BABBEL LABS · PRODUCT PROPOSAL

# Scale the content.

# Keep the promise.

Five AI content moves, an implementation roadmap, and the quality gate that makes them safe — on infrastructure Babbel already owns.

Senior Product Manager, AI Content Creation — Babbel Labs, Berlin · [Your Name] · [Month] 2026

# The moment

The category optimized for engagement. Learners judge outcomes. That gap is Babbel's opening.

**€352M**

**Babbel revenue 2024, growing +6.6%**

Business of Apps [10]

**$1.04B**

**Duolingo revenue 2025 — 3× Babbel, +39%**

Motley Fool [6]

**–80%**

**Duolingo share price vs 2025 peak, after the “AI-first” backlash**

Motley Fool · TechCrunch [6][2]

**6 / 24**

**Pronunciation checks an AI tutor got right in expert testing**

Independent testing [8]

**Babbel owns the trust. It just doesn't have enough content.** Its own AI pipeline already cut a ~35-hr manual process and hit 85%+ editor acceptance [18].

# What learners are saying

> *I finished the level in 2–3 weeks and that was all there was. Very disappointing.*

— B2 German learner — Babbel review comments [13]

> *After A2/B1 the content starts to feel repetitive — ordering food, the weather, booking tickets.*

— Long-form Babbel review, 2025 [14]

> *Babbel dumps a ton of vocabulary on you and doesn't include it in subsequent lessons.*

— App Store review [15]

> *I have a 1,200-day streak in Duolingo and I cannot hold a conversation in Spanish.*

— Most-quoted line, 2026 cross-app analysis [1]

*Live screenshots of each review ship in the written proposal's appendix.*

# The catalog gap

**40%**

of language-specific complaints are about languages outside the top 10 [1] — learners the incumbents consistently short-change.

| 14 | 25 | 40+ | 50+ |
|---|---|---|---|
| Babbel · languages sold | Rosetta Stone · languages sold | Duolingo · languages sold | Pimsleur · languages sold |

*Japanese · Korean · Arabic · Mandarin · Hindi — demand segments Babbel currently converts at zero [13][16][17].*

# Five moves, in order

Ranked by: fixes a live complaint → runs on the existing pipeline → converts to revenue.

**P1** **Topic Packs** Refill the shelf for B1+ — “German for Job Interviews”, “Spanish for Healthcare”. LOW EFFORT · 8 WKS

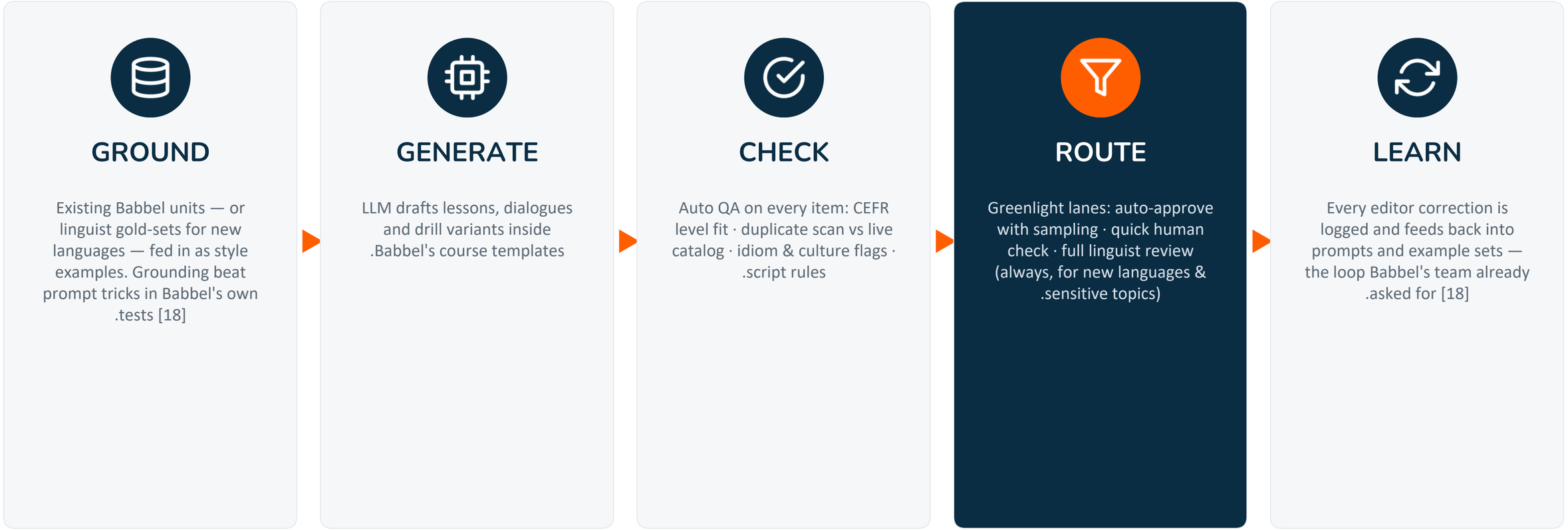
**P2** **Vocabulary Recycling** New sentences deliberately reuse words taught weeks ago. SMALLEST LIFT · WITH P1

**P3** **Honest Progress Check** “You're A2 speaking, B1 reading.” The truth, as a feature. MEDIUM · 1 QUARTER

**P4** **Dialect Packs** Mexican vs Castilian · Brazilian vs European Portuguese. MEDIUM · 2 QUARTERS

**P5** **New Languages** Japanese · Korean · Arabic · Mandarin — via linguist gold-sets. THE PRIZE · 2–3 QTRS

# How the AI is used — one pipeline, five products




**Humans never leave the loop — they move up it.** Experts author gold-sets, hold review authority, and visibly train the system. Zero replacement framing — the exact lesson from Duolingo's backlash [2][3].

# P1 Topic Packs

LOW EFFORT · HIGH IMPACT

## FIXES

*“I finished the level in 2–3 weeks” [13] — top-LTV subscribers churning off an empty shelf.*

## WHAT

AI-generated, human-approved themed packs for B1+ (“German for Job Interviews”) that reuse validated grammar and swap in fresh domains. Lowest-risk content class that exists.

## HOW AI IS USED

- Drafts grounded in existing B1/B2 units as style examples [18]
- Auto-checked: level fit + duplicate scan vs catalog
- Editors approve via Greenlight — 100% review in pilot, sampled at scale

## PROVE

Pack attach rate · renewal delta of exposed cohort · editor-hours vs 35-hr manual baseline [18].

€ `+1 month avg tenure x 100k subs x ~$8.95/mo = ~$0.9M  [12][20] · uplift = pilot ASSUMPTION`

# P2 Vocabulary Recycling

SMALLEST CHANGE IN THIS DECK

## FIXES

*“Dumps a ton of vocabulary… doesn't include it in subsequent lessons” [15].*

## WHAT

Generator receives the course's taught-word list and weaves earlier words into new sentences. “Der Bahnhof” from week 1 returns in week 4's dialogue.

## HOW AI IS USED

- Taught-vocab list injected into generation context
- Model must reuse 2–3 earlier words per new item — naturally
- Greenlight level-fit check blocks forced, awkward sentences

## PROVE

Recall accuracy on recycled vs control words · 30/60-day retention · editor naturalness rating.

€ Better outcomes -> renewal. Rides P1's cohort math; no separate spend. [18]

# P3 Honest Progress Check



### FIXES

*Category's #1 unmet demand: honest proficiency, not gamified levels [1].*

### WHAT

Optional monthly 5-min check scored against CEFR — Babbel's curriculum already is: “Speaking A2 · Reading B1 · here's what moved.” No XP. No confetti.

### HOW AI IS USED

- LLM scores writing & speech against CEFR descriptors
- Benchmarked vs human raters — agreement stats published in-product
- Abstains when unsure instead of guessing (the TalkPal lesson [8])

### PROVE

Check participation · score-delta ↔ renewal correlation · B2B adoption.

€ `+1pt first-renewal x 100k cohort = ~1,000 subs = ~$0.1M/yr [20][21] + B2B: 1,000+ companies [19]`

# P4 Dialect Packs

MARKET-TARGETED

## FIXES

*Dialect mismatch drives negative reviews in Brazil / LatAm [1].*

## WHAT

Topic-Pack machinery pointed at regional variants — Babbel already runs two Spanishes editorially; generation makes broad coverage affordable.

## HOW AI IS USED

- Same generator + regional style guides per variant
- Culture flags tuned per market
- Regional TTS voices — the added complexity that sequences this after P1

## PROVE

Rating & sentiment shift in target markets · variant adoption.

€ Market-specific growth; sized with regional funnel data. [1]

# P5 New Languages



### FIXES

*“Still lacks Japanese, Chinese and Arabic” [13]. Babbel: 14 languages. Rivals: 25–50+ [16][17].*

### WHAT

Per language: 1–2 linguists author a ~20-lesson gold set → pipeline drafts the A1 course from it → 100% human review. Launch A1 only, fronted by the named linguists who approved it. Prove it, then extend.

### HOW AI IS USED

- Few-shot grounding on the linguist gold set — no legacy corpus needed
- Script-aware QA: honorific registers · RTL layout · tone marks
- Full-review Greenlight lane, always — no sampling for new languages

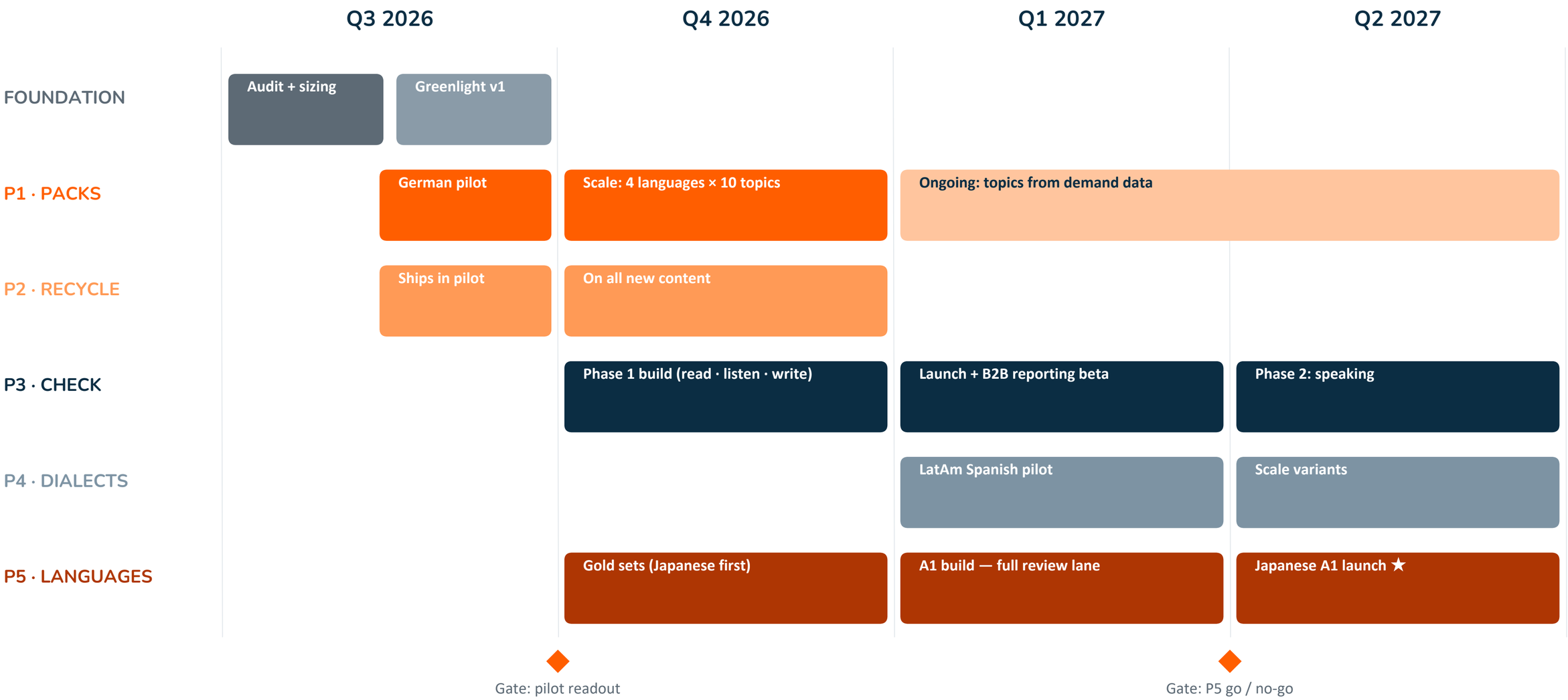
### PROVE

New-subscriber acquisition per language · A1 completion vs benchmarks · editorial rejection trend.

€ Each credible language = a demand segment converting at zero today. Sized with internal data. [16][17]

# Implementation roadmap

Four quarters from audit to a new-language launch — each phase gated by the numbers the previous one produced.



# Inside Q3 — the first 90 days

### DAYS 1–30 · GROUND TRUTH

Audit pipeline acceptance & edit-rate data by content type. Interview editors on where review time really goes. Pull search + ticket data to rank Topic Pack demand and size new-language demand (Lever 4).

**→ Baseline metrics memo**

### DAYS 31–60 · PROVE IT SMALL

Ship the 3-topic German pilot with recycling on, behind Greenlight, to a measured holdout of B1+ learners.

**→ Live pilot + dashboard: completion, editor-time, cohort renewal**

### DAYS 61–90 · DECIDE WITH EVIDENCE

Read out the pilot. Green-light P1/P2 scale-up and P3 phase-1. Present the P5 gold-set plan with the day-1–30 demand sizing.

**→ Roadmap argued from the pilot's own numbers**

# Where the money is

Babbel is private — impact is written as formulas on public anchors, re-runnable with internal numbers.

### Lever 1 · Tenure (P1+P2)

```
+1 month tenure x 100,000 subs x ~$8.95/mo  ≈  $0.9M
```

PUBLIC: >18-mo avg relationship [12]; ~$8.95/mo annual plan [20]. ASSUMPTION: ≥1-month uplift — what the P1 holdout measures.

### Lever 2 · First renewal (P3)

```
+1 pt renewal x 100,000 cohort  ≈  1,000 subs  ≈  ~$0.1M/yr
```

PUBLIC: only ~half of education-app subscribers survive first renewal [21]. ASSUMPTION: honest progress moves renewal low single digits [1].

### Lever 3 · Unit economics (all)

```
100 generated units  ≈  3,500 expert-hours redeployed to review & gold-sets
```

PUBLIC: ~35 hrs/unit manual; 85%+ acceptance on Babbel's pipeline [18]; Duolingo at 20,500 units/qtr [5].

### Lever 4 · New markets (P5)

```
14 languages today vs 25–50+ at rivals  →  segments converting at zero
```

PUBLIC: catalog gap [13][16][17]. Deliberately no point estimate — sizing is a first-30-days task.

### Lever 5 · B2B (P3)

```
CEFR reporting  →  upsell surface on 1,000+ company accounts
```

PUBLIC: Babbel for Business, >100% revenue renewal [19]. Enterprise buyers need training-ROI evidence.

# Risks — and why the plan already handles them

| Risk | It's real | Built-in answer |
|---|---|---|
| Quality regression (“AI slop”) | Duolingo backlash; complaints shifted to “bad AI content” [1][2] | Greenlight lanes; 100% review for new languages; published human-agreement stats (P3) |
| Expert resistance | Contractor framing became Duolingo's crisis [2][3]; Babbel's own team saw early skepticism [18] | Experts own gold-sets & review authority; their edits visibly train the system; zero replacement framing |
| Wrong feedback (P3) | 6-of-24 accuracy in AI-tutor testing [8] | Abstain-when-unsure thresholds; human-rater benchmark before launch |
| New-language risk (P5) | No Babbel corpus to ground on [18] | Linguist gold-sets as grounding; A1-only launch; script-aware QA |

# AI-first in how it's built.

# Human-first in what learners feel.

Duolingo collapsed that distinction and paid in trust and market value [2][6]. This plan is engineered to keep it — and every claim in it is public, footnoted, and re-runnable.

*Accompanying: 11-page written proposal (21 public sources, screenshot evidence log) · Greenlight — clickable review-gate prototype*

# Appendix - sources

All public. Full URLs and the screenshot evidence log ship in the written proposal.


[1] Unstar 2026 — cross-app review analysis
[2] TechCrunch, Aug 2025 — Duolingo AI-memo walk-back
[3] Entrepreneur 2025 — CEO clarifies AI stance
[4] Fortune, Apr 2026 — AI-usage metric dropped
[5] Invezz, May 2026 — 20,500 units in Q1 2026 vs 7,100/qtr 2025
[6] Motley Fool, Mar 2026 — DUOL −80%; FY25 $1.04B rev, $414M profit
[7] TheStreet, May 2026 — gross margin 73% → ~69% on AI costs
[8] Oh Yeah Sarah (Medium) 2025 — TalkPal testing, 6/24 accuracy
[9] LanguaTalk 2026 — Speak scoring leniency (competitor-authored)
[10] Business of Apps 2026 — Babbel €352M 2024 (+6.6%)
[11] Babbel press 2026 — 25M+ subs sold; ~60k lessons; ~200 experts
[12] Babbel press 2020 — >18-mo avg relationship; 1.5 subs/user
[13] Mezzofanti Guild — B2 content exhausted; missing JA/ZH/AR
[14] Padmanabhan (Medium) 2025 — repetitive after A2/B1
[15] JustUseApp — vocabulary dumping; speech-timing complaints
[16] Intellectia 2026 — Duolingo 40+ languages
[17] Test Prep Insight — Rosetta 25; Pimsleur 50+ languages
[18] ZenML LLMOps DB — Babbel AI pipeline: 35 hrs/unit; 85%+ acceptance
[19] Babbel press 2022 — B2B: 1,000+ companies; >100% renewal
[20] Babbel public price lists 2025–26 — ~$8.95/mo annual plan
[21] Business of Apps 2026 — ~half survive first renewal (edu apps)