

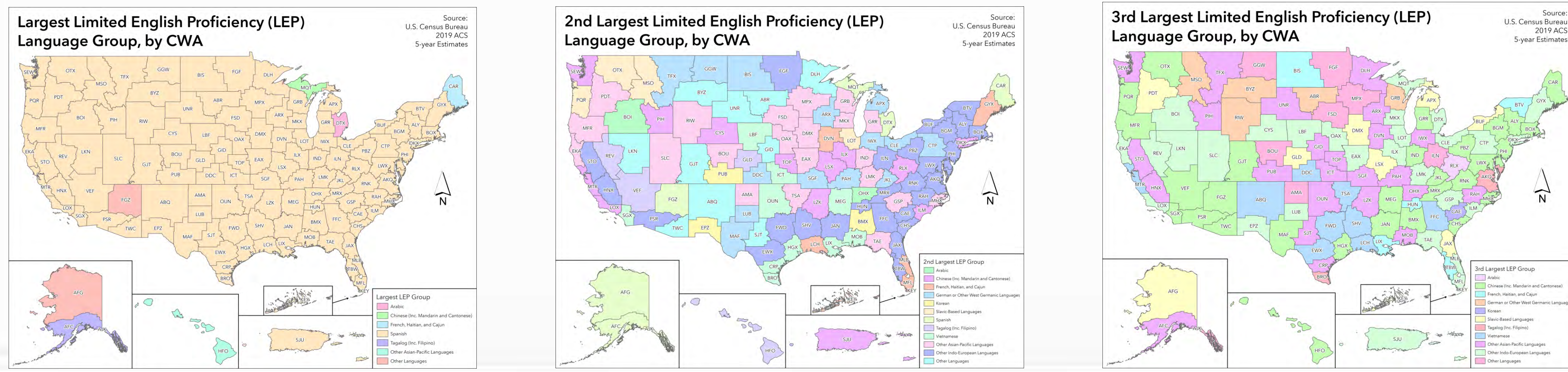


Introducing the National Weather Service Artificial Intelligence Language Translation Program

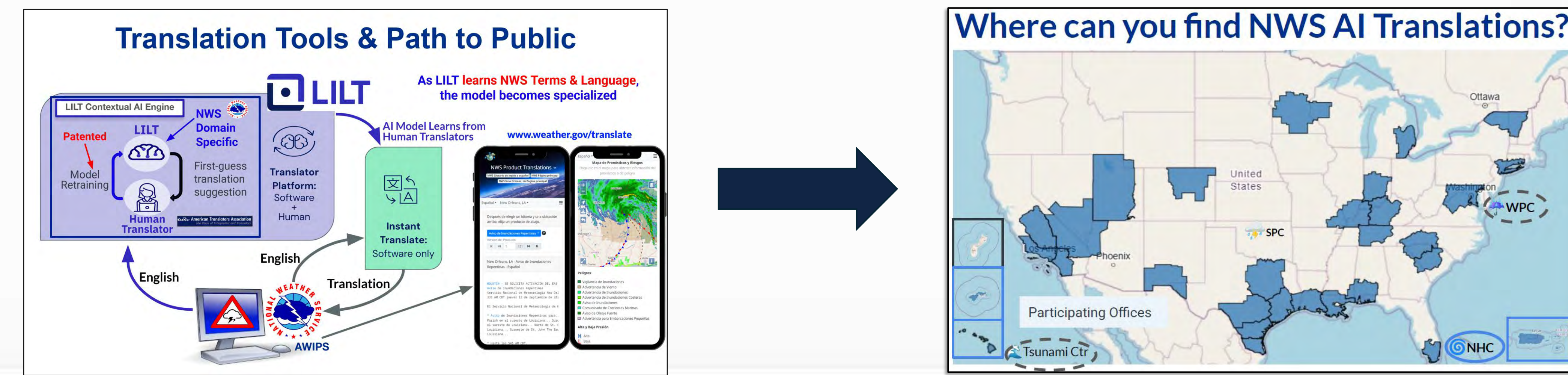
Joseph E. Trujillo-Falcón, Monica L. Bozeman, Liam E. Llewellyn, Samuel T. Halvorson, Meryl Mizzell, Stuti Deshpande, Bob Manning, Chris Rohrbach, Ian Blaylock, Angel Montanez, and Todd Fagin



The Multilingual Challenge (Trujillo-Falcón et al., 2021; Llewellyn et al., 2026)



Developing the NWS AI Language Translation Program (Trujillo-Falcón et al., 2026)



NWS Product Translations and Visualizations

Users in participating offices can access real-time translated text products. Blue text indicates dialectal variations communicators can select. All NWS graphics are available nationwide.

WEATHER IN MORE LANGUAGES WEATHER.GOV/TRANSLATE

Multilingual Weather Safety Infographics

Our team is automating multilingual weather infographics, including social media safety plans with copy-and-paste templates tailored for Facebook, Twitter, and alt text.

Translation Quality and Validation

Every translation is automatically sent through a second, untrained AI to retranslate it back into English. Scoring software then compares that result to the original, flagging meaning gaps. Results feed into a custom NWS dashboard showing which products and languages need retraining. The public can also provide thumbs-up or thumbs-down ratings for products on the website.

Metric	Core Equation	High-Level Explanation
Bilingual Evaluation Understudy Score (BLEU)	$\frac{\sum_{c \in C} \sum_{g \in G} \text{Count}_{\text{BLEU}}(n - \text{gram})}{\sum_{c \in C} \sum_{g \in G} \text{Count}_{\text{BLEU}}(n - \text{gram})}$	Scores how similarly machine translations match human translations using a word-by-word n-gram overlap comparison (Papinen et al. 2002)
Character F-score (ChrF+)	$\text{ChrF}^+ = \frac{\text{chrP} + \text{chrR}}{\beta^2 + \text{chrP}} + \frac{\text{chrR}}{\beta^2 + \text{chrP}}$	Evaluates translation quality at the character level by comparing character overlaps (Popović 2015)
Cross-lingual Optimized Metric for Evaluation of Translation (COMET)	$e_{x_i} = \mu E_{x_i}^\alpha$	A neural framework for machine translation evaluation based on contextual accuracy and fluency (Rei et al. 2020)
Translation Edit Rate (TER)	$\text{TER} = \frac{\# \text{ of edits}}{\text{average \# of reference words}}$	Measures the number of edits needed to make a target translated sentence look the same as the original reference sentence (Snover et al. 2006)
Fuzzy String Matching (Fuzz)	$C_{i,j}^k = \begin{cases} 1 & \text{if } i-j \leq k \\ 0 & \text{otherwise} \end{cases}$	Evaluates the similarity between strings using edit distance (Levenshtein 1966)

	Original English	Spanish Translation	Backtranslation
	Most dry grasses, dead leaves, and other tree litter could easily ignite and spread fire quickly.	La mayoría de las hierbas secas, hojas muertas y otra basura de árboles podrían encenderse fácilmente y propagar fuego rápidamente.	Most dried weeds, dead leaves and other tree debris could easily catch fire and spread fire quickly.
Segment score	BLEU: 34.1	ChrF++: 67.3	COMET: 0.90
"Good score"	> 50	> 50	> 0.7
"Bad score"	< 15	< 30	< 0.3

Practical Applications

NWS AI translations are moving into real-world use as (1) the National Hurricane Center now incorporates Spanish tropical products across Atlantic and Pacific coasts, (2) 31 NWS offices deliver warnings in up to 12 languages, and (3) translated materials for the 2026 FIFA World Cup bring multilingual weather safety to major global sporting events.