

PUBLIC EDITION

The Practical AI Build Guide

What to automate, what to custom-build, and when private infrastructure is actually worth it.

01 WHAT

Choose the build

02 WHERE

**Choose the
environment**

03 WHO

Choose the operator

BUILD THE SMALLEST SOLUTION THAT SOLVES THE PROBLEM.

wedreamlabs.io



HOW TO USE THIS GUIDE

Start with the decision, not the technology.

You do not need to know which model, cloud, or automation platform to choose. First identify the business job, then decide how much ownership and operational responsibility are appropriate.

01	The executive model	03	07	Choose where it runs	11
02	Six service levels	04	08	Technology map	12–13
03	Start: decision and audit	05	09	Ownership and operations	14–15
04	Build: automation to product	06–08	10	Costs and decision gates	16–18
05	Own: private AI platform	09	11	Four realistic examples	19
06	Compare the project types	10	12	Routing scorecard	20

QUICK START

If the opportunity is still unclear, begin on page 20. If the project is clear but deployment is not, begin on page 11.

THE EXECUTIVE MODEL

Every AI project is three decisions.

Keeping these decisions separate prevents a simple business problem from turning into an unnecessarily expensive infrastructure project.

01

What needs to be built?

Advice, an audit, one automation, an assistant, a complete product, or a private platform.

02

Where should it run?

Managed services, cloud accounts your company owns, or a dedicated private environment.

03

Who operates it afterward?

Your team, a care plan, an embedded partner, or a shared operating model.

THE GOVERNING PRINCIPLE

Build the smallest solution that solves the problem - with only the control the business actually needs.

BUSINESS JOB → DELIVERY TYPE → DEPLOYMENT → OPERATING MODEL

THE SERVICE LANDSCAPE

From one answer to a private AI platform.

The levels describe increasing delivery scope. They are not a maturity contest: a \$2,500 automation can be the right answer for a company that does not need custom software.

01

START

AI Decision Session

\$500

FIXED · DELIVERED IN 2 DAYS

02

START

AI Opportunity Audit

\$5,000

FIXED · 5 WORKING DAYS

03

BUILD

Workflow Automation Sprint

From \$2,500

FIXED SCOPE · 1-2 WEEKS

04

BUILD

AI Assistant Sprint

\$10k-\$20k

FIXED SCOPE · UP TO 30 DAYS

05

BUILD

Custom AI Product

From \$25k

PAID SPECIFICATION · 4-8+ WEEKS

06

OWN

Private AI Platform

From \$7.5k

ARCHITECTURE · BUILDS FROM \$40K

FLEXIBLE · ADVISORY

AI Advisory Block

\$1,500

5 hours · valid for 90 days

START

Buy clarity before complexity.

Use a Decision Session for one concrete choice. Use an Opportunity Audit when the business area is known but the best AI project is not.

LEVEL 01

AI Decision Session

\$500

FIXED · DELIVERED IN 2 DAYS

DELIVERED

- Recommended tools and delivery approach
- Approximate budget, risks, and next step
- One-page decision memo within 48 hours

A founder or leader deciding what to build, buy, change, or stop.

LEVEL 02

AI Opportunity Audit

\$5,000

FIXED · 5 WORKING DAYS

DELIVERED

- Three to five opportunities ranked by value, effort, and risk
- Recommended delivery and ownership approach
- 90-day roadmap, written report, and review call

A team that wants to use AI but does not yet have one tightly defined project.

AI ADVISORY BLOCK

\$1,500

Use senior AI and product judgment when you need it, without commissioning a build.

5 hours · valid for 90 days /
Additional time at \$300/hour.

BOOK A FREE OPPORTUNITY CALL >

LEVEL 03 / BUILD

Automate one workflow - completely.

A defined process is usually the fastest path to visible value. We map the workflow, connect the tools, add controls, test failure cases, and hand it over.

From \$2,500

PUBLIC PRICE

fixed scope · 1-2 weeks

DELIVERY MODEL

One

DEFINED BUSINESS OUTCOME

WHAT YOU RECEIVE

- One clearly defined workflow
- Connections to up to three existing tools
- One approval path where needed
- Testing, error alerts, and basic recovery
- Documentation, team walkthrough, and handover
- 14-day defect warranty

GOOD FIT WHEN

- The process has a clear trigger and expected output.
- A person can explain the current steps and exceptions.
- The first version needs no custom application interface.

CHOOSE ANOTHER LEVEL WHEN

- The team needs an AI worker that makes repeated judgments.
- The solution needs users, roles, permissions, or many workflows.
- Private infrastructure is a hard requirement.

SCOPE BOUNDARY

No custom application interface, complex data migration, or unlimited workflow branches.

AUTOMATE ONE WORKFLOW

>

LEVEL 04 / BUILD

One useful AI system. Up to 30 days.

The 30-day promise applies to a focused assistant with one business job, defined integrations, clear approval points, and a measurable outcome.

\$10k-\$20k

FIXED SCOPE

30 days

MAXIMUM DELIVERY WINDOW

14 days

POST-LAUNCH SUPPORT

01	02	03	04
Define	Build	Connect	Validate
Job, users, sources, quality bar	Interface, agent flow, safeguards	Up to two business systems	Real cases, approvals, training

THE SPRINT INCLUDES

- One defined business job and success measure
- Simple interface or existing-channel integration
- Connections to up to two business systems
- Human approval and escalation where needed
- Usage logs and basic quality checks
- Production deployment, training, and 14 days of support

THE IMPORTANT BOUNDARY

The 30-day promise applies only to this focused sprint.

Multiple user roles, billing, an administration layer, or several workflows make it a Custom AI Product.

SCOPE AN ASSISTANT SPRINT

>

LEVEL 05 / BUILD

When AI becomes a real product capability.

A custom product is appropriate when the solution needs its own users, interface, permissions, data model, administration, analytics, or multiple coordinated workflows.

From \$25k

PUBLIC PRICE

paid specification • 4-8+ weeks

DELIVERY MODEL

One

DEFINED BUSINESS OUTCOME

WHAT YOU RECEIVE

- Product and architecture definition
- Custom interface, authentication, and permissions
- Database and multiple workflows or agents
- Business-system integrations
- Administration, analytics, and quality controls
- Production deployment, documentation, and handover

GOOD FIT WHEN

- The experience is used repeatedly by employees or customers.
- The system needs authentication, roles, and a durable data model.
- The business is investing in a product capability, not a quick experiment.

CHOOSE ANOTHER LEVEL WHEN

- One existing workflow can solve the problem.
- A configured assistant with two integrations is enough.
- The scope has not completed a paid specification phase.

DISCUSS A CUSTOM PRODUCT

>

SCOPE BOUNDARY

A paid specification phase is required before a fixed build quote.

LEVEL 06 / OWN

Maximum control - with the operational reality included.

Private deployment is valuable when a real data, regulatory, latency, compute, or control requirement justifies the setup and maintenance burden.

From \$7.5k

ARCHITECTURE PHASE

From \$40k

IMPLEMENTATION

6-12+ weeks

TYPICAL BUILD WINDOW

ACCESS	Identity, permissions, secure service boundaries
APPLICATION	Interfaces, workflows, agents, approvals
MODEL	Managed API or private model serving
DATA	Databases, storage, retrieval, backups
COMPUTE	Cloud, VPS, dedicated GPU, or on-premises

WHAT THE BUILD COVERS

- Client-owned cloud, VPS, dedicated GPU, or on-premises architecture
- Databases, storage, compute, and model-serving design
- Secure service access and permission boundaries
- Monitoring, backups, cost controls, and recovery plan
- Model evaluation and optional adaptation
- Runbook, technical handover, and team training

QUOTED SEPARATELY

Hardware, cloud, model usage, licenses, security certification, and ongoing operations are quoted separately.

PRIVATE IS A REQUIREMENT TO PROVE – NOT A BADGE TO BUY.

[PLAN PRIVATE INFRASTRUCTURE](#)

>

COMPARE THE BUILD TYPES

The right scope is defined by the job - not the tool.

Use this table to separate a workflow automation, focused assistant, full product, and private platform before discussing specific technology.

COMPARE	03 Workflow Automation Sprint	04 AI Assistant Sprint	05 Custom AI Product	06 Private AI Platform
STARTING POINT	Known process	Focused AI job	Product capability	Control requirement
INTERFACE	Existing tools	Simple or embedded	Custom interface	Custom or internal
CONNECTIONS	Up to 3 tools	Up to 2 systems	Multiple systems	Private services
TIMING	fixed scope · 1-2 weeks	fixed scope · up to 30 days	paid specification · 4-8+ weeks	architecture · builds from \$40k
PRICE	From \$2,500	\$10k-\$20k	From \$25k	From \$7.5k
BEST SIGNAL	Repeatable steps	Judgment required	Users and roles	Data or compute constraint

RULE OF THUMB

Add custom software only when repeat use, multiple users, or product economics justify it.

DEPLOYMENT

Choose control deliberately.

The same assistant can run through managed providers, in cloud accounts your company owns, or in a dedicated private environment. Each step adds control and operational responsibility.

01

LOWEST OPERATIONAL LOAD

Managed and fast

Established cloud tools and model providers give you the quickest route to a working solution with the least maintenance.

TECHNOLOGY EXAMPLES

Make / Zapier / Vercel / Anthropic / OpenAI

Fastest and least expensive to launch, with recurring provider costs.

02

MORE OWNERSHIP

Runs in your accounts

The application, workflow, and data are deployed in cloud accounts your company owns and pays for directly.

TECHNOLOGY EXAMPLES

n8n / Railway / PostgreSQL / Docker

More control, with more responsibility for access, billing, and operations.

03

MAXIMUM CONTROL

Private environment

Services run on a dedicated VPS, rented GPU, private cloud, or customer premises when the requirement justifies the added work.

TECHNOLOGY EXAMPLES

Coolify / vLLM / NVIDIA

Highest control and setup cost, plus an ongoing maintenance requirement.

LAUNCH SPEED	Fastest	Balanced	Slowest
CLIENT CONTROL	Core assets	Deployment accounts	Infrastructure
OPS LOAD	Lowest	Medium	Highest

GET A RECOMMENDED ARCHITECTURE >

TECHNOLOGY MAP / 1

Tools are implementation choices - not the offer.

Technology is selected after the business job, deployment requirement, and operating model are clear. These names show the range of components we can combine.

01

AUTOMATION AND INTEGRATIONS

Connect the tools you already use

Move information, trigger actions, request approvals, and keep existing systems in sync.

TOOLS AND COMPONENTS

Make / Zapier / n8n / Airtable / APIs and webhooks

02

AGENT APPLICATIONS

Give your team a focused AI worker

Use a purpose-built interface or configured agent workspace for one clear job and operating boundary.

TOOLS AND COMPONENTS

Hermes Agent / Custom agent runtime / Human approval gates

03

APPLICATION HOSTING

Give the system a reliable place to run

Choose hosting based on interface, API, worker, uptime, region, and ownership requirements.

TOOLS AND COMPONENTS

Vercel / Railway / Cloudflare / Docker / Coolify / Dokploy

04

DATA AND STORAGE

Keep business information organized and recoverable

Design for ownership, backups, retention, permissions, recovery, and appropriate retrieval.

TOOLS AND COMPONENTS

PostgreSQL / Supabase / Redis / S3-compatible storage / pgvector / Qdrant

TECHNOLOGY MAP / 2

Tools are implementation choices - not the offer.

Technology is selected after the business job, deployment requirement, and operating model are clear. These names show the range of components we can combine.

05

MANAGED MODEL APIS

Use strong AI without operating GPUs

The fastest route to capable models, with data processing governed by the selected provider's commercial terms.

TOOLS AND COMPONENTS

Anthropic / OpenAI / Google Gemini

06

PRIVATE AND OPEN-WEIGHT MODELS

Run a model in a controlled environment

Select a model family and serving layer after evaluating quality, license, hardware, latency, and total operating cost.

TOOLS AND COMPONENTS

Hugging Face / Llama / Mistral / Qwen / vLLM / Ollama

07

INFRASTRUCTURE AND GPUS

Use dedicated compute when it is justified

A private cloud, dedicated VPS, rented GPU, and on-premises machine provide different levels of control.

TOOLS AND COMPONENTS

Docker / Coolify / Dokploy / NVIDIA / RunPod / AWS / Google Cloud / Azure

08

RELIABILITY AND CONTROL

Know what happened and keep people responsible

Add logs, quality checks, alerts, access controls, backups, and approval paths appropriate to the risk.

TOOLS AND COMPONENTS

Langfuse / OpenTelemetry / Sentry / Grafana

OWNERSHIP

Know what is yours before the build begins.

Ownership is a contract and account-structure question. A private deployment does not automatically grant rights to a model's weights, and a cloud deployment does not mean the agency owns your work.

DELIVERED TO THE CLIENT

01 Project code

Custom source code created for the engagement, subject to the agreement.

02 Business data

Your source documents, records, prompts, outputs, and approved datasets.

03 Configuration

Workflows, environment definitions, schemas, and documented settings.

04 Documentation

Architecture notes, runbooks, operating procedures, and handover material.

LICENSED OR SHARED

01 Third-party platforms

Cloud tools, SaaS products, and APIs remain licensed services.

02 Model rights

Hosted models and open-weight models each have their own commercial license.

03 Pre-existing components

Reusable libraries and prior intellectual property stay under their original terms.

04 External subscriptions

Provider access continues only while the relevant account and subscription remain active.

HANDOVER CHECKLIST

☐ Accounts and access☐ Code and configuration☐ Data export☐ Runbook and recovery☐ Owner training☐ Open costs and renewals

AFTER LAUNCH

A production system always has an operator.

Before launch, assign ownership for access, provider bills, data quality, incidents, model changes, user support, and business outcomes.

01

Build and hand over

Included

Receive the project files, configuration, documentation, training, and a clean operating handover.

02

AI Care Plan

From \$1,000/month

Monitoring, updates, incident support, and a small allowance for maintenance and minor improvements.

03

Embedded AI Partner

\$8k-\$15k/month

Ongoing strategy, implementation, experimentation, and improvement with a three-month minimum.

NAME THESE OWNERS BEFORE GO-LIVE

Business owner OUTCOME AND ADOPTION	System owner ACCESS AND CONFIGURATION	Data owner QUALITY AND RETENTION
Incident owner ALERTS AND RECOVERY	Budget owner USAGE AND RENEWALS	Change owner UPDATES AND APPROVALS

[DISCUSS AN OPERATING MODEL](#)

COSTS

The project fee is not the total cost of ownership.

Separate implementation from recurring providers, infrastructure, support, and internal operations. Larger control requirements usually shift cost from vendors to engineering and operations.

01

Project fee

Discovery, design, implementation, testing, documentation, and launch.

02

Provider usage

Automation subscriptions, model tokens, email, authentication, and external APIs.

03

Infrastructure

Hosting, databases, storage, networking, monitoring, backups, and GPUs.

04

Operations

Maintenance, incident response, quality review, model updates, and user support.

RELATIVE RECURRING COST PRESSURE

Workflow automation	LOW	Platforms and connected SaaS
AI assistant	LOW-MEDIUM	Model API, data, logs, hosting
Custom product	MEDIUM	Application stack plus usage growth
Private platform	HIGH	Dedicated compute plus operations

QUOTE CHECK

Every proposal should show the build fee, expected third-party costs, who owns each account, and the selected after-launch plan. Hardware, cloud, model usage, and licenses are separate unless explicitly included.

SELF-HOSTING

Private infrastructure is justified by requirements - not fashion.

Running your own services can increase control and portability. It also creates responsibility for security, availability, upgrades, observability, backups, and incident response.

01

Is there a documented data, regulatory, latency, or dedicated-compute requirement?

YES

NO

02

Can managed providers not meet that requirement under acceptable commercial terms?

YES

NO

03

Is there a named technical owner for patching, monitoring, backups, and incidents?

YES

NO

04

Does the expected workload justify the added setup and ongoing operating cost?

YES

NO

SELF-HOSTING DOES NOT REMOVE VENDORS. IT CHANGES WHICH RISKS YOUR TEAM OPERATES.

DECISION

Four clear yes answers

Plan a private architecture phase before quoting the implementation.

Any uncertain or no answer

Start managed or client-owned, then revisit after measuring real constraints.

MODEL ADAPTATION

Fine-tune only after the failure is measurable.

Fine-tuning is one possible optimization step. It is not the default answer to missing knowledge, weak instructions, unreliable source data, or a workflow that has not been defined.

01

Define

Create a real evaluation set and a measurable quality target.

02

Baseline

Test strong existing models with clear prompts and structured outputs.

03

Add context

Use trusted retrieval, tools, examples, and workflow controls.

04

Diagnose

Identify the repeated failure that remains and quantify its business impact.

05

Adapt

Fine-tune only if data, licensing, operations, and expected improvement justify it.

PROCEED ONLY IF

- ☐ The target behavior is stable and can be scored.
- ☐ Representative, permitted training data exists.
- ☐ Prompting, context, and tools have reached a plateau.
- ☐ The improvement is worth evaluation, training, serving, and monitoring costs.

DO NOT USE IT TO FIX

- ☐ Missing or untrusted business knowledge
- ☐ An undefined workflow or weak success metric
- ☐ A one-off prompt that has not been tested
- ☐ A need that strong hosted models already satisfy

FIRST BUILD THE EVALUATION. THEN EARN THE RIGHT TO CHANGE THE MODEL.

EXAMPLES

What a buyer can actually purchase.

These ranges show how the same AI conversation can lead to very different scopes. Final pricing depends on integrations, data readiness, risk, and deployment requirements.

01

Lead routing automation

Qualify inbound leads, update the CRM, notify the right owner, and draft the first follow-up.

Workflow Automation Sprint**\$2.5k-\$5k**

Managed services / 1-2 weeks

02

Company knowledge assistant

Answer employee questions from trusted sources and route uncertain or sensitive questions to a person.

AI Assistant Sprint**\$10k-\$20k**

Managed or client-owned cloud / Up to 30 days

03

Customer-facing AI product

Launch a branded application with users, billing, data, an AI workflow, and an administration layer.

Custom AI Product**\$25k-\$75k+**

Client-owned cloud / 4-8+ weeks

04

Private document intelligence platform

Process sensitive documents with dedicated infrastructure, controlled access, monitoring, and a private model option.

Private AI Platform**\$40k-\$120k+**

Private VPS, GPU, or cloud environment / 6-12+ weeks

ROUTING SCORECARD

Choose the first statement that sounds true.

This is a routing tool, not a diagnosis. A short opportunity call can confirm the level and prevent over-scoping.

☐ I have one concrete question.	AI Decision Session	\$500
☐ We want AI but do not know where to begin.	AI Opportunity Audit	\$5,000
☐ We know the repetitive workflow.	Workflow Automation Sprint	From \$2,500
☐ We need a focused AI worker.	AI Assistant Sprint	\$10k-\$20k
☐ We need users, roles, or a complete product.	Custom AI Product	From \$25k
☐ Infrastructure control is a hard requirement.	Private AI Platform	From \$7.5k

FAST ROUTING RULE

If scope is unclear, start with clarity. If the workflow is clear, go directly to the build.

NEXT STEP

Scope the right AI build.

Book a free AI Opportunity Call. We will identify the likely service level, deployment approach, and next decision.

[BOOK THE OPPORTUNITY CALL](#)calendly.com/fede-wedreamlabs/new-meeting