

# Technical Appendix

This technical document summarizes the methodology, data sources and analytical outputs used across Canary programs. It is intended as a reference for monitoring, reporting and knowledge-sharing within the Canary road map. The Canary system is a digital marketing monitoring and reporting framework developed by Vital Strategies to track and analyze online marketing for harmful products, including tobacco, alcohol and ultra-processed products.

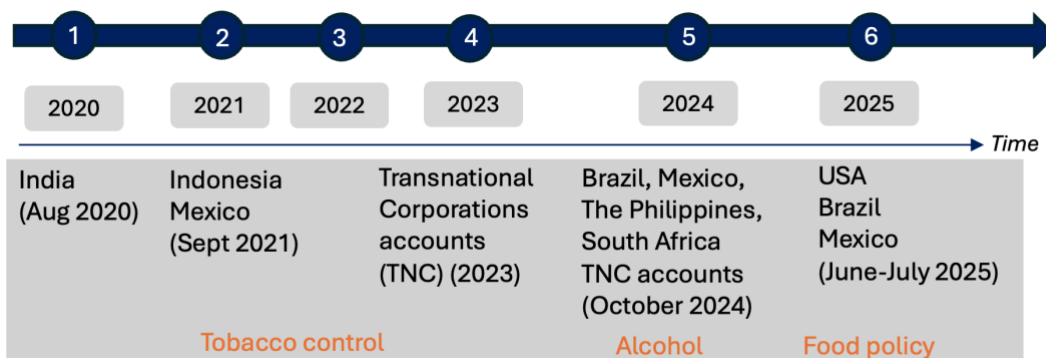
By integrating insights from real-time monitoring with structured data analysis, the service provides indicative evidence on the scale, reach, and narrative strategies employed in digital spaces by industries that manufacture and market these products.

## Canary System

The Canary system systematically monitors, classifies and analyzes online marketing content for harmful products across multiple countries. Formerly known as TERM (Tobacco Enforcement and Reporting Movement), Vital Strategies' Canary service has tracked digital marketing of harmful commercial products since 2020. It covers tobacco, alcohol and ultra-processed products (Figure 1), with the goal of mapping dominant marketing tactics, message framing strategies and brand presence. Accounts were drawn from global corporate brands, national brands and subsidiary accounts to provide broad coverage of observed marketing practices.

Canary has expanded over time, in both product scope and geographic coverage (Figure 1).

**Figure 1: Program-specific timeline and country inclusion**



## Methodology

Canary uses a hybrid approach that integrates human expertise with artificial intelligence or AI-assisted content classification, implemented through a standardized, multi-phase process encompassing (Table 1):

### 1. Systematic account selection

The account identification process follows a structured, iterative process to capture comprehensive coverage of the most relevant and active sources of digital marketing content.

Each program begins with a defined “brand universe” of major product-specific companies and brands operating in each participating country. This list is developed through consultation with subject matter experts and local country-based research partners and validated through the review of national market share and industry intelligence reports (e.g., Euromonitor International). This approach ensures that the selected brands accurately reflected dominant producers and distributors across products (including electronic cigarette and nicotine products, sugary drinks, or types of alcoholic beverages).

Once the brand universe is established, official/commercial and active social media accounts associated with

these brands are identified across Facebook, Instagram, X (formerly Twitter) and YouTube. TikTok is included despite platform restrictions that limit third-party access to public content and despite lower levels of activity among some commercial accounts. Accounts are verified for authenticity—that the account represents an official or brand-linked profile—and assessed based on activity and engagement (e.g., posting frequency, follower counts). Priority is given to verified brand accounts, corporate parent profiles, third-party retailers or sponsorship pages. All accounts are validated with country teams so they are aligned with national product markets and the Canary codebook.

These accounts are configured within a Software as a Service (SaaS)-based consumer and social media analytics platform to enable continuous and automated data collection of all publicly available content.

Keyword-based monitoring: To complement account-based tracking and capture broader harmful product-related marketing content, Boolean keyword and hashtag searches are conducted using structured queries developed in collaboration with national experts. These searches included:

- Generic product terms (e.g., “beer,” “vodka,” “e-cigarette,” “vape,” “energy drink”)
- Brand-specific terms (e.g., “#Coca-Cola,” “#Heineken,” “#Marlboro,” “#BaileysRecipes”, “#Brahma,” “#JackDaniels,” “#JohnnieWalker,” “#CoronaSunsets”)

Boolean operators (AND, OR, NEAR, EXCLUDE) are used to refine searches and reduce false positives, increasing the precision of content identification.

All accounts and keyword search terms are systematically documented for each program and country to ensure transparency and replicability. To maintain relevance and adapt to evolving marketing dynamics, the Canary research team in collaboration with the contracted agency conducts continuous reviews of all configured accounts and keyword queries. This continuous review process enables the Canary system to remain responsive to new product launches, emerging digital platforms, and shifting promotional strategies within the industry. A complete list of monitored accounts can be provided upon request.

## 2. Automated data collection

Data are collected using the Synthesio platform, operated by IPSOS CRED under contract with Vital Strategies. Synthesio is an AI-assisted consumer and social media analytics platform for monitoring digital media content, including social media and online news, with a particular focus on identifying indirect marketing through public relations and media stories. The Synthesio platform captures data daily from social media, online news sites, and video- or photo-sharing platforms. Each captured post or news article includes the original URL or link, relevant text excerpts from conversations or articles, and any associated visual content, such as photos or videos, that may be pertinent for subsequent analysis.

The dataset includes all publicly accessible, organic marketing posts published by configured accounts, as well as posts identified through keyword-based discovery that directly or indirectly promote conventional or new products during the study period. Because of platform restrictions where third-party tools cannot access or extract content from accounts not directly tracked or authorized within the system, the analysis includes only original posts from accounts configured in the system.

For better linguistic and contextual relevance, only posts written in English or in the commonly spoken languages of each participating country are retained. Each post consists of an image or a video, often accompanied by textual content, hashtags or emojis. Duplicates, posts unrelated to alcoholic products, and posts that are inactive or deleted at the time of data cleaning were systematically removed.

## 3. Coding and classification

All posts were coded using the Canary coding framework, which includes type of account, type of product and two additional principal dimensions:

- Marketing tactic based on the mode of brand promotion (e.g., direct ads, price promotion, events and sponsorships, corporate social responsibility, cobranding, community-based marketing, etc.).
- Message framing based on the narrative used to shape audience perception (e.g., glamorization,

informational, entertaining, health claims, secular community celebrations, personal care and wellness, environment-eco-awareness, etc.).

Marketing tactic and message framing are coded through a hybrid human–AI workflow designed to improve efficiency while maintaining accuracy:

*Step 1*—Automated pre-classification: Posts are automatically categorized using the keyword classifiers from the Synthesio platform, designed and supervised by the Canary researchers.

*Step 2*—Human quality control: Trained analysts from the Vital Strategies Canary team review outputs, applying corrections and ensuring label precision, especially in ambiguous or context-specific posts.

*Step 3*—Consensus validation: Discrepancies identified during quality control are discussed in internal review meetings and resolved using pre-defined decision-tree rules documented in the Canary codebook, ensuring methodological consistency across countries.

Quality control is applied at all stages of data processing and coding for accuracy, consistency and reliability.

- Automated validation scripts developed in Python and deployed in Hugging Face were used to identify structural inconsistencies including missing values, duplicate posts and invalid field entries.
- Detected errors were reviewed and corrected before analysis to maintain data integrity.
- In addition to automated checks, a random sample of 10%-20% of posts was manually re-coded by senior analysts to assess inter-coder reliability and evaluate alignment with the established classification syntax.
- Posts with unresolved discrepancies are subsequently reviewed through structured consensus sessions, where analysts apply predefined decision rules to reach agreement.

Insights from these discussions are used to refine the coding syntax and update keyword dictionaries, establishing an iterative feedback loop that continuously improved data quality and enhanced the accuracy of subsequent monitoring cycles.

#### **4. Data analysis and visualization**

Findings are presented through dashboard (Power BI) country-level situation reports and cross-market summaries to inform advocacy, policy engagement, and future monitoring activities. Analysis covers the following dimensions:

- Descriptive analysis:
  - Overall volume of posts and engagement level across countries
  - Frequency and distribution of posts by social media platforms, product brands and types of accounts
- Content and Strategy Analysis
  - Distribution of marketing tactics and message framings by country, platform, and account type
  - Identification of key promotional strategies, and recurring themes in product promotion
- Comparative analysis
  - Cross-country comparisons highlighting common global patterns and context-specific marketing strategies

No statistical weighting or significance testing is applied; findings summarize observed patterns within the monitored dataset.

Table 1: Canary’s core functions and key activities

Module	Core function	Key activities
Module 1: Systematic account selection	Identifies and defines relevant social media accounts for monitoring	Account mapping, keyword and brand search
Module 2: Automated data collection	Acquire digital marketing content systematically through a SaaS-based consumer and social media analytics platform	Application programming interface (API) access, scraping, metadata extraction, quality verification
Module 3: Data classification	Categorize collected content according to standardized codebook	Manual coding, automated classification: SaaS based syntax and large language model (LLM based), quality control and accuracy validation
Module 4: Reporting and dissemination	Generate analytical insights, visualizations (interactive dashboard), and public-facing reports and policy briefs	Dashboard and data updates, internal review, policy translation/insights, report publication, using human insights and generative AI

All analyses provide a descriptive snapshot of publicly available social media marketing activity and do not estimate total market activity or population exposure. Because data are drawn from selected brand accounts and a defined monitoring window, the results should be interpreted as descriptive indicators of marketing presence and framing, rather than as inferential statistics or population-level estimates.

### 5. Ethical considerations

Only publicly available posts are collected and analyzed; no private or user-generated data are accessed at any stage. All analyses comply with Vital Strategies’ data ethics and privacy standards and were conducted in full compliance with platform terms of service.

### 6. Replicability and continuous improvement

Canary’s methodology is designed to be scalable and replicable across different countries and product categories. Its flexible framework allows consistent application while accommodating local contexts and data environments. Continuous refinement is achieved through structured feedback loops with country partners, ensuring the system evolves in response to emerging marketing practices, platform innovations, and analytical needs.

## 7. Limitations

Canary collects and analyzes illustrative examples of harmful product marketing to identify trends and patterns over time. It is not intended to provide a census of all industry activity. All analyses are based on a specific time frame of monitoring data and should be interpreted as an exploratory, descriptive assessment of relative product marketing activity. All findings illustrate observable patterns within the reference period but are not intended for inferential or population-level generalization.

Data were obtained from publicly available social media content through predefined brand accounts and keyword-driven searches on selected platforms. Posts from private profiles, paid advertisements, and influencer-generated content not accessible via public application programming interface (APIs) were not included, which may result in partial coverage of marketing activity. In fact, published content through external or partner accounts (e.g., event pages, media collaborators or influencer pages) rather than directly from the brand's main account is not captured by the Synthesio platform.

By configuration, Canary captures only original posts authored by verified or commercial brand, influencer, news, and magazine accounts. Content generated by individual users - including tagged posts, comments, reposts, and quote reposts - is intentionally excluded, as these constitute secondary engagement or amplification of marketing material rather than primary promotional content. Influencer and commercial accounts that promote products but do not meet the predefined inclusion criteria are not captured by the system.

As product industries and companies increasingly route promotion through such personal or lifestyle accounts, the dataset likely underestimates the full scale and reach of digital harmful product marketing. Moreover, platform-specific access restrictions -- particularly for platforms such as TikTok may have further influenced data completeness and representativeness.

Some content may be missed due to untracked keywords or languages or posts which are image- or video-based without any captions or keywords. Variations in overall marketing activity may reflect short-term promotional efforts by a few highly active brands rather than sustained trends, while others remain less engaged during the same period.

### About Canary



Canary is a real-time media monitoring and reporting service that harnesses AI tools alongside expert analysis to identify emerging threats in the marketing of dangerous substances.

### About Vital Strategies

Vital Strategies is a global health organization that believes every person should be protected by equitable and effective public health systems. We partner with governments, communities and organizations to reimagine public health, and the result is millions of people living longer, healthier lives. Our goal is to build a future where better health is supported across all facets of our lives, in our families, communities, in our environment and our governments.

## Join the movement



To learn more visit [alertsbycanary.org](https://alertsbycanary.org)  
or follow us  [@alertsbycanary](https://www.instagram.com/alertsbycanary)  
 and [alertsbycanary](https://twitter.com/alertsbycanary)