho et al.* [Carvalho et al. 2025] indicam limitações na qualidade da informação sobre medicamentos em plataformas digitais, destacando a baixa menção a efeitos adversos e a presença de conteúdos que incentivam o uso de forma implícita ou direta. Da mesma forma, *Oliveira et al.* [Oliveira et al. 2026], *Silva et al.* [Silva et al. 2026] e *Pombeiro et al.* [Pombeiro et al. 2025] apontam a influência das mídias sociais na adoção de práticas relacionadas ao uso de medicamentos, evidenciando o crescimento do uso *off-label* associado à pressão estética.

No contexto de saúde pública, *Dias et al.* [Dias et al. 2024] analisam conteúdos e interações no YouTube para caracterizar a popularidade e o engajamento em torno do uso de cigarros eletrônicos, enquanto *Teixeira e Reis* [Teixeira and Reis 2023] investigam padrões de discurso de ódio em larga escala a partir de comentários coletados via API da plataforma, no contexto da CPI da COVID-19. Abordagens recentes também exploram o uso de modelos de linguagem para análise em larga escala: *Cavalini et al.* [Cavalini et al. 2025] analisam desinformação sobre vacinas, enquanto *Henz et al.* [Henz et al. 2025] avaliam o uso desses modelos para análise de sentimentos em textos de redes sociais no contexto de saúde mental, evidenciando seu potencial para compreensão de comportamentos e dinâmicas sociais.

Apesar dessas contribuições, observa-se que a maioria dos estudos sobre semaglutida baseia-se em revisões bibliográficas ou em análises com amostras reduzidas e coleta

¹CFM orienta uso *off-label* em prescrição, disponível em: <https://portal.cfm.org.br/noticias/cfm-orienta-uso-off-label-em-prescricao>

²YouTube Data API, disponível em: <https://developers.google.com/youtube/v3>

manual, o que limita a compreensão do fenômeno em escala. Nesse sentido, permanece uma lacuna quanto ao uso de abordagens baseadas em coleta sistemática e automatizada de dados em plataformas digitais, especialmente no contexto brasileiro. Para avançar nesse cenário, este trabalho utiliza a API oficial do YouTube para coleta automatizada de dados, permitindo a captura sistemática e reproduzível de conteúdos relevantes.

3. Metodologia

A construção do corpus e das análises seguiu um *pipeline* estruturado e reproduzível, composto por quatro etapas: (i) coleta de dados; (ii) filtragem e construção da base de dados; (iii) rotulação dos vídeos; (iv) caracterização do corpus por métricas de engajamento e modelagem de tópicos.

3.1. Coleta de dados

Os dados foram coletados por meio da YouTube Data API [Google 2024c], considerando o período entre fevereiro de 2019 e janeiro de 2025, sendo as análises realizadas de forma agregada e anonimizada, sem identificação individual dos usuários. Para cada expressão de busca, foram recuperados os resultados retornados pela API dentro do limite permitido, utilizando paginação para garantir a coleta dos vídeos disponíveis no escopo definido. A base inicial foi composta por 2.677 vídeos, 1.315 canais, 314.586 comentários e 177.634 autores distintos. A estratégia de coleta foi baseada em consultas estruturadas a partir dos principais nomes comerciais da semaglutida (*Ozempic*, *Wegovy* e *Rybelsus*), associadas a diferentes contextos informacionais, incluindo definição, usos terapêuticos, benefícios, efeitos adversos e comparações entre medicamentos. Foram utilizadas expressões como “O que é Ozempic?” e “Ozempic efeitos colaterais”, aplicadas de forma equivalente aos diferentes nomes comerciais do medicamento, buscando reduzir possíveis vieses na coleta.

3.2. Filtragens e construção do corpus

Foram aplicadas etapas de filtragem para garantir consistência temática, foco no público brasileiro e qualidade textual. Inicialmente, foram mantidos apenas vídeos e comentários em português brasileiro, identificados via Google Cloud Natural Language [Google 2024b]. Em seguida, foram removidos vídeos do tipo *shorts* (duração inferior a 60 segundos) e conteúdos sem transcrição, sendo estas obtidas por meio da *YouTube-TranscriptApi* [Depoix 2023]. Utilizou-se o modelo Gemini 2.5 Pro [Google 2024a] para geração de resumos sintéticos, auxiliando a etapa de rotulação. Por fim, foram considerados apenas vídeos de canais classificados como Prata (100 mil inscritos) e Ouro (1 milhão de inscritos) [YouTube 2024]. A amostra final foi composta por 206 vídeos, 90 canais, 28.558 comentários e 20.121 usuários distintos, conforme a Tabela 1.

Tabela 1. Sumário dos dados analisados

Descrição	Valor
Número de Canais	90
Número de Vídeos	206
Número de Comentários	28.558
Número de Usuários	20.121

3.3. Rotulação do corpus

A rotulação foi realizada por dois avaliadores independentes, incluindo um especialista da área da saúde, com base em um instrumento adaptado do Brief DISCERN [Khazaaal et al. 2009]. A análise foi conduzida principalmente sobre as transcrições, com apoio de elementos audiovisuais quando necessário e nos casos de discordância, prevaleceu a avaliação do especialista. O processo de classificação foi estruturado a partir de uma lógica binária baseada em dois eixos independentes: (i) presença de critérios associados à **desinformação** (como omissão de riscos, linguagem sensacionalista, recomendações sem orientação médica, uso exclusivo de depoimentos e promoção comercial); (ii) presença de critérios associados à **não desinformação** (como uso de fontes confiáveis, recomendação profissional, explicação técnica, menção a riscos e reconhecimento de variabilidade individual). Cada vídeo foi avaliado quanto à ocorrência desses critérios, permitindo definir a classificação final a partir da combinação dos dois eixos, resultando em quatro categorias mutuamente exclusivas: *Desinformação* (presença apenas de critérios de desinformação), *Não Desinformação* (presença apenas de critérios de não desinformação), *Informação com Ressalvas* (presença simultânea de ambos) e *Inconclusivo* (ausência de critérios). A confiabilidade da rotulação foi avaliada por meio do coeficiente Kappa de Cohen [Landis and Koch 1977]. Para ilustrar o processo de rotulação, um exemplo de vídeo classificado como *Desinformação* apresenta o título “Estudo mostra que Ozempic, além da diabetes, também ajuda no alcoolismo e consumo de drogas”. Nesse caso, foram identificados critérios como linguagem sensacionalista, recomendações sem orientação médica e uso de depoimento pessoal como evidência, além da apresentação superficial de evidências científicas. Esse conjunto de elementos caracteriza a predominância de critérios de desinformação, justificando sua classificação.

3.4. Caracterização do corpus

A caracterização do corpus foi realizada com o objetivo de compreender o contexto informacional sobre a semaglutida no YouTube, considerando a dinâmica de produção de conteúdo, o comportamento dos usuários e os padrões discursivos associados ao tema. Para isso, foram analisadas métricas agregadas ao longo do tempo, a distribuição das categorias de classificação dos vídeos e indicadores de engajamento. Adicionalmente, foi aplicada modelagem de tópicos com BERTopic [Grootendorst 2022], utilizando embeddings *neuralmind/bert-large-portuguese-cased*, redução de dimensionalidade com UMAP ($n_neighbors=3$, $n_components=10$, métrica cosseno) e clusterização HDBSCAN ($min_cluster_size=10$, $min_samples=10$, métrica euclidiana) para identificação de padrões semânticos nos comentários. Os parâmetros foram definidos de forma a priorizar a formação de agrupamentos coesos e a redução de ruídos.

4. Resultados

Nesta seção, apresentamos nossos resultados, incluindo a caracterização do conteúdo, engajamento e a modelagem de tópicos nos comentários.

4.1. Análise Temporal

A dinâmica temporal das métricas foi analisada por meio de agregações mensais. As métricas de produção de conteúdo, como o número de vídeos por mês (Figura 1(a)) — que representa a quantidade de conteúdos publicados — e de novos canais (Figura 1(b))

— que indica a entrada de novas fontes produtoras — evidenciam um comportamento irregular ao longo do tempo, com picos concentrados principalmente em 2023. Esse padrão indica que a produção não segue uma tendência de crescimento contínuo, mas ocorre de forma episódica, alternando períodos de maior e menor atividade.

As métricas de engajamento, como comentários (Figura 1(c)) — que refletem o volume de interações — e autores únicos (Figura 1(f)) — que indicam a participação de usuários distintos — apresentam crescimento mais acentuado a partir de 2023, com picos expressivos em 2024. Esse comportamento sugere que o engajamento se intensifica independentemente da produção de novos conteúdos, indicando maior interesse e participação do público. Por sua vez, as visualizações e curtidas (Figuras 1(d) e (e)), que representam o alcance e a reação dos usuários, apresentam elevada volatilidade, com picos isolados ao longo do período. Esse padrão sugere que a atenção dos usuários tende a se concentrar em conteúdos específicos, em vez de ocorrer de forma distribuída ao longo do tempo. Tais picos podem estar associados a eventos externos, como medidas regulatórias e alertas sanitários relacionados ao uso da semaglutida³. De forma geral, observa-se intensificação do engajamento a partir de 2023, mesmo sem crescimento contínuo da produção de conteúdo, evidenciando um descompasso entre a geração de conteúdo e a resposta do público, com aumento da participação e do interesse no debate ao longo do tempo.

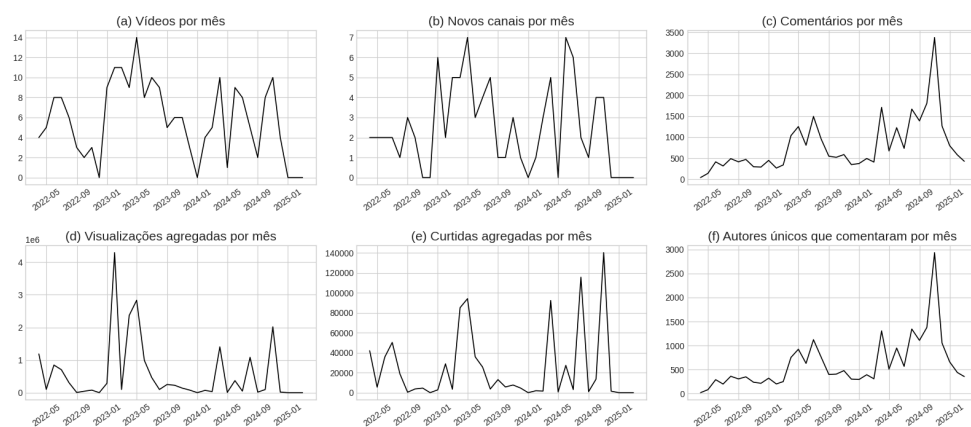


Figura 1. Evolução temporal das métricas de popularidade e engajamento ao longo do período analisado.

4.2. Distribuição das classificações e concordância entre avaliadores

A análise das classificações evidencia predominância da categoria *Informação com Ressalvas* (49,5%), seguida por *Inconclusivo* (22,3%), *Não Desinformação* (16,0%) e *Desinformação* (12,1%). Embora a desinformação explícita seja minoritária, observa-se predominância de conteúdos relacionados à categoria informações com ressalvas, nos quais informações corretas coexistem com omissões ou simplificações. Esse padrão indica que o risco informacional não se restringe a conteúdos falsos, mas à circulação de

³Anvisa determina retenção de receita, disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2025/canetas-emagrecedoras-so-poderao-ser-vendidas-com-retencao-de-receita;>
Apreensão de lotes de medicamentos falsificados, disponível em: [https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2025/anvisa-determina-apreensao-de-lotes-de-medicamentos-falsificados.](https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2025/anvisa-determina-apreensao-de-lotes-de-medicamentos-falsificados)

materiais com credibilidade aparente, porém incompletos. A análise de concordância entre avaliadores resultou em 73,3% de concordância simples e coeficiente de Cohen $\kappa = 0,582$, indicando concordância moderada. Esse resultado é consistente com a natureza interpretativa da tarefa de classificação.

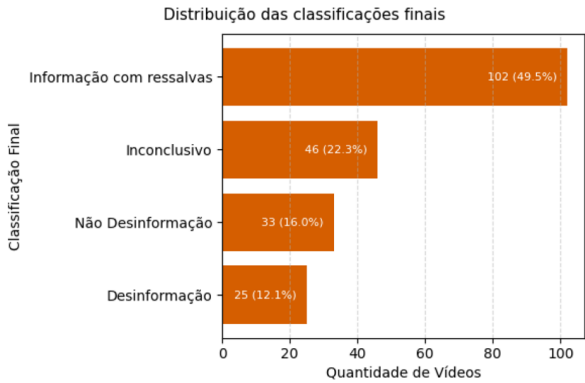


Figura 2. Distribuição das classificações finais dos vídeos analisados.

4.3. Alcance e engajamento

A Tabela 2 apresenta o resumo das métricas absolutas agregadas por categoria de classificação, permitindo comparar o alcance e o nível de interação dos diferentes tipos de conteúdo.

Tabela 2. Métricas agregadas por categoria de classificação

Categoria	Vídeos	Visualizações	Likes	Comentários	Autores
Não Desinformação	34	7.604.643	210.001	11.676	4.438
Informação com ressalvas	102	6.159.489	275.201	14.555	8.788
Desinformação	25	4.775.151	239.845	12.613	6.475
Inconclusivo	46	2.018.940	145.452	6.656	3.384

Os resultados evidenciam que conteúdos classificados como *Informação com Ressalvas* concentram os maiores volumes de interação, especialmente em comentários e participação de autores, indicando papel central na mobilização do debate público. Destaca-se, contudo, um aspecto crítico: conteúdos classificados como *Desinformação*, apesar de representarem a menor quantidade de vídeos na amostra, configuram-se como a segunda categoria com maior volume de interações — incluindo curtidas, comentários e número de autores únicos. Esse comportamento revela não apenas um elevado potencial de alcance, mas, sobretudo, um engajamento expressivo dos usuários com conteúdos potencialmente enganosos. Esse achado reforça uma preocupação relevante no contexto informacional contemporâneo, na medida em que sugere que conteúdos desinformativos não apenas circulam, mas também mobilizam ativamente a audiência, ampliando sua visibilidade e influência no debate público. Tal dinâmica pode contribuir para a disseminação e legitimação de narrativas imprecisas, com possíveis impactos na percepção e tomada de decisão dos usuários. Em conjunto, os resultados apontam que conteúdos pertencentes

às categorias *Informação com Ressalvas* e *Desinformação* desempenham papel relevante na dinâmica informacional, contribuindo para a amplificação de narrativas no ambiente digital.

4.4. Modelagem de tópicos

A modelagem de tópicos identificou 16 agrupamentos temáticos nos comentários, além de um agrupamento residual de ruído (Tópico -1), permitindo analisar a dimensão semântica das interações dos usuários sobre a semaglutida conforme a Tabela 3. Observa-se forte concentração no **Tópico 0**, responsável por 82,07% dos comentários (22.922), associado a relatos de experiência pessoal com foco em perda de peso e resultados percebidos. Termos como “*perdi*”, “*peso*” e “*emagrecer*” indicam que a discussão é predominantemente orientada por narrativas individuais. O **Tópico 1**, segundo mais representativo (6,76%), reúne elogios e agradecimentos ao conteúdo, refletindo um padrão de engajamento social positivo. Já tópicos como **2**, **12** e **15** complementam a dimensão experiencial, abordando descontinuação do uso, efeito rebote e relatos iniciais. Em contraste, tópicos relacionados a aspectos clínicos, como comorbidades (**3**) e efeitos adversos (**4**), apresentam baixa representatividade, sugerindo menor centralidade de informações técnicas. Outros tópicos refletem dimensões complementares, como custo (**9**), alternativas (**8**) e práticas alimentares (**11**), enquanto tópicos de dúvidas (**13**) e comparações (**14**) indicam a presença de incerteza informacional. O **Tópico -1** representa 2,02% dos dados e corresponde a comentários não agrupados pelo modelo. Em conjunto, os resultados evidenciam predominância de conteúdo baseado em experiências pessoais, com menor presença de informações clínicas estruturadas.

Tabela 3. Principais tópicos identificados nos comentários a partir do BERTopic

ID	Breve descrição e principais termos
0	Relatos de experiência pessoal com uso do Ozempic, associados à perda de peso e resultados percebidos (22.922 comentários; 82,07%). <i>Termos:</i> ozempic, peso, perdi, semana, remédio, emagrecer
1	Agradecimentos e elogios ao conteúdo ou ao criador, refletindo engajamento social positivo (1.888 comentários; 6,76%). <i>Termos:</i> parabéns, vídeo, explicação, excelente
2	Discussões sobre descontinuação do uso, efeito rebote e dúvidas sobre continuidade do tratamento (983 comentários; 3,52%). <i>Termos:</i> desmame, rebote, vende, tempo
3	Comentários sobre condições clínicas individuais e comorbidades associadas (314 comentários; 1,12%). <i>Termos:</i> pressão, alta, menopausa, hipotireoidismo
4	Relatos de efeitos adversos e reações colaterais relacionadas ao uso do medicamento (168 comentários; 0,60%). <i>Termos:</i> colateral, sentiu, efeito, colaterais
5	Discussões sobre o uso do Ozempic no contexto do diabetes e controle glicêmico (161 comentários; 0,58%). <i>Termos:</i> diabetes, ozempic, clicks, semana
6	Questionamentos sobre aparência física, flacidez e prática de atividades físicas (150 comentários; 0,54%). <i>Termos:</i> rosto, musculação, atividades, pele
7	Mensagens religiosas e expressões de fé e apoio (149 comentários; 0,53%). <i>Termos:</i> deus, abençoe, continue, família
8	Indicações de produtos naturais e soluções alternativas ao medicamento (140 comentários; 0,50%). <i>Termos:</i> naturais, produtos, casa, vende
9	Discussões sobre custo, preço e acessibilidade do medicamento (125 comentários; 0,45%). <i>Termos:</i> reais, mil, 1000, custa, 200, 00
10	Comentários conversacionais ou genéricos com baixa densidade informacional (97 comentários; 0,35%). <i>Termos:</i> senhor, mulher, barriga, quer
11	Discussões sobre dietas e estratégias alimentares como alternativa ao medicamento (68 comentários; 0,24%). <i>Termos:</i> jejum, intermitente, carboidratos
12	Relatos iniciais e compartilhamento de experiências nas primeiras fases do uso (61 comentários; 0,22%). <i>Termos:</i> experiência, primeiros, compartilhar
13	Perguntas gerais, dúvidas e solicitações de orientação (60 comentários; 0,21%). <i>Termos:</i> gostaria, dúvida, saber
14	Comparações entre o Ozempic e outros medicamentos para emagrecimento (54 comentários; 0,19%). <i>Termos:</i> funcionou, mim, contrave, comigo
15	Discussões sobre ganho de peso, retomada do uso e preocupações após interrupção (25 comentários; 0,09%). <i>Termos:</i> engordou, tomou, novo, emagrecer
-1	Comentários não agrupados pelo modelo, caracterizados como ruído semântico (565 comentários; 2,02%).

4.5. Distribuição das categorias por tópico

A Figura 3 apresenta a distribuição das categorias informacionais nos principais tópicos identificados (Tópicos 0 a 3), selecionados por sua maior representatividade em volume de comentários. Observa-se predominância da categoria *Informação com Ressalvas* em todos os tópicos (38%–42%), indicando um contexto informacional majoritariamente ambíguo. O tópico de **relatos de experiência pessoal** (Tópico 0), responsável por 82,07% dos comentários, concentra 39% de *Informação com Ressalvas* e 28% de *Desinformação*, sugerindo que experiências individuais são frequentemente compartilhadas e validadas a partir de conteúdos com omissões de risco ou baseados em depoimentos (Seção 3.3).

De forma semelhante, o tópico de **elogios e validação social** (Tópico 1) também se associa a conteúdos de menor qualidade informacional (38% de *Informação com Ressalvas* e 24% de *Desinformação*), indicando aceitação do conteúdo mesmo sem evidência clínica direta. Já o tópico de **descontinuação do uso e efeito rebote** (Tópico 2) apresenta 41% de *Informação com Ressalvas* e 25% de *Desinformação*, refletindo incertezas e interpretações variadas sobre os efeitos da interrupção do tratamento. Por sua vez, o tópico de **condições clínicas e comorbidades** (Tópico 3) destaca-se pela elevada proporção de conteúdos *Inconclusivos* (41%), além de 42% de *Informação com Ressalvas*, indicando discussões mais difusas e com menor clareza informacional.

Os tópicos foram selecionados por sua representatividade e por refletirem padrões distintos de engajamento. Em conjunto, os resultados evidenciam que experiências pessoais, validação social e incertezas sobre uso e condições clínicas estão associadas à circulação de informações potencialmente ambíguas, enganosas e inconclusivas.

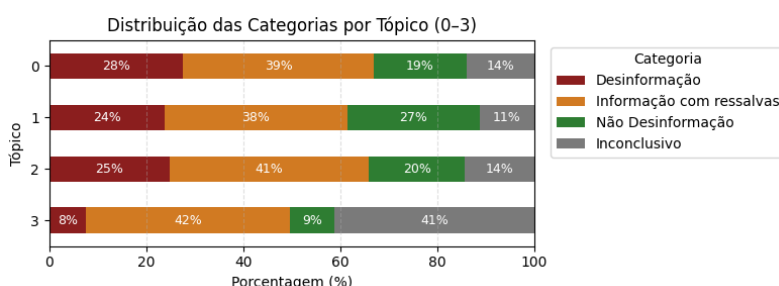


Figura 3. Distribuição das categorias por tópico.

5. Conclusão

Este estudo caracterizou a dinâmica informacional sobre a semaglutida no YouTube por meio da integração entre análise de conteúdo, métricas de engajamento e modelagem de tópicos. Os resultados indicam que o debate público é predominantemente estruturado por conteúdos classificados como *Informação com Ressalvas* (49,5%), que concentram os maiores níveis de interação. Esse achado sugere que o principal vetor de risco informacional não está restrito à desinformação explícita, mas à circulação de conteúdos parcialmente corretos, que combinam informações válidas com omissões relevantes.

Além disso, conteúdos classificados como *Desinformação*, embora minoritários (12,1%), apresentam elevado engajamento, configurando-se como a segunda categoria

com maior volume de interações, o que evidencia seu potencial de amplificação no ambiente digital. A modelagem de tópicos reforça esse cenário ao indicar forte concentração da discussão em narrativas baseadas em experiências pessoais, com predominância do Tópico 0, enquanto tópicos clínicos apresentam baixa representatividade. Esse padrão sugere que a construção do conhecimento ocorre majoritariamente a partir de percepções individuais, o que pode impactar a tomada de decisão dos usuários e representar um risco à saúde pública. Como contribuição, o trabalho demonstra o potencial da integração entre análise de conteúdo, engajamento e modelagem de tópicos para identificar padrões de desinformação em larga escala. Como trabalhos futuros, sugere-se: (i) expandir a análise para outros medicamentos discutidos em plataformas digitais; e (ii) desenvolver modelos automatizados de detecção de desinformação em saúde com base em aprendizado de máquina e processamento de linguagem natural.

Referências

- Carvalho, J. Q. M. d., Caetano, J. D. S., Bokehi, J. R., and Castilho, S. R. d. (2025). Obesidade: qual a qualidade das informações sobre tratamento farmacológico disponíveis no youtube? *RECIIS*.
- Cavalini, A., Turi, L., Pacheco, A., and Comarela, G. (2025). Uma abordagem baseada em llms para análise de desinformação sobre vacinas. In *Simpósio Brasileiro de Computação Aplicada à Saúde*.
- Depoix, J. (2023). Youtube transcript api. <https://github.com/jdepoix/youtube-transcript-api>. Acesso em: 2026.
- Dias, A., Tanure, R. R., Almeida, J., Lima, H., and Ferreira, C. (2024). Análise da percepção do uso de cigarros eletrônicos no brasil por meio de comentários no youtube. In *Brazilian Symposium on Multimedia and the Web*.
- Fantaus, S. S. (2023). Uso irracional de medicamentos: análise do conteúdo veiculado no tiktok sobre medicamentos e suplementos emagrecedores. Trabalho de Conclusão de Curso, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Google (2024a). Gemini api documentation. <https://ai.google.dev>. Acesso em: 2026.
- Google (2024b). Google cloud natural language api. <https://cloud.google.com/natural-language>. Acesso em: 2026.
- Google (2024c). Youtube data api (v3). <https://developers.google.com/youtube/v3>. Acesso em: 2026.
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- Henz, M., Heckler, W., and Barbosa, J. (2025). Uma avaliação da capacidade de modelos de linguagem para análise de sentimentos em um contexto de saúde mental. In *Simpósio Brasileiro de Computação Aplicada à Saúde*.
- International Telecommunication Union (2023). Measuring digital development: Facts and figures 2023. Disponível em: <https://www.itu.int/en/mediacentre/Pages/PR-2024-11-27-facts-and-figures.aspx>. Acesso em: 25 jan. 2025.

- Khazaal, Y., Chatton, A., Cochand, S., Coquard, O., Fernandez, S., Khan, R., Billieux, J., and Zullino, D. (2009). Brief discern, six questions for the evaluation of evidence-based content of health-related websites. *Patient Education and Counseling*, 77(1):33–37.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Oliveira, C. C., Braga, L. H. G., Jordão, P. A., and Silva, C. A. (2026). Uso off label da semaglutida (ozempic®) por indivíduos não diabéticos para perda de peso. *Revista Científica FACS*.
- Pombeiro, F. R. M., Almeida, A. G. d., Oliveira, L. d. R., and Maciel, D. V. (2025). A automedicação de semaglutida para a perda de peso no brasil. *Aurum Revista Multidisciplinar*.
- Silva, E. A. S., Costa, G. S. d., Oliveira, T. B. d., and Conceição, M. S. d. (2026). O uso indiscriminado de semaglutida para fins estéticos: suas causas e consequências em uma revisão bibliográfica. *Pensar Acadêmico*.
- Souza, R. A. d., Santos, A. C. O., and Hott, S. C. (2025). Os efeitos e riscos do uso de fármacos para emagrecimento: uma abordagem científica. *Revista Multidisciplinar Integrada – REMI*, 6(1):1–12.
- Teixeira, M. C. and Reis, J. C. S. (2023). Análise do discurso de Ódio em comentários de vídeos no youtube: Um estudo de caso da cpi da covid-19 no brasil. In *Brazilian Symposium on Databases*.
- World Health Organization (2020). Managing the covid-19 infodemic: promoting healthy behaviours and mitigating the harm from misinformation and disinformation. https://www.who.int/health-topics/infodemic#tab=tab_1. Acesso em: 27 abr. 2026.
- YouTube (2024). Youtube creator awards. <https://support.google.com/youtube/answer/7682560>. Acesso em: 2026.