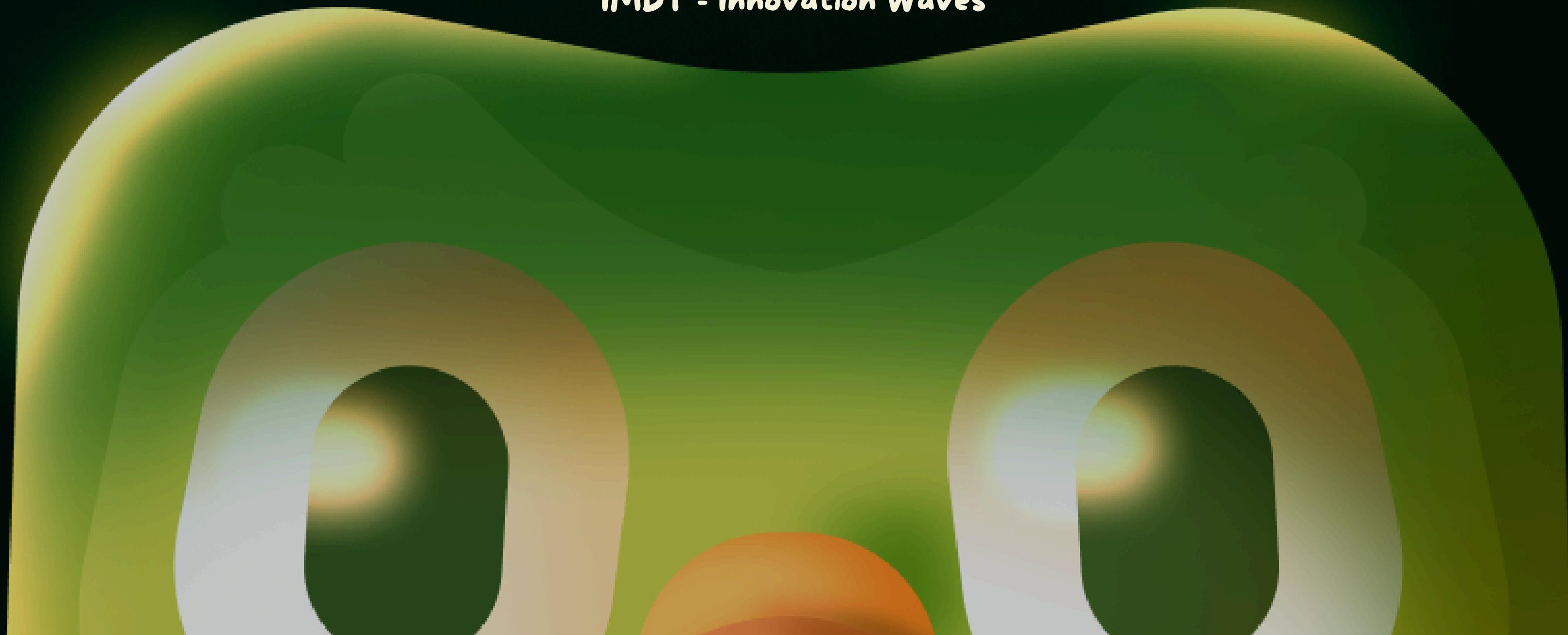


Duolingo & OpenAI

Bora - Carolina - Winter
IMDT - Innovation Waves



Duolingo Max: AI-Powered Language Learning Cost/Benefit Analysis



WHY DUOLINGO:

For our assignment, we chose Duolingo as the reference business case. Duolingo is an ideal example because it has **already integrated AI** models (GPT-4) into its product (Duolingo Max), and the company **publicly communicates the process and the benefits** of this integration.

Goals

- Assess whether Duolingo's AI initiatives are **worth the investment**
- Evaluate **how the Cost/Benefit** criterion **influences** Duolingo's decisions across the following three AI layers:

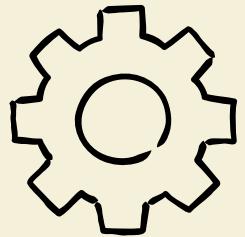
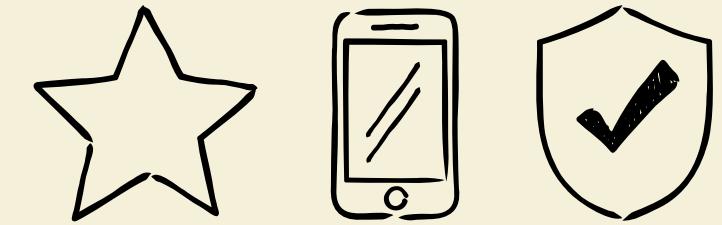
AI Layer	Focus
Definition & Foundations	How Duolingo conceptualizes AI
AI Value Chain	What are the cost and value drivers along Duolingo's AI Value Chain
Make vs Buy	Duolingo's strategic AI acquisition decisions
AI Factory	Duolingo's operational system turning costs into scalable benefits

How Duolingo conceptualizes AI?

AI as a **powerful generative technology** enabling new premium features

“Duolingo Max is a new subscription tier... Both features are powered by the newest, most powerful generative AI technology.”

<https://blog.duolingo.com/duolingo-max/>



AI as a scalable **content creation engine**

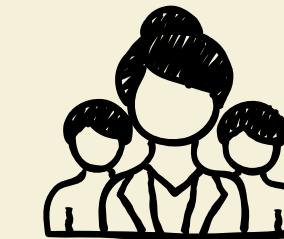
“With a powerful Large Language Model in the hands of Duolingo’s teaching experts, we can generate vast amounts of content for our lessons with the click of a button.”

<https://blog.duolingo.com/large-language-model-duolingo-lessons/>

AI as a **virtual tutor** for personalized learning

“Since its founding in 2011, Duolingo’s mission has been to create a ‘virtual language tutor’ that makes language learning available to anyone, anywhere, at any time.”

<https://www.forbes.com/sites/bernardmarr/2023/04/28/the-amazing-ways-duolingo-is-using-ai-and-gpt-4/>



AI as a way to **scale courses and automate routine work**

“We use AI to accelerate content production, scale courses and automate certain tasks, allowing our human experts to focus on more creative and strategic work.”

<https://www.forbesindia.com/article/news/ai-was-a-factor-for-some-of-our-layoffs-but-not-all-duolingo/92745/1>

Artificial intelligence has always been a huge part of our **strategy**, says principal product manager Edwin Bodge.

“WE HAD BEEN USING IT FOR PERSONALIZING LESSONS AND RUNNING DUOLINGO ENGLISH TESTS. BUT THERE WERE GAPS IN A LEARNER’S JOURNEY THAT WE WANTED TO FILL: CONVERSATION PRACTICE, AND CONTEXTUAL FEEDBACK ON MISTAKES.”

definition

Realistic Cost Estimate for Duolingo's OpenAI Enterprise Contract (2025)

Duolingo's 2025 User Scale

According to Duolingo's 2024–2025 investor reports and public statements:

88 million Monthly Active Users (MAU)
7.4 million paid subscribers, with Max

Duolingo AI Feature Token Costs (Per Active Max User)

A) Explain My Answer

Each explanation = 1,000–2,000 input tokens

With structured pedagogy + examples = \approx 4,000–6,000 output tokens

5k–8k tokens per explanation

B) Roleplay (conversation simulation)

Typical session = 15–25 back-and-forth turns
Each turn \approx 300–500 tokens

5k–12k tokens per session

Combined (daily usage): Most Max users trigger 8k–15k tokens/day.

Assumptions Used for Cost Modeling

These assumptions follow token-usage models published by:

MIT Technology Review (LLM cost modeling, 2024)
Stanford HAI "Foundation Model Cost Index"

OpenAI public token pricing for GPT-4 and Realtime API
Bloomberg interviews with enterprise AI adoption teams

They match what other large-scale enterprise customers (e.g., Stripe, Khan Academy, Morgan Stanley) experience with GPT-4-level models.

2025 Duolingo Max Active User Estimate

Assume a conservative scenario:

Paid users: 7.4M

Max tier share: 15% (industry estimate + past upgrades)

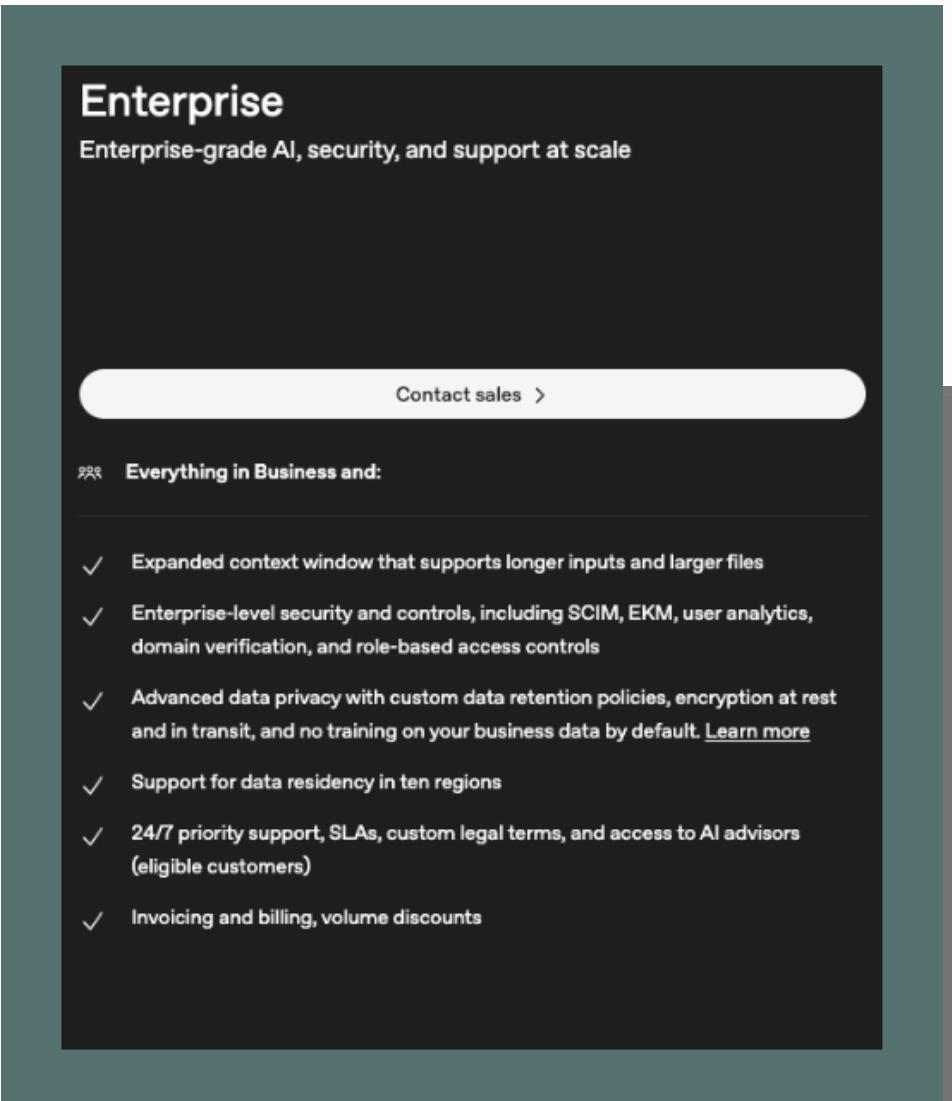
\approx 1.1 million active Duolingo Max users

But daily active Max users \approx 25–35%

\approx 300k–400k daily Max users

We will use 350,000 daily active Max users for modeling.

pricing



Annual Cost Calculation

Total Token Consumption (Realistic Range) Daily tokens

350,000 users × 8k–15k tokens

2.8B – 5.25B tokens per day

Monthly tokens:

≈ 84B – 157B tokens per month

Annual tokens:

≈ 1.0 – 1.8 trillion tokens per year

LOW estimate (best-case contract)

1.0 trillion tokens × \$0.10 per 1M

= \$10 million per year

HIGH estimate (heavier usage + less aggressive discount)

1.8 trillion tokens × \$0.30 per 1M

= \$25–30 million per year

Cost: GPT-4.1 / GPT-4 Turbo Enterprise Pricing

Enterprise pricing is not public, but based on:

OpenAI published API pricing

Bulk discounted enterprise rates

Reported costs from enterprise customers (Stripe, Khan Academy, SAP)

A realistic enterprise token price for GPT-4-class models:

≈ \$0.10 – \$0.30 per 1M tokens (with discount)

(Standard API rates: \$10–\$60 per 1M tokens; enterprise = ~50–100× cheaper)

What Is Included in the Enterprise Contract

According to OpenAI Enterprise documentation and customer reports, Duolingo's contract would cover:

- GPT-4.1 / GPT-4-Turbo inference at discounted enterprise rates
- Realtime / voice API usage (for speaking exercises)
- Custom data retention + encrypted logs
- SLA guarantees (uptime, latency)
- Rate-limit removal
- Expanded context windows
- Dedicated support + technical advisors
- Regional data residency options
- These are standard components of OpenAI Enterprise agreements.

Higher Premium Conversions & Revenue Growth (Duolingo Max)

AI-powered features (Roleplay, Explain My Answer) enable Duolingo to sell a more expensive premium tier.

<https://blog.duolingo.com/duolingo-max/>

Increased Engagement → Higher Customer Lifetime Value

AI-driven feedback and conversations significantly increase time-on-app and user stickiness.

<https://spectrum.ieee.org/duolingo>

Lower Content Production Costs (Faster Course Creation)

AI accelerates lesson generation and grammar explanations, reducing reliance on human writers.

<https://foralink.io/blogs/duolingo-s-ai-innovations-transforming-language-learning-and-beyond>

Operational Efficiency (Reduced Human Labor Costs)

AI automates internal reviewing, QA steps, and support tasks, reducing operational expenses.

<https://raw.studio/blog/the-end-of-duolingo-how-chatgpt-4-is-taking-over-language-learning>

Competitive Advantage

GPT-4-based features differentiate Duolingo from competitors and strengthen its market position.

<https://www.forbes.com/sites/bernardmarr/2023/04/28/the-amazing-ways-duolingo-is-using-ai-and-gpt-4/>

Rapid Scalability

AI enables Duolingo to expand globally while keeping the cost structure lean and scalable. (More Users Without Proportional Staff Costs)

<https://openai.com/index/duolingo/>

New Product Possibilities (AI Tutor, Max Features)

AI makes possible entirely new features that unlock new monetization paths.

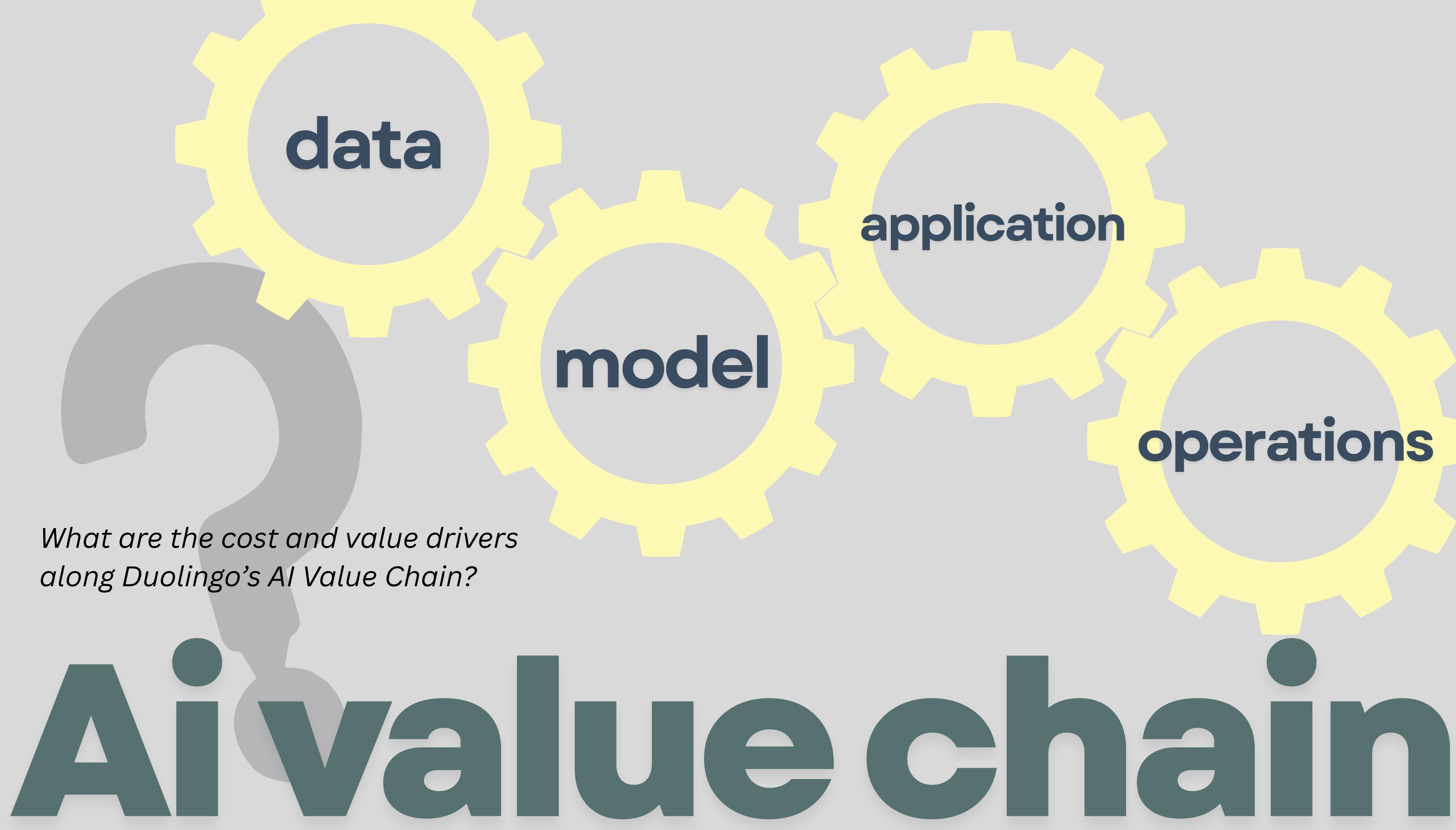
<https://blog.duolingo.com/duolingo-max/>

Brand & PR Boost

AI integration elevates Duolingo's brand as an innovator and attracts media attention.



benefits



The Inputs That Train & Improve Duolingo Max

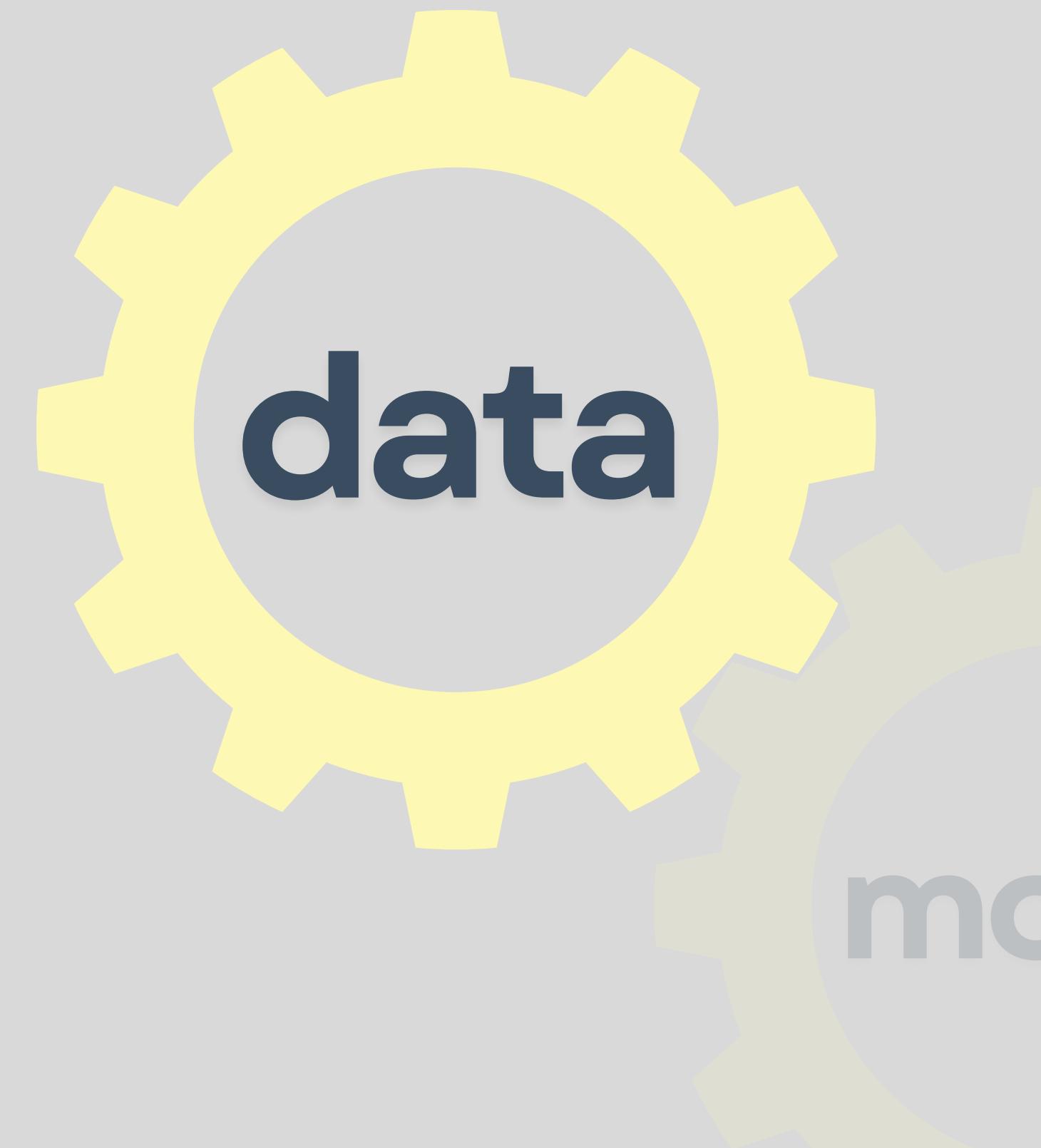
- Learner responses, mistakes, translations, speech samples
- Conversation logs from Roleplay sessions
- User metadata (level, progress, language)
- Curated grammar rules & linguistic content from experts

Value:

- Higher-quality data = more accurate explanations, fewer errors
- Drives personalization → increases retention and satisfaction

Cost drivers:

- Storage of millions of daily interactions
- Data cleaning & linguistic annotation for 40+ languages
- Privacy, governance, and language-specific edge cases



The Intelligence Behind Duolingo Max

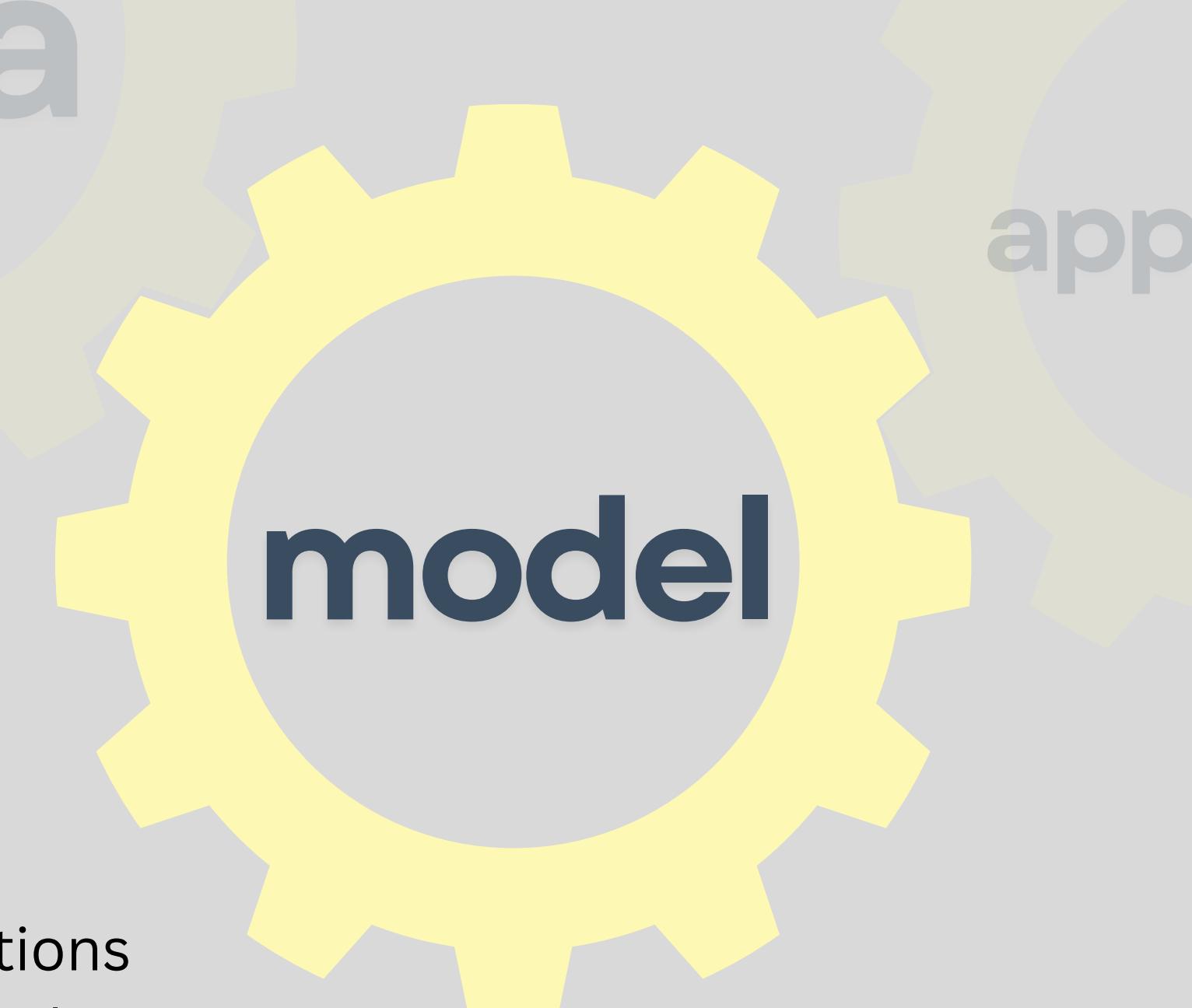
- **GPT-4** Enterprise for core generative tasks
- Smaller **internal models** for speech scoring and personalization
- Prompt engineering + guardrails for safety and pedagogy

Value:

- GPT-4 enables natural conversation and detailed explanations
- Internal models reduce costs for repetitive, predictable tasks

Cost drivers:

- **Token** usage (hundreds of billions per month)
- GPT-4 Enterprise **contract** (\$10M–\$30M/year at current scale)
- **Engineering time** for prompts, fine-tuning, and safety filters



Duolingo Max features

- Roleplay: AI conversation partner adapted to scenarios
- Explain My Answer: instant grammar explanation after errors
- AI conversations / video call practice: near real-time tutoring
- Max-only UI, gamified interactions, lesson flows

Value:

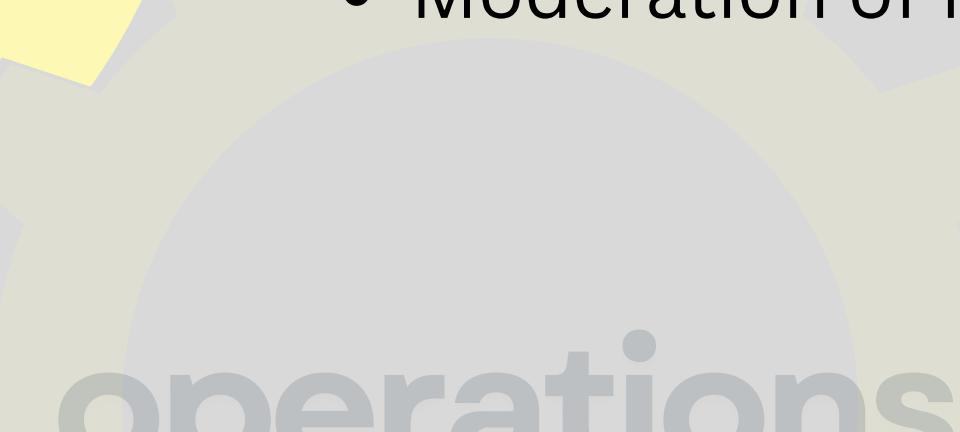
- Unlocks a new premium tier → Duolingo Max
- Increases engagement, time-on-app, and conversion to paid users
- Differentiates Duolingo from competitors

Cost drivers:

- Engineering integration into the app
- UX design, testing, and quality assurance
- Moderation of incorrect or unsafe AI outputs



application



operations



How Duolingo keeps Max reliable and affordable

- Monitoring AI accuracy across 40+ languages
- A/B testing every model update
- Routing requests between GPT-4 and cheaper internal models
- Optimizing prompt lengths to reduce token usage
- Maintaining uptime for millions of learners

Value:

- Scales to millions of Max users without proportional staff growth
- Reduces per-user cost over time
- Ensures consistent quality and safe AI behavior

Cost drivers:

- Cloud compute & inference costs
- Safety pipelines and human review
- Observability & monitoring infrastructure

make vs buy

Buy first: Closed + RAG



Why Duolingo chose to “Buy” (GPT-4 Enterprise)

- Training a large, GPT-4-level model internally would cost \$50M–\$200M and require years of R&D.
- Fine-tuning and maintaining multiple models for 40+ languages would add \$10M–\$20M/year.
- GPT-4 already delivered high-quality Roleplay + Explain-My-Answer → zero development delay.
- Buying allowed Duolingo to launch Max immediately and validate demand without committing to massive fixed costs.

Real usage + cost numbers

- Estimated Max daily active users (2025): ~350,000
- Daily tokens: 2.8B – 5.25B
- Annual tokens: 1.0 – 1.8 trillion
- Enterprise GPT-4 price: \$0.10 – \$0.30 per 1M tokens
- Annual cost: \$10M – \$30M

→ BENEFIT of Buying = lower upfront cost, fast launch, predictable variable spending.

Make later? Open Fine-tuning + RAG

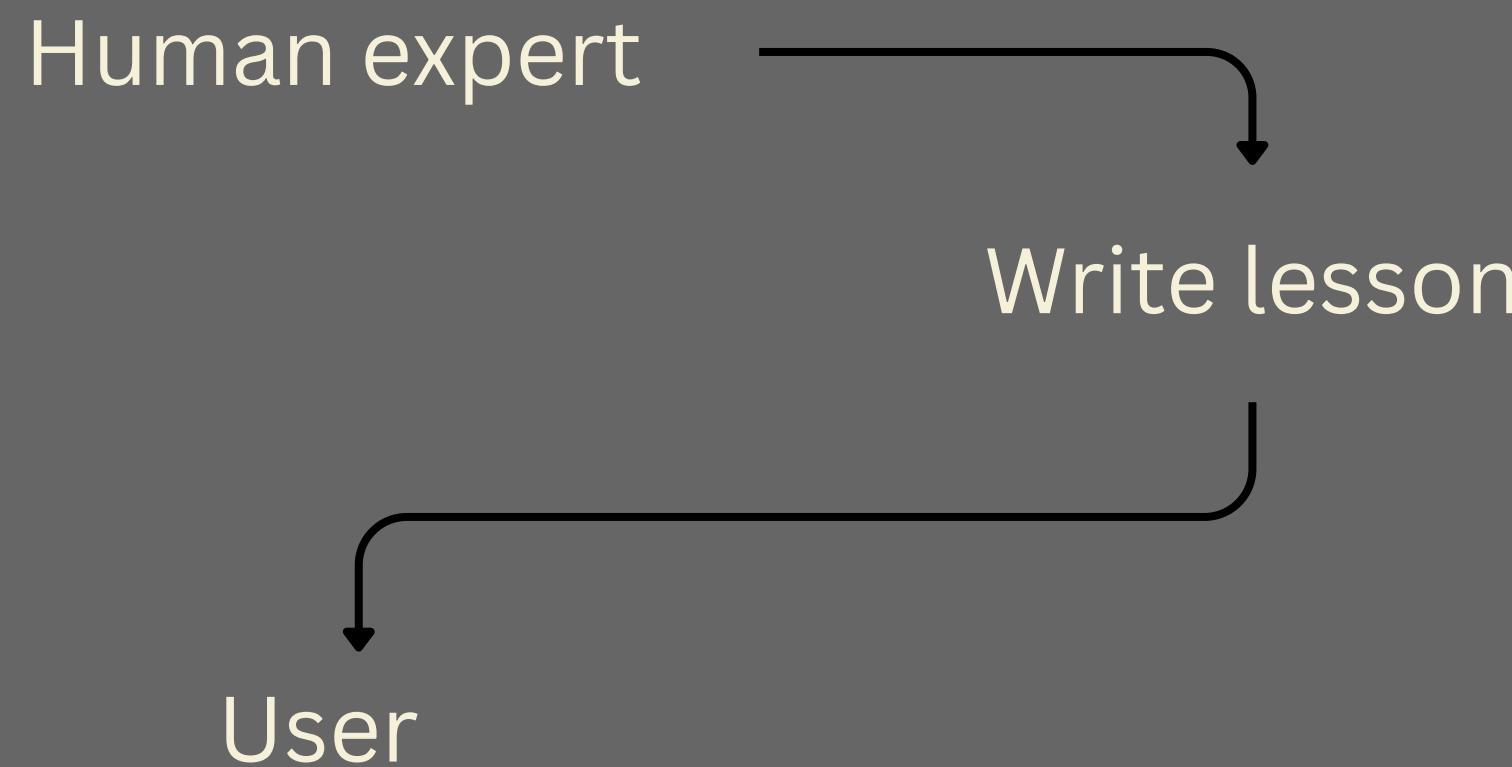
Why shifting to “Make” may become attractive later

- If Duolingo Max adoption grows 2–3×, API costs rise proportionally:
 - 2× users → ~\$50M/year
 - 3× users → ~\$60M–\$80M/year
- A custom 1B–3B model optimized for Duolingo costs:
 - One-time training: ~\$20M–\$40M
 - Annual operating cost: \$5M–\$10M/year
- Internal models can be optimized for language-learning:
 - Shorter, structured outputs → fewer tokens
 - Lower latency + better control over tone & pedagogy
 - No dependency on external pricing or availability

➔ Once Duolingo reaches high scale, API costs outgrow the cost of running their own model.

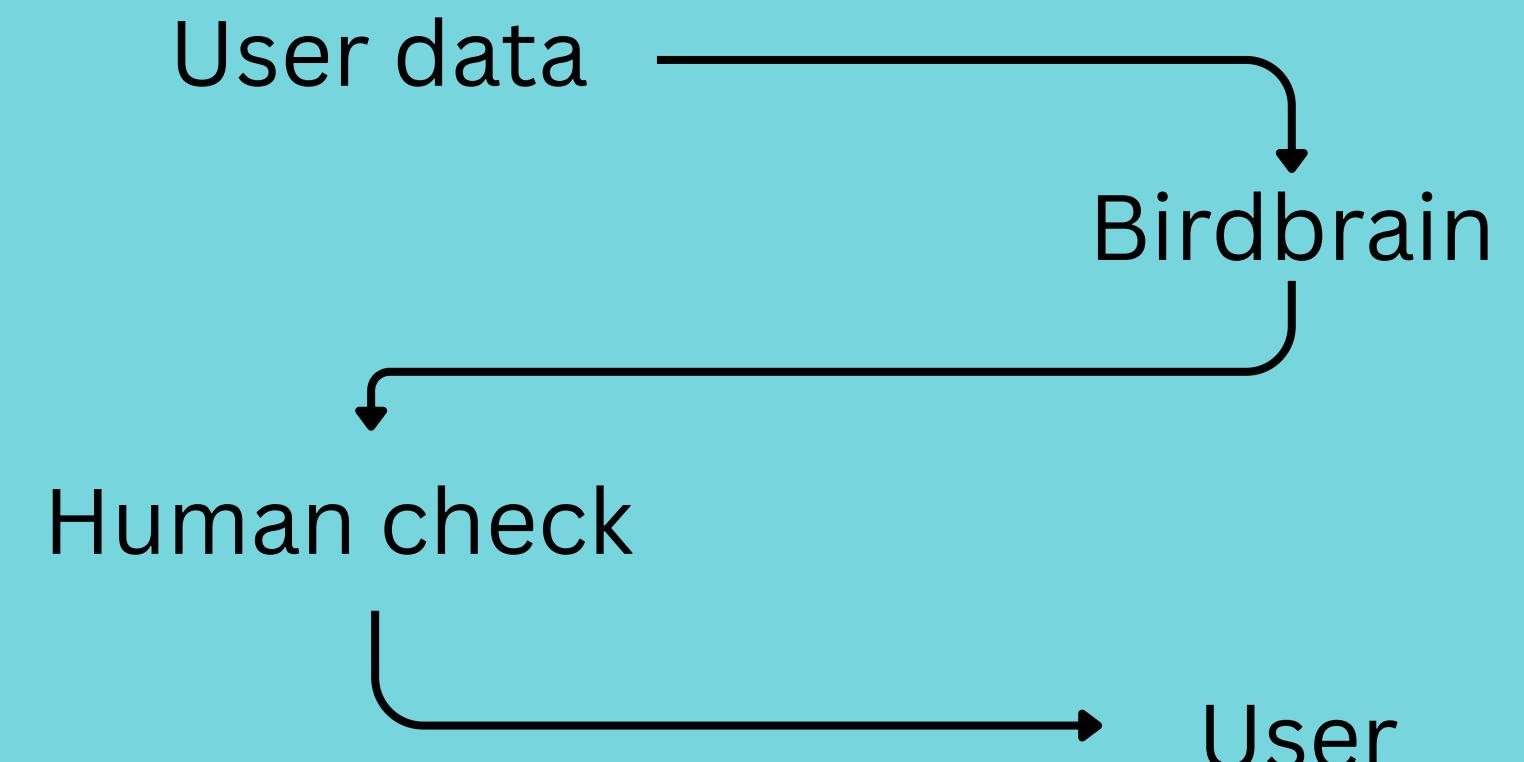
the Ai factory

Craftmanship



First 100 courses = 12 years

AI Factory



150 courses = 1 year

Cost

Workforce: -10% contractor,
Role shifted from creators to Auditors

Quality Risk: AI Hallucinations
need human oversight

Variable cost: High increasing fees
for OpenAI APIs

Summary: Duolingo's **AI Factory** replaces the **artisan** with the **algorithm**, trading the human **cost of creative labor** for the superhuman benefit of **infinite, personalized scale**.

Benefits

Hyper scale: Horizontal scaling of
new niche languages

Margins: 73% gross margins driven
by automation

Revenue: New Max tier monetizes
AI features directly

<https://investors.duolingo.com/static-files/c9bf5861-b19d-4396-b060-c0dbeef34f5>

<https://www.washingtonpost.com/technology/2024/01/10/duolingo-ai-layoffs/>

Thank You for Listening
