

Played Bethel's Yup'ik, the model most often answered across the sea.



META MMS-LID, 4,017 LABELS • KYUK CLIPS

111 to 110.

Played Yup'ik clips from KYUK's broadcasts, Meta's MMS-LID answered Central Siberian Yupik 111 times and Yup'ik 110. That language is spoken on St. Lawrence Island and in eastern Russia.

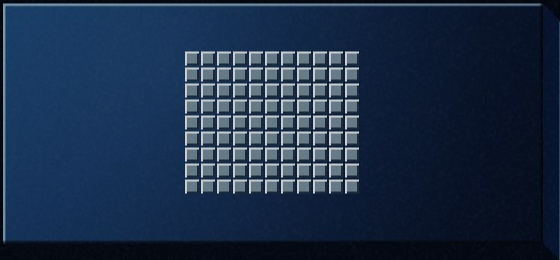


WHAT EACH MODEL CAN NAME

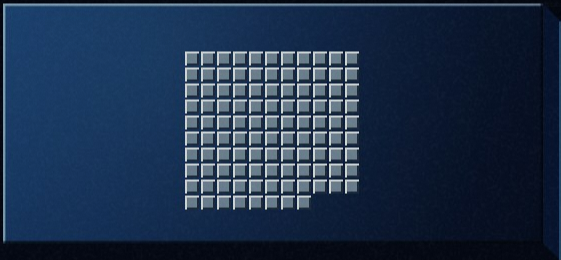
Three big open models list no Yupik at all.

Only Meta's largest model lists Yup'ik, one label in 4,017.
The authors say a label space that big makes
out-of-the-box results poor for languages with little data.

WHISPER LARGE-V3
99 LANGUAGES



VOXLINGUA107
107 LANGUAGES



ESPNET OWSM V4
ABOUT 150 LANGUAGES



META MMS-LID • 4,017 LABELS



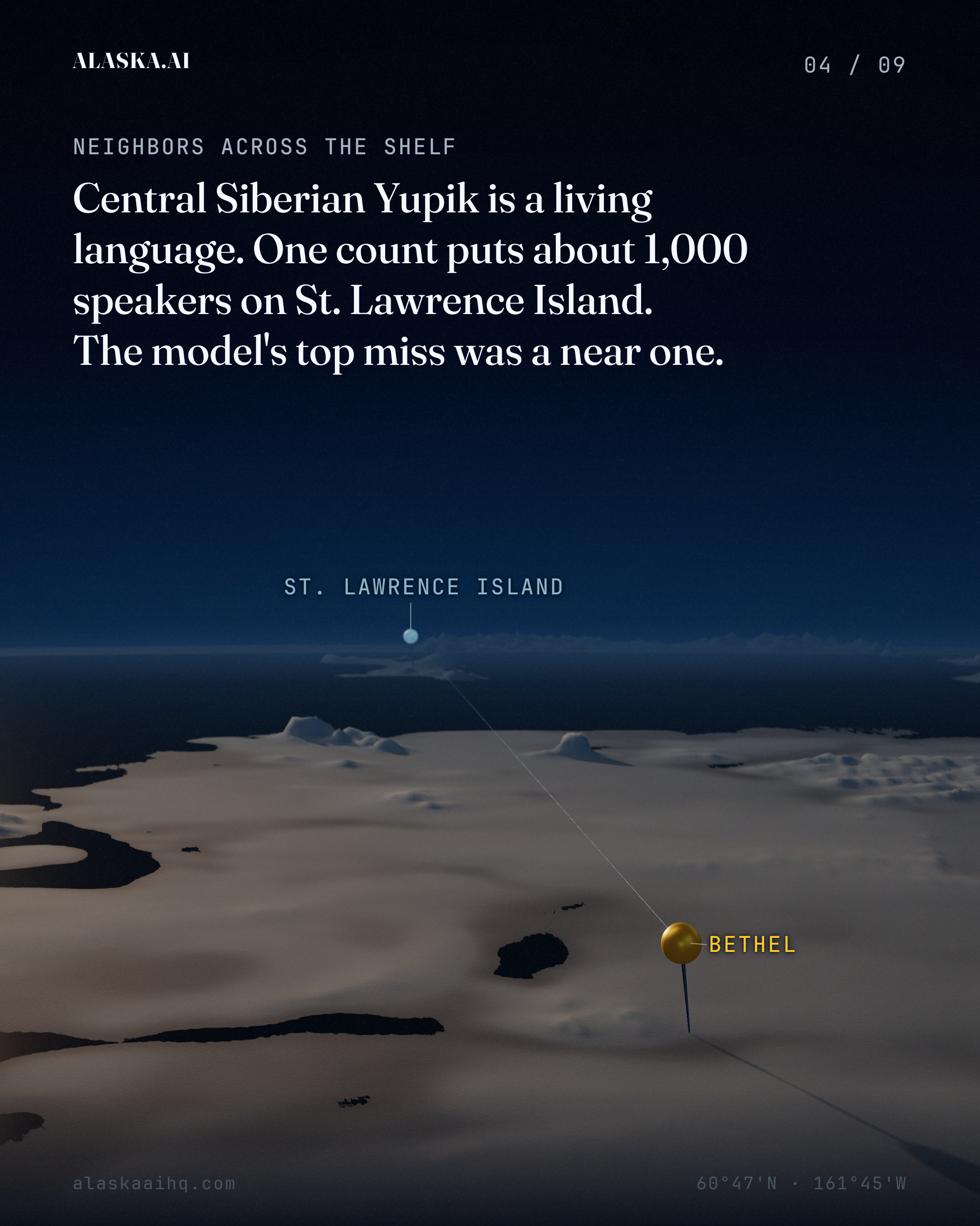
ESS • Central Siberian Yupik

ESU • Yup'ik

ONE TILE = ONE LANGUAGE A MODEL CAN NAME

NEIGHBORS ACROSS THE SHELF

Central Siberian Yupik is a living language. One count puts about 1,000 speakers on St. Lawrence Island. The model's top miss was a near one.



KYUK · BETHEL

The test audio came from Bethel's own station.

KYUK airs about an hour of local news in Yup'ik a day. The researchers labeled 9 of its videos, 5.1 hours, through the American Archive of Public Broadcasting.



A FOUR-LABEL PILOT

The labeled hours lifted recall to 0.92.

Trained on labeled Yup'ik and Samoan clips, a small classifier on a frozen Whisper encoder lifted Yup'ik recall from 0.41 to 0.92. It chooses among four labels, not 4,017. The authors call it pilot-scale.



ST. LAWRENCE ISLAND

PER 100 YUP'IK CLIPS

MODEL	LABELS	RECALL
META MMS-LID	4,017	0.41
PILOT HEAD	4	0.92

WHO LABELS THE NEXT HOURS

The next labels come from speakers.

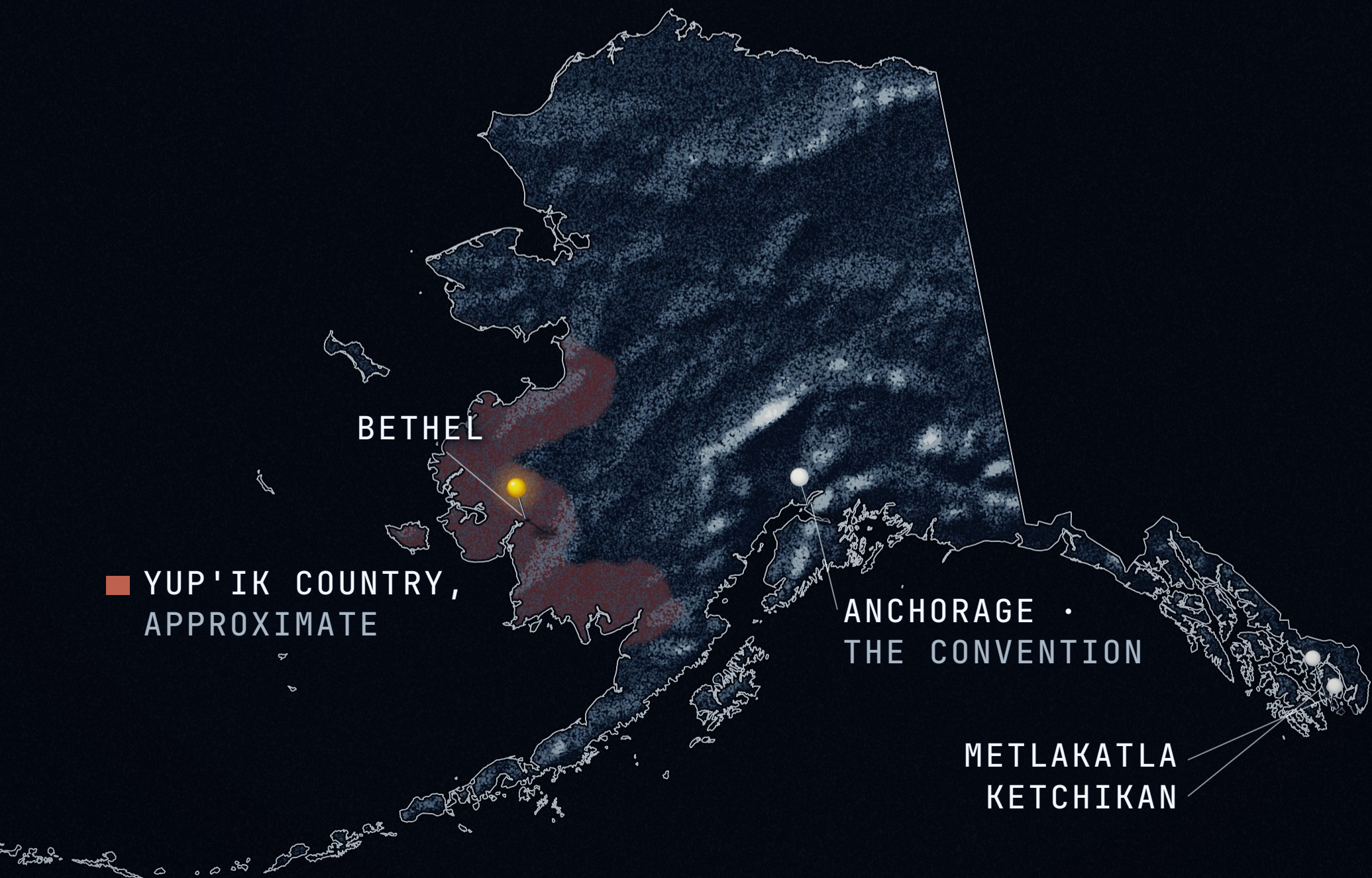
The pilot's labels were written by in-house researchers, not native speakers. The authors plan to hand that work to trained native-speaker annotators, and community language fellows have begun, the paper says.



SATURDAY, OCTOBER 24TH • RESOLUTIONS SET FOR 10:15 A.M.

AFN is set to take up 26-20.

From the Metlakatla and Ketchikan Indian Communities, it seeks permanent, non-competitive federal language funding for every eligible Tribe. Technology and archives are among its uses. It never mentions AI. Committee and board say do pass.





Whoever labels the hours draws the map.

*What does a model builder owe the
speakers who write the next labels?*

AFN SET TO TAKE UP 26-20 SATURDAY, OCTOBER 24TH