

Post-training for Deepfake Speech Detection

Wanying Ge, Xin Wang, Xuechen Liu, Junichi Yamagishi

National Institute of Informatics, Tokyo, Japan

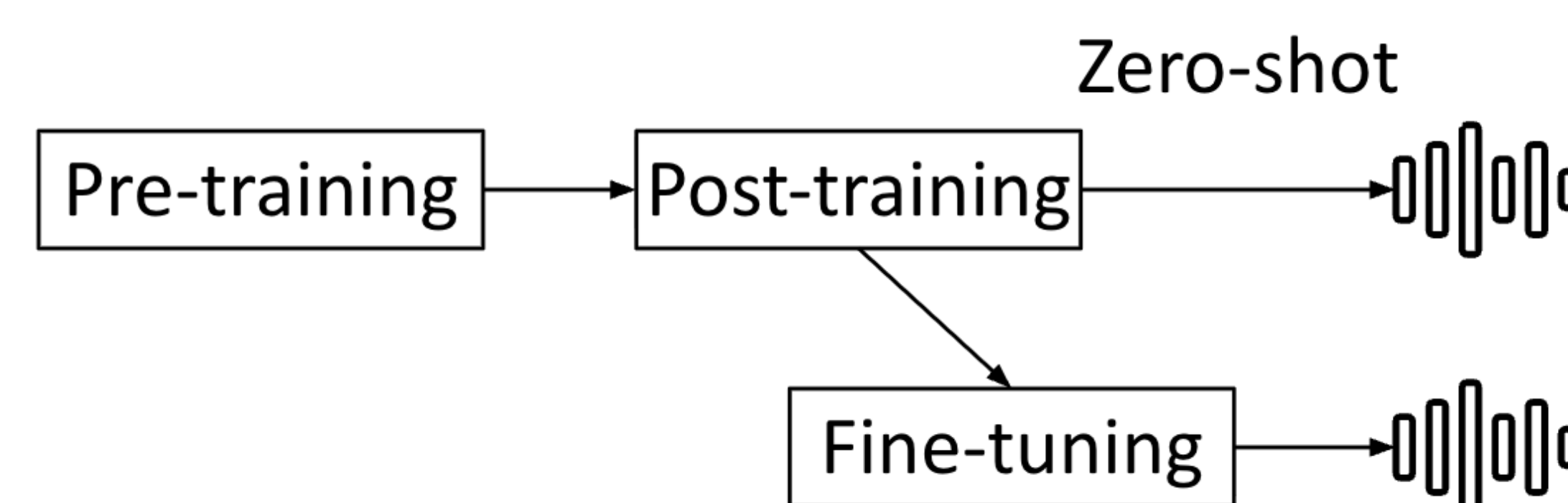


Poster No. SS2-241

TL;DR

AntiDeepfake models are open-source, large-scale foundation models trained on 56,000 hours of real and 18,000 hours of fake speech across 100 languages.

They can be used for zero-shot deepfake detection and feature extraction or fine-tuned for specific tasks and domain.



AntiDeepfake
2.2 B

ZERO-SHOT EER (%) RESULTS

	Model ID	# of params.	ADD 2023	DEEP-VOICE	FakeOrReal		In-the-Wild	Deepfake-Eval-2024
			Track-1.2-R2-Test	Segmented Full Set	original-Test	norm-Test	Full Set	Audio Test
Pre-training + Post-training	HuBERT-XL	964 M	18.91	5.68	2.48	3.18	5.23	34.10
	W2V-Small	95 M	13.02	9.79	22.03	17.89	4.23	33.33
	W2V-Large	317 M	13.25	4.44	0.67	0.97	1.91	33.36
	MMS-300M	317 M	7.93	2.35	1.40	5.92	2.90	32.84
	MMS-1B	965 M	9.06	2.47	1.23	1.73	1.82	27.70
	XLS-R-1B	965 M	5.39	2.47	5.74	12.14	1.35	26.76
	XLS-R-2B	2.2 B	4.67	2.23	2.61	1.64	1.23	27.76

FINE-TUNING EER (%) RESULTS

Model ID	Pre-training + Post-training + Fine-tuning					Pre-training + Fine-tuning				
	4s	10s	13s	30s	50s	4s	10s	13s	30s	50s
W2V-Large	19.56	12.11	10.95	10.52	11.36	24.42	22.46	22.14	21.17	21.53
MMS-300M	17.15	13.38	12.32	11.18	10.79	19.77	13.26	12.72	12.01	12.30
MMS-1B	12.11	10.40	10.02	8.60	9.35	19.85	13.57	11.56	11.04	11.40
XLS-R-1B	11.86	10.00	9.27	8.49	8.28	19.96	17.19	16.31	10.71	11.20
XLS-R-2B	12.14	9.80	9.99	9.46	9.56	12.88	10.76	10.34	9.68	9.84



RawNet2
25.43 M

AASIST
297 K

2021

2022

Release Date

2025