

AI Build-vs-Buy Cost & Decision Worksheet

Price a telemedicine AI feature across all three cost layers, then run it through the decision questions. Engineering guidance, not legal advice — confirm specifics with counsel.

LAYER 1 · INFERENCE (the visible tip)

- Buy: per-provider license, or per-token / per-minute API rate
- Build: GPU rental or amortized hardware + power + hosting
- Worked example: 15-min visit at ~\$0.10/min audio = ~\$1.50/consult
- 2,000 visits/mo x \$1.50 = ~\$3,000/mo all-in for managed inference

LAYER 2 · INTEGRATION (just below the line)

- Capture audio from the call into the feature
- Put a draft in front of the clinician to review and sign
- Write the result back to the record as a draft
- Log every step for audit — required whether you build or buy

LAYER 3 · VALIDATION + MAINTENANCE (the deep mass)

- Accuracy + bias tested on your population (45 CFR 92.210) before launch
- Re-test on a schedule; watch for model drift after launch
- Keep the feature on the supportive side of the FDA decision-support line
- Plan retraining / update governance (FDA PCCP if cleared model)
- On-call + vendor-change handling; ML engineer ~\$200k-\$300k/yr loaded
- Most build-vs-buy analyses miss ~60-80% of this lifetime cost

DECISION QUESTIONS + LOCK-IN CHECK

- Volume high and sustained? Below a few thousand/mo -> buy
- Core differentiator, or commodity? Commodity -> buy it
- Can you staff the upkeep indefinitely? If not -> buy
- Must PHI stay on your infra? If yes -> self-host an open model
- Lock-in: could you leave if the vendor doubled price? Keep a clean seam
- Compliance is identical either way: BAA + no-training clause + PHI boundary

THE ONE-LINE RULE

Every AI feature in a telemedicine product is a build-or-buy decision, and the price you compare first is almost never the price that decides it. Buy means licensing a finished clinical-AI product or calling a vendor's model under a contract; build means assembling the feature yourself, usually on an open model you host or a general API you wrap, and owning the validation, compliance, and upkeep that come with it. The honest comparison is cost per consult over three years including integration, clinical validation, monitoring, and the engineers who keep it working — not the sticker price of a license or the per-token rate of an API. Build wins on per-unit inference and loses on everything around it until volume is very high; the validation tax and the compliance bar are line items on both sides; and a signed BAA with a no-training clause is required wherever patient data flows. Buy now behind a clean integration seam so you can move to build later if your volume justifies owning the model.