

# Curadoria Algorítmica no TikTok: Análise da Estratégia de Engajamento para Novos Usuários

Léo Soares de Oliveira Júnior e Eliane Cristina de Freitas Rocha

**Resumo**—A transição das redes sociais do Grafo Social para o Grafo de Conteúdo, impulsionada por plataformas de vídeos curtos como o TikTok, consolidou a Economia da Atenção. Sob este modelo, a curadoria algorítmica opera como uma caixa-preta, tornando-se crítica no cenário de *Cold Start*, momento em que o sistema necessita reter a atenção de um novo usuário sem dispor de dados comportamentais prévios. Esta monografia investiga empiricamente a estratégia primária de engajamento do TikTok. Adotando a metodologia de Auditoria de Base, foram simuladas 30 contas de navegação passiva em ambiente controlado, coletando os dados referentes aos 200 primeiros vídeos recomendados para cada conta. Os experimentos refutam a aleatoriedade da curadoria inicial, evidenciando a entrega de conteúdos padronizados que transitam progressivamente para a exploração de nichos temáticos. A análise semântica dessa amostra revelou uma dieta informacional com diferentes esferas socioculturais, fortemente ancorada em gatilhos visuais humanos e no fenômeno do Rage Engagement. Conclui-se que o *Cold Start* do TikTok não é composto por conteúdos aleatórios, mas sim conta com um catálogo otimizado para maximizar a retenção através da exploração dos padrões de consumo das massas.

**Index Terms**—Curadoria Algorítmica, TikTok, Cold Start, Sistemas de Recomendação, Economia da Atenção, Modelagem de Tópicos.

## I. INTRODUÇÃO

### A. O Cenário: Do Grafo Social à Economia da Atenção

O ecossistema digital experimentou uma profunda mutação na última década. As primeiras gerações de redes sociais foram fundamentadas no paradigma do "Grafo Social", cujo valor primordial consistia em mapear e digitalizar conexões interpessoais entre amigos, familiares e conhecidos [1]. Contudo, o cenário contemporâneo, impulsionado por plataformas de vídeos curtos como o TikTok, é dominado pelo "Grafo de Conteúdo" [2]. Nesta nova arquitetura, a conexão interpessoal foi substituída pela curadoria automatizada de entretenimento.

Este modelo opera sob as regras da Economia da Atenção, onde a abundância informacional tornou o tempo e o foco do usuário os recursos mais escassos e valiosos do mercado [3], [4]. A lógica de monetização destas plataformas é determinística: a maximização do tempo de tela correlaciona-se diretamente com a exposição a anúncios e, consequentemente, com o aumento da receita [5]. Para se manterem competitivas nesta disputa pela atenção humana, as plataformas precisaram

eliminar atritos na experiência de uso. Portanto, o consumo transformou-se em um processo inteiramente passivo e contínuo. O usuário não necessita mais buscar ativamente por um tema de interesse ou construir uma rede de contatos; o *feed* infinito, alimentado por sistemas de recomendação preditivos, assume o papel de curador absoluto, entregando estímulos otimizados para garantir a imersão na plataforma [6].

### B. O Problema: Opacidade e o Desafio da Partida a Frio

A transferência do poder de escolha informacional do indivíduo para sistemas preditivos introduz um problema metodológico e social crítico: a curadoria torna-se uma "caixa-preta" [7]. Sem acesso às regras de negócios e ao funcionamento dos algoritmos que orquestram a entrega de conteúdo, torna-se inviável para a comunidade científica auditar diretamente como a percepção da realidade dos usuários é moldada.

O ápice desta opacidade algorítmica manifesta-se no desafio técnico conhecido como Partida a Frio (Cold Start). Em sistemas de recomendação, o Cold Start ocorre no exato momento em que um novo perfil ingressa na plataforma [8]. A inteligência artificial depara-se com um cenário de escassez informacional absoluta, devendo reter a atenção de um indivíduo sem dispor de qualquer histórico de interações, sinais de preferência ou rastros de consumo prévio para alimentar algoritmos clássicos de filtragem colaborativa ou baseada em conteúdo.

Do ponto de vista de negócios, os minutos iniciais de navegação configuram a janela de maior vulnerabilidade para a plataforma. Trata-se do momento mais crítico da curadoria automatizada: se o sistema falhar em reter a atenção neste estágio inicial, o risco de abandono do aplicativo (churn) é crítico [8]. Portanto, o Cold Start não representa um período neutro de testes aleatórios, mas sim o momento em que a plataforma é forçada a acionar sua estratégia de engajamento mais segura e eficiente antes de iniciar a hiper-personalização [6].

### C. A Abordagem Metodológica: Auditoria de Base

Para desvendar a estratégia inicial do algoritmo diante da impossibilidade de inspecionar o código-fonte, abordagens investigativas tradicionais frequentemente recorrem à engenharia reversa. Tal prática consiste na tentativa de desconstruir o funcionamento lógico da plataforma através da análise de seus processos internos, buscando isolar as variáveis específicas e os pesos dos modelos de aprendizado que definem a seleção de conteúdo. Contudo, trata-se de um esforço frequentemente frágil frente a sistemas dinâmicos que se adaptam em tempo real. Este trabalho afasta-se desta tentativa de acesso direto à

L. S. Oliveira Júnior está vinculado ao Departamento de Ciência da Computação, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil (e-mail: leosojunior@ufmg.br).

E. C. F. Rocha está vinculada à Escola de Ciência da Informação, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil (e-mail: prof.elianecfr@gmail.com).

Manuscrito elaborado para a disciplina de Monografia em Sistemas de Informação II, 2026.

lógica algorítmica e adota a postura metodológica da Auditoria de Base (*Baseline Auditing*). Ao invés de especular sobre o funcionamento interno da "caixa-preta", a pesquisa concentra-se na observação empírica das suas saídas [7].

A investigação foca estritamente no momento do Cold Start. É relevante notar que este estágio não implica anonimato absoluto: no instante da criação de uma conta, a plataforma capta metadados de contexto, como geolocalização (IP), idioma e dispositivo. Este processo de "datatificação" é o que permite ao sistema iniciar a construção da estratégia de controle sobre o novo perfil [9], [10].

Para isolar variáveis espúrias, o estudo adota um ambiente rigorosamente controlado. Foram simulados perfis homogêneos de jovens adultos (23 anos) navegando de forma totalmente passiva a partir de um mesmo endereço IP e rede em Belo Horizonte, Minas Gerais. Esta abordagem permite registrar como a plataforma performa o que Bhandari e Bimo [11] definem como o "Eu Algoritmizado", onde o sistema projeta uma identidade para o usuário recém-chegado antes mesmo de possuir dados comportamentais. Longe de ditar regras universais, o objetivo é registrar com alta fidelidade as manifestações deste ecossistema sob condições fixas.

#### D. Objetivos e Contribuições

Enquanto grande parte da literatura acadêmica sobre curadoria algorítmica se concentra no estágio avançado da navegação, debatendo fenômenos como o isolamento informacional (onde o usuário é exposto apenas a perspectivas convergentes), as bolhas de filtro e as câmaras de eco (estruturas que amplificam crenças pré-existentes ao blindar o indivíduo contra o contraditório [5]), este trabalho é embasado no pressuposto de que é imprescindível mapear o ponto de partida antes de analisar a trajetória da experiência do usuário.

O objetivo central desta monografia não é analisar a personalização do conteúdo recomendado com base nas interações de um usuário, mas realizar a caracterização semântica e estrutural da "isca inicial" do TikTok. A pesquisa busca responder empiricamente o que o algoritmo considera universalmente eficiente para injetar no feed de um novo usuário antes que ele manifeste qualquer interação que reflita seus interesses ou personalidade, o que Bhandari e Bimo [11] definem como o momento de construção do "Eu Algoritmizado" na plataforma.

Para responder a esta questão, o artigo está estruturado em duas frentes de experimentação: o Experimento 1 afere a homogeneidade do conteúdo recomendado entre diferentes usuários e o ponto de transição entre a entrega de conteúdo de alta performance e o teste de novos nichos semânticos. Enquanto isso, o Experimento 2 emprega Visão Computacional e Processamento de Linguagem Natural (NLP) como instrumentos para extrair os clusters (grupos) temáticos e mapear a dieta informacional de boas-vindas.

Neste contexto, compreende-se a dieta informacional, conforme estabelecido por Johnson [12], como o regime estruturado de estímulos aos quais o usuário é submetido pelo algoritmo. É uma vitrine que, ao limitar o cardápio de conteúdos no Cold Start, atua como um mecanismo de calibração que molda as futuras preferências do novo usuário. Contudo, para

sustentar a arquitetura e a interpretação destes experimentos, faz-se necessário primeiro fundamentar os conceitos que regem esse ecossistema.

## II. FUNDAMENTAÇÃO TEÓRICA

### A. A Delegação da Curadoria Algorítmica

A transição do ecossistema digital contemporâneo é marcada por uma mudança estrutural no modelo de negócios das redes sociais, consolidando a chamada Economia da Atenção. Nesse paradigma, a atenção humana deixa de ser um subproduto da navegação e se torna um recurso mercantilizado, representando a viabilidade comercial destas plataformas [3], [4]. Diferentemente das gerações anteriores, que operavam sob a lógica do Grafo Social, o modelo explorado pelo TikTok prioriza o Grafo de Conteúdo, onde a arquitetura da rede é ditada pela capacidade do item de reter o olhar do usuário pelo máximo de tempo possível [2], [1].

Para garantir a captura contínua desta atenção, a plataforma subverte relações de consumo que historicamente baseavam-se no exercício ativo do indivíduo. Esse esforço cognitivo de busca e seleção foi progressivamente suprimido e delegado a agentes automatizados, permitindo que a Inteligência Artificial assumisse a posição de curadora ininterrupta da realidade digital do usuário [5].

Neste contexto, os sistemas de recomendação transformam o que seria um mero repositório de vídeos em um motor preditivo e agressivo de interesses. A literatura aponta que a plataforma não pode arcar com uma curva de aprendizado lenta e, para evitar o abandono imediato (churn), o sistema atua de forma pragmática [6]. Em vez de tentar adivinhar interesses individuais de um novo usuário, o algoritmo prioriza a entrega de conteúdos que já foram exaustivamente testados e validados pela base global [8].

Ao recorrer a um acervo de conteúdo de sucesso comprovado, a plataforma busca assegurar a retenção antes mesmo de possuir dados suficientes para mapear a identidade do indivíduo. É precisamente essa urgência comercial em reter a atenção a qualquer custo que fundamenta a estratégia inicial de curadoria algorítmica, em que a análise apresenta desafios na fronteira entre a engenharia de software e a psicologia comportamental [6].

### B. Sistemas de Recomendação e Partida a Frio

A urgência em garantir o engajamento imediato esbarra em um problema clássico da computação: o desafio da Partida a Frio (Cold Start). Sistemas de recomendação modernos operam tradicionalmente a partir de dois fundamentos: a filtragem baseada em conteúdo e a filtragem colaborativa, ambas estritamente dependentes de dados estruturados [8]. Enquanto a primeira técnica exige um mapeamento prévio dos interesses específicos do indivíduo para correlacioná-los aos atributos dos itens disponíveis no catálogo (no caso do TikTok, o catálogo de vídeos da plataforma), a segunda necessita de um histórico denso de interações entre múltiplos usuários para extrair padrões de comportamento análogos [8]. No exato momento em que uma conta nova é inaugurada, o sistema depara-se com um cenário de escassez informacional absoluta,

no qual nenhuma dessas técnicas possui insumos suficientes para ser computada de forma eficaz [8].

Esta ausência de parâmetros instaura um gargalo simultaneamente técnico e comercial. Sob a perspectiva de negócios, os minutos iniciais de navegação representam a janela de maior vulnerabilidade para a plataforma. Caso o conteúdo entregue neste momento crítico não possua alta aderência e retenção, quebra-se a promessa de entretenimento fluido, gerando atrito cognitivo e aumentando a possibilidade de abandono da rede (churn) [6]. Tecnicamente, este risco obriga o sistema a equilibrar a recomendação de conteúdos de alta performance, já testados em outros usuários, com a apresentação de conteúdos nichados que podem ou não refletir um interesse específico do indivíduo [6].

Consequentemente, diante de um usuário totalmente desconhecido, a plataforma abstém-se temporariamente do risco de uma hiper-personalização prematura, evitando a recomendação de conteúdos excessivamente segmentados que poderiam resultar em rejeição instantânea. Em vez disso, levanta-se a hipótese de que o sistema recorre a conteúdos padronizados, selecionando vídeos a partir de métricas globais de desempenho. Esta estratégia algorítmica fornece uma dieta informacional de segurança [12], retendo o usuário através de uma vitrine de boas-vindas controlada, até que haja captação suficiente de dados comportamentais explícitos e implícitos para justificar a transição rumo a um feed individualizado [6].

### C. A Semântica da Retenção Inicial

Considerando a adoção desta rede de segurança nos instantes iniciais de navegação, a proposta deste trabalho possui um escopo bem delimitado. Tradicionalmente, grande parte da literatura e do debate sobre curadoria algorítmica e sistemas de recomendação de conteúdo concentra-se no estágio avançado da experiência do usuário, debatendo os riscos da hiper-personalização, das bolhas de filtro e das câmaras de eco [9], [7]. A investigação convencional, contudo, negligencia o momento da constituição do perfil e parte do cenário em que o algoritmo já possui informações suficientes para confinar o usuário em um ambiente altamente aderente aos seus interesses e traços comportamentais previamente registrados.

No entanto, para compreender como um indivíduo é eventualmente confinado em um ecossistema informacional gerado especificamente para ele, é necessário, primeiramente, mapear qual foi a porta de entrada. É virtualmente impossível auditar de forma rigorosa o viés de uma trajetória algorítmica sem antes registrar e isolar com precisão o seu ponto de partida [10].

Por essa razão, esta pesquisa não foca em analisar como o algoritmo se adapta às interações do usuário na rede, mas sim se concentra especificamente na caracterização da isca inicial. O interesse da investigação é desvendar, qualitativa e quantitativamente, as características dos conteúdos que o algoritmo considera universalmente eficientes e seguros para injetar no feed de um usuário antes que ele tenha a oportunidade de manifestar qualquer ação sobre o sistema, consequentemente construindo seu "Eu Algoritmizado" [11].

## III. METODOLOGIA

**Nota de Transparência:** Durante o desenvolvimento desta pesquisa, ferramentas de Inteligência Artificial foram utilizadas para apoiar a implementação do scraper de dados, a formatação visual dos gráficos e a estruturação lógica do relatório. Todo o material gerado passou por curadoria e validação humana e rigorosa.

Para garantir o rigor da coleta de dados e analisar o comportamento algorítmico no estágio de Cold Start, a pesquisa foi estruturada em um ambiente de observação estritamente controlado. A metodologia afasta-se de abordagens baseadas em APIs oficiais, que frequentemente filtram ou limitam os dados entregues a pesquisadores, optando pelo desenvolvimento de uma infraestrutura própria que simula a navegação orgânica de um usuário real, que é uma prática consolidada na literatura de auditoria de algoritmos para contornar a opacidade das plataformas [7], [13].

### A. A Arquitetura da Ferramenta de Coleta

A extração de dados em plataformas digitais contemporâneas apresenta um desafio de estabilidade, dada a constante mudança de seus elementos visuais. Para contornar essa barreira, foi desenvolvida uma ferramenta de automação na linguagem Python, utilizando a biblioteca Selenium. A arquitetura do sistema foi modularizada em gerenciamento de sessão, captura de dados e paralelização de execução.

O grande diferencial do scraper desenvolvido é o método de captura que, diferentemente de bots convencionais que operam raspando o código visual da página (HTML), esta ferramenta monitora ativamente as requisições de rede (XHR) em segundo plano através do protocolo CDP (Chrome DevTools Protocol). A aplicação intercepta os pacotes de dados originais (em formato JSON - JavaScript Object Notation) que o servidor do TikTok envia para o navegador instantes antes de o vídeo ser renderizado na tela.

Esta interceptação de tráfego de rede garante acesso à totalidade dos dados dos vídeos, permitindo a extração de métricas que não são expostas na interface visual do aplicativo [13]. Para cada vídeo entregue pelo feed, o sistema armazenou um JSON individual contendo os seguintes pacotes de dados:

- **Dados do Autor:** Identificador único, nomenclatura, status de verificação e métricas de influência (volume total de seguidores, curtidas acumuladas e contagem de vídeos).
- **Estatísticas de Desempenho do Vídeo:** Número de visualizações (playCount), curtidas, compartilhamentos e quantidade de comentários.
- **Metadados Semânticos:** Descrição, comentários, hashtags, dados do áudio (título e originalidade) e a imagem de capa (thumbnail) em formato jpg.

### B. Condições do Experimento

Para assegurar que os dados não fossem contaminados por variáveis externas, o experimento foi realizado sob condições rígidas de controle. O processo envolveu a criação simultânea

de 30 novas contas de teste, seguindo protocolos de auditoria para mitigar vieses de interação [13].

Para viabilizar a automação e contornar mecanismos de segurança antibot (como CAPTCHAs) que inviabilizariam ou dificultariam a coleta, o login na plataforma não foi executado pelo scraper em tempo real. Em vez disso, a autenticação inicial foi realizada manualmente, utilizando a infraestrutura da ferramenta apenas para extrair e armazenar os cookies de sessão de cada conta. Posteriormente, ao iniciar da coleta, a automação injetou os respectivos cookies em cada instância do navegador, simulando a navegação orgânica de um usuário já previamente autenticado na plataforma.

Com o intuito de isolar o comportamento do sistema de recomendação, o protocolo obedeceu às seguintes restrições paramétricas:

- **Perfil Demográfico e Geográfico:** Todas as contas foram configuradas com a mesma data de nascimento (30/07/2002), simulando indivíduos jovens adultos. As conexões foram roteadas a partir de um único endereço IP (Belo Horizonte, MG), isolando a variável geográfica [13].
- **Passividade:** O comportamento simulado foi de navegação absolutamente passiva. A automação foi programada exclusivamente para rolar a tela (scroll) assim que os dados do vídeo fossem capturados no buffer de rede. Nenhuma curtida, abertura de comentários, visualização de perfis ou busca manual foi realizada durante o experimento, capturando exatamente os primeiros 200 vídeos recomendados para cada conta e evitando, assim, a geração de sinais que poderiam induzir a personalização do feed [11].
- **Variável Temporal e Paralelização:** Para investigar se o algoritmo altera o conteúdo recomendado dinamicamente com base no momento do dia, a variável temporal precisou ser controlada. Em vez de executar as coletas sequencialmente, utilizou-se multiprocessamento para rodar múltiplas instâncias de navegadores simultâneos. As 30 contas foram divididas em três grupos de 10 contas cada: Grupo A com execuções realizadas às 09:00h, Grupo B às 16:00h e Grupo C às 23:00h. Assim, garantimos que as variações no conteúdo refletissem decisões globais do algoritmo e não vieses de interação prévia [13].

Esta arquitetura assegurou que as variações no conteúdo entregue entre as contas refletissem puramente as decisões globais do algoritmo do TikTok, livres de qualquer viés de interação humana prévia.

### C. Tratamento de Dados e Estruturação da Amostra

A etapa seguinte ao término das rotinas de automação consistiu na consolidação dos dados brutos coletados em um formato adequado para análise estatística e exploração semântica. Os dados interceptados do tráfego de rede da plataforma foram armazenados em JSON, estrutura caracterizada pela hierarquia de atributos aninhados. Embora altamente eficiente para o transporte de dados via web, essa estrutura é inadequada para o processamento analítico direto.

Para contornar esta limitação, esses dados foram tratados utilizando a biblioteca *pandas* em um ambiente *Jupyter Notebook*. O algoritmo de tratamento realizou o "achatamento" (flattening) de cada arquivo JSON individual, convertendo os nós da estrutura hierárquica em colunas de um único *DataFrame*.

Durante este processo de conversão, atributos irrelevantes para as hipóteses da pesquisa (como dados residuais e links internos de CDN) foram descartados. O esforço computacional concentrou-se na extração limpa e na tipagem correta de três grandes blocos de variáveis:

- **Influência do autor:** Status de verificação do autor e número de seguidores.
- **Variáveis de Engajamento:** Número de visualizações, de curtidas, de compartilhamentos e de salvamento.
- **Variáveis Textuais e Semânticas:** Descrição do vídeo, hashtags e comentários.

O processamento das coletas paralelas resultou em um conjunto de dados (dataset) de 6.000 vídeos (200 vídeos por conta), acompanhados por 18.119 hashtags e 111.455 comentários de usuários. Para garantir a otimização da leitura (I/O) nas fases de análise estatística e de aprendizado de máquina, este conjunto final de dados foi exportado no formato *Apache Parquet*, acelerando significativamente as consultas realizadas nas etapas dos experimentos subsequentes.

## IV. EXPERIMENTO 1: A DINÂMICA DA ENTREGA

O primeiro bloco de experimentação desta pesquisa buscou compreender a estrutura de distribuição primária do algoritmo do TikTok frente ao desafio absoluto do Cold Start. A hipótese central investigada (H1) postula que a dieta informacional de boas-vindas não é aleatória, mas sim uma "Vitrine de Sucessos" estruturada, cuja composição é calibrada dinamicamente com base no contexto temporal do usuário recém-chegado.

### A. O Teste de Homogeneidade e a Métrica de Sobreposição

Para aferir a homogeneidade do conteúdo recomendado entre os perfis simulados, adotou-se o Índice de Similaridade de Jaccard. Esta métrica matemática calcula a sobreposição entre dois conjuntos (a interseção dividida pela união), fornecendo um valor de 0 (nenhuma similaridade) a 1 (conjuntos idênticos), sendo amplamente utilizada na mineração de dados para quantificar a proximidade entre coleções [14]. A análise foi aplicada cruzando os catálogos dos 200 vídeos iniciais entregues a cada uma das 30 contas de teste, resultando em uma matriz de 900 comparações.

O primeiro achado empírico é a refutação da aleatoriedade no "grau zero" da recomendação. Considerando que o catálogo global da plataforma hospeda bilhões de vídeos ativos, a probabilidade estatística de dois novos usuários, sem qualquer histórico de navegação cruzada, receberem os mesmos vídeos de forma orgânica é praticamente nula [14]. No entanto, o experimento revelou que 100 dos cruzamentos entre as 30 contas apresentaram sobreposição de conteúdo. Este fenômeno atesta de forma inequívoca que o sistema não opera explorando o catálogo de forma aleatória ao receber um perfil novo, mas insere o usuário em uma esteira de recomendação de conteúdo predefinida.

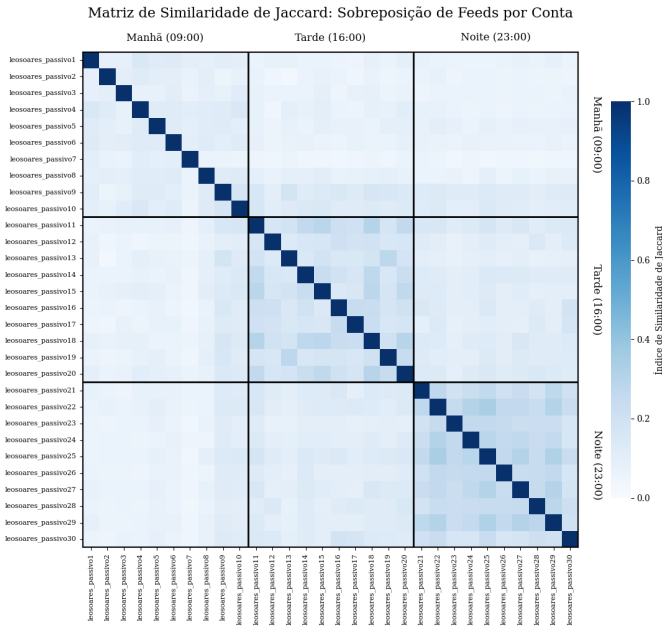


Figura 1. Matriz de Similaridade de Jaccard demonstrando a sobreposição de conteúdo entre as 30 contas, agrupadas por turno de coleta.

### B. Calibragem Dinâmica e o Fator Temporal

A matriz de Jaccard evidencia que o conteúdo recomendado para um novo usuário não é estático ou engessado, uma vez que a paralelização das contas revelou que a variável temporal exerce um papel determinante na curadoria.

Ao agrupar os índices de similaridade, observa-se disparidade entre contas criadas no mesmo horário e contas criadas em horários distintos. A similaridade média intra-grupo (entre contas navegando no mesmo turno) foi de 0,1835. Na prática, isso significa que perfis criados simultaneamente recebem, em média, conteúdos altamente redundantes. O grau máximo de confinamento ocorreu no turno da Noite (23:00h), em que o cruzamento intra-grupo atingiu um índice de Jaccard de 0,2444, representando aproximadamente 78 vídeos idênticos em uma janela de 200 entregas (quase 40% de repetição absoluta do feed).

Em contrapartida, quando as comparações são realizadas no cenário inter-grupo (cruzando contas da Manhã com as da Tarde ou Noite, por exemplo), a similaridade média sofre uma queda abrupta para 0,0936. A sobreposição entre os turnos da Manhã e da Noite, especificamente, resultou no menor índice registrado (0,0771), denotando apenas cerca de 28 vídeos em comum.

Para validar rigorosamente esta observação, aplicou-se o teste estatístico de Mann-Whitney (Estatística  $U = 35065,5$ ), frequentemente utilizado em estudos de mineração de dados para comparar distribuições de populações distintas [14]. O resultado obteve um  $p$ -valor de  $1,28 \times 10^{-34}$ , muito inferior ao limiar de significância de 0,05. Este valor estatístico rejeita a hipótese nula, comprovando matematicamente que a taxa de repetição de vídeos dentro de um mesmo turno é significativamente maior do que entre turnos diferentes.

A conclusão metodológica deste experimento indica que

o algoritmo de recomendação não possui apenas uma única "isca" global. O algoritmo calibra sua linha de base de engajamento dinamicamente, reagindo à janela temporal do dia e, possivelmente, aos padrões de consumo em tempo real da base de usuários ativa [6]. Esse comportamento dinâmico reafirma a natureza da plataforma como um sistema adaptativo que ajusta o conteúdo recomendado conforme o comportamento dos usuários utilizando o aplicativo.

### C. O Ponto de Ruptura

A observação de que a plataforma submete novos usuários a conteúdos padronizados levanta um questionamento sobre a duração dessa estratégia. Na literatura de sistemas de recomendação, este desafio é classicamente tratado como o dilema entre Exploitation (exploração de itens de alta performance já conhecidos) e Exploration (descoberta de novas preferências do usuário) [8]. Para investigar em que momento o algoritmo abandona a homogeneidade do conteúdo inicial e inicia a fase de hiper-personalização, foi aplicada uma análise de janela deslizante (rolling window) de 40 vídeos sobre os índices de Jaccard ao longo dos 200 scrolls de cada conta.

Os dados empíricos mostram que a vitrine inicial possui uma validade curta. A fase primária de navegação (vídeos 0 a 39) registrou o ápice da homogeneidade, com uma similaridade média de 0,0730. À medida que o experimento avançou para as janelas finais (vídeos 160 a 199), a sobreposição caiu para 0,0128, representando uma redução de 82,48% no grau de repetição do feed.

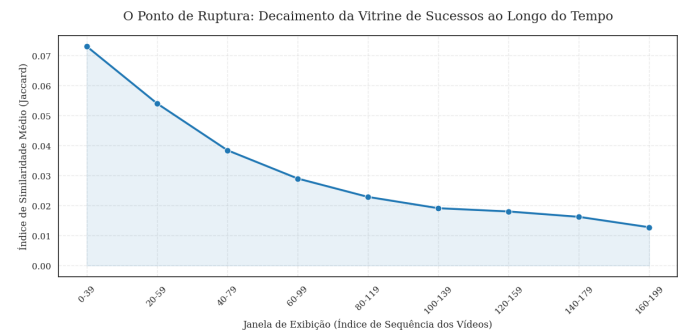


Figura 2. Decaimento da sobreposição de conteúdo (Índice de Jaccard) ao longo da navegação

Para confirmar que esta redução não era fruto do acaso, aplicou-se o teste de correlação de Spearman sobre a distribuição, que identificou uma tendência de decaimento inversamente proporcional com significância estatística ( $\rho = -0,5103$ ,  $p < 0,001$ ) [14]. Adicionalmente, a aplicação do teste de Mann-Whitney entre o bloco inicial e o final que confirmou uma diferença de distribuição conclusiva ( $U = 176135,0$ ,  $p = 7,93 \times 10^{-110}$ ).

A análise das métricas de desempenho também revela aspectos interessantes sobre o conteúdo inicial. Comparando a fase primária (0 a 39) com a fase de encerramento da coleta (160 a 199), a quantidade média de visualizações dos vídeos apresentados caiu de 3,32 milhões para 2,57 milhões (-22,56%). O impacto foi ainda maior nas curtidas, caindo de

438 mil para 280 mil (-35,99%) e nos comentários de 1.872 para 1.404 interações (-25,03%).

A robustez desta queda também foi submetida a novos testes estatísticos (Mann-Whitney e Spearman). Os resultados atestaram a significância da queda em todas as frentes observadas: visualizações ( $U = 802297,5$ ,  $p < 0,001$ ), curtidas ( $U = 788981,5$ ,  $p < 0,001$ ) e comentários ( $U = 753720,0$ ,  $p = 0,023$ ). Além de validar a diferença isolada entre o início e o fim da amostra, os coeficientes de Spearman comprovam uma tendência de queda contínua do engajamento correlacionada ao aumento da profundidade de navegação no contexto do experimento.

Em resumo, os resultados do Experimento 1 validam empiricamente as hipóteses acerca do funcionamento do Cold Start. O TikTok inaugura a experiência do novo usuário, injetando artificialmente os vídeos de maior retenção global naquele dado instante de tempo e, somente após garantir o engajamento inicial, o motor preditivo reduz a segurança do conteúdo recomendado, arriscando conteúdos de menor alcance na tentativa de isolar, por tentativa e erro, o viés semântico do indivíduo [8], [6].

## V. EXPERIMENTO 2: CARACTERIZAÇÃO SEMÂNTICA DA ISCA INICIAL

A observação empírica de que o TikTok submete novos usuários a uma dieta informacional padronizada (conforme demonstrado no Experimento 1) clarifica o estágio de Cold Start do TikTok. No entanto, compreender o que é entregue pelo algoritmo levanta uma questão fundamental sobre a natureza do Grafo de Conteúdo: o que compõe, exatamente, essa vitrine inicial?

Se a recomendação primária não é aleatória, os vídeos selecionados para cativar um usuário recém-chegado devem possuir aspectos semânticos, estéticos e culturais bem definidos. Para compreender esses aspectos, o segundo experimento desta pesquisa abandona a análise das métricas de engajamento e foca na extração semântica do material audiovisual. Esta abordagem de auditoria que utiliza modelos de Inteligência Artificial para decodificar o conteúdo visual e textual em larga escala é uma técnica emergente na literatura para contornar a opacidade de sistemas de recomendação [13], permitindo que o pesquisador "enxergue" o conteúdo sob uma perspectiva próxima a do algoritmo.

### A. Auditoria Visual: Decodificando Imagens de Capa

O primeiro ponto de contato entre o usuário e o conteúdo em plataformas de vídeos curtos é estritamente visual. A imagem de capa (thumbnail), que no TikTok consideramos como o primeiro frame de um vídeo, funciona como o principal atrativo do conteúdo, servindo como um chamariz cognitivo que decide, em uma fração de segundo, se o usuário continuará assistindo ou realizará o scroll [6]. Portanto, para entender a semântica do conteúdo inicial, foi necessário interpretar a semântica dessas capas.

Em face da inviabilidade de categorização humana manual em larga escala e buscando eliminar qualquer viés de interpretação, a primeira etapa deste experimento adotou uma aborda-

gem baseada em Inteligência Artificial Multimodal. Utilizou-se o modelo Large Language Model (LLM) Google Gemini 2.5 Flash para processar os arquivos de imagem associados aos vídeos capturados na coleta.

O modelo foi instruído através de engenharia de prompt a operar como um classificador e extrator de atributos visuais. Para cada capa, o Gemini estruturou as seguintes variáveis:

- **Presença Humana:** Identificação booleana sobre a presença de figuras humanas (rostos, corpos) na imagem.
- **OCR:** Detecção de texto sobreposto na imagem e sua respectiva transcrição
- **Emoção Predominante:** Classificação da emoção predominante expressa pelos sujeitos ou pela composição da cena.
- **Suspeita IA:** Avaliação de indícios de imagens produzidas por IA generativa.
- **Transcrição:** Uma descrição textual completa da cena para traduzir o elemento visual em formato legível para análises posteriores.

Os resultados do processamento das imagens (totalizando milhares de capas únicas analisadas) forneceram as primeiras pistas semânticas da estratégia de retenção. A análise revelou que 62,1% das capas entregues no estágio de Cold Start possuem presença humana explícita.

Distribuição da Presença Humana nas Capas (Cold Start)

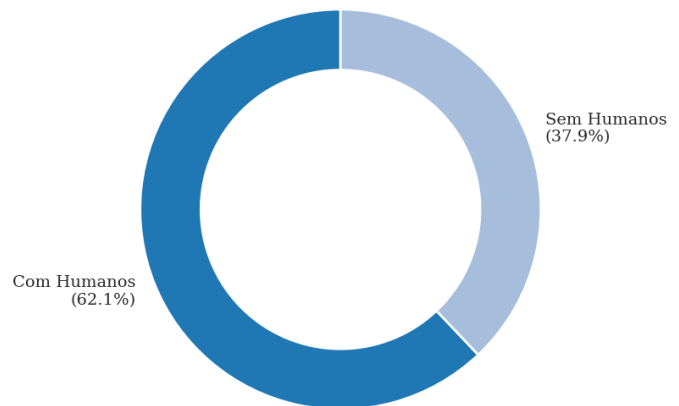


Figura 3. Distribuição de presença humana nas capas de vídeos

Considerando que os vídeos entregues no estágio inicial de navegação compõem a primeira estratégia de retenção do sistema, infere-se que eles representam os conteúdos de mais alta aprovação estatística global da plataforma. O fato de que a maioria absoluta desses conteúdos de alto desempenho exhibe rostos em destaque evidencia que a humanização do conteúdo funciona como um atalho cognitivo eficaz para gerar empatia e prender a atenção. Isso corrobora a literatura sobre Economia da Atenção, que aponta a face humana como um estímulo de alto impacto, capaz de acionar mecanismos empáticos que reduzem o atrito cognitivo e maximizam a retenção do olhar [3].



Assim, o motor de recomendação, otimizado estritamente para maximizar o tempo de tela, assimila esse padrão orgânico de consumo da massa de usuários e o reproduz em escala na dieta informacional inicial. Este resultado reflete a lógica do "Eu Algoritmizado" discutida por Bhandari e Bimo [11], onde a plataforma favorece a presença humana não apenas por uma afinidade estética, mas pela sua eficácia comprovada em reter audiência. A face humana atua, portanto, como a viabilizadora do modelo de negócio de plataformas de redes sociais no contexto da Economia da Atenção [5].

Paralelamente, os dados revelaram que 66,1% das imagens de capa continham sobreposição textual. O uso de textos na tela de abertura cumpre um papel estratégico. Por um lado, atua como uma ferramenta de contextualização rápida, fornecendo a premissa do vídeo de forma imediata para instigar a curiosidade. Por outro, uma parcela expressiva desse volume reflete a prática estrutural de incluir legendas dinâmicas em conteúdos com fala humana. A adoção de legendas superou a função de acessibilidade para se tornar uma estratégia de retenção [6]: ela previne a fuga do usuário em cenários de navegação com o áudio silenciado e engaja a atenção visual tanto no primeiro segundo de exibição quanto no decorrer do vídeo.

Distribuição de Sobreposição Textual nas Capas (Cold Start)

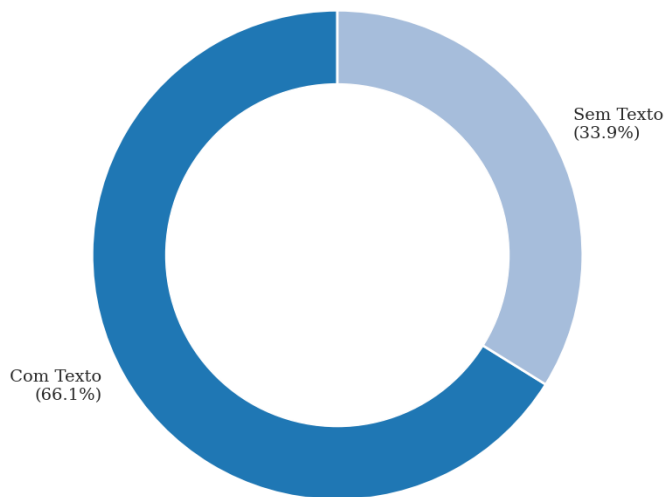


Figura 4. Distribuição de sobreposição textual nas capas de vídeos

Por fim, a auditoria visual quantificou a presença de conteúdos sintéticos na curadoria primária do aplicativo. Os resultados indicam que 11,2% (671 exibições) do material analisado apresentou indícios de geração ou manipulação por Inteligência Artificial.

A aplicação do modelo neste escopo não tem o compromisso de ser 100% assertiva sobre a presença de IA, mas sim de atuar como uma simulação da percepção visual humana. O objetivo é reproduzir a análise de um usuário comum, para o qual a diferenciação entre o real e o sintético nem sempre é

Presença de Conteúdo Sintético nas Capas (Cold Start)

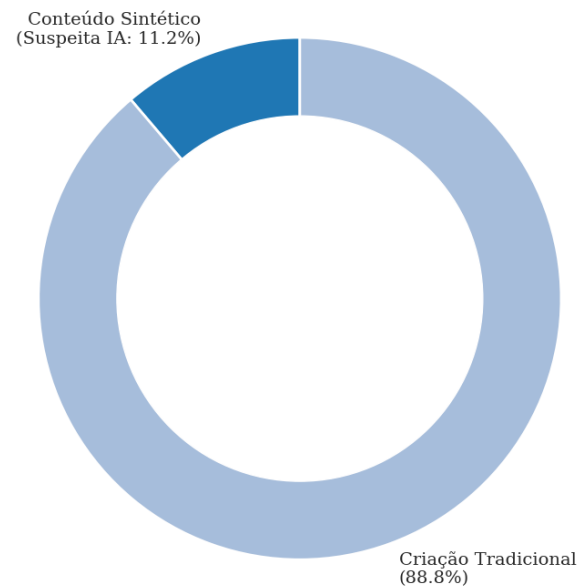


Figura 5. Proporção de conteúdo com indícios de geração por Inteligência Artificial nas capas dos vídeos de Cold Start.



Figura 6. Exemplo comparativo: Conteúdo com indícios de geração por IA (esquerda) versus criação humana tradicional (direita).

evidente durante uma rolagem rápida de tela.

Embora a criação humana tradicional ainda prevaleça de forma majoritária, correspondendo a 88,8% da amostra, a presença de uma fatia superior a 10% de conteúdo sintético no estágio Cold Start é uma observação notável. Esse volume aponta para uma crescente industrialização na produção de conteúdo, cenário no qual criadores delegam a geração de vídeos a ferramentas generativas para escalar a frequência de postagens e testar conteúdos rapidamente [11]. A aceitação desse formato pelo motor de recomendação sugere que, do ponto de vista do algoritmo, a origem sintética do vídeo não funciona como um detrator expressivo de engajamento, desde que o conteúdo cumpra os requisitos de retenção.

## B. Análise de Sentimento

Aprofundando a análise das imagens de capa em que a inteligência artificial identificou presença humana e expressões faciais legíveis, buscou-se categorizar o caráter emocional do conteúdo. A hipótese é a de que o motor de recomendação privilegia emoções específicas para capturar a empatia do usuário nos primeiros segundos da navegação.

A distribuição absoluta revelou um predomínio de emoções de baixa intensidade e positivas. A expressão "Neutra" foi a mais frequente, compondo 47,7% (1.235 aparições) da amostra, seguida pela "Alegria", presente em 31,7% (821 aparições). Emoções de alta excitação e valência negativa apareceram com frequência menor na composição do feed de partida a frio: "Tristeza" representou 3,2% e "Raiva" apenas 1,6% das capas analisadas.

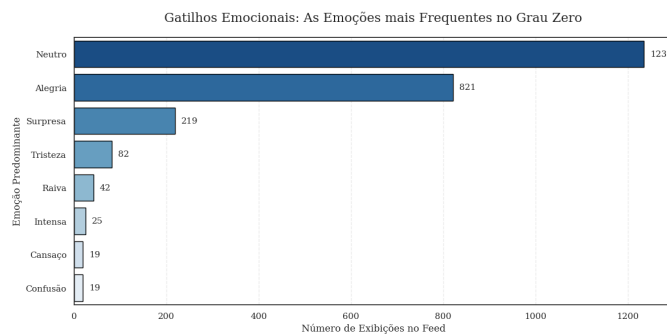


Figura 7. Distribuição do sentimento expresso em capas com presença humana no estágio de Cold Start.

Essa prevalência não indica uma escolha intencional do sistema por uma recepção "amigável", ao contrário, reflete a própria natureza do conteúdo de consumo massivo e de fácil absorção. Expressões neutras e de alegria são inerentes a vídeos casuais, como humor, estilo de vida e narrativas cotidianas, que estatisticamente retêm a atenção do público médio com baixo atrito cognitivo.

A presença ínfima de emoções negativas na vitrine de Cold Start demonstra que conteúdos altamente polarizados (como nichos políticos ou de conflito) não compõem a linha de base universal. O algoritmo apenas reproduz o padrão casual de maior aceitabilidade inicial, reservando a exploração de nichos negativos para etapas posteriores da personalização.

Contudo, ao cruzar as emoções predominantes das capas com a mediana de visualizações globais dos respectivos vídeos, a análise revelou um paradoxo de engajamento. Embora as emoções "Neutra" e "Alegria" sejam as mais abundantes no conteúdo, elas apresentam as menores taxas de retorno em alcance global (1,50 milhão e 1,90 milhão de visualizações medianas, respectivamente). Enquanto isso, as emoções de alta excitação dominam o topo da performance. Vídeos cujas capas transmitem "Surpresa" atingem uma mediana de 2,20 milhões de visualizações, enquanto a "Raiva", campeã de engajamento, registra 2,95 milhões de visualizações.

Esta disparidade entre a frequência de exibição e as métricas de desempenho corrobora a literatura sobre a economia da atenção e o fenômeno do *Rage Engagement* [15]. Conteúdos

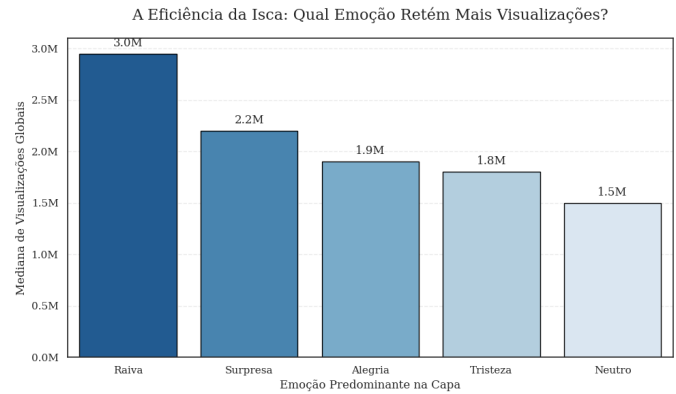


Figura 8. Performance global dos vídeos cruzada com a emoção predominante da imagem de capa, evidenciando a correlação de alta excitação.

que evocam indignação ou surpresa ativam respostas neurológicas de urgência, elevando organicamente o tempo de tela. Como o sistema atua como um otimizador dessas métricas, ele escala a distribuição dos vídeos que capturam a atenção de forma mais agressiva.



Figura 9. Amostras representativas das emoções predominantes identificadas nas capas pelo modelo multimodal.

O fato de a "Raiva" representar apenas 1,6% do feed inicial, mas registrar o maior desempenho global, ilustra a assimetria do Grafo de Conteúdo que fundamenta a estratégia de Cold Start: enquanto o conteúdo inicial é majoritariamente inofensivo para garantir a retenção inicial e evitar o churn, a hiperestimulação gera mais engajamento para a fase subsequente. Assim, a isca inicial atua como o estágio embrionário do "Eu Algoritmizado" [11], onde o algoritmo, ao transitar do modo



de Exploitation para o de Exploration [8], começa a substituir a neutralidade do conteúdo de boas-vindas pela personalização agressiva baseada nas interações do usuário.

### C. Caracterização Semântica e Análise Temática

Para mapear a semântica dos conteúdos recomendados no estágio de Cold Start, foi necessária uma abordagem capaz de ir além da análise isolada de métricas de engajamento ou de imagens estáticas. A proposta desta etapa consistiu em fundir os metadados nativos dos vídeos com a tradução visual realizada previamente, criando uma base de conhecimento unificada para descobrir, de forma automatizada e sem viés de categorização humana, quais são as grandes "iscas" temáticas da plataforma.

A extração semântica foi conduzida utilizando o BERTopic, uma arquitetura de Machine Learning não-supervisionada de estado da arte para modelagem de tópicos [16]. Diferente de abordagens tradicionais (que apenas contam a frequência de palavras), o BERTopic utiliza redes neurais baseadas em Transformers para compreender o contexto e o significado das palavras e frases, e é amplamente reconhecido pela robustez em revelar padrões em grandes volumes de texto [14]. Esta técnica permite mapear não apenas hashtags isoladas, mas os clusters temáticos que compõem a estratégia de retenção do TikTok, contornando a opacidade algorítmica através de uma auditoria fundamentada em evidências semânticas [13]. O processamento analítico foi estruturado nas seguintes etapas metodológicas:

- 1) **Fusão Multimodal e Higiênização dos Dados:** Para cada vídeo, foram concatenadas a descrição original do autor, as hashtags e a descrição visual sintética gerada pelo modelo Gemini. Para evitar que o algoritmo agrupasse vídeos com base em ruídos do sistema, aplicou-se uma higienização, removendo stopwords (em português e inglês), jargões algorítmicos da rede (como "fyp", "foryou", "viral") e termos do prompt de visão computacional (como "imagem", "mostra", "cena").
- 2) **Vetorização Semântica:** O conteúdo textual limpo foi submetido ao modelo SentenceTransformer pré-treinado em múltiplos idiomas (paraphrase-multilingual-MiniLM-L12-v2). Esta rede converteu os textos em embeddings (vetores densos multidimensionais), mapeando a proximidade de significado entre os vídeos.
- 3) **Redução e Extração via c-TF-IDF:** O BERTopic aglomerou os vídeos com maior similaridade vetorial e extraiu as palavras mais representativas de cada nicho utilizando a métrica c-TF-IDF (Class-based Term Frequency-Inverse Document Frequency). O modelo foi parametrizado para filtrar a base e retornar exatamente os 20 clusters temáticos de maior densidade, isolando e descartando os outliers de baixa correspondência.

A definição do número de tópicos (parametrizada para 20 clusters) foi estabelecida mediante um processo iterativo de fine-tuning. A calibração deste hiperparâmetro foi guiada pelo equilíbrio entre a cobertura semântica e a coesão interna dos agrupamentos seguindo diretrizes metodológicas de modelagem de tópicos neurais [16]: enquanto reduções mais acentuadas resultavam em uma simplificação excessiva que ocultava

Tabela I  
TAXONOMIA DO COLD START: OS 20 PRINCIPAIS CLUSTERS TEMÁTICOS E SEUS RESPECTIVOS TERMOS DESCRITORES EXTRAÍDOS PELO MODELO BERTOPIC.

| Cluster    | Representatividade | Principais Termos (c-TF-IDF)                         |
|------------|--------------------|------------------------------------------------------|
| Cluster 0  | 9,8%               | homem, cabelo, olhando, jovem, camisa                |
| Cluster 1  | 9,5%               | mulher, cabelos, jovem, longos, vestindo             |
| Cluster 2  | 7,9%               | céu, sol, água, praia, pôr                           |
| Cluster 3  | 5,3%               | lyrics, songs, song, music, raindance                |
| Cluster 4  | 5,2%               | carro, bmw, car, rua, porsche                        |
| Cluster 5  | 3,8%               | aula, motivação, caderno, disciplina, escola         |
| Cluster 6  | 3,8%               | gato, cat, catsoftiktok, catcontent, rigbycat        |
| Cluster 7  | 3,7%               | anime, personagem, jujutsukaisen, azul, edit         |
| Cluster 8  | 3,2%               | futebol, neymar, jogador, campo, edit                |
| Cluster 9  | 2,9%               | cachorro, cachorros, dog, pet, filhote               |
| Cluster 10 | 2,5%               | chocolate, bolo, receita, leite, nutella             |
| Cluster 11 | 2,0%               | jesus, deus, evangelho, gospel, versiculo            |
| Cluster 12 | 1,8%               | frutas, tomate, historia, personagens, morango       |
| Cluster 13 | 1,6%               | dicas, creatorsearchinsights, viralizar, vídeos, tik |
| Cluster 14 | 1,4%               | bebê, criança, hospitalar, adulto, penteados         |
| Cluster 15 | 1,3%               | motocicleta, bicicleta, estrada, motorcycle, biker   |
| Cluster 16 | 1,2%               | reforma, construção, casa, obra, diy                 |
| Cluster 17 | 1,2%               | cama, deitado, dormir, bed, cobertor                 |
| Cluster 18 | 1,1%               | iphone, tela, dinheiro, smartphone, luck             |
| Cluster 19 | 1,0%               | castelo, darkfantasy, goldenbrown, pedra, gótico     |

nichos temáticos relevantes, a expansão do número de tópicos introduzia ruído analítico, gerando clusters redundantes ou sem coerência semântica. Desta forma, a escolha por 20 grupos demonstrou ser ideal, preservando a variabilidade temática do feed de entrada e assegurando que cada cluster mantivesse um conjunto de termos interpretáveis e distintos.

A aplicação desta arquitetura consolidou a vitrine inicial, provando que a plataforma constrói o feed do novo usuário a partir de um conjunto diverso de estímulos psicológicos de sucesso previamente validados. A estrutura detalhada na Tabela I sintetiza a taxonomia deste estágio de Cold Start, organizando os principais clusters identificados conforme sua representatividade no conjunto de vídeos.

O mapeamento provou que a plataforma atua de forma estruturada, construindo o feed do novo usuário com base em categorias de atenção cientificamente calibradas para maximizar a retenção [6]. Essa categorização gerada permitiu o agrupamento dos clusters nas seguintes macro-categorias de atenção:

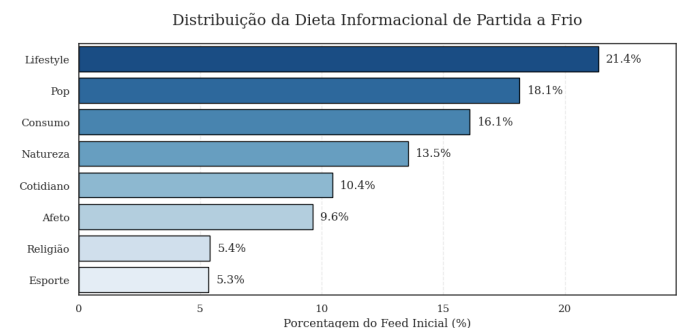


Figura 10. 20 principais clusters do Cold Start, agrupados em macro-categorias.

- **Lifestyle e Identidade (21,4%):** O maior bloco de retenção foca na autoimagem e no apelo estético visual, centralizado em conteúdos que exploram a identidade pessoal através da estética masculina e feminina.
- **Cultura Pop e Entretenimento (18,1%):** A plataforma investiga subculturas através de cliques musicais, lyrics, fandom de animes e estéticas digitais variadas, como a Dark Fantasy, que geram alto engajamento inicial.
- **Consumo e Status (16,1%):** Iscas direcionadas ao materialismo. Este grupo integra vídeos automobilísticos, motociclismo e conteúdos de ostentação tecnológica que apresentam estilos de vida desejáveis.
- **Natureza e Contemplação (13,5%):** Demonstra a recomendação de estímulos visuais de natureza, como praias e paisagens, utilizando o apelo contemplativo como âncoras de retenção para o feed de entrada.
- **Cotidiano e Relatabilidade (10,4%):** Exploração da identificação imediata, testando a rotina estudantil, motivação, tutoriais práticos de DIY e o compartilhamento de experiências íntimas do dia a dia.
- **Afeto e Conforto (9,6%):** Foco em gatilhos empáticos universais, liderados pelo conteúdo de animais de estimação (gatos e cachorros), acompanhados por culinária (comfort food) e conteúdo de cuidado e maternidade.
- **Religiosidade e Esportes (10,7%):** Segmentos ancorados em elementos culturais locais de alta conversão. Engloba o nicho religioso cristão/evangélico (5,4%) e o nicho de esportes de massa, como o futebol (5,3%), que carregam forte carga identitária.

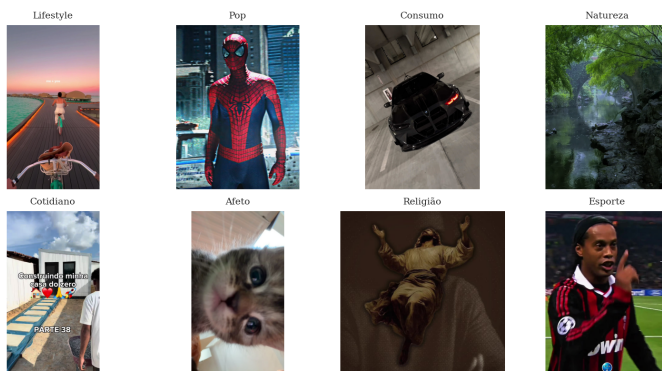


Figura 11. Exemplos visuais das macro-categorias de atenção identificadas no Cold Start.

## VI. CONCLUSÃO E TRABALHOS FUTUROS

Esta pesquisa investigou a mecânica de curadoria primária do TikTok, com o objetivo de decodificar a arquitetura de engajamento do estágio de Cold Start. Diante da transição da estratégia do Grafo Social para o Grafo de Conteúdo, levantou-se a hipótese de que a plataforma não poderia depender da exploração aleatória para reter novos usuários no cenário da Economia da Atenção. Os experimentos conduzidos comprovaram que a dieta informacional de boas-vindas é, na realidade, uma estrutura bem definida, dinâmica e calculada para maximizar a retenção passiva.

A partir do monitoramento simultâneo de 30 contas simuladas, o Experimento 1 refutou a hipótese de aleatoriedade no fornecimento inicial de conteúdo. O índice de similaridade de Jaccard revelou alta sobreposição entre os catálogos entregues, calibrada pelo contexto temporal (turnos de navegação). Mais importante, a análise ao longo da navegação demonstrou como o algoritmo soluciona o dilema entre Exploração e Exploração [8]. Ao longo da navegação, o sistema lentamente abandona a segurança dos conteúdos de altíssima viralidade e inicia a fase exploratória, marcada por uma queda estatisticamente significativa na sobreposição de vídeos e nas métricas globais de engajamento.

O Experimento 2, por sua vez, auxiliou na identificação da semântica deste conteúdo. A auditoria multimodal conduzida com o auxílio de Inteligência Artificial generativa, apontou que a curadoria primária se apoia em elementos cognitivos universais: a humanização do conteúdo (rostos) e o uso de textos sobrepostos para contextualização acelerada. Ao aplicar a modelagem de tópicos via BERTopic, a pesquisa revelou que o Cold Start do TikTok funciona como um retrato sociocultural. O feed pôde ser decomposto em 20 categorias que vão desde estética e lifestyle até subculturas e religião, permitindo que a IA mensure a atenção do usuário em diferentes nichos para construir o seu perfil e, com isso, recomende conteúdo aderente.

Além disso, a análise da emoção predominante trouxe uma importante sugestão acerca do conteúdo. Embora o feed inicial seja majoritariamente composto por vídeos inofensivos de expressões casuais, há engajamento predominante em conteúdos que evocam alta excitação. Isso reforça a literatura sobre o Rage Engagement e a economia da atenção [15], demonstrando que o algoritmo atua como um otimizador que reflete, escala e retroalimenta os padrões de consumo da própria sociedade [5].

Em suma, conclui-se que o sucesso do TikTok não é explicado apenas pela sua capacidade de personalizar conteúdo a longo prazo, mas na sua estratégia de engajamento nos primeiros minutos de uso. O aplicativo tenta capturar a atenção do novo usuário antes mesmo de conhecê-lo, utilizando-se do acúmulo estatístico do comportamento das massas para simular intimidade e relevância, consolidando a construção do "Eu Algoritmizado" [11].

## VII. AGRADECIMENTOS

A realização desta monografia só foi possível graças ao apoio de pessoas e instituições que caminharam ao meu lado. À minha orientadora, Professora Eliane Cristina de Freitas Rocha, agradeço pela confiança e pelo apoio ao longo do desenvolvimento desta pesquisa.

Aos meus pais e familiares, expresse minha profunda gratidão pelo apoio incondicional, tanto material quanto moral. Vocês foram o alicerce que me permitiu dedicar-me com foco à minha formação, sempre acreditando na minha trajetória pessoal, acadêmica e profissional. Um agradecimento especial aos meus sobrinhos que, ainda no início da caminhada, trouxeram tanta alegria e entusiasmo para o futuro e para as etapas que se aproximam. A confiança que todos vocês depositaram em

mim foi mais do que motivação para o fechamento deste ciclo universitário.

À minha namorada, Júlia, agradeço por ser a parceira que caminhou ao meu lado em todos os instantes. Sua dedicação, paciência e compreensão foram essenciais, transformando cada desafio em aprendizado e cada celebração em uma conquista ainda mais significativa. Iluminando o presente, sua companhia torna ainda mais empolgante o caminho que tenho pela frente e tudo o que ainda está por vir.

Por fim, agradeço aos amigos e colegas da UFMG. A convivência com vocês transformou este ciclo em uma experiência coletiva, e as alegrias compartilhadas tornaram esses anos uma experiência inesquecível.

## REFERÊNCIAS

- [1] Y. Zhang, J. Huang, M. Li, and J. Yang, “Contrastive graph learning for social recommendation,” *Frontiers in Physics*, vol. 10, p. 830805, 2022.
- [2] L. Nowacki, “The old king is dead, long live the algorithmic king – the decline of facebook and the rise of tiktok: comparative study of algorithmic design of social media platforms,” *International Social Science Journal*, 2026.
- [3] M. H. Goldhaber, “The attention economy and the net,” *First Monday*, vol. 2, no. 4, Apr 1997.
- [4] C. Celis Bueno, *The Attention Economy: Labour, Time, and Power in Cognitive Capitalism*. London, U.K.: Rowman & Littlefield International, 2017.
- [5] J. van Dijck, T. Poell, and M. de Waal, *The Platform Society: Public Values in a Connective World*. New York, NY, USA: Oxford University Press, 2018.
- [6] Z. Chen and C. Shi, “Analysis of algorithm recommendation mechanism of tiktok,” *International Journal of Education and Humanities*, vol. 4, no. 1, pp. 12–14, 2022.
- [7] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.
- [8] F. Ricci, L. Rokach, and B. Shapira, Eds., *Recommender Systems Handbook*, 3rd ed. New York, NY, USA: Springer, 2022.
- [9] J. van Dijck, “Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology,” *Surveillance & Society*, vol. 12, no. 2, pp. 197–208, 2014.
- [10] J. Cheney-Lippold, “A new algorithmic identity: Soft biopolitics and the modulation of control,” *Theory, Culture & Society*, vol. 28, no. 6, pp. 164–181, 2011.
- [11] A. Bhandari and S. Bimo, “Why’s everyone on tiktok now? the algorithmized self and the future of self-making on social media,” *Social Media + Society*, vol. 8, no. 1, 2022.
- [12] C. A. Johnson, *The Information Diet: A Case for Conscious Consumption*. Sebastopol, CA: O’Reilly Media, 2012.
- [13] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort, “Auditing algorithms: Research methods for detecting discrimination on internet platforms,” *Data and Discrimination: A Research Symposium*, 2014.
- [14] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*. Cambridge University Press, 2020.
- [15] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel, “Emotion shapes the diffusion of moralized content in social networks,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 114, no. 28, 2017.
- [16] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based tf-idf procedure,” *arXiv preprint arXiv:2203.05794*, 2022.