THE CONTENT ENGINE

SUBAGENT 1

Awareness Content

Automatically scrape buyer questions from Reddit, Quora, and Google's People Also Ask — then generate blog posts, LinkedIn carousels, and video scripts that position your brand as the answer.

A step-by-step technical implementation guide.

What This Subagent Does

The Awareness subagent answers the questions your ideal buyers are already asking — before they know your product exists. It automates three things: discovering what your market is asking, filtering for relevance to your product, and generating multi-format content that positions your brand as the expert answer.

The Workflow at a Glance

1. Firecrawl (or SerpAPI) scrapes forums and People Also Ask results
2. Claude categorizes questions by topic cluster and search intent
3. Claude writes content matched to format: blog post, carousel, video script
4. n8n routes drafts to your content calendar for review and publishing

What You Need Before Starting

Required Accounts & API Keys

You will need active accounts and API keys for each of the following services. Most offer free tiers sufficient for initial setup and testing.

Tool	Purpose	Notes
n8n	Workflow orchestration	Self-hosted (free) or n8n Cloud. This is the backbone that connects everything.
Claude API	Content generation	Anthropic API key. Sonnet is cost-effective for volume; Opus for higher quality.
Firecrawl	Web scraping	Scrapes Reddit, Quora threads. Free tier: 500 pages/month. firecrawl.dev
SerpAPI	Google PAA scraping	Returns structured People Also Ask data. serpapi.com. Optional alternative to Firecrawl for PAA.
Google Sheets	Content calendar	Where all drafts land for review. Airtable or Notion work as alternatives.

Estimated Costs

At the API level, running this subagent weekly on 20–30 seed keywords costs roughly $30–60/month. Firecrawl's free tier covers early usage. n8n self-hosted is free. The primary ongoing cost is the Claude API for content generation.

Step 1: Define Your Inputs

Before building anything, you need three things documented. These will be referenced throughout every prompt and workflow node.

1A. Seed Keywords (10–20)

These are the core terms your ideal buyer would search when experiencing the problem your product solves. Not your product name or category — the actual language they use when looking for help.

Example

If you sell project management software to agencies:

- "how to manage client projects"
- "agency project tracking"
- "client deliverable deadlines"
- "scope creep prevention"

If you sell compliance software to fintech:

- "SOC 2 audit prep"
- "fintech regulatory requirements"
- "compliance automation"

1B. ICP Pain Points (3–5)

Write out the specific pain points your product addresses. Claude will use these to filter scraped questions for relevance. Be specific — not "saves time" but "manually reconciling data across 3+ tools wastes 10 hours/week."

1C. Product Positioning

One paragraph describing what your product does, who it’s for, and the core point of view your brand holds. This prevents Claude from writing generic content. It should include your main differentiator and the belief system that drives your product.

Step 2: Build the Scraping Layer

This step sets up automated data collection from three sources: Reddit, Quora, and Google People Also Ask. Each source captures different types of buyer questions.

2A. Scraping Reddit with Firecrawl

Reddit threads contain unfiltered buyer language — the exact phrasing your ICP uses when describing their problems to peers. Firecrawl's API can extract thread content from subreddit URLs.

n8n Workflow Nodes

1. **Schedule Trigger** — Set to run weekly (e.g., every Monday at 6am). This keeps your question pipeline fresh without overwhelming your content calendar.
2. **HTTP Request Node (Firecrawl)** — POST request to Firecrawl's scrape endpoint. Configure one node per subreddit, or use a loop to iterate through a list.

Firecrawl API Call

```
POST https://api.firecrawl.dev/v1/scrape

Headers:
  Authorization: Bearer {{$credentials.firecrawlApi}}
  Content-Type: application/json

Body:
{
  "url":
"https://www.reddit.com/r/YOUR_SUBREDDIT/search?q=KEYWORD&sort=relevance&t=month",
  "formats": ["markdown"],
  "onlyMainContent": true
}
```

Finding the Right Subreddits

Search Reddit directly for your seed keywords and note which subreddits appear most. You want 3–5 active subreddits where your ICP asks questions and discusses problems. Look for subreddits with 10k+ members and daily post activity.

Pro Tip: Deduplication

Add a Code node in n8n after scraping that hashes each question (use a simple MD5 or SHA-256 of the question text) and checks it against a Google Sheet of previously processed questions. This prevents generating duplicate content for the same question across weekly runs.

2B. Scraping Google People Also Ask with SerpAPI

People Also Ask boxes reveal the exact questions Google considers most relevant to a search term. These are high-intent questions that often map directly to bottom-of-funnel blog content.

n8n Workflow Nodes

1. **Loop Node** — Iterates through your list of 10–20 seed keywords.
2. **HTTP Request Node (SerpAPI)** — GET request for each keyword. SerpAPI returns structured PAA data, so no parsing is needed.

SerpAPI Call

```
GET https://serpapi.com/search.json

Query Parameters:
  engine: google
  q: {{$json.keyword}}
  api_key: {{$credentials.serpApi}}

Response (relevant section):
{
  "related_questions": [
    {
      "question": "How do I automate client reporting?",
      "snippet": "...",
      "link": "..."
    }
  ]
}
```

Extract the "related_questions" array from the response. Each object contains the question text, a snippet, and the source URL. You only need the question text for the next step.

2C. Scraping Quora with Firecrawl

Quora captures a different question format — typically longer, more specific, and often framed as "what is the best way to..." questions. These map well to comparison and how-to content.

Firecrawl API Call

```
POST https://api.firecrawl.dev/v1/scrape
```

```
Body:
{
  "url": "https://www.quora.com/search?q=YOUR_KEYWORD",
  "formats": ["markdown"],
  "onlyMainContent": true
}
```

The response will include question titles and excerpts from top answers. Extract the question titles — these become your content prompts in Step 3.

Step 3: Filter and Classify with Claude

Raw scraped data is noisy. Most questions won't be relevant to your product. This step sends all scraped questions to Claude for relevance scoring, deduplication, and topic clustering before any content is generated.

3A. The Filtering Prompt

Add an HTTP Request node in n8n that calls the Anthropic Messages API. Use this system prompt:

```
System Prompt:
You are a content strategist for a B2B SaaS company.

Product: [YOUR PRODUCT DESCRIPTION]
ICP: [YOUR IDEAL CUSTOMER PROFILE]
Core pain points we address:
  1. [PAIN POINT 1]
  2. [PAIN POINT 2]
  3. [PAIN POINT 3]

User Message:
Here is a raw list of questions scraped from Reddit, Quora, and
Google People Also Ask:

{{$json.allQuestions}}

For each question:
1. Score relevance 1-5 based on how directly it maps to our pain points
2. Deduplicate — group near-identical questions and pick the best phrasing
3. Classify by topic cluster (group related questions together)
4. Classify search intent: informational, navigational, or commercial

Output JSON:
{
  "clusters": [
    {
      "topic": "cluster name",
      "questions": [
        {
          "question": "the question",
          "relevance_score": 4,
          "intent": "informational",
          "source": "reddit|quora|paa"
        }
      ]
    }
  ]
```

```
}

Only include questions with relevance score >= 3.
Return valid JSON only. No commentary.
```

3B. The API Call in n8n

```
POST https://api.anthropic.com/v1/messages

Headers:
  x-api-key: {{$credentials.anthropicApi}}
  anthropic-version: 2023-06-01
  Content-Type: application/json

Body:
{
  "model": "claude-sonnet-4-20250514",
  "max_tokens": 4096,
  "messages": [
    {
      "role": "user",
      "content": "[THE FULL PROMPT ABOVE WITH VARIABLES FILLED IN]"
    }
  ]
}
```

3C. Parse the Response

Add a Code node after the Claude HTTP request to extract and parse the JSON:

```
// n8n Code Node (JavaScript)
const response = $input.first().json;
const text = response.content[0].text;

// Strip markdown code fences if present
const clean = text.replace(/```json|```/g, '').trim();
const parsed = JSON.parse(clean);

// Flatten to individual questions for the next step
const questions = parsed.clusters.flatMap(cluster =>
  cluster.questions.map(q => ({
    ...q,
    topic_cluster: cluster.topic
  }))
);
```

```
return questions.map(q => ({ json: q }));
```

Step 4: Generate Multi-Format Content

This is the core content generation step. For each qualifying question, Claude produces three content formats: a 600-word blog post, a 7-slide LinkedIn carousel, and a 60-second video script.

4A. The Content Generation Prompt

This prompt runs once per question (or in batches of 3–5 questions to reduce API calls). Use Claude Sonnet for volume runs or Opus for higher-quality output on priority topics.

```
System Prompt:
You are a content strategist for a B2B SaaS company.

Product: [YOUR PRODUCT DESCRIPTION]
ICP: [YOUR IDEAL CUSTOMER PROFILE]
Brand POV: [YOUR CORE BELIEF / DIFFERENTIATOR]

User Message:
Write content for this question: "{{$json.question}}"
Topic cluster: {{$json.topic_cluster}}
Search intent: {{$json.intent}}

Generate all three formats below:

FORMAT 1 — BLOG POST (600 words)
- Answer the question with tactical depth in the first 100 words
- Include specific, actionable steps or frameworks (not generic advice)
- Weave in our POV naturally without being salesy
- End with a bridge to our solution that feels earned, not forced
- Include a meta title (under 60 chars) and meta description (under 155 chars)

FORMAT 2 — LINKEDIN CAROUSEL (7 slides)
- Slide 1: Hook headline (pattern interrupt or bold claim)
- Slides 2-6: One key point per slide, 2-3 sentences max
- Slide 7: CTA slide
- Each slide should stand alone but build a narrative arc

FORMAT 3 — VIDEO SCRIPT (60 seconds)
- Hook (first 5 seconds): state the problem
- Body (40 seconds): deliver the insight with one concrete example
- CTA (15 seconds): what to do next
- Written for spoken delivery — short sentences, conversational

Tone: direct, opinionated, zero filler. Write like a practitioner, not a marketer.

Output as JSON:
```

```
{
  "question": "...",
  "blog": {
    "meta_title": "...",
    "meta_description": "...",
    "body": "... (markdown)"
  },
  "carousel": {
    "slides": ["slide 1 text", "slide 2 text", ...]
  },
  "video_script": "..."
}
```

4B. Batching for Efficiency

Sending one API call per question works but is slow and expensive. Instead, batch 3–5 questions per API call by wrapping them in an array in the user message. Adjust the prompt to say "For each question in this list" and return an array of content objects.

In n8n, use a SplitInBatches node set to batch size 5 before the Claude HTTP Request node. After the response, use another Code node to flatten the array back into individual items.

4C. Model Selection

Use Case	Recommended Model	Why
Weekly volume runs	Claude Sonnet	Fast, cost-effective, good quality for awareness content
Priority topics	Claude Opus	Higher quality reasoning and more nuanced positioning
Filtering (Step 3)	Claude Sonnet	Classification tasks don’t need max intelligence

Step 5: Route to Your Content Calendar

The final step writes all generated content to a structured content calendar where you can review, edit, and approve before publishing.

5A. Google Sheets Schema

Create a Google Sheet (or Airtable base) with the following columns. This becomes your single source of truth for all awareness content.

Column	Type	Description
Date Created	Date	Auto-populated by n8n
Question	Text	The original scraped question
Topic Cluster	Text	From Claude’s classification in Step 3
Search Intent	Dropdown	Informational / Commercial / Navigational
Relevance Score	Number	1–5 from Claude’s filtering
Source	Dropdown	Reddit / Quora / PAA
Blog Draft	Long text	The 600-word blog post
Meta Title	Text	SEO title (under 60 chars)
Meta Description	Text	SEO description (under 155 chars)
Carousel Draft	Long text	7-slide carousel content
Video Script	Long text	60-second video script
Status	Dropdown	Draft / In Review / Approved / Published
Assigned To	Text	Who’s reviewing this piece
Publish Date	Date	Target publication date

5B. n8n Google Sheets Node

Use n8n’s built-in Google Sheets node ("Append Row" operation) to write each content piece as a new row. Map the JSON output from Step 4 to the corresponding columns.

Connect your Google account to n8n via OAuth, select the spreadsheet and sheet tab, then map fields:

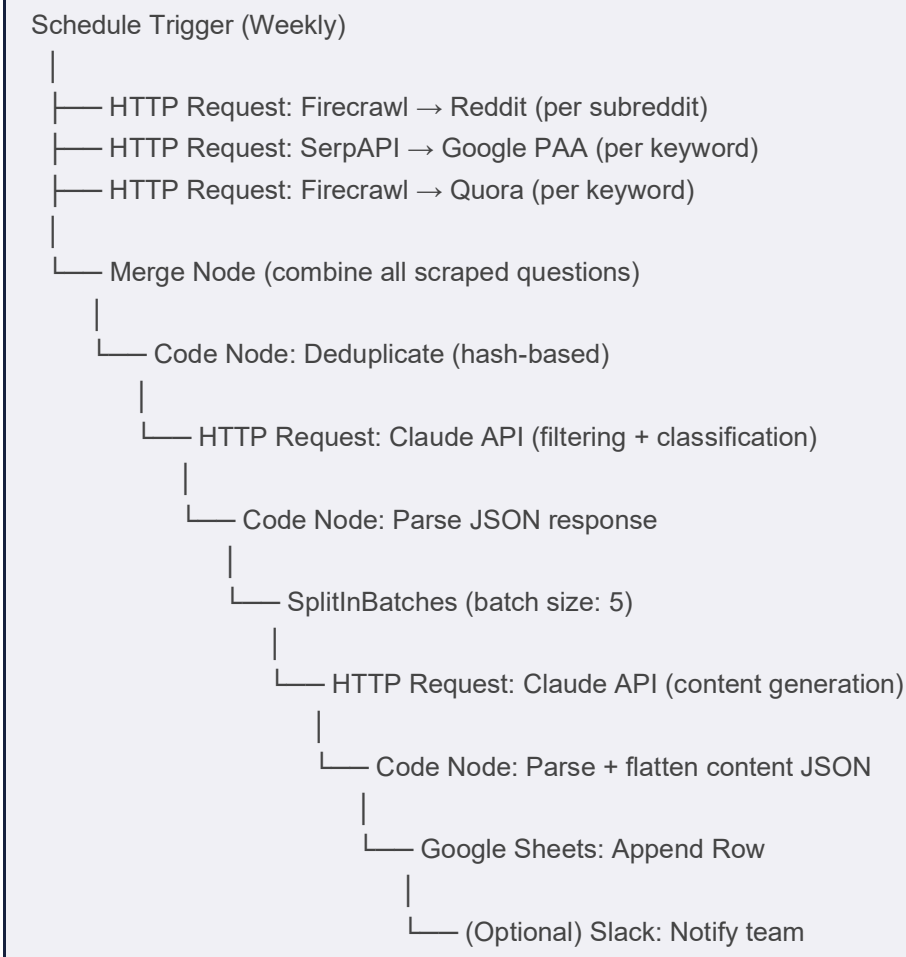
```
Field Mapping (n8n Google Sheets Node):

Date Created    →  {{ $now.toISODate() }}
Question        →  {{ $json.question }}
Topic Cluster   →  {{ $json.topic_cluster }}
Search Intent   →  {{ $json.intent }}
Relevance Score →  {{ $json.relevance_score }}
Source          →  {{ $json.source }}
Blog Draft      →  {{ $json.blog.body }}
Meta Title      →  {{ $json.blog.meta_title }}
Meta Description→  {{ $json.blog.meta_description }}
Carousel Draft  →  {{ $json.carousel.slides.join('\n\n') }}
Video Script    →  {{ $json.video_script }}
Status          →  Draft
```

Step 6: The Complete n8n Workflow

Here is the full workflow assembled end-to-end. Each box represents an n8n node. Build this as a single workflow with the Schedule Trigger as the entry point.

Workflow Architecture



Node Count and Estimated Build Time

Metric	Value
Total n8n nodes	10–14 (depending on subreddit/keyword count)
Build time (first time)	2–4 hours
Build time (experienced)	45–90 minutes
Weekly run time	5–15 minutes (API dependent)

Output per run	15–40 content pieces across 3 formats

Step 7: Human-in-the-Loop Review

This is not optional. Automated content generation produces drafts, not published pieces. Every piece needs a human pass before going live. Here's the review workflow:

1. **Voice check.** Does this sound like your brand, or does it sound like AI? Edit for your specific tone, add personal anecdotes or customer examples that Claude doesn't have access to.
2. **Accuracy check.** Verify any claims, statistics, or technical details. Claude will occasionally hallucinate specifics — particularly around competitor capabilities or industry data.
3. **Bridge check.** Is the connection from the content to your product earned and natural? If it feels forced, rewrite the last 1–2 paragraphs. The best awareness content builds trust without making the reader feel sold to.
4. **SEO check.** Review the meta title and description. Make sure the target keyword appears naturally. Add internal links to related content on your site.
5. **Format check.** Carousel slides should each make one point clearly. Video scripts should sound natural when read aloud — read them out loud before approving.

> **Time Investment**
>
> Plan for 10–15 minutes per content piece for review and editing. If your weekly run generates 20 pieces, that's roughly 3–5 hours of editorial work per week. This is the trade-off: the engine handles 80% of the work, you handle the 20% that requires judgment and brand knowledge.

Troubleshooting Common Issues

Problem	Solution
Claude returns invalid JSON	Add "Return valid JSON only. No markdown fences, no commentary." to the end of your prompt. In your Code node, strip ```json and ``` fences before parsing.
Too many irrelevant questions pass filtering	Raise the relevance threshold from 3 to 4. Make your pain point descriptions more specific in the system prompt.
Content sounds generic	Add 2–3 examples of your brand's writing style to the system prompt. Include a "do not" list: no jargon, no corporate speak, no filler phrases.
Firecrawl returns empty results	Some subreddits block scraping. Try different subreddits or use Reddit's own API (requires an app registration at reddit.com/prefs/apps).
Duplicate content across weeks	Implement the hash-based deduplication from Step 2A. Store all processed question hashes in a persistent Google Sheet.
API rate limits	Add Wait nodes (5–10 seconds) between batched API calls. SerpAPI has 100 searches/month on the free tier — upgrade if needed.
Carousel text is too long per slide	Add a constraint to the prompt: "Each carousel slide must be under 40 words." Claude follows explicit word counts well.

What's Next

This subagent handles the top of your funnel — capturing demand that already exists by answering the questions your buyers are actively searching for.

In Subagent 2 (Interest Content), you'll build a trend detection system that monitors industry conversations in real time and generates POV content that inserts your brand into the topics your buyers are already discussing.

Up next: Subagent 2 — Interest Content (Trend Detection → POV Takes)