

Video Call for every Super subscriber: run it on the mini model with a monthly allowance

To Duolingo's CFO (written from public data only)
From Gary Wu
Date 24 October 2026
Basis Duolingo's 10-K, 10-Qs and shareholder letters to Q2 2026; OpenAI's price lists as archived

Recommendation. Extend Video Call to existing Super subscribers on a mini real-time model with a 20-minute monthly allowance, about 5 calls of 4 minutes each, and keep the flagship model and unlimited calls as what Max adds. In the base case that holds gross margin at 71.37%, above the 71.0% Duolingo guided for Q3 2026; giving Super the calls Max has today would take it to 64.91%.

What is being decided. In February Duolingo said it would "move Video Call to be part of the Super Duolingo tier", reaching "about ten times as many learners", and in August that it expects "to extend that access to existing Super subscribers later this year". At the Q2 2026 run-rate of \$298.45m revenue and 12.60m average paid subscribers, that is 11.34m Super subscribers; if 30% call in a month, 3.40m callers. Margin was 72.78% in Q4 2025, before Super had the feature, so the room down to the target is \$5.31m a quarter, or \$0.52 per caller a month.

Option	AI cost per caller a month	Cost a quarter	Margin, base	Bear	Bull
A · Flagship, unlimited	\$2.30	\$23.48m	64.91%	53.0%	69.4%
B · Mini, unlimited	\$0.72	\$7.32m	70.33%	66.7%	71.7%
C · Flagship, 20 minutes	\$1.32	\$13.47m	68.27%	63.3%	70.5%
D · Mini, 20 minutes (recommended)	\$0.41	\$4.20m	71.37%	69.9%	72.1%
E · Flagship, unlimited, price up	\$2.30	\$23.48m	65.68%	54.9%	69.7%

Why the mini model, and why an allowance. One call of 4 minutes costs \$0.3771 on OpenAI's GPT-Realtime-2 list and \$0.1175 on gpt-realtime-mini. About 38% of the flagship cost is the model reading the conversation again before each reply from Lily, the character learners talk to, a cost that grows with the square of the call's length. Even with every Super call on the mini model, unlimited use costs \$7.32m a quarter and leaves margin at 70.33%, because the heaviest 15% of callers use 80 minutes a month at \$0.0294 a minute. The allowance cuts the average caller from 24.4 minutes to 14.0 minutes and the cost to \$4.20m.

The options I rejected. The same allowance on the flagship model leaves margin at 68.27%; to hold the target the flagship needs an allowance of 5.5 minutes, less than two calls, which is not a feature. Raising the Super price does not fix a margin rate: a rise of 13.9% keeps gross profit dollars whole but leaves margin at 65.68%, and holding the rate would take 440%, because each extra dollar carries its own costs. Unlimited calls on the mini model fall short unless Duolingo pays 72.6% of list or less, and its contract price is not public.

Does the cost model match Duolingo's own numbers. Gross margin fell 1.91 points from Q1 2024 to Q1 2025, \$4.41m, "mainly due to lower subscription gross margin as a result of increased generative AI costs" for Max. On the GPT-4o list in force then (\$0.5340 a call) and at 50% of 2026's use per caller (Duolingo has since "more than doubled the average number of words spoken per user in Video Call"), the model puts Max's Video Call at \$2.90m, 66% of the drop; the model should come in below the drop, since it leaves out Max's other AI features such as Explain My Answer and Roleplay. In February Duolingo expected to give up 3.78 points for "expanding access to AI-powered features for all users". The model gives 2.45 points on the mini model and 7.87 points on the flagship, so at base-case usage the outlook, which also covers voice answers for all learners, is much closer to a mini-class cost than to the flagship.

What would change the view. The plan holds the target until more than 37.9% of Super subscribers call in a month, or until less than 65.5% of the re-read conversation is served from the cache (base cases 30% and 80%); these are the two drivers that move the margin most. Read both monthly from the first rollout cohort. If either is crossed, step the allowance down to 15 minutes, which holds up to 46.2% adoption. The model leaves out voice answers for all learners

and any change to Max, and Duolingo does not disclose Max subscribers, adoption or call length, so those are judgments, stated in the workbook.

Sources: Duolingo's filings on SEC EDGAR (FY2025 10-K 0001628280-26-012494; 10-Qs to 0001628280-26-053603; shareholder letters filed as EX-99.2 to the earnings 8-Ks); OpenAI's API pricing page and announcements as archived by the Internet Archive (snapshots of 1 October 2024 to 1 June 2026). Every number is read from the workbook Duolingo_Video_Call_Unit_Economics.xlsx after Excel recalculates it (All 22 checks OK). Duolingo is not involved; nothing here is inside information or investment advice.