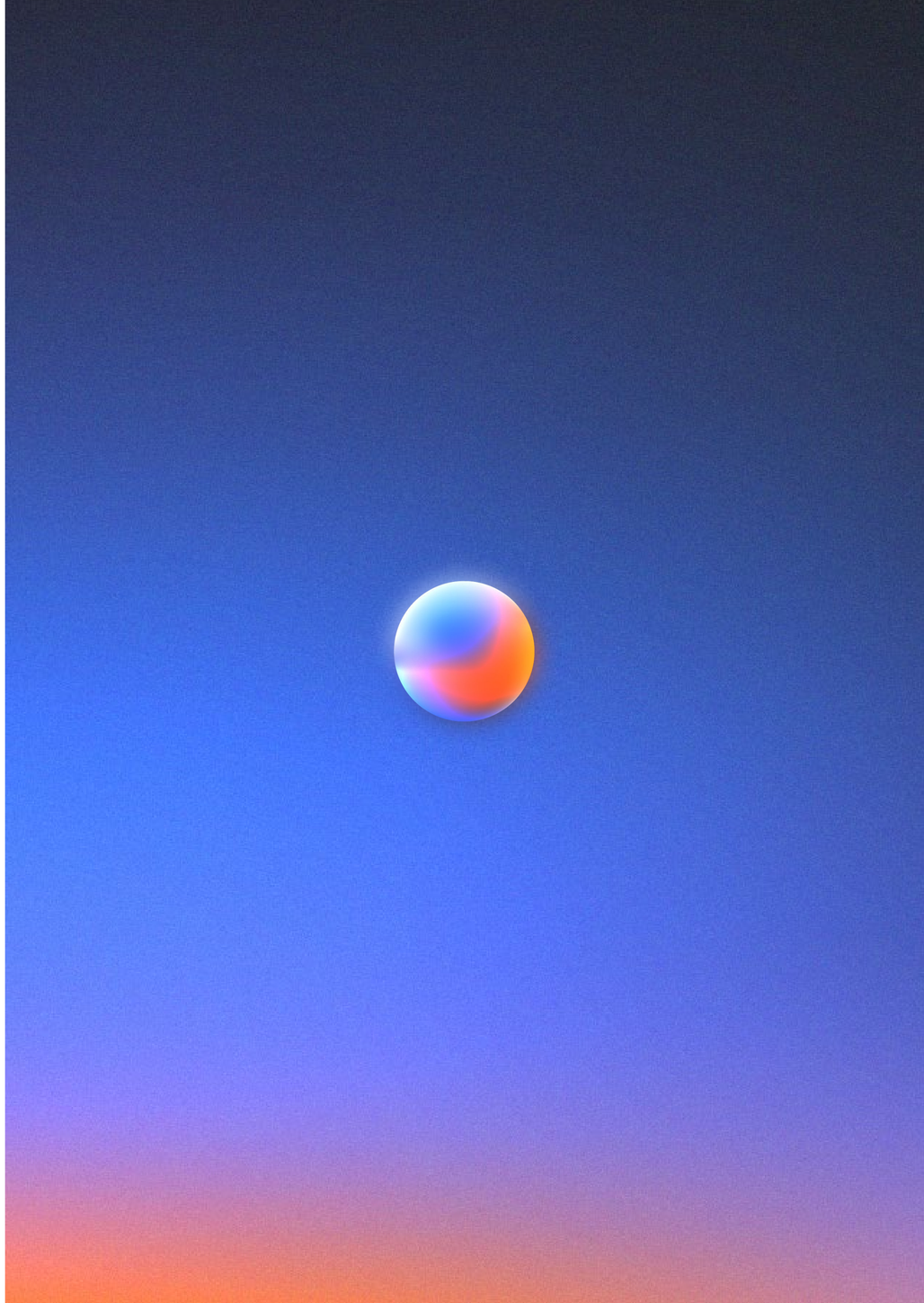
Going Agentic:



The GTM Field Kit.

PUNCH

Three tools to build, run and test your first agentic GTM workflow.

From the Going Agentic roundtable, with Zeno Farhat, Sumair Shakir and Ben Robinson.

Tool 1: The Stack

Set up five tools and one loop in a weekend



Zeno Farhat
AI Growth Engineer,
keragon

Tool 2: The Scoring Rubric

Decide who to contact first, and on which channel



Sumair Shakir
Principal GTM Strategist,
ClayVengers

Tool 3: The stress test

Check a workflow is ready before it reaches a buyer



Ben Robinson
Chief Commercial Officer,
PUNCH

One rule runs through all three: agents draft, people decide.

Ready to build? Go to the stack. Want to know who to call first? Start with the scoring rubric. About to hand a workflow to an agent? Use the stress test.

Why people still make the call

“Some household robots learn by watching people do tasks through camera glasses. They have the maths, cameras and sensors, and they still need extensive training to pick something up.

A child doesn't. We start grabbing things at two and are good at it by ten or twelve, with no training programme at all. That kind of skill comes from learning gradually over years, and AI hasn't found a shortcut for it.

There's something in learning over a period of time which AI hasn't been able to shortcut. And that's going to show up in all the knowledge work as well. The same gap shows up in GTM. Sorting, scoring and categorising have a right answer, so agents learn them fast. Offer design, strategy and copy rely on judgement built over years of experience. That's why every tool in this kit keeps a person signing off.”

Sumair Shakir

Tool 1.

The agentic GTM stack



Zeno Farhat
AI Growth Engineer, Keragon

01 / The agentic GTM stack

Use it when:	Why it matters:	You'll finish with:	Time and skill:
You want outbound running without an agency, and someone on the team can work with APIs.	Five tools for about $280 a month, and you can swap any of them out later. You learn what works before you commit to a bigger platform.	A leads table, an agent that scores and drafts every day, and warmed domains ready to send.	One weekend, plus two weeks of domain warmup. Basic coding, or a technical teammate.

– Let's get to it.

Five tools one person can run

Tool	Pre-setup	Price
Clay Finds and enriches accounts	Your own API keys and one narrow ICP	$167 - $185 Launch
Supabase Stores leads, signals and replies	One leads table with a status column	$20 (Pro)
Claude Code Builds your agents and scripts	A CLAUDE.md file with your ICP, tone and rules	$20 (in Claude Pro)
Vercel Hosts agents and runs them on a schedule	A hard spend limit	$20 (Pro, includes $20 usage)
Instantly Sends outreach and warms inboxes	Two or three secondary sending domains	$47 (Growth) plus domains

When to drop Clay: Once you know which two or three sources find your buyers, your ICP is stable, and the bill is growing faster than pipeline. Then ask Claude Code to rebuild those steps as a script using your own keys.

From $280/m
plus domains and Anthropic API usage.

Five tools one person can run.
Each does one job in the loop:
find, store, build, host, send.

HotSauce by Punch! COMING SOON

Signal intelligence built for this stack. HotSauce surfaces the accounts showing real buying signals, so your agents start from evidence instead of a cold list.

FIND OUT MORE

Weekend build checklist

You can have a working outbound loop by Sunday night. The two-week wait before sending is deliberate: it gives your new domains time to warm up, you time to check the drafts.

Friday

- [ ] Buy two or three secondary domains, connect them to Instantly, turn on warmup
- [ ] Sign up for Claude Pro, Vercel, Supabase and Clay
- [ ] Install Claude Code on your computer and sign in with your Claude account

Saturday

- [ ] Create the leads table in Supabase (file 1 on the next page)
- [ ] Build one Clay table for one industry, one company size, one job title: 100 to 200 rows
- [ ] Push the rows into Supabase

Sunday

- [ ] Save CLAUDE.md (file 2 below) in a new folder and open Claude Code there
- [ ] Build one agent: read new leads, score them, draft a first line for top scorers, write both back
- [ ] Paste into Claude Code:

```
Read CLAUDE.md. Build a script that reads rows in
leads where status = new, scores each one with
the rubric, writes account_score, contact_score,
channel, evidence_gaps and routing_reason, drafts
first_line for rows scoring 60 or more, and sets
status to scored. Show me 5 sample outputs before
running on the full table.
```

- [ ] Deploy to Vercel on a daily schedule, spend limit set
- [ ] Review the first 20 drafts. Fix the instructions file, not the drafts
- [ ] Approve leads: in the Supabase table editor, change status to approved and add your name in reviewed_by. Nothing goes to Instantly without this step.

Two weeks

- [ ] Connect approved leads to Instantly at 20 to 30 emails per inbox per day
- [ ] Export rows where status = approved as a CSV and upload them, or ask Claude Code to build a script that pushes approved rows to Instantly through its API. Then set status to sent.

Two code scripts you want to copy and paste

01 **Leads table**
Paste into Supabase SQL editor

02 **CLAUDE.md**
Save in your project folder

File 1: Leads table (paste into the Supabase SQL editor)

```
create table leads (
  id               uuid primary key default gen_random_uuid(),
  created_at       timestamptz default now(),
  company          text not null,
  domain           text,
  first_name       text,
  last_name        text,
  title            text,
  email            text unique,
  phone            text,
  linkedin_url     text,
  source           text,
  signal           text,
  signal_source    text,
  signal_date      date,
  account_score    int check (account_score between 0 and 60),
  contact_score    int check (contact_score between 0 and 40),
  icp_score        int generated always as (account_score + contact_score)
stored,
  channel          text check (channel in ('phone','linkedin','email','hold')),
  evidence_gaps    text,
  routing_reason   text,
  first_line       text,
  status           text default 'new'
    check (status in
('new','scored','held','suppressed','approved','sent','replied','meeting','dead')),
  reviewed_by      text,
  notes            text
);

create index on leads (status);
create index on leads (icp_score desc);

-- Keeps lead data private. Only your server-side agent (service role key)
can read or write.
alter table leads enable row level security;
```

To avoid any breaks in the code, use the text files through these links.

FILE ONE

FILE TWO

File 2: CLAUDE.md (save in your project folder)

```
# Outbound agent: project rules

## Who we sell to
- Industry: [e.g. B2B SaaS]
- Company size: [e.g. 50 to 500 employees]
- Buyer: [e.g. VP Sales, Head of Revenue Operations]
- Strong signals: [e.g. hiring SDRs, new sales leader in last 90 days]
- Disqualify: [e.g. agencies, competitors, under 20 employees]

## What we offer, in one sentence
[Your offer]

## How we score
- Account (60): business fit 20, use-case relevance 20, relevant change 20.
- Contact (40): role remit 20, influence 10, scope alignment 10.
- Full points for direct, current evidence; half points for partial evidence;
zero if the evidence does not support the criterion.
- Missing evidence: award no points and record the gap in evidence_gaps.
- Channel: 80 to 100 phone, 60 to 79 linkedin, 40 to 59 email, below 40 hold.

## How we write
- First lines are under 25 words and mention one specific, checkable fact.
- No flattery, no "I hope this finds you well", no exclamation marks.
- If there is no real signal, skip the lead. Never invent one.

## Data rules
- All leads live in the Supabase `leads` table.
- Only touch rows at the status you were asked to process.
- Write account_score and contact_score. icp_score is calculated
automatically.
- Write criterion-level points and a cited reason for every score in routing_
reason.
- Record missing evidence in evidence_gaps.
- Never send email. Drafting only; a person approves before anything goes to
Instantly.

## Before you finish
- Show me 5 sample outputs before running on the full table.
```

Tool 2.

The Scoring Rubric



Sumair Shakir
Principal GTM Strategist,
ClayVengers

02 / The Scoring Rubric

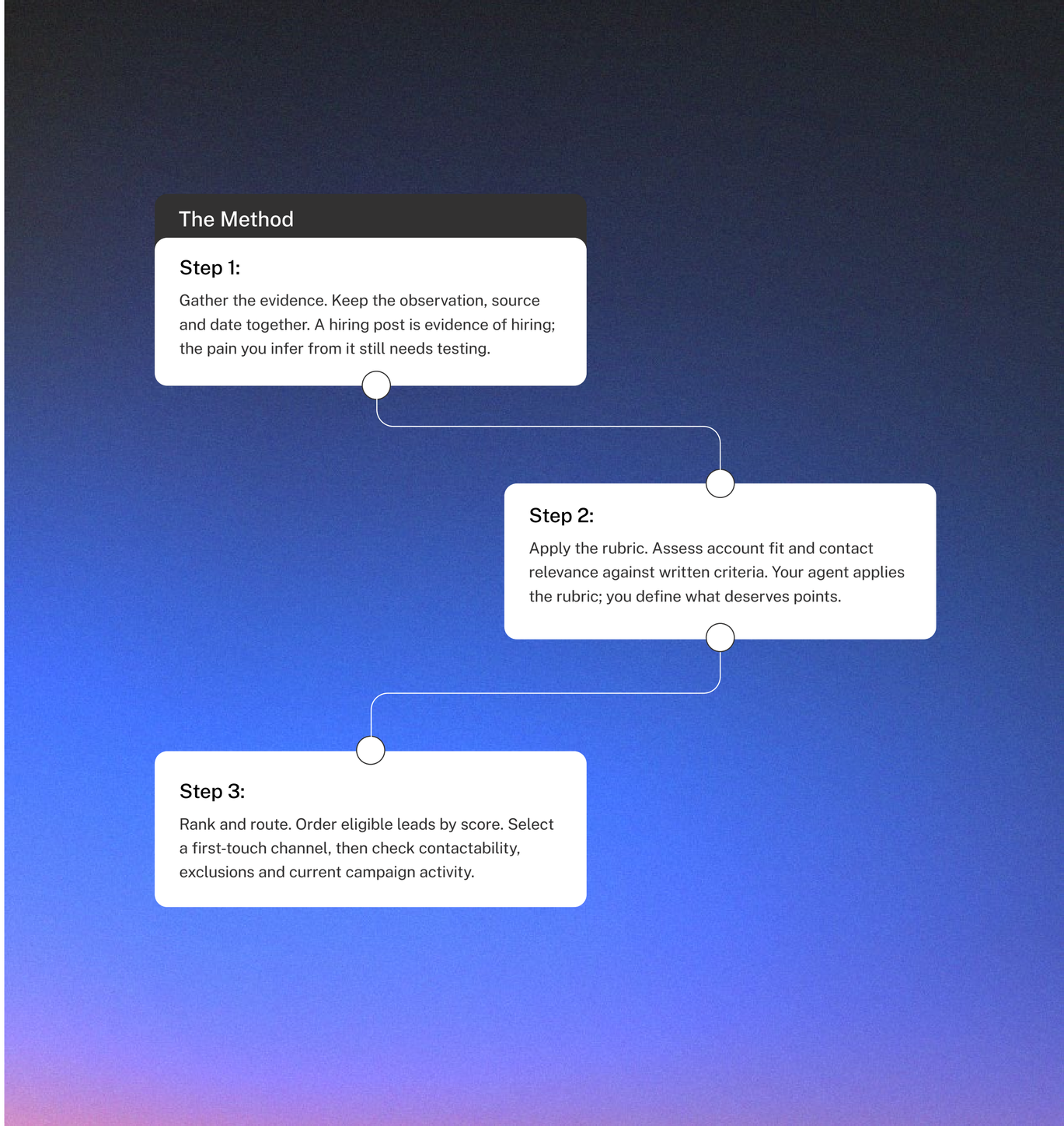
Use it when:	Why it matters:	You'll finish with:	Time and skill:
You have more leads than calling time, and reps choose who to contact by gut feel.	A list says who you could contact. A rubric says who to contact first, and shows the evidence behind each score so you can check it.	A 100-point rubric adapted to your campaign, routing rules for phone, LinkedIn and email, and agent instructions.	An afternoon. No coding.

– Into the action

Use AI to score accounts and contacts, then turn the highest-priority leads into an outreach queue.

A list tells you who you could contact. A rubric helps you decide who deserves attention first.

Score the company and the person separately. Combine those scores into a lead priority, then use that priority to allocate attention across phone, LinkedIn and email.



The Starter Rubric: Score the account, then score the person

60
Account points

40
Contact points

Replace the criteria, weights and thresholds with your campaign rubric.

Before Scoring:

01: Define mandatory ICP requirements, exclusions, the target problem and the relevant roles.

02: Suppress known exclusions.

03: Hold records with missing mandatory evidence for review.

Criterion	Points
Account	**/60**
Business Fit Matches the target segment and operating profile.	20
Use-case relevance Evidence connects the account to the problem you solve.	20
Relevant change A current change makes that problem worth investigating.	20
CONTACT	**/40**
Role remit Responsibilities overlap with the problem or workflow.	20
Influence Evidence of involvement in evaluation or implementation.	10
Scope alignment Their team, region or business unit matches the use case.	10

Scoring anchors:

- FULL POINTS — Direct, current evidence
- HALF POINTS — Partial evidence
- ZERO POINTS — Does not support the criterion

Define campaign-specific examples before running the batch.

Missing evidence:

Award no points yet and flag the gap. Missing data is not a failed fit check. Keep account score, contact score and evidence gaps visible; the total is a priority score, not a buying probability.

Route the first touch: Put the score to work

These score bands are proposed defaults, not a record of your existing routing rules. Agree your thresholds before deployment.

80-100	60-79	40-59	0-39
PHONE FIRST	LINKEDIN FIRST	EMAIL FIRST	Hold for more evidence, or exclude if clearly unsuitable

Three fictional leads, same rubric

Lead One:

Account: 50/60
- Strong segment fit (20)
- Relevant change (20)

Contact: 40/40
- Partial use-case evidence (10)
- Scope (10)
- Contact has supported remit (20)
- Influence (10)

90 /100
Total score

Recommended route: PHONE FIRST

Added to reps task list. Routing reason here

Lead Two:

Account: 40/60
- Strong segment fit (20)
- Use-case (20)

Contact: 30/40
- Qualifying change (0)
- Scope (5)
- Contact has supported remit (20)
- Influence (5)

70 /100
Total score

Recommended route: LINKEDIN FIRST

Added to reps task list. Routing reason here

Lead Three:

Account: 30/60
- Strong segment fit (20)
- Partial use-case (10)

Contact: 20/40
- Qualifying change (0)
- Scope (5)
- Contact has supported remit (10)
- Influence (5)

50 /100
Total score

Recommended route: EMAIL FIRST

A score recommends a route. It does not authorise a send. Check channel availability and eligibility, active sequences, ownership and suppression again. If the preferred channel is unavailable, apply an agreed fallback or route for review.

Run, inspect, improve: Give your agent a decision contract

Stage	What to set up
Inputs	Provide one record per account-contact pair, with stable IDs, company and role evidence, source dates, exclusions, existing campaign activity and available channels.
Scoring	Require criterion-level points and cited reasons. Calculate account and contact subtotals in code, then sum them. Separate missing evidence, conflicting evidence and confirmed non-fit.
Output	Produce a ranked queue containing both scores, total score, proposed first-touch channel, evidence gaps, routing reason and review status. Keep held and suppressed records separately.
Review	Inspect examples from every channel band and cases near each threshold. Check whether a high company score is hiding a weak contact, or incomplete research is depressing the score.
Measurement	Track agreement with your review, time spent correcting, cost per accepted lead and results by score band and channel. Channel results alone do not prove the scoring caused an improvement.

Try it on one campaign

- Define the rubric
- Score a reviewable batch
- Correct the decisions
- Freeze the revised rules and test on fresh records before expanding

Use the instruction file (CLAUDE.md in Tool 1) with your approved rubric and evidence export. Start with a reviewable output.

Tool 3.

The Stress Test



Ben Robinson
Chief Commercial Officer,
Punch!

03 / The Stress Test

Use it when:	Why it matters:	You'll finish with:	Time and skill:
Before an agent touches anything a buyer sees.	A faster workflow can still create more work. This checks the agent really saves time once you count review, corrections and mistakes.	A one-page pilot brief agreed with the person accountable for the result.	One to two hours with that person. No coding.

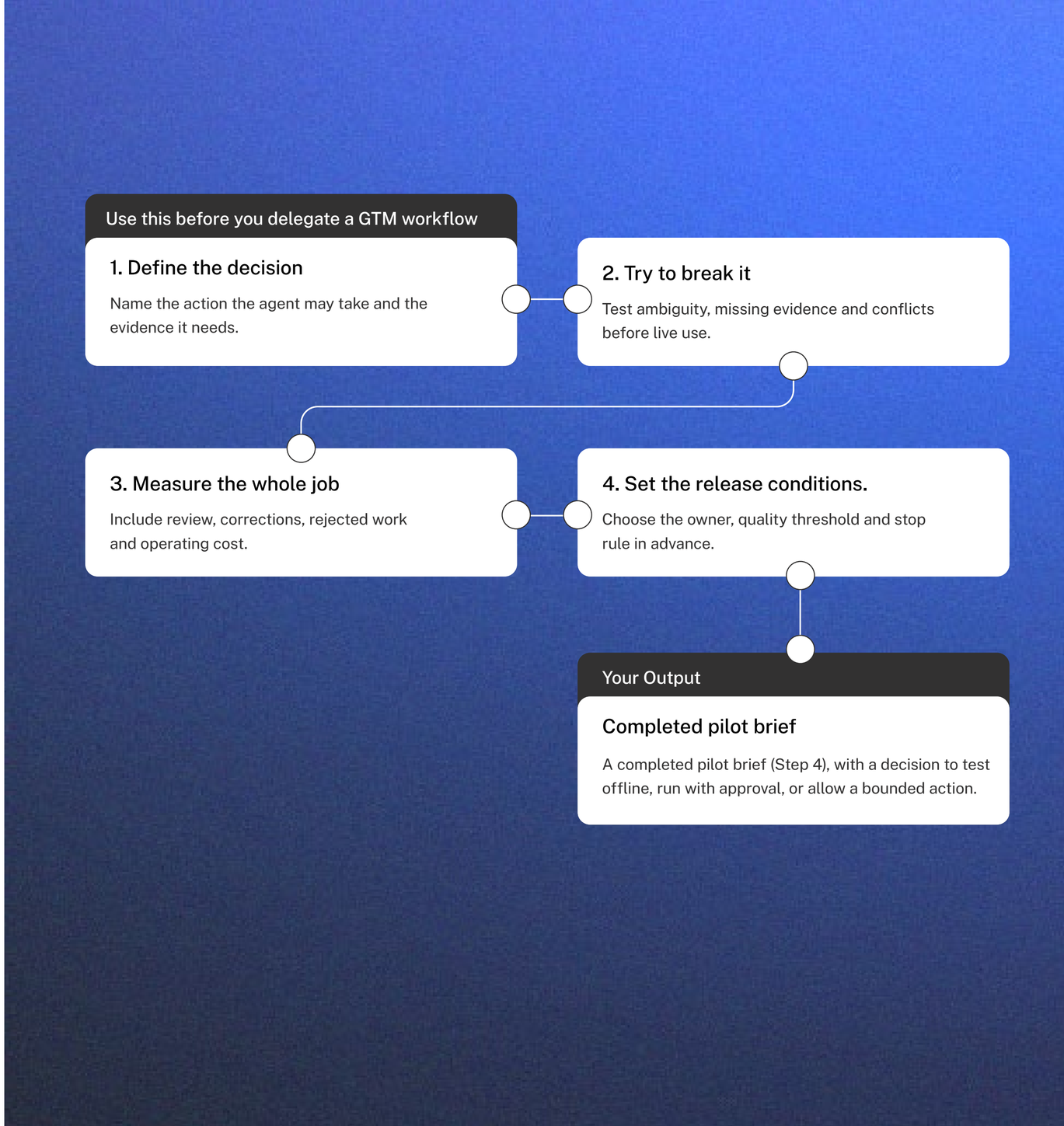
– Let’s put it to work

Choose one workflow.

Find its weak points. Set the conditions for a credible pilot.

A faster workflow can still create more work.

An agent can research more accounts or draft more messages while increasing false positives, correction time and prospect-facing mistakes. Count what your team can actually use.



Define the boundary. What exactly are you delegating?

	Check	Question	If no
	Evidence	Can a reviewer inspect the source, date and reason for the recommendation?	Test offline. Require source capture before relying on the output.
	Decision rule	Can two people apply the same written acceptance criteria?	Agree the criteria and label disputed examples first.
	Uncertainty	Does it stop or request review when evidence is missing or contradictory?	Add an explicit unresolved route. Do not force every case into yes/no.
	Action control	Can an owner inspect, edit and stop the next action?	Keep external actions behind approval.
	Recovery	Can you trace an error, stop repeat actions and repair affected records?	Limit the pilot to outputs that do not change live systems.

Choose a boundary: Offline evaluation > draft with human approval > a narrow action with monitoring and stop controls. Passing these checks supports a pilot; it does not establish accuracy.

Assess one decision at a time. A workflow may be ready to draft while still needing approval to act.

HotSauce by Punch! COMING SOON

Signal intelligence built for this stack. HotSauce surfaces the accounts showing real buying signals, so your agents start from evidence instead of a cold list.

FIND OUT MORE

Try to break it. Stress test signal to outreach

Scenario	What can go wrong	Required behaviour
A job ad mentions Salesforce	Supported: the advert mentions Salesforce. Unsupported: the company has broken CRM handoffs or plans to buy a new tool.	Require the source and date; frame the possible pain as a question.
The right title belongs to the wrong remit	A senior title alone does not prove ownership of the problem. A regional leader may not own a global process.	Require role-scope evidence or route the contact for review.
The source is stale or conflicts with another source	An old hiring post may describe a team that no longer exists. Conflicting evidence should remain visible.	Apply your freshness rule; unresolved cases go to a reviewer.
The account is excluded or already being contacted	A relevant signal does not override a do-not-contact rule or justify overlapping outreach.	Apply suppression before selection and recheck before sending.

Test Pack:

Include clear passes, clear failures, missing data, contradictory sources and exclusions. Record expected decision, actual decision, evidence and correction. Keep a separate representative sample to estimate everyday performance.

Worked example: an agent finds a hiring signal, selects a contact and proposes an outreach angle. Illustrative only.

Measure the whole job. Count usable work. Include the cleanup.

Compare the current process and the assisted process on the same work, using the same acceptance criteria.

Illustrative batch: 100 records per process

Measure	Current	Agent-assisted
Production / setup and operation	6.0 hours	1.0 hour
Review and corrections	1.0 hours	3.0 hours
Total human time	7.0 hours	4.0 hours
Accepted outputs	90	80
Human minutes per accepted output	4.7	3.0
Labour cost at an assumed $50/hour	$350	$200
Incremental tool / run cost	$0	$100
Cost per accepted output	$3.89	$3.75

The decision is not obvious.

43%
Less human time

90→80
accepted outputs

4%
cost improvement

Quality determines the trade-off

Include setup allocation and rework consistently. Track serious errors separately: a low average cost cannot cancel out an unacceptable failure.

Cost per accepted output =
(all human hours × hourly cost + attributable tool cost)
÷ accepted outputs.

Define a usable output before testing.

For this example: a record that passes account and role criteria, includes a current cited signal, clears exclusions and contains no unsupported claim. Approval to send remains a separate decision.

Set the release conditions. Your one workflow pilot brief.

Pilot brief	Fillable PDF fields
Workflow, trigger and intended output	
Agent may do / human must approve	
Required evidence and acceptance criteria	
Test set: representative cases plus difficult and excluded cases	
Pass thresholds: accepted-output rate, error limits, time and cost	
Stop rule and person who can pause / repair the workflow	
Review date, accountable owner and decision: stop / revise / expand	

Fill in the fields, save a copy and use it to agree the test with the person accountable for the result.

Take one disputed decision to your team.

Which step would you trust an agent to take, what evidence would change your mind, and who owns the mistake? Use that answer to choose your next test.

Build, buy or blend?

"It’s a blend. It will always be a blend.

There are core platforms it would be irresponsible to move away from: data tools like Clay, FullEnrich and AirScale, Claude Code as the interface, and campaign tools like Smartlead, Instantly and EmailBison. Those are household tools, and you should buy them.

Then you blend them with what you build in-house: internal dashboards, skills you build in Claude, and applications for monitoring, management and the client interface.

I wouldn’t put node-based automation tools like n8n at the centre. What works is Claude Code, the household platforms, and your own tools for management and client-facing work.

We blend everything at Punch!, and there isn’t one number that captures it. A single outbound campaign that gets 100 responses is easy to measure. A whole go-to-market engine is much harder to report on."

Ben Robinson, CCO, Punch!

Next step

Watch the roundtable.

Going Agentic: Inside the Shift in GTM, recorded 5 November 2026.

WATCH ROUNDTABLE

Or...

Skip the build.

Talk to Punch! about HotSauce and get the agents running for you.

BOOK A DISCOVERY CALL

punchb2b.com emea@punchb2b.com usa@punchb2b.com