

THE COMPLETE PLAYBOOK

# Scrape Any Website In Two Clicks With **Thunderbit**

The AI Chrome extension that turns any website into a structured table — leads, content ideas, product research, competitor data — and feeds it straight into Claude for the part where it becomes useful.

USE IT YOURSELF · OR SELL IT AS A SERVICE

---

Setup · Three workflows · Prompt playbook · Claude integration · 7-day plan

# What's Inside

A working playbook for turning any public website into structured business data — and feeding that data into Claude to do the part that actually moves the needle. Either for your own business, or as a service you sell to clients.

**01** The Day-Long Job You Can Now Do in 2 Clicks

---

**02** What Thunderbit Actually Is

---

**03** Install in 2 Minutes

---

**04** The 2-Click Workflow

---

**05** The Killer Feature: Subpage Scraping

---

**06** Workflow 1: Content Research from LinkedIn

---

**07** Workflow 2: A Lead List That Fills Itself

---

**08** Workflow 3: Product Gap Research

---

**09** Connecting Thunderbit to Claude

---

**10** The Full Prompt Playbook

---

**11** Pre-Built Templates: The 1-Click Library

---

**12** AI Data Enhancement Beyond Scraping

---

**13** Pricing & Plans

---

**14** The Service Angle: Sell Lead Lists, Research, and Monitoring

---

**15** Power Moves: 12 Non-Obvious Tactics

---

**16** Your 7-Day Mastery Plan

---

**17** Pre-Scrape Checklist

---

# The Day-Long Job You Can Now Do in 2 Clicks

---

Finding content ideas, qualified leads, or profitable product ideas used to take days. Going page by page, copying everything into a spreadsheet by hand — that's hours, every time. The proof of what works is already public. You just couldn't get to it without burning your week.

Content with high engagement is sitting on LinkedIn, Twitter, and YouTube right now. Products with high review counts and ratings are listed on Amazon, Etsy, and Gumroad. Businesses you could sell to are visible on Yelp, Google Maps, and every industry directory ever built. Their reviews prove they're active and real.

The data exists. The problem was extraction — going one page at a time, copy-pasting into spreadsheets, then trying to make sense of the mess. Most people just gave up and worked from gut feeling instead.

This is the gap Thunderbit closes. Click once, AI looks at whatever page you're on and figures out how to organize the data into a table. Click again, it scrapes the whole thing. Then you hand the table to Claude, and Claude turns the raw data into content drafts, qualified leads with contact info, or a market gap analysis you can build a product around.

**2**

CLICKS TO SCRAPE

**200k+**

USERS WORLDWIDE

**40+**

1-CLICK TEMPLATES

**Free**

FOREVER TIER

## What this playbook actually delivers

By the end you'll know how to install Thunderbit, run a 2-click scrape on any page, use Subpage Scraping to enrich every row with extra data, export to Google Sheets, and hand the result to Claude with prompts that turn raw data into ready-to-use content, leads, and product ideas.

You'll also see the service angle — how the same workflow lets you charge real money offering lead generation, market research, and competitor monitoring to clients.

# What Thunderbit Actually Is

---

Thunderbit is an AI-powered Chrome extension that turns any website into a structured table — with no code, no CSS selectors, no Python scripts, and no configuration. It's been ranked #1 Product of the Week on Product Hunt and is trusted by 200,000+ users including teams at Harvard, BCG, Adidas, Red Bull, Wix, MIT, and Patagonia.

The core mechanic is what makes it different from every previous-generation scraper. Where old tools made you specify CSS selectors, write XPath, or drag-and-drop fields, Thunderbit just looks at the page with AI and figures out the structure for you. You click "AI Suggest Columns" and it returns a sensible table layout — names, prices, dates, ratings, links — based on what's actually on the page.

## What Thunderbit can pull

- **Text** — headlines, descriptions, reviews, articles, long-form content
- **Links and URLs** — for following up or enriching with Subpage Scraping
- **Images** — product photos, profile pictures, thumbnails
- **Contact info** — emails and phone numbers visible on the page
- **Numbers** — prices, ratings, view counts, follower counts, review counts
- **Dates** — post dates, publication dates, event dates
- **Structured listings** — directories, search results, marketplace pages

Thunderbit works on sites behind a login (when you're logged in) and on public sites alike. It handles pagination, supports infinite-scroll pages, and can even work with screenshots and uploaded files when a page is hard to capture cleanly.

## Why pair it with Claude?

Thunderbit gives you the raw material. Claude turns the raw material into something useful — content drafts, enriched lead lists, market analyses, product positioning. Used together, you get a complete workflow that used to require a researcher, a writer, and a virtual assistant.

### THE MENTAL MODEL

Think of Thunderbit as the hands that go grab the data. Claude is the brain that figures out what to do with it. Each one is useful alone, but the leverage is in chaining them: Thunderbit extracts, Google Sheets stores, Claude analyzes and acts. One repeatable system across leads, content, and research.

## Install in 2 Minutes

---

Thunderbit is a Chrome extension. Free to install, no account required to start scraping. The entire setup takes two minutes.

### 1 Go to [thunderbit.com](https://thunderbit.com)

Open [thunderbit.com](https://thunderbit.com) and click the "Add to Chrome" button. You can also search "Thunderbit" in the Chrome Web Store directly.

### 2 Add to Chrome and approve permissions

Chrome will ask for the standard extension permissions. Approve them. Thunderbit needs page access to read content for the AI to structure.

### 3 Pin it to your toolbar

Click the puzzle-piece extensions icon in your Chrome toolbar, find Thunderbit, and click the pin icon. This puts the Thunderbit button in your browser permanently, one click away on every page.

### 4 Verify it works

Open any list-style page — Amazon search results, a directory, LinkedIn search — and click the Thunderbit icon. The "AI Web Scraper" panel should open on the side. If it does, you're ready.

#### NO ACCOUNT REQUIRED TO TEST

You don't need to sign up to try the basic 2-click workflow. The free tier is free forever — sign up later only when you hit the limits, or when you want export and AI features the paid plans unlock.

# The 2-Click Workflow

---

The core Thunderbit workflow is two buttons. Once you've used it on three different pages, you'll understand why it eliminates 90% of manual data collection.

## 1 Navigate to the page you want data from

Could be anything: a LinkedIn search result, a Yelp directory page, a Gumroad category, a Reddit post list, an Amazon search. Whatever has the data you want.

## 2 Open Thunderbit

Click the Thunderbit icon in your toolbar. The AI Web Scraper panel opens.

## 3 Click "AI Suggest Columns" (click 1)

Thunderbit's AI reads the page and suggests a column structure — name, price, rating, link, date, whatever the page contains. You can edit the column names, add your own, or remove ones you don't need. Most of the time the default suggestion is exactly right.

## 4 Click "Scrape" (click 2)

Thunderbit extracts every row on the page that matches the column structure. Watch the table populate. That's it — the page is now a structured table you can use.

## 5 Export to Google Sheets, Notion, Airtable, or Excel

Pick your destination. Thunderbit pushes the table into your tool of choice, ready to share, filter, sort, or hand to Claude.

### THE HIDDEN SUPERPOWER

You can *edit the suggested columns before scraping* to add columns Thunderbit didn't think of — like "category I want to assign" or "notes." Define them in plain English ("AI-generated summary of this listing") and Thunderbit will use AI to fill them in. This turns scraping into structuring in one pass.

# The Killer Feature: Subpage Scraping

---

If you only learn one Thunderbit feature beyond the basic 2-click workflow, learn this one. Subpage Scraping is the difference between "I have a list of names" and "I have a fully enriched lead database with details from every individual page."

## What it does

A standard scrape only pulls what's visible on the page you're on. So if you scrape a Yelp directory of 30 businesses, you get 30 names, ratings, and links — but not their full address, hours, or anything else hidden inside each individual listing.

Subpage Scraping fixes that. Thunderbit visits every linked subpage from your scraped list, reads what's inside each one, and adds that information into new columns of your table. Same as hiring an intern to go click into every listing one by one and write down the missing details for you.

## When to use it

- **Directory enrichment** — Yelp, Google Maps, industry directories where listings have detail pages
- **Lead research** — pulling addresses, hours, services, "about" content from individual business pages
- **Product research** — going from a search results page into each individual product to pull descriptions, features, reviews
- **Article aggregation** — scraping a list of article headlines, then pulling each article's full body, author, publish date
- **Profile detail enrichment** — going from a list of LinkedIn profiles into each one to extract bio, current role, location

## How to use it

1. Do a standard scrape first to get your base table (with the link/URL column included).
2. Open the Subpage Scraping option in Thunderbit.
3. Click into one of the listings to set up the subpage column structure — Thunderbit's AI suggests fields based on what those individual pages contain.
4. Run the scrape with Subpage Scraping enabled.
5. Watch the enriched table populate.

### **ONE THING TO KNOW ABOUT TIMING**

Subpage Scraping on a page with many listings can take a few minutes, since it's visiting each subpage one by one. On a big directory it's not instant — good time to grab a coffee while it runs. The payoff is enormous: instead of 30 names, you walk away with 30 fully detailed records ready for outreach.

### **Why this changes everything**

Before Subpage Scraping, building a lead list from a directory meant clicking into each listing yourself. With it, the directory becomes a one-shot extraction job. Combined with Claude (Chapter 9) you can then enrich each row with email addresses from the underlying websites — completing the full pipeline from "I have a directory URL" to "I have a usable, enriched, contact-ready lead list."

## Workflow 1: Content Research from LinkedIn

---

The first high-value Thunderbit workflow — finding content ideas based on what's actually performing on LinkedIn right now, then turning those proven structures into your own posts using Claude. Total time: 5–10 minutes.

### Step-by-step

#### 1 Go to LinkedIn and search your niche

Use the search bar at the top of LinkedIn. Type the keyword most relevant to your space (e.g. "AI automation," "B2B sales," "real estate marketing"). After search results load, click the **Posts** tab to filter for actual posts (not profiles or jobs).

#### 2 Open Thunderbit and click "AI Web Scraper"

The panel opens. Click **AI Suggest Columns**. Thunderbit reads the search results and suggests columns: post text, author, likes, comments, date.

#### 3 Click Scrape

The table populates with recent posts and their actual engagement numbers — the social proof of what's working in your space right now.

#### 4 Export to Google Sheets

Click Export → Google Sheets. The data lands in a new sheet you can immediately hand to Claude.

#### 5 Hand the sheet to Claude

Open Claude with the Google Sheets connector active. Paste the prompt from the playbook in Chapter 10 (Prompt 1) — point it at your sheet, ask it to identify the top 5 posts by combined likes + comments, explain why each worked, and rewrite each one for your voice and topic.

#### WHY THIS BEATS "GENERATE ME LINKEDIN IDEAS"

Asking Claude for LinkedIn post ideas blind gives you generic suggestions Claude has no way to validate. Asking Claude to model your post on five proven, real-time, top-performing posts in your niche gives you content with a track record before you even publish. The signal lives on LinkedIn already. Thunderbit just hands it to you.

## **What you walk away with**

Five proven post structures from what's actually working right now in your niche — already rewritten for your voice and your topic. Content planning for the week, done in minutes.

The same workflow works on X/Twitter, Reddit, YouTube comments, and any other social platform where engagement numbers are visible.

## Workflow 2: A Lead List That Fills Itself

---

This is the workflow most worth learning if you sell anything to local businesses, or any niche with a public directory of potential clients. The combination of Thunderbit's Subpage Scraping + Claude with Firecrawl MCP turns a directory page into a complete, contact-ready lead list — without you opening a single tab manually.

### Why this works

Public directories — Yelp, Google Maps, industry directories — list businesses with their websites. They almost never list business emails directly, because the directories don't want you bypassing their contact forms. But the businesses themselves usually have email addresses on their own websites — on the contact page, the about page, or in the footer.

Thunderbit scrapes the directory and pulls the websites. Claude (with the Firecrawl MCP connected) then visits each website and pulls the email addresses for you, marking each row as processed so it never repeats work.

### Step-by-step

#### 1 Open your directory search

Yelp, Google Maps, an industry directory — wherever your prospects publicly list themselves. Search by category + location. Example URL pattern: `yelp.com/search?find_desc=physical+therapy&find_loc=Austin,+TX`.

#### 2 Open Thunderbit → AI Web Scraper → AI Suggest Columns

Thunderbit picks up business name, category, rating, and the website link for each one.

#### 3 Click Scrape

Two clicks, and every business on the page is in a table. You now have names, ratings, and website links — but not yet emails.

#### 4 Enable Subpage Scraping

Open the Subpage Scraping option. Click into one listing to let AI suggest additional columns from inside that subpage — full address, hours, services, any extra detail visible there. Run the scrape with Subpage Scraping enabled.

## 5 Export to Google Sheets

You now have an enriched lead table — name, category, rating, address, hours, website, and any other detail you wanted. Still missing: contact emails.

## 6 Connect Firecrawl MCP to Claude

This is a one-time setup. Get a Firecrawl API key from [firecrawl.dev/app/api-keys](https://firecrawl.dev/app/api-keys). In Claude.ai, go to **Settings** → **Connectors** → **Add custom connector**. Set the URL to:

```
https://mcp.firecrawl.dev/YOURAPIKEY/v2/mcp
```

Replace `YOURAPIKEY` with your actual key. Leave OAuth Client ID and Secret blank. Click Add. Then in a Claude chat, click the + button, open Connectors, and enable Firecrawl.

## 7 Hand the sheet to Claude

Paste the email-enrichment prompt from Chapter 10 (Prompt 2). Claude will visit every website in your sheet — homepage, contact page, about page — find the business email, write it into the Email column, and mark each row as processed.

## What you walk away with

A complete lead list with names, addresses, websites, AND contact emails — built from a directory URL in roughly 15–30 minutes, depending on list size. The same workflow handles 30 leads or 300. You never opened a single tab yourself.

### A NOTE ON RESPECT AND COMPLIANCE

Thunderbit only scrapes what's visible on the page — the same data anyone visiting the site can see. That doesn't mean every use is permitted. Check each site's terms of service before bulk scraping. Respect robots.txt. For business email outreach, follow CAN-SPAM, GDPR, and equivalent rules in your region. The tool is powerful — use it for legitimate outreach, not spam.

## Workflow 3: Product Gap Research

---

If you're thinking about building a course, a template pack, a digital product, or any niche offering — this is how you figure out what to build before you build it. Use it on Gumroad, Etsy, the Apple App Store, the Notion Template gallery, or any marketplace where competing products are listed publicly.

### Why this matters

Most failed products fail not because of bad execution, but because they're product number 47 covering exactly the same thing 46 other people are already selling. The hard part isn't building — it's picking what to build. This workflow tells you exactly which angles, topics, and formats are missing or underrepresented in a market, so you build something that stands out instead of something that drowns.

### Step-by-step

#### 1 Open the marketplace search for your topic

Examples: Gumroad search for "notion template," Etsy for "wedding planner printable," App Store for "habit tracker," Reddit for "indie hacker tools." Anywhere your potential competitors list themselves publicly.

#### 2 Open Thunderbit → AI Web Scraper → AI Suggest Columns

Thunderbit picks up the title, description, rating, review count, and price for every listing on the page.

#### 3 Click Scrape and export to Google Sheets

You now have a sheet of every product in your target space with its positioning (title + description) and its market signal (rating + review count).

#### 4 Hand the sheet to Claude for gap analysis

Paste the product-gap prompt from Chapter 10 (Prompt 3). Claude reads through every existing product and identifies what topics, angles, or formats are underrepresented or completely missing — then suggests 3 product ideas that would stand out from everything already listed.

### What you walk away with

A clear map of the competitive landscape and three positioning angles your product could occupy that nobody else is. Instead of guessing what to build, you know exactly where the gap is.

### **PRO VARIATION: USE IT FOR KEYWORD POSITIONING, NOT JUST PRODUCT IDEAS**

The same scrape + Claude pattern works for: finding underused angles in YouTube content (scrape top videos for your keyword and ask Claude where the gaps are), spotting positioning gaps in a SaaS category (scrape a competitor list), or finding niche audiences nobody is targeting (scrape Reddit subreddits in adjacent spaces).

# Connecting Thunderbit to Claude

---

Thunderbit's job ends when the data is in a Google Sheet. Claude's job begins from there. For the full workflows in this playbook you need two connectors set up — Google Sheets (so Claude can read your scraped data) and optionally Firecrawl MCP (so Claude can visit websites to enrich rows).

## Google Sheets connector (essential)

This lets Claude read and write to your Google Sheets directly. Without it, Claude can't see your scraped data.

1. In Claude.ai, go to **Settings** → **Connectors**.
2. Find Google Sheets in the list and click **Connect**.
3. Authorize Claude to access your Google account (OAuth flow).
4. Once connected, every Claude chat can read sheets you share with it by URL.

## Firecrawl MCP (optional, for email enrichment)

This lets Claude visit any URL on the web, read the content, and pull information from it — essential for the Workflow 2 email-enrichment pattern.

1. Get your Firecrawl API key at [firecrawl.dev/app/api-keys](https://firecrawl.dev/app/api-keys) (Firecrawl has a free tier to get started).
2. In Claude.ai, go to **Settings** → **Connectors** → **Add custom connector**.
3. Set the URL to:

`https://mcp.firecrawl.dev/YOURAPIKEY/v2/mcp`

1. Replace `YOURAPIKEY` with your actual key.
2. Leave OAuth Client ID and Secret blank.
3. Click **Add**.
4. In a Claude chat, click the **+** button, open Connectors, and enable Firecrawl.

## The chain of tools

TOOL	JOB IN THE CHAIN
<b>Thunderbit</b>	Extracts structured data from any webpage in 2 clicks
<b>Google Sheets</b>	Stores the data; readable and writable by Claude
<b>Claude</b>	Analyzes the data and generates the output you actually wanted
<b>Firecrawl MCP</b>	(Optional) Lets Claude visit websites to enrich each row further

### ONE-TIME SETUP, PERMANENT LEVERAGE

Once both connectors are set up, you don't touch them again. Every future workflow inherits them automatically. The 5 minutes of setup pays back the first time you save 3 hours on a lead list.

# The Full Prompt Playbook

---

Copy-paste prompts for the workflows in Chapters 6–8, plus adapted patterns for other common Thunderbit + Claude pipelines. Everything in `[BRACKETS]` is your specific information — swap it in and send.

## Prompt 1 — LinkedIn content research

Use after scraping LinkedIn posts (Workflow 1).

### CONTENT · TOP POSTS → MY VERSIONS

Open the Google Sheet `[PASTE SHEET LINK]`. Find the 5 posts with the highest combined likes and comments.

For each one, briefly explain why it worked – the hook, the structure, the angle they used.

Then write me a new version of each post, adapted to my voice and my topic: `[YOUR NICHE/ TOPIC]`. Keep the structure that made the original work, but make the content and examples mine.

## Prompt 2 — Lead list email enrichment

Use after scraping a directory with Subpage Scraping enabled (Workflow 2). Requires Firecrawl MCP connected.

### LEADS · ENRICH WITH EMAILS FROM WEBSITES

Open the Google Sheet `[PASTE SHEET LINK]`. For every row where the "processed" column is empty: use Firecrawl to visit the website listed in that row. Check the homepage, the contact page, and the about page. If you find a business email address, write it into the "Email" column. Then mark that row as "processed" in the "processed" column, whether or not an email was found. Skip any row that's already marked as processed.

## Prompt 3 — Product gap analysis

Use after scraping a marketplace (Workflow 3).

#### RESEARCH · FIND MARKET GAPS

Open the Google Sheet [PASTE SHEET LINK]. This is a list of existing products in the topic [YOUR TOPIC].

Based on the titles and descriptions, identify what topics, angles, or formats are underrepresented or missing entirely. Then suggest 3 product ideas that would stand out from everything already listed here.

### Prompt 4 — Outreach email drafts from a lead list

Use after completing Workflow 2 to draft personalized outreach for each lead.

#### OUTREACH · DRAFTS FOR EACH LEAD

Open the Google Sheet [PASTE SHEET LINK]. For each row with an email address, write a short outreach email (under 80 words) introducing my service: [DESCRIBE YOUR OFFER]. Personalize each one using the business name, category, and any context from the address or about-page info in the sheet. No "I hope this finds you well." Lead with a specific reason for reaching out. End with an easy yes – a quick reply, not a meeting request. Write the email into a new "Outreach Draft" column.

### Prompt 5 — Competitor monitoring summary

Use after scraping competitor blog posts, product pages, or social media regularly.

#### MONITORING · COMPETITOR SUMMARY

Open the Google Sheet [PASTE SHEET LINK]. This is fresh data scraped from [COMPETITOR / SOURCE] – [DESCRIBE WHAT WAS SCRAPPED]. Compare it to the previous version (sheet tab: [OLD SHEET TAB NAME]). Tell me: (1) what's new since last time, (2) what they've removed or stopped emphasizing, (3) any pricing or positioning shifts, (4) the three most important things I should know.

### Prompt 6 — Review sentiment analysis

Use after scraping product or service reviews from Amazon, Yelp, G2, Trustpilot, or similar.

### RESEARCH · REVIEW THEMES & OBJECTIONS

Open the Google Sheet [PASTE SHEET LINK]. This is a list of reviews for [PRODUCT/CATEGORY].

Identify: (1) the 5 most common positive themes, (2) the 5 most common complaints or pain points, (3) the language customers use repeatedly (their actual phrases, not paraphrased), (4) any feature requests or unmet needs that come up multiple times.

Format as a research brief I could hand to a copywriter or product team.

## Prompt 7 — Trending content discovery

Use after scraping a Reddit subreddit, Twitter search, or any topic-specific feed.

### CONTENT · FIND WHAT'S RESONATING

Open the Google Sheet [PASTE SHEET LINK]. This is recent content from [SOURCE – e.g. r/Entrepreneur, Twitter search for "AI automation"]. Sort by engagement (upvotes/likes + comments).

Identify the top themes that are driving engagement right now. For each theme, suggest 2 content angles I could create for [MY PLATFORM/AUDIENCE] that would tap into the same energy without copying the original posts.

## Prompt 8 — Talent/hiring research

Use after scraping LinkedIn profiles or job listings.

### HIRING · CANDIDATE SHORTLIST

Open the Google Sheet [PASTE SHEET LINK]. This is a list of [LinkedIn profiles / job listings] in [LOCATION/INDUSTRY].

I'm looking for [ROLE/SKILLS]. Score each candidate on a scale of 1–10 based on the visible profile signals – title progression, company quality, relevant experience. Add the score in a new "Match Score" column and a one-line reason in a "Notes" column. Then list the top 10 names I should reach out to first.

### NOTICE THE PATTERN

Every prompt above follows the same shape: "Open the Google Sheet [LINK]. Here's what this data is. Here's what I want you to do with it." Once you've used three of these, you'll write the rest from instinct. The hard part isn't the prompt — it's deciding what raw data to scrape in the first place.

## Pre-Built Templates: The 1-Click Library

For popular sites you'd scrape regularly — Amazon, eBay, Zillow, LinkedIn, Apollo, TikTok, Etsy, and dozens of others — Thunderbit ships pre-built scraper templates. You don't even need AI Suggest Columns for these. One click, the right structure is pre-loaded, and you scrape.

### Categories of pre-built templates

CATEGORY	NOTABLE TEMPLATES
<b>E-commerce</b>	Amazon, eBay, Etsy, AliExpress, Shopify stores, Coupang, Coles, Lidl, Sainsbury's, Galaxus, Tradera, Carrefour UAE
<b>Real Estate</b>	Zillow, Redfin, Realtor.com aggregation, general property listings
<b>Lead Generation</b>	Apollo, RocketReach, TruePeopleSearch, FastPeopleSearch, Thatsthem, Tracxn, Zocdoc
<b>Social Media</b>	Twitter (posts and profiles), TikTok (videos, profiles, comments), Pinterest, Snapchat, Reddit posts, Patreon
<b>Content / News</b>	News Scraper, Wikipedia, SlideShare, Naver
<b>Marketplaces &amp; Classifieds</b>	Craigslist, OpenSooq
<b>Workspace tools</b>	Slack messages (export)

### When to use a template vs. the AI scraper

- **Use a pre-built template** when you scrape that site regularly, when the data structure is well-known, and when you want the cleanest possible output without thinking.
- **Use the AI scraper (AI Suggest Columns)** for sites without a template, custom searches, or whenever you want a different column structure than the template provides.

### How to find templates

Browse the full library at [thunderbit.com/template](https://thunderbit.com/template). Each template lists a star rating from real users and the use cases it's best for. When you're on a site that has a template (e.g. amazon.com), Thunderbit can also auto-suggest the relevant template when you open the extension.

## THE 80/20 OF TEMPLATES

For most solo founders, four templates cover 80% of real work: **LinkedIn** (content + leads), **Apollo or RocketReach** (leads), **Amazon or Etsy** (product research), and **Reddit or Twitter** (content trends). Master those four; reach for AI Suggest Columns for everything else.

# AI Data Enhancement Beyond Scraping

---

Thunderbit isn't just an extractor. It can clean, transform, and enrich data *while* it's scraping — letting you skip post-processing steps that would normally happen in Claude or Excel later.

## Built-in AI transformations

### CAPABILITY 1

#### Summarize

Define a column like "Summary of this product" or "1-sentence description of this article" and Thunderbit's AI generates the summary inline as it scrapes. Long descriptions, articles, or reviews become readable bullets in your table directly.

### CAPABILITY 2

#### Categorize

Add a column like "Category: budget / mid / premium" or "Tone: casual / professional / aggressive" and Thunderbit's AI assigns the right tag to each row as it goes. Manual sorting becomes unnecessary.

### CAPABILITY 3

#### Translate

Define a column like "English translation" or "German version" and Thunderbit translates the extracted content automatically. Scrape a foreign-language site and walk away with the localized version in your sheet.

### CAPABILITY 4

#### Format / Calculate

Reformat phone numbers into a consistent style, convert prices to a single currency, calculate discount percentages on the fly. Define the transformation in plain English in the column description and Thunderbit's AI applies it.

### WHY THIS SAVES REAL TIME

Most "scrape → clean → analyze" workflows waste 30–60% of the time on the cleaning step. Doing the transformation inline during the scrape collapses it into one operation. The result lands in Google Sheets already structured, summarized, categorized, and translated — ready for Claude or direct use, no cleanup pass needed.

## Pricing & Plans

---

Thunderbit offers a free tier that's free forever plus three paid plans for higher volume and advanced features. Annual billing gets a 20% discount with all credits upfront.

### The plan tiers

PLAN	BEST FOR
<b>Free</b>	Trying it out, occasional scrapes, small projects. Free forever — no credit card required to start.
<b>Basic</b>	Regular use for personal projects or a solo business. Higher monthly credit allowance, more AI features unlocked.
<b>Professional</b>	Power users — sales, marketing, or research professionals running multiple workflows weekly. Generous credit allowance, full AI features.
<b>Business</b>	Teams running scraping at scale or agencies offering scraping as a service to multiple clients.

### The credit model

Thunderbit runs on credits. Each scrape, Subpage Scraping operation, and AI transformation consumes credits based on volume and which AI model is used (Basic vs. Advanced). The free plan includes enough monthly credits to run several real workflows; paid plans scale credits up to match the work you do.

### Two ways to consume Thunderbit

- **Chrome Extension** — the main product, what this playbook is built around. Free tier and all paid plans.
- **Web Scraper API** — programmatic scraping for your own apps or internal workflows. Pay-per-use, no UI.
- **Managed Scraping Service** — you describe what you need; Thunderbit's team builds and runs it for you. Custom quote.

### How to make your credits go further

- **Use pre-built templates when available** — they're optimized for credit efficiency.

- **Don't enable Subpage Scraping unless you need it** — it consumes more credits because it visits each subpage. Use it where the value justifies the cost (lead enrichment, product detail extraction).
- **Define exactly the columns you need** — fewer columns = less AI inference = lower credit cost.
- **Use the right AI tier** — Basic model for simple extractions; Advanced for complex inference, summarization, or categorization. Don't pay for Advanced when Basic suffices.
- **Annual billing saves 20%** — if you've decided Thunderbit is part of your stack, the annual plan pays back fast.

### WHERE TO START

Start on Free. Run one real workflow end to end — pick LinkedIn content research or product gap research, whichever maps to a current need. If the workflow saves you real hours and you find yourself reaching for it weekly, upgrade to Basic or Professional. Don't pay for capacity you haven't earned yet.

## The Service Angle

---

Lead lists, market research, and competitor monitoring are services real businesses pay real money for. With Thunderbit + Claude, one person can deliver in hours what agencies used to charge thousands for over weeks.

### The market reality

Companies pay \$500–\$5,000+ for one-off lead lists. They pay \$1,000–\$10,000+ for competitor and market research reports. They pay ongoing retainers of \$500–\$3,000/month for continuous competitive monitoring. The data has always been public; the labor was the constraint. Thunderbit removes the labor constraint.

### Three offers you can productize today

#### OFFER 1

##### Lead List as a Service

Client gives you a niche + geography + ideal customer profile. You deliver a clean, enriched list of 100–500 qualified leads with names, websites, addresses, and contact emails. Typical price: **\$500–\$2,500** per list. Time invested with this workflow: ~2–4 hours.

#### OFFER 2

##### Market Research Report

Client wants to know what's selling in a niche, what's missing, and where the opportunities are. You scrape the relevant marketplaces, run Claude analysis, and deliver a 5–10 page report with top performers, gaps, positioning angles, and recommended product directions. Typical price: **\$1,500–\$5,000**.

#### OFFER 3

##### Competitor Monitoring Retainer

Monthly retainer to scrape your client's 5–10 main competitors — their pricing, product launches, content, ads, and reviews. Deliver a monthly briefing summarizing what changed and what it means. Typical retainer: **\$500–\$2,000/month**.

## The simple positioning

Don't market as a "web scraping service." Pick a niche and own the outcome:

- "I build lead lists for B2B SaaS companies targeting mid-market accounts"
- "I run competitive research for indie SaaS founders before they launch"
- "I deliver monthly competitor reports for DTC e-commerce brands"

Niche positioning beats generalist positioning every time. Same skill, higher rates, easier referrals.

## The starter offer

If you've never sold a research or lead-gen service: offer one lead list at \$500 to three target clients. Use those three as case studies. From client four, raise to \$1,500. From client ten, charge \$3,000+ and stack ongoing monitoring retainers on top.

### THE RECURRING-REVENUE ANGLE

One-off lead lists are valuable but transactional. The real long-term play is the competitor monitoring retainer — you scrape the same sources on a schedule (weekly or monthly), Claude summarizes what changed, and the client pays you every month for the briefing. Build five of those at \$1,500/mo each and you have a \$7,500 MRR service business running off your laptop.

## Power Moves: 12 Non-Obvious Tactics

---

Once you've run your first three workflows, these are the moves that separate good scraping output from great — and casual users from people getting real leverage.

1. **Edit the AI's column suggestions before scraping, don't accept them blindly.** Add the columns you actually want — including AI-generated ones like "1-sentence summary" or "category tag" — so the scrape outputs structured, analyzed data in one pass.
2. **Always include the link/URL column.** Even if you don't think you need it now, you'll want it later for Subpage Scraping or for Claude to visit each source.
3. **Use Subpage Scraping selectively.** Brilliant for lead enrichment; overkill for a quick list scan. The credit cost is higher — match the cost to the value.
4. **Chain scrapes across pages of pagination.** For a directory with 20 pages of results, scrape page by page and append to the same Google Sheet. Claude can then analyze the full dataset, not just the first page.
5. **Build a sheet template per workflow.** Pre-set columns, formatting, and a "processed" column in advance. Reuse the same template every time you scrape that source.
6. **Combine pre-built templates with AI columns.** Use the Amazon template for product data, then add a custom AI column like "Why this is selling well" — best of both worlds.
7. **Run regular re-scrapes for time-series data.** Scrape a competitor's pricing page once a week, append to the same sheet with a date column. Claude can then analyze trends, not just a snapshot.
8. **Use the "processed" column pattern across all multi-step workflows.** It lets you stop and resume Claude jobs without redoing work — essential for large lead lists where token limits or interruptions can break a single Claude run.
9. **Translate scraped content during the scrape.** If your audience is in a different language than your data source, save the post-processing step by defining a translation column during scraping.
10. **Hand Claude both the scraped data AND your context.** Don't just paste "analyze this." Tell Claude who you are, what you sell, what your voice is, and what decision the output should support. Specific context produces specific output.
11. **Schedule a weekly "fresh data" routine.** Pick one regular scrape (LinkedIn content, competitor pricing, niche subreddit) and run it on the same day every week. Compound insight beats one-off curiosity.
12. **Productize one workflow before scaling.** If you want to sell this as a service, pick ONE offer (e.g. monthly competitor reports for SaaS founders), sell it five times, then add the next. Productize by service, not by tool.

## Your 7-Day Mastery Plan

---

A concrete week to go from "Thunderbit installed" to "running a full workflow end-to-end and walking away with something usable in your business." One focused block per day.

### DAY 1

#### Install and verify

Go to [thunderbit.com](https://thunderbit.com). Install the extension. Pin it to your toolbar. Open any page with a list (Amazon search, LinkedIn search, a directory) and click the icon to confirm the AI Web Scraper panel opens.

### DAY 2

#### Your first 2-click scrape

Pick a low-stakes page — your favorite news site, a product category on Amazon, a Reddit subreddit. Run AI Suggest Columns → Scrape → Export to Google Sheets. Look at what came back. Don't optimize yet; just see how clean the default output is.

### DAY 3

#### Connect Claude to Google Sheets

In Claude, enable the Google Sheets connector. Open the sheet from Day 2 and ask Claude a real question about the data: "Summarize the top 5 things in this sheet" or "What patterns do you see?" Feel the chain working.

### DAY 4

#### Run Workflow 1 — LinkedIn content research

Pick a keyword from your niche. Search LinkedIn Posts. Scrape with Thunderbit. Export. Run Prompt 1 from Chapter 10. Walk away with 5 proven post structures, already rewritten for your voice.

## DAY 5

### **Run Workflow 3 — product gap research**

Pick a marketplace relevant to a product idea you've been considering. Scrape the search results. Run Prompt 3 from Chapter 10. Walk away with three product positioning angles nobody else is occupying.

## DAY 6

### **Run Workflow 2 — the full lead list**

Connect Firecrawl MCP to Claude (Chapter 9). Pick a directory, scrape with Subpage Scraping enabled, and run Prompt 2 to enrich emails. This is the workflow that pays for the playbook. Walk away with a real, usable lead list.

## DAY 7

### **Productize or repeat**

Pick one outcome from the week that you'd like to do regularly. Build it into a routine — weekly content research, monthly competitor monitoring, or a packaged lead-list service to sell. The skill compounds when you turn it into a habit, not a one-off.

## Pre-Scrape Checklist

---

Run through this before you start any meaningful scrape — your own or a client's. 60 seconds of preparation prevents wasted runs and bad data.

- Thunderbit is pinned to your Chrome toolbar
- You're on the actual page with the data you want, not a parent page or summary
- You're logged in if the data requires it
- You've confirmed the page contains the data type you need (not just navigation/menu items)
- You've decided whether you need Subpage Scraping (worth the extra credit cost) or just a standard scrape
- You've reviewed the AI-suggested columns and added any custom AI-generated ones you want (summary, category, translation)
- Your Google Sheets connector is connected in Claude (if you'll hand to Claude after)
- Firecrawl MCP is set up if the next step involves visiting external websites (lead email enrichment)
- You have a clear question or output you want from Claude — not just "analyze this"
- You've checked the source site's terms of service for compliance
- For email outreach, you're prepared to follow CAN-SPAM, GDPR, or your region's rules

### Start scraping at [thunderbit.com](https://thunderbit.com)

Free to install, free forever on the base tier. Pin it once, then point it at whatever page has the data you need.

The hard part of business research, lead generation, and competitive intelligence was never the analysis — it was getting the data into a usable form in the first place. Thunderbit removes that blocker. What's left is the only part that actually matters: which data, which question, which decision.

[Get Thunderbit at thunderbit.com](https://thunderbit.com) →

## **Final word**

The proof of what works is already public. Content with high engagement is already published. Products with high reviews are already listed. Businesses worth selling to are already in directories. The signal is everywhere — it was always the extraction that defeated you.

Two clicks per scrape. One prompt per analysis. One workflow per outcome. That's the entire system.

Pick one page today. Scrape it. See what you couldn't see before.