

01 / WHAT I DO

I ship *AI* features into your *business*.

Freelance AI integration for law firms, agencies, ops teams, and B2B SaaS. Fixed scope. Fixed price. Shipped in 1–3 weeks. Handed over with a full learning kit so your team can own it.

> **PRODUCTION, NOT DEMOS**

I built and run my own B2B SaaS. I know what AI work actually costs — and how it breaks.

> **START WHERE YOU ARE**

Roadmap if you're stuck. Sprint if you've scoped it. Either way, fixed price.

> **YOU OWN IT AFTER**

Code, prompts, AI agents — handed over so your team runs it without me.

02 / WHO I AM

I'm Harry.

I build the *AI features* that ship.

Seven years as a software engineer across finance, regulated industries, analytics, and telecoms. I work in Python (FastAPI, Django, Flask) and TypeScript (Node, Next.js), and I've spent the last year specialising in AI features, agents, and workflows that ship into production — not demos.

Most of my clients don't have a dedicated AI team. They're law firms drowning in document review, agencies running outbound by hand, e-commerce ops teams triaging support tickets, accountants categorising invoices, founders shipping their first AI feature. They know AI should help. They don't know where to start. That's the gap I sit in — translating “we should do something with AI” into a working tool their team uses on Monday.

I also built and run Filemender, a B2B SaaS for post-production teams, end to end and solo. The growth stack (next page) runs my marketing function. I know what production AI work actually costs, where it breaks, and how to engineer around the failure modes — because I've hit them in my own product.

I work remotely across European business hours, take on one client at a time, and I'd rather turn down work than over-promise on a timeline. If we're a fit, I'll tell you. If we're not, I'll point you at someone who is.

> MODELS I BUILD WITH

Claude is the default for production work (evals, long-context, tool use). I benchmark against GPT, Gemini, and Grok per project, and route to open-source Llama or DeepSeek for on-prem or cost-sensitive workloads. Day-to-day I work inside Cursor. Model choice is part of the deliverable, not an assumption.

03 / FILEMENDER CASE STUDY

Eight Claude agents run the *growth function*.

Filemender is a media file validation tool for post-production studios. The interesting part isn't the product — it's the AI agents and workflows running on top of it. Eight scheduled Claude agents handle SEO content, lead research, cold email drafting, LinkedIn posting, and Twitter monitoring. The whole marketing function runs as engineering.

8

scheduled Claude agents in production, each with a single narrow job

10×

cost reduction after migrating routine work from Opus to Haiku

0

fabricated leads after adding the verification layer

> THREE LESSONS I CARRY INTO CLIENT WORK

01 Model selection matters more than model choice

The first version ran Opus across every agent. 70% of the work was being done at premium prices when Haiku would do it indistinguishably. For clients, I build cost monitoring in from day one.

02 LLM output is a candidate, not a fact

The lead researcher once invented contact details. Now every claim the model makes — email, role, company — is verified against an external source before it lands downstream. Trust nothing that can break the world.

03 Templated personalisation fools no one

V1 of cold email used “{{first_name}}” mad-libs and recipients noticed. V2 reads the prospect's site and recent posts first, then drafts from that context. Reply rates improved meaningfully.

Full 12-minute case study: harrybenham.dev/case-studies/filemender

04 / OTHER WORK

Smaller builds, same *playbook*.

Short engagements where I shipped a single AI feature end-to-end. Scope pinned, price pinned, handed over with evals.

DOCUMENT REVIEW · 10 DAYS

Contract-clause extraction for a law firm

Replaced a paralegal workflow that took 3 hours per contract. Claude reads the PDF, extracts 18 named clauses into a structured review table, and flags anything unusual. Now runs in under 90 seconds per doc.

CLAUDE SONNET

PDF PARSING

POSTGRES

CLASSIFICATION · 10 DAYS

Inbound-lead triage for a B2B agency

Replaced a hand-written rules engine with a two-model classifier: Haiku for cheap triage, Sonnet for ambiguous cases. Cut ops time per lead from 4 min to 18 seconds.

CLAUDE HAIKU

CLAUDE SONNET

ZAPIER

AGENTS · ONGOING

Outbound agent for a pre-seed founder

Daily agent that researches ICP accounts, writes a first-draft cold email, and queues it for the founder's review. Runs for under £30/mo of model spend.

CLAUDE SONNET

PLAYWRIGHT

CRON

05 / THREE PACKAGES. FIXED SCOPE.

Start where you are.

Most AI work that fails in real businesses fails for the same reasons: no clear scope, no evals, runaway costs, no handover. I package the work to remove those failure modes upfront.

<p>START HERE</p> <p>AI ROADMAP</p> <h2>£1,500</h2> <p>1 WEEK · SHIP ONE SMALL AGENT</p> <p><i>I ship one Small agent off the menu by Friday. If you genuinely don't know what you need yet, swap it for a strategy plan instead.</i></p> <ul style="list-style-type: none"> › 1 Small agent shipped end-to-end (default) › OR a 5–8 page strategy plan with 3 ranked use cases › Eval suite + handover doc included › 14 days of bug-fix support › 100% credited toward a Sprint if you proceed <hr/> <p>+ £50/MO PER AGENT · CARE</p>	<p>MOST POPULAR</p> <p>AI SPRINT</p> <h2>£3,000</h2> <p>2 WEEKS · 1 MEDIUM OR 2 SMALL</p> <p><i>One Medium agent OR two Small agents off the menu, shipped end-to-end into your stack. Most of the menu lives here.</i></p> <ul style="list-style-type: none"> › 1 Medium agent OR 2 Small agents shipped › Eval suite covering the critical paths › Cost monitoring with per-feature spend caps › Plain-English handover doc — model, prompts, failure modes › 30 days of bug-fix support <hr/> <p>+ £50/MO PER AGENT · CARE</p>	<p>BIGGER SCOPE</p> <p>AI BUILD-OUT</p> <h2>£6,000</h2> <p>3–4 WEEKS · 2 MEDIUM OR 1 LARGE</p> <p><i>Two Medium agents bundled, or one Large agent (account health scoring, anomaly detection). Narrow and deep, not sprawling.</i></p> <ul style="list-style-type: none"> › 2 Medium agents OR 1 Large agent shipped › Evals at every junction, not just the output › Cost dashboards your finance team can read › Full handover: code, infra, runbooks, training session › 60 days of bug-fix support <hr/> <p>+ £50/MO PER AGENT · CARE</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

> HOW A SPRINT RUNS

WEEK 0

Scope, success criteria, and the eval set locked in writing.

DAY 1-5

Feature on staging. Evals running green.

DAY 6-10

Cost caps, monitoring, plain-English handover doc.

DAY 11+

Production. 30 days of bug-fix support on me.

06 / THE LEARNING KIT

No black box.

Your team *learns how it works.*

Every Sprint and Build-out ships with a full learning kit so your team can run, debug, and extend the system without me. The handover isn't a README — it's six concrete artifacts your team gets on day one.

> 01

Architecture walkthrough

30-minute Loom recorded during handover. Every decision explained — model choice, routing, fallbacks, why this and not that.

> 02

Annotated prompt library

Every prompt in the system, committed to your repo with comments on why it's structured that way. When it drifts, you know which knob to turn.

> 03

Plain-English runbook

Step-by-step for the ops your team will actually run: re-triggering a job, checking spend, rolling back a prompt, adding a new case.

> 04

Failure-mode playbook

What breaks, what the signal looks like, what to do. Every hallucination class and edge case hit during build — documented, not hidden.

> 05

Eval suite, runnable

The test harness lives in your repo. Change a prompt or swap a model? Re-run in one command, see the delta, ship with confidence.

> 06

Post handover support

30 days of bug-fix support on a Sprint, 60 on a Build-out. If something breaks in that window, I fix it for free.

07 / BEFORE THE CALL

Questions *I get a lot.*

The answers most people want before they'll spend 15 minutes with a freelancer. If yours isn't here, just ask it on the call.

Q · 01

We don't know what to build yet. Too early to talk?

No — that's exactly what the Roadmap is for. Or, if anything in the Agent Menu (harrybenham.dev/menu) caught your eye, we can ship that as the Roadmap deliverable instead. 100% credited toward a Sprint if you proceed.

Q · 02

How much will the AI cost to run each month?

Most agents on the menu run on £20–£100/mo of API + hosting. Per-feature spend caps go into every Sprint so a runaway loop can't run away with your card. You get a costed estimate per use case on the Roadmap.

Q · 03

Where does the agent live after we ship?

Default for most clients is Care (£50/month per agent): I host the agent on my standardised stack, monitor 24/7, fix bugs free, you never log into a hosting dashboard. New requests are quoted separately as a mini-Sprint. Cancel any month. If you have an in-house tech team and prefer Self-host, that's free — you own everything from day 30.

Q · 04

What does “fixed price” mean if scope changes?

Scope is locked in writing on Day 0 — what's in, what's out, the eval set, success criteria. New features mid-build get quoted as a small add-on Sprint, never billed surprise-style. If I under-quoted, that's my problem, not yours.

Q · 05

What if the output is wrong or the model hallucinates?

The single most important question — which is why evals are non-negotiable in every package. I build a test suite covering the critical paths. For high-stakes outputs (legal, financial, customer-facing) I add a verification layer or a human-in-the-loop step.

Q · 06

Do you need access to our codebase?

Almost never. Most menu agents run in standard channels (web chat, email, WhatsApp, Slack, your CRM via API) and don't touch your product code. If something does need deeper integration, I work alongside your engineers — you stay in control of your codebase.

Q · 07

What if our team has zero technical capability?

That's the most common case. The Care tier is built for it: I host, I monitor, I tweak, I send a monthly status email. You never log into a hosting dashboard. The agent just works — same way you don't manage your own email server.

Q · 08

Are our documents and data safe? NDA?

Yes on both. Happy to sign an NDA before discovery. Client data never leaves your infrastructure unless you explicitly want it to. I route model calls through enterprise endpoints with zero data retention where requested. UK GDPR DPA available on request.

08 / LET'S TALK

Pick a time that *works*.

Fifteen minutes, no prep needed. Tell me what you're trying to build (or that you don't know yet), and I'll tell you whether I'm a fit, what it would cost, and how long it would take.

> WHAT TO EXPECT

No slide deck

Just a conversation. Camera optional.

No hard sell

If we're not a fit, I'll say so on the call.

No follow-up spam

One email after with notes. Then it's on you.

BOOK A 15-MIN INTRO CALL

calendly.com/harrybenham173/15min

OR EMAIL

harry@harrybenham.dev

> WANT THE BUILD MENU?

See the AI Agent Menu at harrybenham.dev/menu — 50+ agents I can ship, scoped and priced.

Looking forward to hearing from you.

Harry Benham