Progressive Layer-wise Supervision: Deep-to-Shallow Supervision Annealing for Efficient and Robust Speech Deepfake Detection

Hoan My Tran
Université de Rennes, CNRS, IRISA
Lannion, France
hoan.tran@irisa.fr

Xin Wang
National Institute of Informatics
Tokyo, Japan
wangxin@nii.ac.jp

Xuechen Liu
Xi'an Jiaotong-Liverpool University
Taicang, China
xuechen.liu@xjtlu.edu.cn

GitHub

Abstract

Speech foundation models pretrained via self-supervised learning provide powerful representations for deepfake detection, yet we find that standard fine-tuning exploits only the deepest transformer layers — leaving the majority of the network's capacity discriminatively inert. Layer-wise probing of *XLS-R-128* reveals that the top two-thirds of the transformer stack contribute negligibly to anti-spoofing performance after standard fine-tuning, a phenomenon we term shallow-layer dormancy. We connect this observation to the representational redundancy of pretrained transformers: adjacent layers encode highly similar information, and fine-tuning without explicit intermediate supervision reinforces rather than breaks this redundancy. To address this, we propose Progressive Layer-wise Supervision (PLS), a training framework that attaches a shared linear classifier to every transformer layer and schedules layer-wise loss contributions via an exponentially annealed power weighting. Early training concentrates supervision on deep layers to preserve the pretrained representational hierarchy; later training progressively redistributes supervision to shallower layers, converting dormant earlier representations into discriminative ones. Competitive detection ($\leq 9.25\%$ average EER) is achievable on *XLS-R-128* at layer 7, which reduces the parameter count from 315M to 101M and exemplifies an early exit mechanism. Experiments across twelve benchmarks spanning in-domain, and out-of-domain deepfake scenarios show PLS achieves a new state-of-the-art with 239M parameters, outperforming prior methods with larger and more complex architectures.

1 Introduction

The rapid advancement of neural speech synthesis has made AI-generated voices increasingly indistinguishable from genuine recordings [1]. Text-to-speech (TTS) and voice conversion (VC) technologies, while enabling beneficial applications [2], are increasingly misused for impersonation, fraudulent robocalls, and identity theft [3]. Detecting synthetic speech has become a pressing forensic and societal challenge. State-of-the-art (SOTA) countermeasures leverage speech foundation models (SFMs) — large transformers pretrained on vast speech corpora via self-supervised learning (SSL) — as front-end feature extractors [4, 5, 6, 7, 8, 9, 10]. These models are subsequently

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

fine-tuned for the binary *bona fide* (real) / *spoofed* (fake) classification task. Despite strong in-domain performance, they generalize poorly to out-of-domain (OOD) scenarios involving unseen synthesizers [11, 12, 13, 14, 15, 16, 17, 18], limiting their real-world utility.

The shallow-layer dormancy problem. We identify a specific and underexplored failure mode of standard SFM fine-tuning. Layer-wise probing of *XLS-R-128* 300M (Figure 2, Table 1) reveals that layers 0–15 of 24 consistently yield average equal error rate (EER) exceeding 20% across all twelve evaluation benchmarks after standard fine-tuning — competitive performance (EER $< 12\%$) emerges only from layer 16 onward. Two-thirds of the network contributes nothing to deepfake detection, despite containing over 200M parameters. This observation is not merely about efficiency. We hypothesize that these dormant layers represent untapped generalization capacity: if earlier representations could be made discriminative, they would provide the model with multiple independent detection pathways, reducing its brittle dependence on the final-layer representation and improving robustness to distribution shift.

Connection to representational redundancy. Prior work has extensively documented that pre-trained SFMs exhibit high representational similarity across adjacent layers [19, 20, 21]. This redundancy arises from SSL pretraining objectives that optimize for smooth, hierarchical speech representations — not for task-specific discrimination. Standard fine-tuning, which applies supervision only at the final layer, fails to disrupt this redundancy in earlier layers: they remain anchored to the pretrained representational geometry [22, 23, 24, 25], which was not designed for anti-spoofing. Explicit supervision is necessary to realign these layers toward the detection task.

The ordering problem. A response is to supervise all layers simultaneously with uniform weights. We show empirically, however, this is counterproductive: applying strong supervision to shallow layers before deep layers have adapted to the anti-spoofing task degrades final-layer performance by 1.2–3.5% EER. This is because the pretrained representational hierarchy is valuable — high-level semantic and acoustic features are organized progressively across depth [19, 20, 21], and disrupting shallow layers before deep layers have stabilized collapses this structure prematurely [24, 25].

Our approach. We propose Progressive Layer-wise Supervision (PLS), which supervises all layers through a shared linear classifier while scheduling layer-wise loss contributions via an exponentially annealed power weighting. The schedule concentrates supervision on deep layers early in training, allowing high-level representations to stabilize first, then progressively redistributes supervision toward shallower layers, awakening dormant earlier representations in a controlled, bottom-up order. An even/odd epoch alternation further anchors the primary classification objective, preventing drift from accumulated intermediate supervision. We make the following specific contributions:

- **Progressive Layer-wise Supervision (PLS).** We introduce an exponentially annealed power weighting for layer-wise losses that activates the full transformer stack in a deep-to-shallow order, converting redundant earlier representations into discriminative ones.
- **Efficient early-exit inference.** The shared classifier enables inference to exit at layer 7 of 24, reducing parameters by 70% with competitive performance under 10% of average EER.
- **State-of-the-art OOD generalization.** PLS achieves 6.03% average EER across eight datasets, outperforming all prior methods with larger and more complex backends.

2 Related work

Deepfake generation and anti-spoofing countermeasures. Recent advances in TTS and VC have enabled the synthesis of speech that is nearly indistinguishable from genuine human recordings [1]. Modern TTS systems typically follow a multi-stage pipeline: textual input is converted into linguistic representations, which are then mapped to acoustic features and synthesized into waveforms using neural vocoders. VC systems preserve the linguistic content of a source utterance while transforming its speaker characteristics to match a target speaker, often from reference speech. Together, these technologies raise both opportunities and security concerns [26, 27, 28].

Early countermeasures relied on handcrafted features such as CQCCs [29] and MFCCs [30] with classical classifiers. The field shifted to end-to-end neural architectures [31, 32, 33], and more recently,

to SFM-based pipelines using *Wav2Vec2* [34] or *WavLM* [35] as front-end extractors followed by increasingly complex backends — graph attention networks [32], Transformer [7], Mamba [8], Kolmogorov–Arnold networks [9, 36], and feature gating [10]. We improve how the SFM encoder itself is trained, and achieve competitive results with a simple linear backend.

Representational redundancy in pretrained speech models. Pretrained transformer encoders exhibit high representational similarity between adjacent layers [19, 20, 21]. Pasad et al. [19, 20] show via probing that SFMs encode phonetic and semantic information in a graduated hierarchy, with considerable redundancy particularly in middle layers. Dorszewski et al. [21] show that substantial portions of the transformer stack in SFMs can be pruned with minimal performance loss, confirming functional redundancy. This redundancy is connected theoretically to layer-wise saturation in deep networks [25, 24, 23] and to the neural collapse phenomenon [22], where class-conditional representations converge toward an equiangular tight frame geometry at the output layer. Our work leverages this redundancy constructively: rather than pruning redundant layers, we activate them for the downstream detection task that they can be pruned after training.

Multi-exit and layer-wise training. Multi-exit inference has been explored in NLP [37, 38], where auxiliary classifiers at intermediate layers enable adaptive computation. Jiang et al. [24] show in vision transformers that shared classifiers trained across all layers with aligned objectives can reduce representational redundancy and improve multi-exit performance. For audio deepfake detection, Pimentel et al. [39] attach classifier at fixed intermediate layers of *WavLM* for early exit. Our approach uses a shared classifier across all layers, which is both parameter-efficient and encourages representational alignment across depth while introducing the annealing schedule that we show is essential for preserving pretrained representational structure while activating intermediate layers.

3 Method

Preliminaries. SFMs consist of a convolutional feature extractor $f : \mathcal{X} \to \mathcal{Z}$ mapping raw waveform to latent frames $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_T\}$, followed by L transformer layers. Let $\mathbf{H}^{(l)} = \{\mathbf{h}_1^{(l)}, \ldots, \mathbf{h}_T^{(l)}\}$ denote hidden states at layer $l \in \{0, \ldots, L\}$, where $l = 0$ corresponds to the output of the convolutional feature extractor and $l \in \{1, \ldots, L\}$ corresponds to the outputs of the subsequent transformer layers. Global average pooling over $\mathbf{H}^{(L)}$ yields a fixed-dimensional embedding; a shared linear classifier c produces logits $\hat{\mathbf{y}}^{(l)}$ with binary cross-entropy loss $\mathcal{L}_L$.

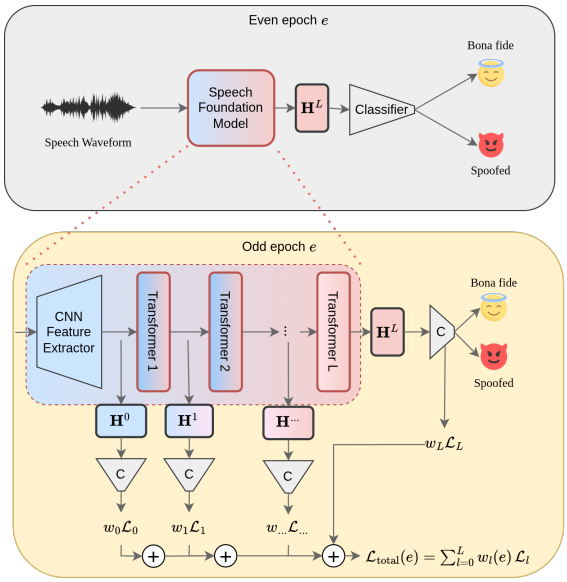


Figure 1: PLS training objective. During even epochs, the model is anchored by standard supervision on the final transformer layer $\mathbf{H}^{(L)}$. In odd epochs, a deep-to-shallow supervision is applied, where the total loss $\mathcal{L}_{\text{total}}$ is a weighted sum of all layer-wise losses $\mathcal{L}_l$. The layer weights w_l evolve via a decaying power parameter $\rho(e)$, transitioning from high-level to uniform representation.

Characterizing shallow-layer dormancy. We begin with an empirical characterization of the problem. After standard fine-tuning of *XLS-R-128* 300M, we attach the trained linear classifier to each layer separately and evaluate on all twelve benchmarks (Table 1). When classifier head is attached to layers 0–7, it yields average EERs above 40% on every benchmark, including in-domain sets. Competitive detection (EER < 12%) emerges only from layer 16 onward, with the optimum at layer 21 (6.13% Avg. EER). The shallow 15 of the 24 transformer layers, i.e., $l < 16$, carry negligible discriminative information for anti-spoofing.

Why does standard fine-tuning fail to activate earlier layers? While this is not surprising, we argue that the answer lies in representational redundancy across the layers and the way of fine-tuning. SSL pretraining minimizes objectives that reward smooth, consistent representations across adjacent layers — producing representations highly similar to each other [19, 20]. In standard fine-tuning, gradients originate exclusively from the final classifier and decay through successive transformer blocks. Consequently, deeper layers receive stronger task-specific adaptation signals, while earlier layers remain close to their pretrained state. This produces a representational bottleneck in which discriminative information is concentrated near the output layers, leaving a substantial fraction of the network effectively underutilized for downstream deepfake detection task.

Is uniform multi-layer supervision insufficient? The immediate remedy appears to be applying supervision to all layers simultaneously with uniform weights from the first epoch. We show empirically in Figure 2 that this is counterproductive: final-layer EER degrades by 2% relative to standard fine-tuning, despite improving early-layer EER. We attribute this to pretraining structure disruption. The pretrained representational hierarchy — where progressively deeper layers encode progressively more abstract speech properties — is valuable for the downstream task [40, 41]. When strong supervision is applied to shallow layers before the model has adapted its deeper representations to anti-spoofing, the resulting gradient updates propagate through the full stack and destabilize the deep-layer representations that are still in the process of being fine-tuned. Ablation study is found in Appendix C (Table 32).

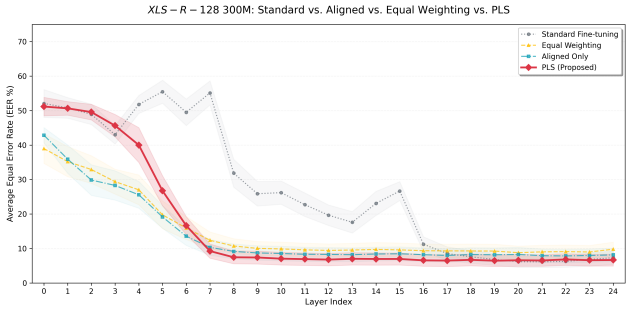


Figure 2: Proposed PLS approach comparing to standard, aligned, and equal weighting fine-tuning.

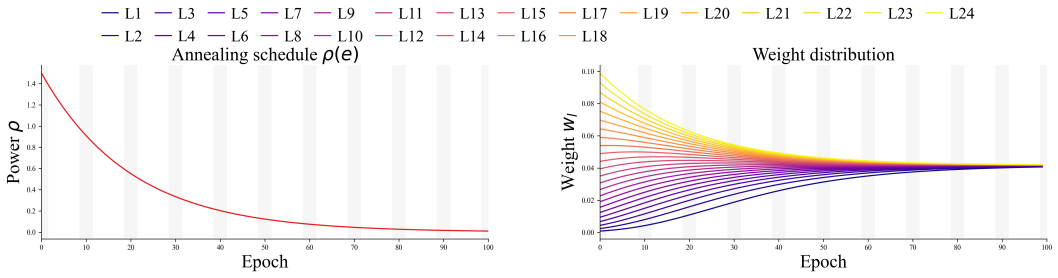


Figure 3: Illustration of annealing schedule, and weight distribution through epochs.

Progressive Layer-wise Supervision. We define a scheduled multi-layer training objective. Let $\mathcal{L}_l$ denote the binary cross-entropy loss when the classifier head is attached to the l-th layer. At epoch e, the total training loss is:

$$\mathcal{L}_{\text{total}}(e) = \sum_{l=0}^{L} w_l(e)\,\mathcal{L}_l. \tag{1}$$

Layer weights are defined by an epoch-dependent power parameter $\rho(e)$:

$$w_l(e) = \frac{(l+1)^{\rho(e)}}{\sum_{j=0}^{L}(j+1)^{\rho(e)}}, \tag{2}$$

where $\rho(e) = M \cdot \exp(-ke)$, with $M > 0$ controlling initial emphasis on deep layers and $k > 0$ controlling the annealing rate. PLS can be interpreted as encouraging a more gradual redistribution of

task information across the transformer hierarchy, rather than concentrating discriminative capacity exclusively in the final layers.

Training dynamics. This schedule induces two natural phases (Figure 3):

- **Stabilization phase (large ρ, early epochs)**: Weights concentrate on $l \approx L$, The objective reduces approximately to standard final-layer fine-tuning, allowing deep representations to stabilize for the anti-spoofing task while shallow layers are only lightly updated.
- **Activation phase ($\rho \to 0$, late epochs)**: Weights converge toward $w_l \to \frac{1}{L+1}$, distributing supervision uniformly across all layers. Intermediate and shallow layers now receive direct supervision signals and progressively acquire discriminative structure.

Epoch alternation for primary objective anchoring. Even with annealing, accumulated multi-layer supervision over many odd-numbered epochs can drift the primary classification objective. To prevent this, we adopt an alternating strategy [24]: for even-numbered epochs, we restrict supervision to the final layer only ($\mathcal{L}_{\text{total}}(e) = \mathcal{L}_L$), ensuring the primary classification objective remains well-anchored, as shown in Figure 1; for odd-numvered epochs, we apply the full scheduled objective $\mathcal{L}_{\text{total}}(e)$ from Equation 1. The complete training algorithm is listed in Appendix A.

4 Experimental setup

All models are trained exclusively on the `ASVspoof 2019` logical access (`19LA`) training partition [42], with model selection on the `19LA` development set. The `19LA` corpus contains bona fide speech from `VCTK` [43] and spoofed samples generated by 17 TTS/VC systems — 6 seen during training, 13 reserved for evaluation. Weighted cross-entropy addresses class imbalance. RawBoost augmentation [44] is applied. Audio inputs are dynamically padded.

In Appendix 5.2, we also trained the models on other corpora to evaluate the impact of the training dataset [16, 17, 45]. Detailed dataset descriptions and statistics are provided in Appendix B.1.

For OOD evaluation, we use a diverse suite of datasets. `ASVspoof5` (`25DF`) [16] contains utterances synthesized with advanced TTS/VC systems and novel adversarial attacks (ATs). `In-the-Wild` (`ITW`) [11] contains audio collected from social media platforms, `Fake-or-Real` (`FoR`) [46] comprises samples from diverse open datasets and real-world recordings, and `SpoofCeleb` (`SC`) [17] includes TTS-generated speech with diverse speakers under noisy and reverberant conditions. Robustness to vocoder artefacts is assessed via `LibriSeVoc` (`LSV`) [13], which covers autoregressive, diffusion-based, and GAN-based vocoders. Diffusion and flow-matching scenarios are further evaluated on `DFADD` [14] and `DIFSSD` [18]. Finally, we also evaluated on `DECRO` [12] and `MLAAD` [15]. All results are reported as EER (%); the lower, the better. The code are available here[1].

5 Results and analysis

5.1 Aligned training enabling early-exit

Shallow-layer dormancy. Table 1 confirms the dormancy problem for *XLS-R-128* 300M under standard fine-tuning. Layers 0–7 yield average EER > 40% — near-random performance. Competitive performance (EER < 12%) emerges only from layer 16 onward, with best performance at layer 21 (5.80% Avg. EER). The final layer (layer 24, 7.33%) is outperformed by layer 21, suggesting mild overfitting to the training distribution at the topmost layer.

Aligned training compared with standard fine-tuning. Aligned training ($\rho = 1$ [24]) shifts the layer-wise dynamics of *XLS-R-128* 300M, significantly improving early and intermediate representation utility. Notably, aligned training achieves competitive performance by layer 7 (9.56%), whereas standard fine-tuning requires 16 layers (11.23%)—a 9-layer inference saving. It also stabilizes performance: while standard fine-tuning fluctuates (5.80–31.84% across layers 8–24), aligned training remains robust (6.33–7.72%), enabling early-exit.

[1] RawBoost repository

Table 1: *XLS-R-128* 300M under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer; Avg. is the unweighted mean across 12 datasets.

Layer	Params.	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
		19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	Avg.	Pool.
Standard Fine-tuning															
0	13.13M	69.59	56.03	58.48	52.93	15.34	47.48	52.91	58.94	43.23	45.20	73.13	51.46	52.06	50.64
1	25.72M	45.19	38.82	47.51	52.59	67.20	44.00	56.63	45.96	51.95	31.70	66.58	61.33	50.79	45.00
2	38.32M	47.90	41.00	46.62	50.62	64.22	46.68	50.64	40.63	44.08	28.62	68.07	57.97	48.92	45.95
3	50.92M	32.37	36.16	34.92	45.02	61.21	53.09	49.72	38.80	39.09	32.28	37.79	55.44	42.99	43.92
4	63.51M	46.00	41.66	42.01	47.33	72.86	51.55	54.17	61.71	56.28	45.51	43.50	58.82	51.78	48.28
5	76.11M	46.43	42.26	51.47	51.87	72.58	74.38	60.32	64.27	62.87	37.66	41.73	59.83	55.47	53.47
6	88.71M	38.95	31.09	41.09	49.56	75.34	49.33	64.25	55.07	60.39	31.30	37.64	60.05	49.51	48.54
7	101.30M	45.94	44.69	53.85	51.65	71.23	72.44	62.34	49.14	66.25	38.95	36.72	68.30	55.13	58.95
8	113.90M	11.76	13.13	14.24	23.25	47.92	44.78	51.29	34.17	47.06	41.39	22.07	30.96	31.84	29.13
9	126.49M	15.64	15.04	9.13	9.99	33.59	20.33	37.26	42.39	42.07	41.78	25.71	17.67	25.88	20.65
10	139.09M	18.09	17.45	10.29	11.70	37.35	16.62	42.73	36.69	39.42	40.15	25.83	18.01	26.19	21.63
11	151.69M	13.50	12.08	6.93	9.09	28.22	10.64	34.98	41.86	31.45	35.51	31.38	16.80	22.70	16.23
12	164.28M	7.26	9.53	5.76	8.67	28.88	10.55	36.40	28.72	24.87	31.19	23.34	20.57	19.65	17.49
13	176.88M	3.70	4.77	4.77	10.56	29.25	12.58	34.57	21.98	31.98	28.20	11.84	16.97	17.60	16.24
14	189.48M	6.77	9.05	8.69	16.96	35.83	17.67	44.55	28.47	37.24	34.19	16.79	20.82	23.09	20.11
15	202.07M	18.83	19.51	11.40	17.38	31.95	21.26	33.95	35.88	37.53	40.61	34.00	17.65	26.66	20.35
16	214.67M	1.55	2.76	2.93	5.57	15.91	6.09	22.64	14.44	21.93	19.62	6.76	14.54	11.23	12.88
17	227.26M	1.17	2.96	1.88	2.30	8.29	2.44	18.73	12.43	16.73	18.74	3.30	13.35	8.53	12.38
18	239.86M	0.73	2.42	1.77	2.01	6.40	1.99	18.74	9.53	14.90	17.06	2.25	13.07	7.57	12.31
19	252.46M	0.34	2.49	1.53	2.28	5.69	2.35	17.97	8.62	12.34	14.56	1.16	12.53	6.82	12.20
20	265.05M	0.11	2.11	1.41	1.67	5.22	1.23	17.63	7.13	10.33	12.80	1.39	12.87	6.16	12.12
21	277.65M	0.07	2.07	1.42	1.67	5.09	1.01	17.59	6.89	10.63	11.57	2.32	13.18	5.80	11.99
22	290.25M	0.07	2.22	1.43	1.70	4.82	1.01	18.65	6.63	10.51	11.91	2.13	14.01	6.26	11.74
23	302.84M	0.10	2.31	1.49	1.67	4.59	0.88	20.62	7.81	10.24	12.31	1.86	17.81	6.81	11.99
24	315.44M	0.10	2.36	1.46	1.75	4.69	0.93	19.57	17.51	10.16	12.44	3.48	13.47	7.33	11.70
Aligned Fine-tuning															
0	13.13M	21.98	22.08	35.17	53.23	49.12	82.24	49.20	23.57	54.53	34.74	23.37	50.51	41.65	43.52
1	25.72M	31.84	31.85	34.56	55.51	45.51	80.25	44.75	22.28	47.42	46.41	27.38	52.08	43.32	51.35
2	38.32M	24.16	27.40	28.89	48.88	34.11	83.92	42.20	17.62	39.30	41.14	20.06	49.29	38.08	45.60
3	50.92M	20.16	24.84	25.02	45.77	26.48	73.20	40.90	20.00	40.16	40.68	16.84	42.10	34.68	41.65
4	63.51M	19.17	19.20	19.92	32.97	24.41	50.00	43.29	18.39	35.50	36.08	16.79	42.73	29.87	37.96
5	76.11M	7.08	9.82	10.73	11.66	15.39	22.23	35.25	13.37	39.15	24.54	7.94	46.32	20.29	33.46
6	88.71M	1.17	6.03	3.52	3.33	8.40	7.90	23.89	11.76	18.97	20.42	3.72	41.16	12.52	27.81
7	101.30M	0.29	3.96	1.94	1.82	4.87	3.22	19.95	7.94	17.96	14.71	2.53	35.58	9.56	23.80
8	113.90M	0.23	3.91	1.86	1.14	4.12	2.52	19.33	6.38	14.07	16.10	1.33	21.63	7.72	16.07
9	126.49M	0.11	3.62	1.76	0.99	4.07	2.33	18.69	6.91	13.80	15.28	1.35	22.24	7.60	16.36
10	139.09M	0.10	3.38	1.67	1.02	4.00	2.07	18.58	7.03	13.89	15.00	1.55	21.23	7.46	15.76
11	151.69M	0.07	3.02	1.60	0.95	3.97	2.11	18.57	7.13	13.51	14.09	1.51	24.56	7.59	18.23
12	161.28M	0.07	2.98	1.52	0.99	4.05	2.24	18.47	7.29	13.60	14.08	1.40	25.31	7.67	18.40
13	176.88M	0.07	2.78	1.55	0.92	4.01	2.07	18.33	7.54	13.39	13.48	1.53	19.56	7.10	14.61
14	189.48M	0.08	2.73	1.55	0.91	4.20	2.07	18.39	7.67	13.70	13.37	1.69	20.09	7.20	15.07
15	202.07M	0.08	2.67	1.61	1.02	4.17	1.90	18.58	7.79	14.00	12.80	1.92	19.18	7.14	14.41
16	214.67M	0.07	2.44	1.56	1.06	4.09	1.77	18.38	7.43	12.94	12.66	1.93	17.00	6.78	12.87
17	227.26M	0.09	2.42	1.62	1.14	4.36	1.68	18.38	7.67	13.89	12.79	1.74	13.83	6.63	12.48
18	239.86M	0.07	2.00	1.58	1.36	4.37	1.64	18.24	6.89	13.96	11.53	1.88	13.81	6.44	12.55
19	252.46M	0.08	1.92	1.54	1.40	4.45	1.14	18.38	6.63	13.90	10.89	1.67	14.29	6.36	13.42
20	265.05M	0.05	1.59	1.54	1.55	4.32	1.36	18.37	5.93	12.90	10.64	2.49	15.21	6.33	13.81
21	277.65M	0.05	1.59	1.45	1.63	4.42	1.94	19.27	6.50	12.52	10.01	2.92	16.31	6.55	13.86
22	290.25M	0.07	1.59	1.44	1.59	4.44	2.74	20.33	7.43	12.20	9.51	2.62	16.77	6.73	13.69
23	302.84M	0.08	1.71	1.49	1.59	3.98	2.44	18.60	9.29	12.28	9.57	2.72	15.76	6.63	13.21
24	315.44M	0.10	1.96	1.51	1.63	3.62	2.57	18.52	11.40	12.28	9.50	3.83	14.97	6.82	12.58

Colour coding by Avg. EER: red > 40%, grey 12–40%, green < 12%.

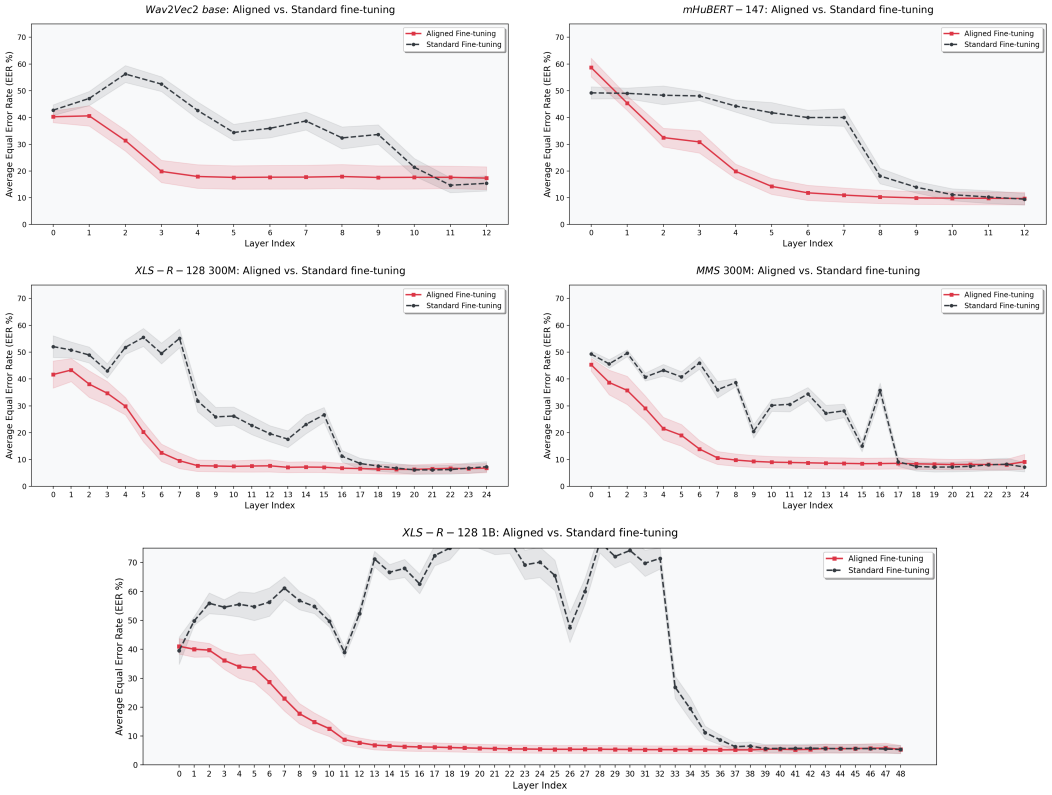


Figure 4: Comparison (Average EER %) of *standard* and *aligned* training across SFMs.

Effect of model scale and pretraining language. Figure 4 highlights two consistent trends across architectures. First, scaling model capacity substantially improves robustness, especially for multilingual models. Under standard fine-tuning, the average EER of multilingual models decreases steadily from 9.46% for *mHuBERT-147* (base) to 5.80% for *XLS-R-128* (large) and further to 5.36% for *XLS-R-128* (XL), indicating that larger models produce more transferable representations for spoofing detection. Second, aligned fine-tuning improves layer efficiency by shifting the optimal decision layer earlier, while preserving competitive final-layer performance. For example, the best-performing layer of *XLS-R-128 XL* moves from layer 48 under standard fine-tuning to layer 36 under aligned fine-tuning, while the average EER improves slightly from 5.36% to 5.23%. This suggests that alignment encourages discriminative representations to emerge earlier in the network, reducing inference depth without sacrificing performance.

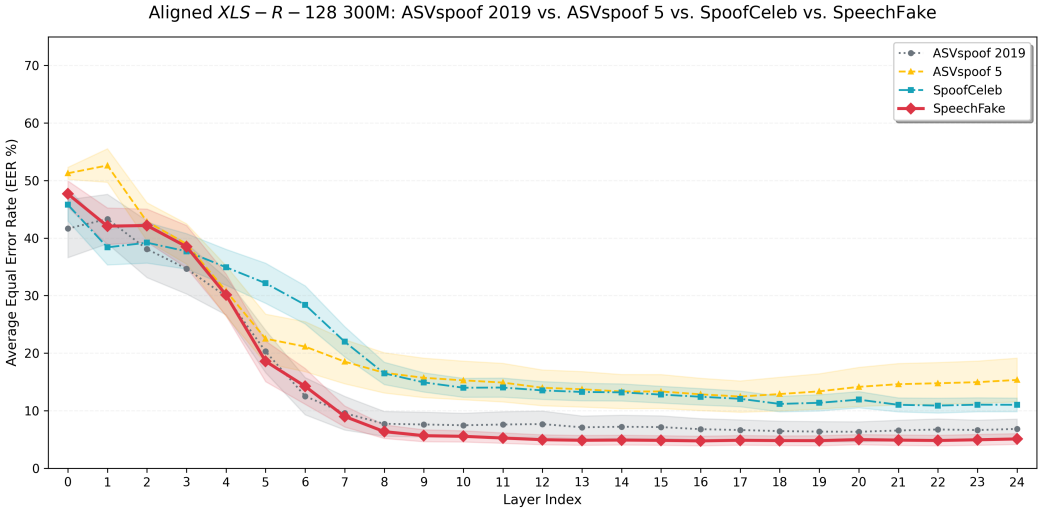


Figure 5: Performance (Average EER %) of aligned fine-tuning trained on different datasets.

5.2 Impact of training datasets on layer-wise generalization

We next analyze how the choice of training dataset affects the layer-wise behavior of the model across diverse evaluation benchmarks. Figure 5 reports the EER obtained by probing each encoder layer when the model is trained separately on four datasets with different characteristics: `ASVspoof 2019`, `ASVspoof 5`, `SpoofCeleb`, and `SpeechFake`. Detail results are described in Appendix C.1.

These results reveal that training data diversity and size of the training corpus has a larger effect on robustness. Conversely, datasets such as `SpeechFake`, which contain broader acoustic and synthesis variability, enable the model to learn more invariant spoof-related cues, leading to substantially better cross-dataset generalization. Overall, these results suggest that model scale and multilingual pretraining are the primary drivers of generalization, while aligned fine-tuning primarily improves computational efficiency by making strong representations emerge earlier. Based on its superior trade-off between robustness and early-layer efficiency, *XLS-R-128* is utilized as our main backbone. Further analysis is conducted via PLS on the `19LA` training set, chosen specifically for its smaller scale in comparison to the more expansive `SpeechFake` corpus.

5.3 Effect of PLS on earlier representation

Effect of power parameter ρ. We ablate the layer-weight allocation in the aligned supervision objective (Eq. 2) using fixed exponents $\rho \in \{1, 1.5, 2\}$. This isolates the impact of weighting bias before introducing the dynamic schedule (Sec. 3). As shown in Table 2, ρ governs the trade-off between depth specialization and distributed representation learning:

- **Low bias ($\rho = 1$):** Provides relatively uniform supervision, allowing discriminative capacity to emerge early. Average EER falls below 12% by layer 7 (11.11%), though final-layer performance (7.36%) lags behind higher exponents.
- **Moderate bias ($\rho = 1.5$):** Strikes the best balance. It optimizes intermediate layers smoothly (7.23% EER at layer 9) while achieving the best overall Avg. EER (6.40%) at layer 19. This suggests that a moderate deeper-layer preference guides high-level abstraction without starving shallow layers.
- **High bias ($\rho = 2$):** Concentrates optimization near the output, resembling standard fine-tuning. Early layers remain under-optimized (28.82% EER at layer 5 vs. 22.33% for $\rho = 1.5$), effectively wasting the representational capacity of the intermediate backbone.

Table 2: Ablation of static power parameter ρ (no exponential decay) on *XLS-R-128* 300M with aligned fine-tuning. **Layer**: best-performing layer for each ρ. EER (%, ↓).

ρ	Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Avg.
		19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	
1	23	**0.08**	**1.71**	1.49	1.59	**3.98**	2.44	18.60	9.29	12.28	**9.57**	2.72	15.76	7.36
1.5	19	0.09	2.29	**1.33**	**1.44**	4.88	2.20	17.83	**7.41**	**10.26**	14.06	1.79	13.27	**6.40**
2	19	0.10	1.84	1.43	1.51	5.41	**1.73**	**16.99**	8.21	10.58	14.48	**2.16**	**13.19**	6.47

Bold indicates best results, the second-best are underlined.

These results indicate that while deeper-layer emphasis is necessary for peak performance, over-concentration hinders the "early-exit" utility of the model. This motivates a dynamic supervision that relaxes from a high ρ toward a more uniform distribution.

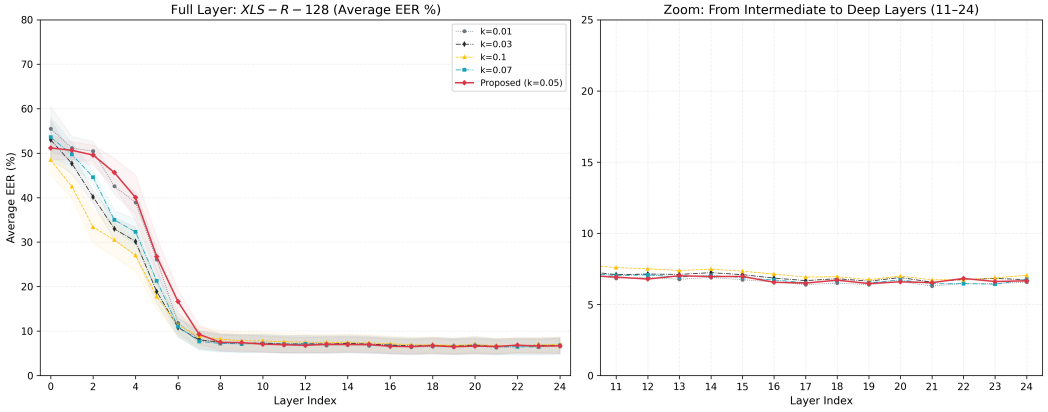


Figure 6: Effect of PLS with $M = 1.5$ and k decay.

Effect of decay rate k. We further evaluate the PLS dynamics by varying the decay rate $k \in \{0.01, 0.03, 0.05, 0.07, 0.1\}$ with fixed $M = 1.5$ (Fig. 6). The decay rate controls the transition from deep-layer initialization to holistic supervision. A slow decay ($k = 0.01$) fixes the model's focus on deep layers too long, delaying the optimization of earlier features. Conversely, a rapid decay ($k = 0.1$) collapses the supervision prematurely into uniform weighting, losing the stability benefits of the initial deep-layer bias. We find $k = 0.05$ provides the optimal trade-off, allowing the model to stabilize high-level features before progressively refining the entire encoder for better generalization.

The results clearly highlight the advantage of the proposed PLS (Table 3) over both standard and aligned fine-tuning (Table 1), particularly when considering the behavior of earlier layers. Unlike other approaches, PLS introduces a gradual and controlled supervision mechanism that better leverages the representational hierarchy of the model. This progressive strategy allows early layers to contribute meaningfully without being prematurely optimized, leading to a much more stable and consistent performance across diverse conditions. As a result, PLS avoids the large performance variance observed in the first layers of both baselines and achieves significantly improved generalization, especially on OOD and challenging scenarios. Moreover, while aligned fine-tuning tends to delay

Table 3: *XLS-R-128* 300M under *PLS* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
0	50.48	49.39	50.62	50.89	40.74	51.06	47.09	59.07	36.46	53.00	76.06	49.12	51.16	46.50
1	54.88	61.57	49.11	47.06	43.94	53.50	41.52	55.23	44.15	49.65	63.21	43.88	50.64	47.58
2	49.98	57.82	43.82	51.95	42.56	46.46	43.98	59.73	47.57	66.30	37.37	47.25	49.57	51.83
3	35.15	49.74	35.46	52.74	52.93	70.33	44.47	42.26	37.83	55.79	25.71	45.15	45.63	47.20
4	21.68	27.52	29.42	51.53	60.82	77.77	55.63	25.05	40.42	25.47	20.27	44.50	40.01	41.02
5	7.19	10.14	12.88	27.98	39.38	61.93	35.78	22.39	41.45	23.20	14.21	24.25	26.73	29.44
6	2.86	6.13	5.77	13.02	15.64	31.81	29.49	18.01	23.00	25.44	11.22	17.11	16.62	19.34
7	0.24	3.27	2.11	2.61	6.57	8.87	22.29	9.93	14.54	20.99	4.43	15.09	9.25	15.14
8	0.16	3.13	1.40	1.33	4.76	3.71	19.53	8.74	11.35	18.21	2.99	14.08	7.45	13.48
9	0.14	3.36	1.44	1.25	4.59	3.09	18.86	9.27	12.37	18.24	2.20	13.82	7.39	13.07
10	0.10	3.14	1.35	1.17	4.41	2.44	18.60	8.46	11.71	17.19	2.28	13.69	7.04	12.84
11	0.11	2.96	1.32	1.06	4.35	2.16	18.26	8.87	11.10	17.10	2.32	13.32	6.91	12.64
12	0.10	3.25	1.37	1.10	4.37	1.81	18.00	8.87	10.44	16.85	1.88	13.29	6.78	12.65
13	0.10	2.93	1.39	1.10	4.44	2.29	17.94	9.93	11.32	16.94	2.39	13.30	7.01	12.57
14	0.11	2.83	1.37	1.14	4.32	2.61	17.91	10.08	11.25	16.50	2.09	13.24	6.95	12.53
15	0.08	2.50	1.35	1.14	4.27	3.00	17.94	9.93	11.81	15.71	2.51	13.26	6.96	12.56
16	0.08	2.46	1.30	1.21	4.15	2.29	17.75	8.90	10.49	14.78	2.15	13.07	6.55	12.41
17	0.10	2.65	1.28	1.18	4.16	2.07	17.68	8.22	10.21	15.27	2.06	12.99	6.49	12.25
18	0.08	2.58	1.33	1.21	4.27	2.29	17.93	9.17	11.07	15.14	2.16	13.18	6.70	12.30
19	0.10	2.74	1.35	1.21	4.13	2.61	18.35	7.97	10.39	14.06	1.70	12.91	6.46	12.20
20	0.11	3.16	1.35	1.11	4.01	2.65	17.93	8.21	10.58	15.74	1.09	13.07	6.58	12.22
21	0.10	2.67	1.30	1.18	4.55	3.45	18.35	8.51	9.68	13.31	1.70	13.47	6.52	12.41
22	0.07	2.24	1.21	1.33	5.48	4.99	19.38	8.47	11.09	10.72	2.11	14.74	6.82	12.99
23	0.10	2.68	1.32	1.33	4.88	3.22	18.52	8.47	11.21	11.18	1.44	14.87	6.60	12.68
24	0.11	2.88	1.39	1.29	4.20	3.04	19.73	7.96	10.74	11.07	2.13	15.31	6.65	12.03

useful learning signals to deeper layers, PLS ensures a smoother transition across layers, yielding competitive—and often superior—performance not only in early layers but also maintaining strong results throughout the entire network. This demonstrates that PLS is not merely a training heuristic, but an effective paradigm for exploiting layer-wise learning dynamics in SFMs.

Table 4: Overall performance (EER %, ↓) compared to SOTA approaches. **Bold** indicates best results, the second-best are underlined. † denotes results reported from [47]. ⋄ denotes results computed by using official released checkpoints.

Model	#Params. (Inference)	In-domain			Out-of-domain					Average
		19LA	21LA	21DF	25DF	ITW	DFADD	LSV	SC	
SLS[6]†⋄	340M	0.23	2.87	1.91	18.76	7.46	7.54	1.97	24.51	8.16
Conformer[7]†⋄	319M	0.19	1.03	2.06	18.85	7.79	8.89	2.35	38.15	9.91
MultiConv[10]⋄	317M	**0.08**	2.77	1.43	15.19	4.44	**6.60**	1.70	18.71	6.37
KAN[9]⋄	317M	0.11	2.29	1.49	14.10	5.31	15.63	1.52	22.06	7.81
Mamba[8]⋄	317M	0.11	1.78	1.51	13.58	5.12	8.62	1.82	17.87	6.30
Nes2NetX[48]†⋄	317M	0.12	1.88	1.49	22.06	5.52	11.15	2.88	54.28	12.42
AASIST[4]†⋄	317M	0.22	**0.82**	2.85	16.25	11.20	11.93	11.21	26.72	10.15
Baseline - FT	315M	0.10	2.36	1.46	13.47	4.69	17.51	1.75	19.57	7.61
Layer 17 (Ours)	239M	0.10	2.65	**1.28**	**12.99**	**4.16**	8.22	1.18	**17.68**	**6.03**
Layer 11 (Ours)	152M	0.11	2.96	1.32	13.32	4.35	8.87	**1.06**	18.26	6.28

Detection performance versus SOTA. Table 4 reports the EER of the compared methods on both in-domain and OOD evaluation sets. Our proposed PLS models consistently improve over the standard fine-tuning baseline, especially in OOD conditions. For instance, *Aligned - Layer 17* reduces the average EER from 7.61% to 6.03%, corresponding to a relative improvement of 20.8%, while also outperforming the baseline on all OOD datasets. Notably, it achieves lower EERs on 25DF (12.99%), ITW (4.16%), DFADD (8.22%), LSV (1.18%), and SC (17.68%), indicating stronger robustness to unseen spoofing conditions. Compared with prior SOTA methods, *PLS - Layer 17* achieves the best overall average EER of 6.03%, outperforming *Mamba* while using fewer trainable parameters (239M vs. 317M). Moreover, deeper pruning levels preserve strong detection capability: *PLS - Layer 11* attains a competitive 6.28% average EER with only 152M parameters, showing that the proposed alignment strategy enables efficient model compression without sacrificing generalization. These results suggest that the proposed PLS alignment improves the transferability of spoof-discriminative representations, leading to better cross-domain robustness with reduced model complexity.

6 Discussion and limitation

We attribute OOD gains to increased representation diversity by distributing discriminative power across layers, PLS prevents the final-layer vulnerability to distribution shifts seen in standard models. This approach yields a 70% reduction in inference parameters (101M vs. 315M) without the architectural overhead or teacher-dependency of distillation. However, these benefits are most evident in large multilingual models, as smaller backbones may lack the capacity to absorb intermediate supervision without performance trade-offs. Furthermore, while PLS optimizes inference, it still requires training the full model to convergence prior to pruning.

7 Conclusion

We introduced Progressive Layer-wise Supervision (PLS) to resolve shallow-layer dormancy in fine-tuned speech foundation models (SFMs). By utilizing exponentially annealed power weighting and shared classification, PLS activates early layers, enabling reliable early-exit inference nine layers sooner than standard method and achieving state-of-the-art out-of-domain generalization (6.03% EER) comparing to prior approaches with fewer inference parameters. While PLS significantly reduces inference parameters, future work will focus on decreasing training time by analyzing relevant SFM hidden features to prune redundant parameters during the training phase itself.

Acknowledgements

This work was granted access to the HPC/AI resources of IDRIS under the allocation made by GENCI, as well as the GPU resources provided by the Expression team at IRISA. This work was partially supported by JST PRESTO, Japan (Grant Number JPMJPR23P9).

References

[1] Xu Tan, Tao Qin, Frank Soong, and Tie-Yan Liu. A survey on neural speech synthesis. *arXiv preprint arXiv:2106.15561*, 2021.

[2] Joao Medeiros. How intel gave stephen hawking a voice. *Wired Retrieved from https://www. wired. com/2015/01/intel-gave-stephen-hawking-voice*, 2015.

[3] Femi Folorunsho and Benedicta Frema Boamah. Deepfake technology and its impact: ethical considerations, societal disruptions, and security threats in ai-generated media. *International journal of information technology and management information systems*, 16(1):1060–1080, 2025.

[4] Hemlata Tak, Massimiliano Todisco, Xin Wang, Jee weon Jung, Junichi Yamagishi, and Nicholas Evans. Automatic speaker verification spoofing and deepfake detection using wav2vec 2.0 and data augmentation. In *The Speaker and Language Recognition Workshop (Odyssey 2022)*, pages 112–119, 2022. doi: 10.21437/Odyssey.2022-16.

[5] Yi Zhu, Surya Koppisetti, Trang Tran, and Gaurav Bharaj. SLIM: Style-linguistics mismatch model for generalized audio deepfake detection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=fymr0CBDHZ`.

[6] Qishan Zhang, Shuangbing Wen, and Tao Hu. Audio deepfake detection with self-supervised XLS-r and SLS classifier. In *ACM Multimedia 2024*, 2024. URL `https://openreview.net/forum?id=acJMIXJg2u`.

[7] Duc-Tuan Truong, Ruijie Tao, Tuan Nguyen, Hieu-Thi Luong, Kong Aik Lee, and Eng Siong Chng. Temporal-channel modeling in multi-head self-attention for synthetic speech detection. In *Interspeech 2024*, pages 537–541, 2024. doi: 10.21437/Interspeech.2024-659.

[8] Hoan My Tran, Damien Lolive, David Guennec, Aghilas Sini, Arnaud Delhay, and Pierre-François Marteau. Leveraging SSL Speech Features and Mamba for Enhanced DeepFake Detection. In *Interspeech 2025*, pages 5323–5327, 2025. doi: 10.21437/Interspeech.2025-1703.

[9] Hoan My Tran, Damien Lolive, David Guennec, Aghilas Sini, Arnaud Delhay, and Pierre-François Marteau. Kan you hear the truth? audio deepfake detection with kolmogorov–arnold networks. In *2025 33rd European Signal Processing Conference (EUSIPCO)*, pages 561–565, 2025. doi: 10.23919/EUSIPCO63237.2025.11226534.

[10] Hoan My Tran, Damien Lolive, Aghilas Sini, Arnaud Delhay, Pierre-François Marteau, and David Guennec. Multi-level ssl feature gating for audio deepfake detection. In *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, page 11766–11775, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720352. doi: 10.1145/3746027.3754568. URL https://doi.org/10.1145/3746027.3754568.

[11] Nicolas Müller, Pavel Czempin, Franziska Diekmann, Adam Froghyar, and Konstantin Böttinger. Does audio deepfake detection generalize? In *Interspeech 2022*, pages 2783–2787, 2022. doi: 10.21437/Interspeech.2022-108.

[12] Zhongjie Ba, Qing Wen, Peng Cheng, Yuwei Wang, Feng Lin, Li Lu, and Zhenguang Liu. Transferring audio deepfake detection capability across languages. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 2033–2044, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394161. doi: 10.1145/3543507.3583222. URL https://doi.org/10.1145/3543507.3583222.

[13] Chengzhe Sun, Shan Jia, Shuwei Hou, and Siwei Lyu. Ai-synthesized voice detection using neural vocoder artifacts. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 904–912, 2023. doi: 10.1109/CVPRW59228.2023.00097.

[14] Jiawei Du, I-Ming Lin, I-Hsiang Chiu, Xuanjun Chen, Haibin Wu, Wenze Ren, Yu Tsao, Hung-Yi Lee, and Jyh-Shing Roger Jang. Dfadd: The diffusion and flow-matching based audio deepfake dataset. In *2024 IEEE Spoken Language Technology Workshop (SLT)*, pages 921–928, 2024. doi: 10.1109/SLT61566.2024.10832250.

[15] Nicolas M. Müller, Piotr Kawa, Wei Herng Choong, Edresson Casanova, Eren Gölge, Thorsten Müller, Piotr Syga, Philip Sperl, and Konstantin Böttinger. Mlaad: The multi-language audio anti-spoofing dataset. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2024. doi: 10.1109/IJCNN60899.2024.10650962.

[16] Xin Wang, Héctor Delgado, Hemlata Tak, Jee weon Jung, Hye jin Shim, Massimiliano Todisco, Ivan Kukanov, Xuechen Liu, Md Sahidullah, Tomi Kinnunen, Nicholas Evans, Kong Aik Lee, Junichi Yamagishi, Myeonghun Jeong, Ge Zhu, Yongyi Zang, You Zhang, Soumi Maiti, Florian Lux, Nicolas Müller, Wangyou Zhang, Chengzhe Sun, Shuwei Hou, Siwei Lyu, Sébastien Le Maguer, Cheng Gong, Hanjie Guo, Liping Chen, and Vishwanath Singh. Asvspoof 5: Design, collection and validation of resources for spoofing, deepfake, and adversarial attack detection using crowdsourced speech. *Computer Speech & Language*, 95:101825, 2026. ISSN 0885-2308. doi: https://doi.org/10.1016/j.csl.2025.101825. URL https://www.sciencedirect.com/science/article/pii/S0885230825000506.

[17] Jee-weon Jung, Yihan Wu, Xin Wang, Ji-Hoon Kim, Soumi Maiti, Yuta Matsunaga, Hyejin Shim, Jinchuan Tian, Nicholas Evans, Joon Son Chung, Wangyou Zhang, Seyun Um, Shinnosuke Takamichi, and Shinji Watanabe. Spoofceleb: Speech deepfake detection and sasv in the wild. *IEEE Open Journal of Signal Processing*, 6:68–77, 2025. doi: 10.1109/OJSP.2025.3529377.

[18] Kratika Bhagtani, Amit Kumar Singh Yadav, Paolo Bestagini, and Edward J. Delp. Diffssd: A diffusion-based dataset for speech forensics. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025. doi: 10.1109/ICASSP49660.2025.10889450.

[19] Ankita Pasad, Ju-Chieh Chou, and Karen Livescu. Layer-wise analysis of a self-supervised speech representation model. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 914–921, 2021. doi: 10.1109/ASRU51503.2021.9688093.

[20] Ankita Pasad, Bowen Shi, and Karen Livescu. Comparative layer-wise analysis of self-supervised speech models. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10096149.

[21] Teresa Dorszewski, Albert Kjøller Jacobsen, Lenka Tětková, and Lars Kai Hansen. How redundant is the transformer stack in speech representation models? In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025. doi: 10.1109/ICASSP49660.2025.10887866.

[22] Vardan Papyan, XY Han, and David L Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117 (40):24652–24663, 2020.

[23] Andrey Gromov, Kushal Tirumala, Hassan Shapourian, Paolo Glorioso, and Dan Roberts. The unreasonable ineffectiveness of the deeper layers. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=ngmEcEer8a.

[24] Jiachen Jiang, Jinxin Zhou, and Zhihui Zhu. Tracing representation progression: Analyzing and enhancing layer-wise similarity. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=vVxeFSR4fU.

[25] Peter Súkeník, Christoph H. Lampert, and Marco Mondelli. Neural collapse is globally optimal in deep regularized resnets and transformers. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=8WKOk4U9R4.

[26] Sefik Emre Eskimez, Xiaofei Wang, Manthan Thakker, Canrun Li, Chung-Hsien Tsai, Zhen Xiao, Hemin Yang, Zirun Zhu, Min Tang, Xu Tan, Yanqing Liu, Sheng Zhao, and Naoyuki Kanda. E2 tts: Embarrassingly easy fully non-autoregressive zero-shot tts. In *2024 IEEE Spoken Language Technology Workshop (SLT)*, pages 682–689, 2024. doi: 10.1109/SLT61566.2024.10832320.

[27] Zeqian Ju, Yuancheng Wang, Kai Shen, Xu Tan, Detai Xin, Dongchao Yang, Eric Liu, Yichong Leng, Kaitao Song, Siliang Tang, Zhizheng Wu, Tao Qin, Xiangyang Li, Wei Ye, Shikun Zhang, Jiang Bian, Lei He, Jinyu Li, and sheng zhao. Naturalspeech 3: Zero-shot speech synthesis with factorized codec and diffusion models. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=dVhrnjZJad.

[28] Yushen Chen, Zhikang Niu, Ziyang Ma, Keqi Deng, Chunhui Wang, JianZhao JianZhao, Kai Yu, and Xie Chen. F5-TTS: A fairytaler that fakes fluent and faithful speech with flow matching. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6255–6271, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.313. URL https://aclanthology.org/2025.acl-long.313/.

[29] Massimiliano Todisco, Héctor Delgado, and Nicholas Evans. Constant q cepstral coefficients: A spoofing countermeasure for automatic speaker verification. *Computer Speech & Language*, 45:516–535, 2017. ISSN 0885-2308. doi: https://doi.org/10.1016/j.csl.2017.01.001. URL https://www.sciencedirect.com/science/article/pii/S0885230816303114.

[30] Howard Lei and Eduardo Lopez. Mel, linear, and antimel frequency cepstral coefficients in broad phonetic regions for telephone speaker recognition. In *Interspeech 2009*, pages 2323–2326, 2009. doi: 10.21437/Interspeech.2009-389.

[31] Hemlata Tak, Jee weon Jung, Jose Patino, Madhu Kamble, Massimiliano Todisco, and Nicholas Evans. End-to-end spectro-temporal graph attention networks for speaker verification anti-spoofing and speech deepfake detection. In *2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge*, pages 1–8, 2021. doi: 10.21437/ASVSPOOF.2021-1.

[32] Jee-weon Jung, Hee-Soo Heo, Hemlata Tak, Hye-jin Shim, Joon Son Chung, Bong-Jin Lee, Ha-Jin Yu, and Nicholas Evans. Aasist: Audio anti-spoofing using integrated spectro-temporal graph attention networks. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6367–6371, 2022. doi: 10.1109/ICASSP43922.2022.9747766.

[33] Yujie Chen, Jiangyan Yi, Jun Xue, Chenglong Wang, Xiaohui Zhang, Shunbo Dong, Siding Zeng, Jianhua Tao, Zhao Lv, and Cunhang Fan. RawBMamba: End-to-End Bidirectional State Space Model for Audio Deepfake Detection. In *Interspeech 2024*, pages 2720–2724, 2024. doi: 10.21437/Interspeech.2024-698.

[34] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33:12449–12460, 2020.

[35] Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, et al. Wavlm: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, 16(6):1505–1518, 2022.

[36] Hoan My Tran, Aghilas Sini, David Guennec, Arnaud Delhay, Damien Lolive, and Pierre-François Marteau. Kan we make models simpler for audio deepfake detection with kolmogorov–arnold networks? In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 18057–18061, 2026. doi: 10.1109/ICASSP55912.2026.11460320.

[37] Ji Xin, Raphael Tang, Jaejun Lee, Yaoliang Yu, and Jimmy Lin. DeeBERT: Dynamic early exiting for accelerating BERT inference. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2246–2251, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.204. URL `https://aclanthology.org/2020.acl-main.204/`.

[38] Wangchunshu Zhou, Canwen Xu, Tao Ge, Julian McAuley, Ke Xu, and Furu Wei. Bert loses patience: Fast and robust inference with early exit. *Advances in Neural Information Processing Systems*, 33:18330–18341, 2020.

[39] Arthur Pimentel, Yi Zhu, Heitor R. Guimarães, and Tiago H. Falk. Efficient audio deepfake detection using wavlm with early exiting. In *2024 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, 2024. doi: 10.1109/WIFS61860.2024.10810699.

[40] Shu wen Yang, Po-Han Chi, Yung-Sung Chuang, Cheng-I Jeff Lai, Kushal Lakhotia, Yist Y. Lin, Andy T. Liu, Jiatong Shi, Xuankai Chang, Guan-Ting Lin, Tzu-Hsien Huang, Wei-Cheng Tseng, Ko tik Lee, Da-Rong Liu, Zili Huang, Shuyan Dong, Shang-Wen Li, Shinji Watanabe, Abdelrahman Mohamed, and Hung yi Lee. SUPERB: Speech Processing Universal PERformance Benchmark. In *Interspeech 2021*, pages 1194–1198, 2021. doi: 10.21437/Interspeech.2021-1775.

[41] Jiatong Shi, Shih-Heng Wang, William Chen, Martijn Bartelds, Vanya Bannihatti Kumar, Jinchuan Tian, Xuankai Chang, Dan Jurafsky, Karen Livescu, Hung yi Lee, and Shinji Watanabe. ML-SUPERB 2.0: Benchmarking Multilingual Speech Models Across Modeling Constraints, Languages, and Datasets. In *Interspeech 2024*, pages 1230–1234, 2024. doi: 10.21437/Interspeech.2024-2248.

[42] Xin Wang, Junichi Yamagishi, Massimiliano Todisco, Héctor Delgado, Andreas Nautsch, Nicholas Evans, Md Sahidullah, Ville Vestman, Tomi Kinnunen, Kong Aik Lee, et al. Asvspoof 2019: A large-scale public database of synthesized, converted and replayed speech. *Computer Speech & Language*, 64:101114, 2020.

[43] Junichi Yamagishi, Christophe Veaux, and Kirsten MacDonald. Cstr vctk corpus: English multi-speaker corpus for cstr voice cloning toolkit (version 0.92). *The Rainbow Passage which the speakers read out can be found in the International Dialects of English Archive:(http://web.ku. edu/˜ idea/readings/rainbow. htm).*, 2019.

[44] Hemlata Tak, Madhu Kamble, Jose Patino, Massimiliano Todisco, and Nicholas Evans. Rawboost: A raw data boosting and augmentation method applied to automatic speaker verification anti-spoofing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6382–6386. IEEE, 2022.

[45] Huang et al. SpeechFake: A large-scale multilingual speech deepfake dataset incorporating cutting-edge generation methods. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9985–9998, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.493. URL https://aclanthology.org/2025.acl-long.493/.

[46] Ricardo Reimao and Vassilios Tzerpos. For: A dataset for synthetic speech detection. In *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 1–10, 2019. doi: 10.1109/SPED.2019.8906599.

[47] Sandipana Dowerah, Atharva Kulkarni, Ajinkya Kulkarni, Hoan My Tran, Joonas Kalda, Artem Fedorchenko, Benoit Fauve, Damien Lolive, Tanel Alumäe, and Mathew Magimai.-Doss. Speech df arena: A leaderboard for speech deepfake detection models. *IEEE Open Journal of Signal Processing*, 7:73–81, 2026. doi: 10.1109/OJSP.2026.3652496.

[48] Tianchi Liu, Duc-Tuan Truong, Rohan Kumar Das, Kong Aik Lee, and Haizhou Li. Nes2net: A lightweight nested architecture for foundation model driven speech anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 20:12005–12018, 2025. doi: 10.1109/TIFS.2025.3626963.

[49] Xuechen Liu, Xin Wang, Md Sahidullah, Jose Patino, Héctor Delgado, Tomi Kinnunen, Massimiliano Todisco, Junichi Yamagishi, Nicholas Evans, Andreas Nautsch, et al. Asvspoof 2021: Towards spoofed and deepfake speech detection in the wild. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:2507–2522, 2023.

[50] Jaime Lorenzo-Trueba, Junichi Yamagishi, Tomoki Toda, Daisuke Saito, Fernando Villavicencio, Tomi Kinnunen, and Zhenhua Ling. The Voice Conversion Challenge 2018: Promoting Development of Parallel and Nonparallel Methods . In *The Speaker and Language Recognition Workshop (Odyssey 2018)*, pages 195–202, 2018. doi: 10.21437/Odyssey.2018-28.

[51] Zhao Yi, Wen-Chin Huang, Xiaohai Tian, Junichi Yamagishi, Rohan Kumar Das, Tomi Kinnunen, Zhen-Hua Ling, and Tomoki Toda. Voice Conversion Challenge 2020 — Intra-lingual semi-parallel and cross-lingual voice conversion —. In *Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020*, pages 80–98, 2020. doi: 10.21437/VCCBC.2020-14.

[52] Vineel Pratap, Qiantong Xu, Anuroop Sriram, Gabriel Synnaeve, and Ronan Collobert. Mls: A large-scale multilingual dataset for speech research. In *Interspeech 2020*, pages 2757–2761, 2020. doi: 10.21437/Interspeech.2020-2826.

[53] Heiga Zen, Viet Dang, Rob Clark, Yu Zhang, Ron J. Weiss, Ye Jia, Zhifeng Chen, and Yonghui Wu. LibriTTS: A Corpus Derived from LibriSpeech for Text-to-Speech. In *Interspeech 2019*, pages 1526–1530, 2019. doi: 10.21437/Interspeech.2019-2441.

[54] Changhan Wang, Morgane Riviere, Ann Lee, Anne Wu, Chaitanya Talnikar, Daniel Haziza, Mary Williamson, Juan Pino, and Emmanuel Dupoux. VoxPopuli: A large-scale multilingual speech corpus for representation learning, semi-supervised learning and interpretation. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 993–1003, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.80. URL https://aclanthology.org/2021.acl-long.80/.

[55] Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 29:3451–3460, October 2021. ISSN 2329-9290. doi: 10.1109/TASLP.2021.3122291. URL https://doi.org/10.1109/TASLP.2021.3122291.

[56] Marcely Zanon Boito, Vivek Iyer, Nikolaos Lagos, Laurent Besacier, and Ioan Calapodescu. mHuBERT-147: A Compact Multilingual HuBERT Model. In *Interspeech 2024*, pages 3939–3943, 2024. doi: 10.21437/Interspeech.2024-938.

[57] Sanyuan Chen, Yu Wu, Chengyi Wang, Zhengyang Chen, Zhuo Chen, Shujie Liu, Jian Wu, Yao Qian, Furu Wei, Jinyu Li, and Xiangzhan Yu. Unispeech-sat: Universal speech representation learning with speaker aware pre-training. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6152–6156, 2022. doi: 10.1109/ICASSP43922.2022.9747077.

[58] Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. Xls-r: Self-supervised cross-lingual speech representation learning at scale. In *Interspeech 2022*, pages 2278–2282, 2022. doi: 10.21437/Interspeech.2022-143.

[59] Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. Unsupervised Cross-Lingual Representation Learning for Speech Recognition. In *Interspeech 2021*, pages 2426–2430, 2021. doi: 10.21437/Interspeech.2021-329.

[60] Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, Alexei Baevski, Yossi Adi, Xiaohui Zhang, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. Scaling speech technology to 1,000+ languages. *J. Mach. Learn. Res.*, 25(1), January 2024. ISSN 1532-4435.

[61] ASR Omnilingual, Gil Keren, Artyom Kozhevnikov, Yen Meng, Christophe Ropers, Matthew Setzler, Skyler Wang, Ife Adebara, Michael Auli, Can Balioglu, et al. Omnilingual asr: Open-source multilingual speech recognition for 1600+ languages. *arXiv preprint arXiv:2511.09690*, 2025.

[62] Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pages 3519–3529. PMlR, 2019.

A Algorithm of computing training loss

The algorithm, which returnes the training loss for the e-th epoch based on the Progressive Layer-wise Supervision and the training dynamics, is listed below.

Algorithm 1: Progressive Layer-wise Supervision (PLS)

Input: $\{\mathcal{L}_l\}_{l=0}^{L}$: cross-entropy losses for L layers;
current epoch e;
initial emphasis M;
decay rate k.
Output: Training loss $\mathcal{L}_{\text{total}}(e)$

```
// Primary objective anchoring (even epochs)
```
if $e \equiv 0 \pmod 2$ **then**
  return $\mathcal{L}_L$ `// Standard final-layer fine-tuning`
end
```
// Compute scheduled power parameter (Eq. 2)
```
$\rho(e) \leftarrow M \cdot \exp(-k \cdot e)$
```
// Compute depth-aware layer weights
```
$W \leftarrow 0$
for $l = 0$ ***to*** L **do**
  $\tilde{w}_l \leftarrow (l+1)^{\rho(e)}$
  $W \leftarrow W + \tilde{w}_l$
end
```
// Normalize and compute weighted objective (Eq. 1)
```
$\mathcal{L}_{\text{total}} \leftarrow 0$
for $l = 0$ ***to*** L **do**
  $w_l \leftarrow \tilde{w}_l \,/\, W$
  $\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{total}} + w_l \cdot \mathcal{L}_l$
end
return $\mathcal{L}_{\text{total}}$

B Experiment details

B.1 Datasets

This section summarizes the datasets used in our experiments, covering both benchmark anti-spoofing corpora and recent deepfake datasets designed to evaluate robustness under diverse synthesis conditions. The details are also listed in Table 5.

ASVspoof 2019 LA. The `ASVspoof 2019` LA dataset (`19LA`) [42] is a standard benchmark for evaluating spoofing countermeasures against synthesized and voice-converted speech. It contains bona fide utterances from the VCTK corpus [43] and spoofed samples generated using 17 TTS and VC systems. Six attack methods are available during training and development, while thirteen additional unseen attacks are reserved for evaluation, making the benchmark suitable for studying generalization to unknown spoofing methods in controlled conditions.

ASVspoof 2021 LA and DF. The `ASVspoof 2021` LA dataset (`21LA`) [49] extends `19LA` by introducing more realistic transmission effects, including codec distortions and channel variability, to better simulate telephony and VoIP scenarios. The `ASVspoof 2021` DF (`21DF`) track focuses on general audio deepfake detection beyond ASV-specific vulnerabilities. It includes bona fide and spoofed utterances processed under various codec conditions to evaluate robustness against compression artifacts [50, 51].

ASVspoof 5. `ASVspoof 5` (`25DF`) [16] is the latest edition of the `ASVspoof` challenge and presents a substantially more realistic evaluation setting. Bona fide speech is collected from approximately

2,000 speakers under diverse acoustic environments, primarily derived from the Multilingual LibriSpeech corpus [52]. Spoofed samples are generated using modern TTS/VC systems and adversarial attacks, while the evaluation set introduces unseen codecs and compression conditions to assess generalization in realistic scenarios.

SpoofCeleb. SpoofCeleb (SC) [17] is a large-scale in-the-wild speech deepfake benchmark built from VoxCeleb1 celebrity recordings. It contains over 2.5 million utterances from 1,251 speakers, with spoofed speech generated using 23 TTS systems. The evaluation protocol emphasizes noisy and reverberant environments, high speaker diversity, and realistic synthetic attacks, making it challenging for existing spoofing detectors.

In-the-Wild. The In-the-Wild (ITW) dataset [11] contains approximately 38 hours of short audio clips collected from public speeches and podcasts involving 58 English-speaking public figures. Spoofed samples are obtained from publicly available synthetic audio sources and are matched with bona fide utterances in speaking style, duration, and background noise. ITW serves as a realistic out-of-domain benchmark, highlighting the degradation of models trained only on controlled datasets.

Fake-or-Real. The Fake-or-Real (FoR) dataset [46] contains over 195,000 English utterances, including both real and synthetic speech produced by commercial and research TTS systems. Bona fide recordings cover a wide range of speakers, accents, recording devices, and acoustic environments. Several dataset variants are provided; in this work, we use the original version to evaluate generalization under diverse real-world conditions.

Multi-Language Audio Anti-Spoofing Dataset. The Multi-Language Audio Anti-Spoofing Dataset (MLAAD) [15] is a multilingual benchmark designed to improve cross-lingual generalization in spoofing detection. It contains more than 570 hours of spoofed speech generated by 119 TTS systems across 40 languages. Evaluation data includes additional noise, music, and codec perturbations, making MLAAD a valuable benchmark for testing robustness to linguistic diversity and unseen synthesis methods.

Deepfake Cross-Lingual Evaluation. The Deepfake Cross-Lingual (DECRO) dataset [12] is a bilingual benchmark targeting cross-lingual robustness in audio deepfake detection. It includes English and Chinese subsets, with spoofed samples generated using both TTS and VC systems. The dataset is specifically designed to measure the impact of language mismatch between training and testing conditions.

LibriSeVoc. LibriSeVoc (LSV) [13] focuses on detecting artifacts introduced by neural vocoders rather than complete TTS pipelines. Built from LibriTTS [53], it contains over 92,000 utterances generated by six neural vocoders representing autoregressive, GAN-based, and diffusion architectures. This dataset is useful for analyzing vocoder-specific artifacts in synthetic speech.

Diffusion and Flow-Matching Audio Deepfake Dataset. The Diffusion and Flow-Matching Audio Deepfake Dataset (DFADD) [14] is designed to evaluate spoofing countermeasures against speech generated by diffusion and flow-matching models. Built on VCTK [43], it contains approximately 180 hours of data with over 160,000 spoofed utterances, posing challenges due to the naturalness of modern generative methods.

Diffusion-based Synthetic Speech Dataset. The Diffusion-based Synthetic Speech Dataset (DIFFSSD) [18] targets the detection of highly realistic synthetic speech generated by diffusion-based TTS systems. It includes around 200 hours of speech from 74 speakers, paired with 70,000 spoofed utterances generated by both open-source and commercial diffusion models. DIFFSSD is particularly useful for evaluating the robustness of spoofing detectors to SOTA diffusion synthesis.

SpeechFake. The SpeechFake (SF) dataset [45] is a massive repository of synthetic audio containing over 3 million deepfake utterances and exceeding 3,000 hours of playtime. The dataset is strategically divided into a bilingual focus on English and Chinese, alongside a broader multilingual collection spanning 46 languages. To ensure technical diversity, samples were produced using 40

different tools—ranging from 30 open-source projects to 10 commercial APIs—utilizing technologies like TTS, VC, and neural vocoders. By integrating cutting-edge synthesis methods with high speaker variability and linguistic breadth, SF is designed to simulate authentic deepfake threats and improve the ability of detection models to generalize across different languages.

Table 5: Statistics of datasets used in the study.

Dataset	Language	# Bona fide	# Spoofed	Attack
Training				
ASVspoof 2019 logical access (19LA) [42]	English	2,580	22,800	TTS, VC
ASVspoof5 (25DF) [16]	English	18,797	163,560	TTS, VC, AT
SpoofCeleb (SC) [17]	English	230,948	230,9473	TTS
SpeechFake (SF) [45]	English/Chinese	75,708	633,354	TTS, VC, Vocoded
Development				
ASVspoof 2019 logical access (19LA) [42]	English	2,548	22,296	TTS, VC
ASVspoof5 (25DF) [16]	English	31,334	109,616	TTS, VC, AT
SpoofCeleb (SC) [17]	English	7,963	47,778	TTS
SpeechFake (SF) [45]	English/Chinese	12,618	105,544	TTS, VC, Vocoded
Evaluation				
ASVspoof 2019 logical access (19LA) [42]	English	7,355	63,882	TTS, VC
ASVspoof 2021 logical access (21LA) / deepfake (21DF) [49]	English	14,816 / 14,869	133,360 / 519,059	TTS, VC
ASVspoof5 (25DF) [16]	English	138,688	542,086	TTS, VC, AT
Fake-or-Real (FoR) [46]	English	2,264	2,370	TTS
In-the-Wild (ITW) [11]	English	19,963	11,816	Unknown
Diffusion-based Synthetic Speech Dataset (DFADD) [14]	English	755	3,000	TTS
Librisevoc (LSV) [13]	English	2,641	15,846	Vocoded
SpoofCeleb (SC) [17]	English	9,113	82,017	TTS
Deepfake Cross-Lingual Evaluation (DECRO-EN) [12]	English	4,306	14,884	TTS, VC
Multi-Language Audio Anti-Spoofing Dataset (MLAAD-EN) [15]	English	28,233	36,000	TTS
Diffusion-based Synthetic Speech Dataset (DIFFSSD) [18]	English	12,113	42,500	TTS

AT: Adversarial attack.

B.2 Speech foundation models

We evaluate a diverse set of SFMs spanning different SSL objectives, model scales, and language coverage, in order to assess how these factors influence spoofing detection performance. Table 9 summarizes the best-performing hidden layer for each model under standard and aligned fine-tuning.

Wav2Vec2. *Wav2Vec2* [34] is a transformer-based self-supervised speech encoder trained using masked latent prediction with a contrastive objective. It learns contextualized representations directly from raw audio by predicting quantized latent targets for masked time steps. We evaluate both English-only versions and *Wav2Vec2 Vox.*, a multilingual variant pretrained on multilingual VoxPopuli corpus [54]. These models provide a strong baseline for assessing their impact on spoofing robustness.

HuBERT. *HuBERT* [55] extends the masked prediction framework by predicting offline-generated cluster assignments instead of contrastive targets. This hidden-unit prediction objective encourages the model to learn phonetic and acoustic structure from unlabeled speech. We evaluate the standard English *HuBERT* model and *mHuBERT-147* [56], a compact multilingual extension trained on 147 languages, enabling analysis of cross-lingual generalization for spoofing detection.

UniSpeech-SAT. *UniSpeech-SAT* [57] augments self-supervised speech learning with speaker-aware contrastive objectives, improving speaker discriminability while preserving phonetic information. We evaluate both the standard *UniSpeechSAT* model and *UniSpeechSAT+*, a variant version pretrained on more diverse English speech data. Their speaker-aware representations are particularly relevant for anti-spoofing, where preserving speaker cues can help identify synthesis artifacts.

WavLM. *WavLM* [35] builds upon *HuBERT*-style masked prediction by introducing gated relative position bias and utterance mixing during pretraining, allowing the model to capture both content

and speaker information more effectively. We evaluate *WavLM* and *WavLM+*, where the latter is pretrained on a larger corpus and offers improved representation quality. These models are expected to be particularly robust to channel and speaker variability.

XLS-R. *XLS-R* [58] extends the *Wav2Vec2* framework to massively multilingual speech representation learning using data from over 100 languages. We evaluate *XLS-R-53* [59] and *XLS-R-128*, which differ primarily in parameter count and pretraining scale. Their multilingual pretraining is designed to improve transfer across languages and recording conditions, making them strong candidates for out-of-domain spoofing detection.

MMS. *MMS* [60] is a massively multilingual speech representation model trained on over one thousand languages using self-supervised objectives similar to *Wav2Vec2*. Its large linguistic diversity allows it to capture speech characteristics beyond language-specific patterns, potentially improving robustness to unseen spoofing conditions and cross-lingual distribution shifts.

Omnilingual. *Omnilingual* [61] is a multilingual speech foundation model designed for universal language representation across a broad set of languages and domains. By leveraging large-scale multilingual pretraining, it aims to learn language-agnostic acoustic features that generalize well across diverse speech tasks. This makes it particularly suitable for evaluating whether universal representations improve robustness against unseen spoofing attacks.

Model scale and language scope. To analyze the role of model capacity, we group SFMs into *Base*, *Large*, and *XL* categories according to their parameter count. We also distinguish between English-only and multilingual pretraining. English-only models are trained primarily on English corpora and may learn more specialized acoustic representations, whereas multilingual models are exposed to broader phonetic and acoustic diversity, potentially improving robustness to out-of-domain and multilingual spoofing scenarios. This model suite enables a systematic comparison of how SSL objectives, parameter scale, and multilingual exposure affect the transferability of hidden-layer representations for synthetic speech detection.

Performances of each SFM in the standard and the proposed training scheme ablation study are presented as tables. To navigate across the tables, Table 6 lists them with their corresponding hyperlinks.

Table 6: Summary of result tables.

Speech foundation model	Description	Hyperlink
Base		
Wav2Vec2	Aligned and standard fine-tuning	Table 10
Wav2Vec2 Vox.	Aligned and standard fine-tuning	Table 11
HuBERT	Aligned and standard fine-tuning	Table 12
UniSpeech-SAT	Aligned and standard fine-tuning	Table 13
UniSpeech-SAT Plus	Aligned and standard fine-tuning	Table 14
WavLM	Aligned and standard fine-tuning	Table 15
WavLM Plus	Aligned and standard fine-tuning	Table 16
mHuBERT-147	Aligned and standard fine-tuning	Table 16
Large		
Wav2Vec2	Aligned and standard fine-tuning	Table 18
Wav2Vec2 Vox.	Aligned and standard fine-tuning	Table 19
HuBERT	Aligned and standard fine-tuning	Table 20
UniSpeech-SAT	Aligned and standard fine-tuning	Table 21
WavLM	Aligned and standard fine-tuning	Table 22
XLS-R-53	Aligned and standard fine-tuning	Table 23
XLS-R-128	Aligned and standard fine-tuning	Table 24
MMS	Aligned and standard fine-tuning	Table 25
Omnilingual	Aligned and standard fine-tuning	Table 26
X-Large		
XLS-R-128	Aligned and standard fine-tuning	Table 27
MMS	Aligned and standard fine-tuning	Table 28
Omnilingual	Aligned and standard fine-tuning	Table 29
SFM aligned fine-tuning ablation study		
All	Best layer with lowest average EER	Table 9
Impact of training data		
XLS-R-128 300M	Aligned fine-tuning on different datasets	Table 8
PLS ablation study		
XLS-R-128 300M	k decay	Table 30
XLS-R-128 300M	ρ scheduling	Table 31
XLS-R-128 300M	Equal-weighting and aligned-only fine-tuning	Table 32

B.3 Hyperparameters and computation resources

Hyperparameters and computation resources used for the experiments are summarized in Table 7.

Table 7: Hyperparameters and architecture details of the models.

Training configuration	Value
Epochs	15
Audio length	Dynamic batching
Optimizer	Adam
Learning rate	2.5×10^{-6}
Weight decay	$1e^{-4}$
Weighted cross-entropy loss	0.9 for real, 0.1 for fake
Early-stop patience	3
Data augmentation	RawBoost
GPU resource	1 NVIDIA A100

C Supplementary experiments results

C.1 Impact of training datasets on layer-wise generalization

We next analyze how the choice of training dataset affects the layer-wise behavior of the model across diverse evaluation benchmarks. Table 8 reports the EER obtained by probing each encoder layer when the model is trained separately on four datasets with different characteristics: `ASVspoof 2019`, `ASVspoof 5`, `SpoofCeleb`, and `SpeechFake`.

A first clear observation is that the optimal operating layer depends strongly on the training corpus. When trained on `ASVspoof 2019`, performance improves steadily from shallow to middle layers, reaching the best average EER at layer 23 (7.36%), while pooled EER is minimized at layer 17 (12.48%). This indicates that training on a controlled benchmark encourages the model to rely on progressively higher-level artifacts that become increasingly discriminative in deeper representations. However, despite strong in-domain performance (e.g., 0.08% on `19LA`), generalization remains limited on more realistic datasets such as `SpoofCeleb` (18.60%) and `25DF` (15.76%), suggesting over-specialization to benchmark-specific synthesis traces.

Training on `ASVspoof 5` yields a different behavior. While the best average EER (10.88%) is higher than that of `ASVspoof 2019`, the resulting detector generalizes more consistently across modern out-of-domain datasets. For instance, the model achieves lower EER on `25DF` (4.52%), `DFADD` (4.40%), and `SpoofCeleb` (12.01%), with the optimal layer shifted earlier (layer 22). This suggests that the broader acoustic diversity and stronger codec variability of `ASVspoof 5` encourage the encoder to learn more robust representations.

The effect is even more pronounced when training on in-the-wild corpora. Models trained on `SpoofCeleb` and `SpeechFake` achieve good results on in-the-wild datasets. Notably, `SpeechFake` yields strong performance across nearly all benchmarks, reaching 5.57% on `SC`, and 7.22% on `DIFSSD`, 3.73% on `MLAAD-EN`, with more robust results on `ITW` (1.50%) and `DFADD` (4.37%). Its pooled EER also reaches the overall best value of 7.83% at layer 16. This indicates that exposure to large-scale and acoustically diverse synthetic data leads to more transferable representations, particularly in intermediate layers.

Interestingly, the layer at which performance saturates also varies with the training corpus. For `ASVspoof 2019`, the best performance emerges in deeper layers (17–23), whereas for `SpeechFake`, performance peaks earlier (12–16) and then plateaus. This suggests that training on diverse and realistic datasets shifts spoof-discriminative information toward intermediate representations, reducing the need for deeper task-specific adaptation. In contrast, benchmark-centric datasets require deeper layers to separate bona fide and spoofed speech.

These results reveal that training data diversity and size of the training corpus has a larger effect on robustness. Although `ASVspoof 2019` produces excellent in-domain results, it underperforms under domain shifts. Conversely, datasets such as `SpeechFake`, which contain broader acoustic and synthesis variability, enable the model to learn more invariant spoof-related cues, leading to substantially better cross-dataset generalization. This finding highlights that the most transferable spoofing cues emerge when the training data covers realistic synthesis conditions, and that such cues are encoded primarily in the middle layers of the pretrained encoder.

Table 8: *XLS-R-128* 300M under *aligned* fine-tuning on different training datasets: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
ASVspoof 2019														
0	21.98	22.08	35.17	53.23	49.12	82.24	49.20	23.57	54.53	34.74	23.37	50.51	44.83	43.52
1	31.84	31.85	34.56	55.51	45.51	80.25	44.75	22.28	47.42	46.41	27.38	52.08	48.01	51.35
2	24.16	27.40	28.89	48.88	34.11	83.92	42.20	17.62	39.30	41.14	20.06	49.29	40.75	45.60
3	20.16	24.84	25.02	45.77	26.48	73.20	40.90	20.00	40.16	40.68	16.84	42.10	39.70	41.65
4	19.17	19.20	19.92	32.97	24.41	50.00	43.29	18.39	35.50	36.08	16.79	42.73	35.96	37.96
5	7.08	9.82	10.73	11.66	15.39	22.23	35.25	13.37	39.15	24.54	7.94	46.32	23.93	33.46
6	1.17	6.03	3.52	3.33	8.40	7.90	23.89	11.76	18.97	20.42	3.72	41.16	14.79	27.81
7	0.29	3.96	1.94	1.82	4.87	3.22	19.95	7.94	17.96	14.71	2.53	35.58	11.11	23.80
8	0.23	3.91	1.86	1.14	4.12	2.52	19.33	6.38	14.07	16.10	1.33	21.63	9.36	16.07
9	0.11	3.62	1.76	0.99	4.07	2.33	18.69	6.91	13.80	15.28	1.35	22.24	8.69	16.36
10	0.10	3.38	1.67	1.02	4.00	2.07	18.58	7.03	13.89	15.00	1.55	21.23	8.68	15.76
11	0.07	3.02	1.60	0.95	3.97	2.11	18.57	7.13	13.51	14.09	1.51	24.56	8.47	18.23
12	0.07	2.98	1.52	0.99	4.05	2.24	18.47	7.29	13.60	14.08	1.40	25.31	8.40	18.40
13	0.07	2.78	1.55	0.92	4.01	2.07	18.33	7.54	13.39	13.48	1.53	19.56	7.97	14.61
14	0.08	2.73	1.55	0.91	4.20	2.07	18.39	7.67	13.70	13.37	1.69	20.09	8.02	15.07
15	0.08	2.67	1.61	1.02	4.17	1.90	18.58	7.79	14.00	12.80	1.92	19.18	8.01	14.41
16	0.07	2.44	1.56	1.06	4.09	1.77	18.38	7.43	12.94	12.66	1.93	17.00	7.73	12.87
17	0.09	2.42	1.62	1.14	4.36	1.68	18.38	7.67	13.89	12.79	1.74	13.83	7.72	12.48
18	0.07	2.00	1.58	1.36	4.37	1.64	18.24	6.89	13.96	11.53	1.88	13.81	7.51	12.55
19	0.08	1.92	1.54	1.40	4.45	1.14	18.38	6.63	13.90	10.89	1.67	14.29	7.46	13.42
20	0.05	1.59	1.54	1.55	4.32	1.36	18.37	5.93	12.90	10.64	2.49	15.21	7.50	13.81
21	0.05	1.59	1.45	1.63	4.42	1.94	19.27	6.50	12.52	10.01	2.92	16.31	7.56	13.86
22	0.07	1.59	1.44	1.59	4.44	2.74	20.33	7.43	12.20	9.51	2.62	16.77	7.44	13.69
23	0.08	1.71	1.49	1.59	3.98	2.44	18.60	9.29	12.28	9.57	2.72	15.76	7.36	13.21
24	0.10	1.96	1.51	1.63	3.62	2.57	18.52	11.40	12.28	9.50	3.83	14.97	7.39	12.58
ASVspoof 5														
0	33.82	34.38	43.30	57.02	45.14	54.42	49.68	29.42	45.57	39.63	60.57	56.66	45.80	43.09
1	25.55	22.70	26.05	52.71	46.85	40.85	43.90	25.85	45.89	36.85	39.43	54.02	38.39	37.96
2	24.91	24.30	25.89	54.68	44.45	57.64	40.06	23.17	46.44	41.92	32.49	54.40	39.20	38.77
3	27.49	25.93	26.39	55.66	37.61	44.30	36.86	20.65	45.30	45.10	34.79	52.37	37.70	38.19
4	28.92	26.74	25.82	57.97	32.67	37.76	31.08	13.90	40.05	40.38	36.11	47.69	34.92	36.21
5	31.42	28.54	24.85	57.63	24.81	36.75	25.43	5.17	32.39	36.06	41.53	41.27	32.15	32.72
6	29.42	26.57	21.69	54.91	22.45	25.31	22.28	3.74	29.78	33.47	36.95	34.17	28.40	28.80
7	24.10	22.52	16.62	42.26	14.44	19.57	18.61	1.99	23.60	28.32	27.66	24.06	21.98	22.16
8	19.61	18.95	13.11	30.21	9.65	13.79	15.28	1.99	19.60	23.63	15.75	16.25	16.48	17.15
9	18.00	17.60	11.80	25.52	8.83	11.52	13.56	3.17	18.29	22.11	16.44	12.13	14.92	14.75
10	16.94	16.39	10.99	24.73	7.97	10.77	13.40	3.04	17.37	21.53	14.51	10.17	13.98	13.77
11	16.87	16.13	10.91	25.44	7.97	10.42	12.82	3.70	17.33	21.25	15.88	9.48	14.02	13.31
12	16.36	15.85	10.51	24.66	8.01	9.95	12.62	4.63	16.23	19.96	14.95	8.49	13.52	12.57
13	16.18	15.69	10.41	24.23	7.90	9.15	12.62	4.77	15.62	19.39	14.86	8.15	13.25	12.24
14	16.06	15.45	10.26	23.93	7.49	9.54	12.56	5.17	15.51	19.02	15.35	7.82	13.18	11.94
15	15.84	15.23	10.08	23.13	7.32	9.71	12.47	4.92	14.72	18.58	14.47	7.04	12.79	11.51
16	15.70	14.92	9.83	22.42	6.96	8.65	12.23	5.43	14.48	18.28	13.93	6.18	12.42	11.05
17	14.90	14.33	9.45	22.04	6.99	9.28	12.19	5.18	13.76	17.07	13.56	5.97	12.06	10.68
18	14.40	13.55	9.14	21.08	6.58	9.15	12.04	3.95	12.72	16.54	9.76	5.13	11.17	10.34
19	14.56	13.77	8.91	22.30	6.66	7.86	12.08	3.57	13.02	15.97	11.38	6.28	11.36	10.53
20	15.74	15.22	9.91	22.07	6.15	7.81	11.95	6.33	12.96	17.02	12.65	5.16	11.92	10.66
21	13.80	13.62	8.94	19.51	6.49	8.91	11.70	5.81	12.57	15.88	10.47	4.48	11.02	10.04
22	13.60	13.28	8.48	20.48	6.33	9.45	12.01	4.40	11.79	15.47	10.80	4.52	10.88	9.71
23	13.03	12.91	8.46	18.89	6.26	9.88	12.19	6.61	12.00	16.68	12.07	3.45	11.04	9.58
24	13.11	12.59	8.74	18.86	6.46	9.24	12.31	6.51	12.54	15.78	12.44	3.61	11.02	9.49
SpoofCeleb														
0	53.64	49.33	48.94	51.46	54.17	47.84	45.83	58.94	50.43	47.77	55.79	51.00	51.26	50.00
1	60.77	56.46	56.74	46.12	57.90	57.29	26.88	62.64	43.39	54.38	63.89	44.97	52.62	51.76
2	48.21	45.45	44.38	47.59	50.41	55.57	17.37	21.86	46.03	55.16	36.46	45.63	42.84	51.81
3	45.02	43.24	39.50	36.88	35.87	55.52	11.81	20.00	41.66	54.36	47.80	36.25	38.99	44.61
4	35.85	37.31	27.08	21.24	20.98	53.84	6.20	9.25	31.40	48.86	53.83	23.76	30.80	31.05
5	28.48	29.75	15.02	11.28	7.80	45.62	2.77	5.28	25.01	36.11	48.31	14.72	22.51	20.33
6	28.92	30.14	14.40	9.88	5.46	43.96	1.21	4.37	22.21	30.25	49.68	13.08	21.13	17.71
7	26.12	28.21	13.59	7.23	4.13	31.14	0.66	5.53	19.85	25.69	48.00	12.05	18.52	15.95
8	24.96	26.53	12.92	5.53	3.59	25.36	0.46	5.18	17.83	21.30	44.26	11.03	16.58	14.91
9	24.39	24.66	11.86	5.03	3.58	23.93	0.30	5.03	16.87	18.90	43.94	10.04	15.71	14.05
10	24.00	24.13	11.33	4.81	3.69	23.41	0.24	4.90	16.16	17.63	43.10	9.79	15.27	13.75
11	22.32	22.21	9.93	4.67	3.63	25.53	0.24	4.53	16.43	16.68	42.89	9.51	14.88	13.25
12	20.75	20.86	9.25	4.96	3.37	23.76	0.23	3.44	15.83	16.21	39.62	9.43	13.98	13.10
13	19.74	20.26	8.98	4.99	3.31	23.11	0.23	3.31	15.69	15.91	39.81	9.33	13.72	12.77
14	19.76	20.16	8.75	5.19	3.25	22.31	0.24	2.91	14.96	15.36	37.90	9.17	13.33	12.74
15	19.67	20.09	8.49	5.04	3.23	23.76	0.24	3.17	14.86	15.13	37.16	9.18	13.33	12.65
16	18.91	19.20	8.47	5.26	3.18	23.09	0.24	2.39	14.36	14.71	35.09	9.11	12.83	12.48
17	18.02	18.58	8.23	4.97	3.15	23.11	0.22	2.67	13.83	13.93	33.95	8.80	12.46	12.17
18	18.38	18.17	7.95	4.73	3.31	23.20	0.21	3.31	14.89	13.82	37.74	8.92	12.89	12.12
19	20.23	19.70	8.37	4.96	3.23	24.17	0.23	3.17	14.36	13.87	38.97	9.00	13.36	12.20
20	21.59	20.92	8.71	5.19	3.06	26.59	0.22	2.77	14.51	14.48	42.78	8.85	14.14	12.25
21	22.67	20.83	8.87	5.19	3.03	27.82	0.20	2.26	15.33	14.74	45.28	9.02	14.60	12.37
22	22.83	20.64	9.01	5.19	3.19	27.88	0.20	2.66	15.58	15.15	45.73	9.10	14.76	12.16
23	23.35	20.84	9.08	5.07	3.21	28.10	0.22	2.79	15.74	15.61	46.54	8.88	14.95	12.12
24	24.43	21.65	9.59	5.42	3.28	30.71	0.19	2.49	15.96	14.49	47.05	8.80	15.34	12.40
SpeechFake														
0	36.42	37.20	40.90	50.62	57.59	49.03	57.28	46.74	56.85	36.56	54.81	48.09	47.67	41.61
1	30.51	33.78	36.36	53.46	42.47	48.94	55.47	28.64	58.31	36.11	26.73	54.02	42.07	43.95
2	34.85	36.12	40.39	55.58	28.85	49.42	51.15	38.70	58.52	29.60	32.33	50.65	42.18	43.25
3	28.99	30.14	37.30	55.02	21.71	47.67	50.83	28.74	58.43	23.72	27.55	51.94	38.50	42.33
4	26.72	28.98	27.80	18.55	14.21	48.23	38.19	19.07	56.58	16.80	26.50	39.81	30.12	32.49
5	16.96	20.51	12.90	1.45	3.97	36.49	17.15	14.05	44.63	6.09	22.64	26.14	18.58	19.65
6	11.39	14.61	8.25	0.53	2.56	28.01	10.01	9.92	39.28	5.06	22.92	18.23	14.23	15.45
7	5.43	9.63	4.96	0.27	1.86	14.57	7.20	5.66	20.77	4.53	18.30	14.76	9.00	12.13
8	3.81	8.01	3.50	0.15	1.56	7.47	6.39	4.37	10.45	4.14	13.65	12.78	6.36	10.11
9	3.56	7.61	3.13	0.08	1.62	5.87	6.21	3.84	7.65	4.02	11.89	12.33	5.65	9.38
10	3.52	7.59	2.93	0.08	1.73	5.65	6.07	4.67	7.34	4.00	10.96	11.91	5.54	9.06
11	3.27	7.40	2.72	0.11	1.64	5.31	5.71	4.50	7.23	3.97	9.73	11.31	5.24	8.65
12	2.68	6.87	2.69	0.14	1.62	5.11	5.64	3.99	7.17	3.86	8.48	11.17	4.95	8.42
13	2.57	6.74	2.64	0.15	1.56	4.77	5.62	3.84	7.11	3.86	8.20	11.10	4.85	8.17
14	2.63	6.78	2.72	0.19	1.62	4.77	5.61	3.97	7.32	3.85	8.13	11.09	4.89	8.21
15	2.50	6.65	2.72	0.19	1.67	4.77	5.51	3.97	7.36	3.86	7.81	10.97	4.83	8.12
16	2.37	6.26	2.55	0.19	1.62	4.68	5.57	4.10	7.22	3.73	7.69	10.93	4.74	7.83
17	2.54	6.52	2.64	0.19	1.66	4.47	5.52	4.39	7.43	3.86	7.92	11.09	4.85	7.99
18	2.34	6.26	2.62	0.19	1.67	4.55	5.56	4.37	7.46	3.98	7.24	11.27	4.79	8.03
19	2.59	6.49	2.70	0.19	1.60	3.32	5.45	4.12	7.93	3.96	7.90	11.20	4.79	7.98
20	2.80	6.82	2.76	0.19	1.64	3.71	5.50	4.37	8.58	4.02	7.85	11.41	4.97	8.32
21	2.88	7.02	2.72	0.19	1.60	2.57	5.62	4.10	8.10	3.94	8.36	11.43	4.88	8.24
22	2.57	6.55	2.54	0.23	1.50	3.26	5.41	4.00	8.41	3.71	8.08	11.48	4.81	7.99
23	2.27	6.12	2.84	0.27	1.50	3.26	5.63	4.37	10.36	3.59	7.71	11.34	4.94	7.90
24	2.30	6.44	2.89	0.30	1.63	4.10	5.93	4.53	9.82	3.71	7.59	11.76	5.08	8.17

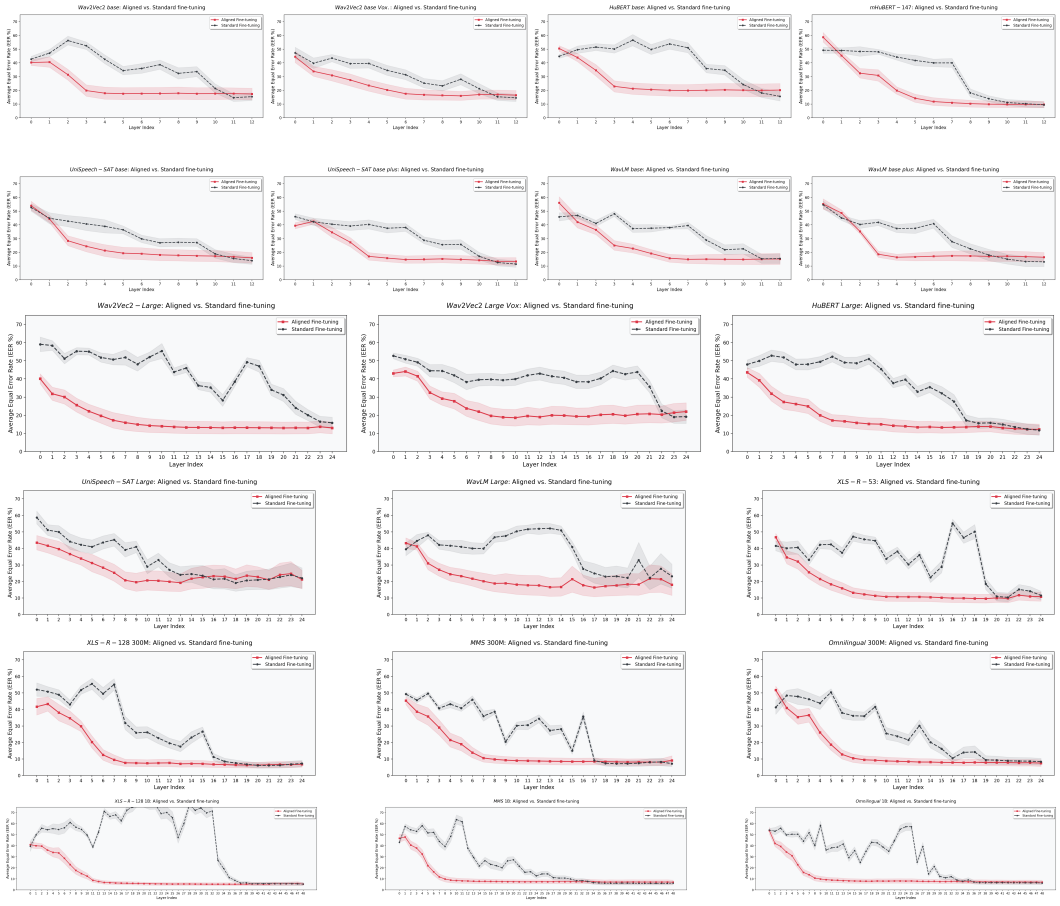


Figure 7: Comparison (Average EER %) of *standard* and *aligned* training across SFMs.

C.2 Impact of model scale and pretraining language

Table 9 and Figure 7 provide a comprehensive view of how model scale, pretraining language scope, and fine-tuning strategy interact to determine both the optimal inference depth and OOD generalization.

Scaling consistently improves OOD robustness, especially for multilingual models. Under standard fine-tuning, average EER decreases steadily with model size within the multilingual family: *mHuBERT-147* achieves 9.46%, *XLS-R-128* 300M reaches 5.80%, and *XLS-R-128* 1B further reduces this to 5.36%. This trend is particularly pronounced on challenging OOD benchmarks such as SC, DFADD, and DIFSSD, where larger multilingual models reduce EER by more than half relative to comparably-sized English-only systems. For instance, *XLS-R-128* 300M achieves 17.59% on SC and 10.63% on DIFSSD, compared to 55.12% and 56.78% for *WavLM Large*, despite both being large-scale models. This suggests that multilingual pretraining exposes the encoder to broader acoustic and phonetic diversity, producing representations that transfer more robustly to unseen synthesis conditions regardless of language or vocoder type.

English-only models show greater variability and are more sensitive to OOD shift. Across both base and large configurations, English-only models exhibit higher EER on OOD datasets, particularly on SC, DIFSSD, and DFADD — benchmarks that involve noisy conditions, diffusion-based synthesis, or diverse speaker characteristics. *UniSpeechSAT Large* performs with the best average EER of 19.14% driven largely by poor performance on SC (30.42%) and DIFSSD (53.29%). In contrast, multilingual large models — *XLS-R-128*, *MMS*, and *Omnilingual* achieve more balanced profiles across all dataset categories.

Table 9: Best-layer EER (%, ↓) under *standard* and *aligned* fine-tuning across all evaluated SFMs. Models are grouped by scale and pretraining language scope (**EN** = English-only, **Multi** = multilingual). **Layer**: transformer layer index at which the minimum Avg. EER is attained.

Size	Model	Lang.	Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Avg.
				19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	
Standard Fine-tuning																
BASE	Wav2Vec2	EN	11	1.44	6.32	7.11	7.57	14.02	10.42	32.92	17.51	29.67	23.16	5.72	20.38	14.69
	HuBERT	EN	12	1.23	5.96	7.56	10.80	15.55	4.14	40.30	11.10	26.61	27.98	8.18	28.20	15.63
	UniSpeechSAT	EN	12	1.14	3.83	6.46	9.59	17.04	5.07	27.77	16.31	29.30	22.92	7.59	20.32	13.95
	UniSpeechSAT+	EN	12	0.37	1.92	6.40	7.65	12.55	9.41	28.26	14.97	23.23	15.57	3.22	14.34	11.49
	WavLM	EN	11	3.41	4.80	9.85	12.54	12.46	3.26	22.97	16.16	45.52	26.06	8.48	18.43	15.33
	WavLM+	EN	12	0.52	2.04	5.21	3.10	11.62	10.73	23.72	14.15	45.68	19.32	4.65	16.57	13.11
	Wav2Vec2 Vox.	Multi	12	0.36	3.37	9.09	13.59	11.36	12.82	30.10	22.11	27.22	18.11	4.37	22.07	14.55
	mHuBERT-147	Multi	12	0.44	1.41	4.34	2.61	10.84	2.20	20.70	13.77	19.67	18.38	3.69	15.53	9.46
LARGE	Wav2Vec2	EN	24	0.85	2.48	10.60	29.38	14.28	7.77	35.60	14.58	28.46	21.13	4.74	19.99	15.82
	HuBERT	EN	24	0.65	3.57	5.95	7.42	14.55	13.08	37.22	10.07	21.34	7.47	3.30	16.83	11.79
	UniSpeechSAT	EN	18	0.84	3.62	5.86	6.17	12.18	48.45	30.42	10.35	53.29	35.39	8.18	14.93	19.14
	WavLM	EN	20	1.93	4.16	4.55	13.18	11.40	8.65	55.12	13.77	56.78	41.33	5.57	49.52	22.16
	Wav2Vec2 Vox.	Multi	23	1.39	4.27	8.94	19.65	17.72	24.90	36.45	16.98	36.77	35.48	5.20	20.87	19.05
	XLS-R-53	Multi	21	0.16	1.79	3.82	7.84	10.49	4.14	33.07	6.51	19.71	13.03	4.86	19.69	10.43
	XLS-R-128	Multi	21	0.07	2.07	1.42	1.67	5.09	1.01	17.59	6.89	10.63	11.57	2.32	13.18	5.80
	MMS	Multi	19	0.15	3.33	2.56	2.16	7.11	1.47	20.06	5.95	12.83	13.94	2.28	14.24	7.17
	Omnilingual	Multi	24	0.16	1.31	3.13	6.11	5.96	0.48	20.22	16.54	16.80	15.34	2.81	12.13	8.42
XL	XLS-R-128	Multi	48	0.14	1.38	1.68	1.98	5.73	1.73	15.69	4.63	10.58	6.72	1.42	12.62	5.36
	MMS	Multi	44	0.14	1.45	2.07	2.55	4.77	1.06	17.73	6.50	11.30	12.92	2.09	10.20	6.06
	Omnilingual	Multi	48	0.10	1.16	2.63	3.03	4.22	1.23	15.09	9.80	12.89	9.66	2.63	12.45	6.24
Aligned Fine-tuning																
BASE	Wav2Vec2	EN	12	0.61	4.97	8.64	7.00	12.19	6.24	37.52	20.63	49.71	30.53	6.01	24.12	17.35
	HuBERT	EN	7	0.53	3.62	9.06	10.46	15.24	17.72	43.86	20.63	53.47	20.59	11.45	31.71	19.86
	UniSpeechSAT	EN	12	1.05	4.76	9.17	6.85	12.01	1.55	32.88	18.04	47.57	24.14	10.61	23.57	16.02
	UniSpeechSAT+	EN	12	1.20	2.59	6.99	7.69	14.17	5.78	30.00	19.07	27.74	24.56	5.64	15.84	13.44
	WavLM	EN	10	0.94	5.78	6.04	5.53	15.85	3.41	25.15	13.41	40.58	37.49	5.32	19.07	14.88
	WavLM+	EN	4	1.91	12.76	8.04	11.62	10.02	21.69	28.05	21.18	23.54	25.18	11.70	21.60	16.44
	Wav2Vec2 Vox.	Multi	9	0.90	7.70	7.98	11.59	11.80	11.57	36.80	14.60	32.92	26.01	2.81	27.23	15.99
	mHuBERT-147	Multi	12	0.16	3.89	3.02	1.25	9.75	2.70	24.81	15.23	18.61	17.74	2.20	17.45	9.74
LARGE	Wav2Vec2	EN	20	1.03	2.16	6.08	5.08	12.08	2.03	36.31	16.83	31.97	15.87	5.48	21.09	13.00
	HuBERT	EN	24	0.60	3.39	6.16	3.41	12.94	15.06	39.09	13.77	20.94	7.94	5.14	18.13	12.21
	UniSpeechSAT	EN	13	0.78	3.21	6.60	2.42	11.79	53.93	26.25	12.71	55.75	34.25	7.06	16.28	19.25
	WavLM	EN	13	0.42	4.88	2.68	2.84	7.47	1.55	46.85	11.79	47.64	18.23	2.68	51.63	16.55
	Wav2Vec2 Vox.	Multi	10	0.63	4.14	8.04	12.61	16.83	6.97	36.33	13.37	47.86	46.60	5.83	24.35	18.63
	XLS-R-53	Multi	19	0.12	1.81	3.86	4.21	11.91	2.48	31.48	6.50	20.91	11.48	3.16	18.03	9.66
	XLS-R-128	Multi	23	0.08	1.71	1.49	1.59	3.98	2.44	18.60	9.29	12.28	9.57	2.72	15.76	7.36
	MMS	Multi	22	0.15	2.40	3.46	3.07	8.77	3.67	25.43	13.62	11.48	9.51	2.34	13.71	8.27
	Omnilingual	Multi	24	0.10	1.96	3.01	1.86	5.40	0.76	20.83	9.67	16.89	12.98	2.88	14.00	7.53
XL	XLS-R-128	Multi	36	0.07	1.67	1.93	1.97	4.68	3.88	16.12	4.39	7.36	5.17	1.98	13.55	5.23
	MMS	Multi	48	0.12	0.88	2.30	3.60	6.72	2.07	24.17	5.95	11.84	11.27	2.81	13.05	7.06
	Omnilingual	Multi	48	0.23	1.28	2.46	2.96	4.43	0.93	16.94	11.53	13.37	15.13	2.20	13.59	7.09

Aligned fine-tuning shifts the optimal decision layer earlier without sacrificing final-layer accuracy, but its benefit depends on model capacity. Across large and XL multilingual models, aligned fine-tuning consistently moves the best-performing layer earlier in the network while maintaining or slightly improving the best average EER. *XLS-R-128* 1B moves from layer 48 to layer 36 (5.36% → 5.23%), and for *Omnilingual* 1B, the optimal layer remains at 48 in both settings, though aligned fine-tuning slightly degrades performance (6.24% → 7.09%), suggesting this model's representations are already well-organized under standard fine-tuning and less amenable to realignment.

The benefit of aligned fine-tuning is reversed for smaller English-only models. In the base setting, aligned fine-tuning frequently increases average EER relative to standard fine-tuning: *HuBERT-base* worsens from 15.63% to 19.86%, *WavLM+* from 13.11% to 16.44%, and *UniSpeechSAT* from 13.95% to 16.02%. We attribute this to insufficient representational capacity: when the backbone has fewer parameters and less diverse pretraining, enforcing early discriminative alignment disrupts the specialization already concentrated in deeper layers without providing enough intermediate capacity to compensate. This is consistent with the observation that smaller models tend to have sharper, more concentrated discriminative profiles, making them more sensitive to the auxiliary supervision introduced during aligned training. Conversely, large multilingual models appear to have sufficient redundant representational capacity to absorb intermediate supervision while still maintaining strong final-layer performance, making them the primary beneficiaries of the alignment strategy.

C.3 CKA representational similarity analysis

To complement the EER-based layer-wise probing in Table 3, we use Centered Kernel Alignment (CKA) [62] to directly measure how representations evolve across transformer layers and how PLS

reshapes the representational geometry of *XLS-R-128* 300M. We compute linear CKA between all pairs of layer representations from the FoR evaluation set. We compute separate heatmaps for bona fide and spoofed samples under standard fine-tuning and under PLS, yielding four matrices. High CKA values (close to 1.0) indicate that two layers encode similar information; low values indicate representational divergence.

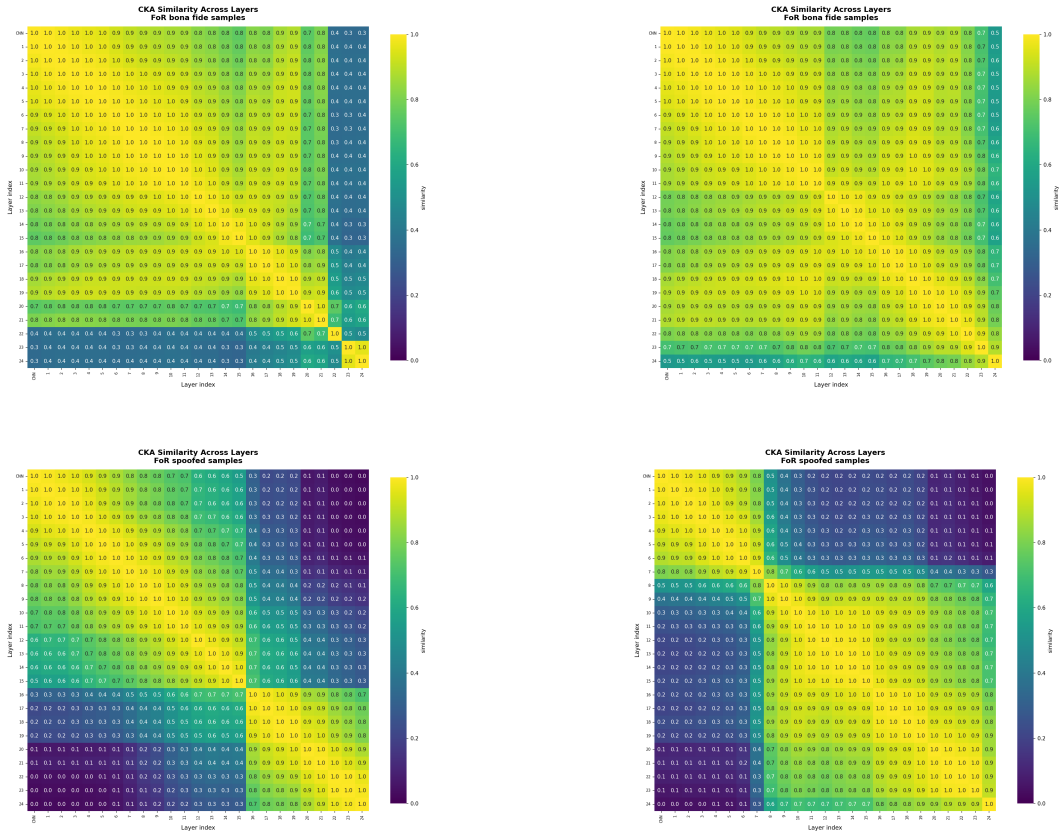


Figure 8: Centered Kernel Alignment analysis across the transformer layers of *XLS-R-128* using the FoR evaluation set. The left column displays standard fine-tuning, while the right column shows the results for PLS.

The four heatmaps (Fig. 8) reveal several consistent patterns. Standard fine-tuning produces high inter-layer similarity in the lower stack. Under standard fine-tuning, layers 0–15 form a densely interconnected block with pairwise CKA values consistently above 0.8 for both bona fide and spoofed samples. This confirms the representational redundancy hypothesis stated in Section 2: SSL pretraining produces smooth, slowly varying representations across adjacent layers, and final-layer supervision fails to disrupt this structure in shallower layers. The upper block (layers 16–24) exhibits somewhat lower similarity, indicating that only the topmost layers have been meaningfully redirected toward anti-spoofing. PLS breaks the lower-block homogeneity. Under PLS, the high-similarity block contracts. Layers 0–6 retain high inter-layer similarity (consistent with the persistent EER above 15% in Table 3 for these layers). This result corresponds precisely where PLS produces competitive EER (below 12%), indicating that the progressive supervision successfully differentiates representations that were previously redundant. Bona fide and spoofed representations diverge earlier under PLS. Comparing the bona fide and spoofed heatmaps, we observe that under standard fine-tuning the similarity structures are nearly identical up to layer 15 — consistent with those layers encoding class-agnostic speech properties. Under PLS, the two similarity matrices diverge from approximately layer 7 onward, with the inter-layer similarity structure of spoofed samples showing stronger vertical banding (suggesting spoof-discriminative features are encoded more stably across depth). This provides indirect evidence that PLS activates earlier layers to encode genuinely discriminative representations rather than merely standard supervision.

Table 10: *Wav2Vec2 Base* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	39.20	48.26	43.54	50.41	43.81	39.71	43.90	44.25	48.44	36.95	24.31	49.98	42.73	50.02
1	43.02	58.67	50.41	49.68	37.33	63.03	52.38	51.25	42.57	38.73	28.66	49.73	47.12	56.13
2	64.01	64.39	61.58	45.13	40.61	79.74	51.28	67.71	47.87	48.10	56.83	47.78	56.25	55.35
3	63.36	62.34	57.85	40.47	41.65	66.75	37.21	47.16	57.22	57.43	53.88	44.39	52.47	51.29
4	39.20	41.41	27.94	28.97	30.58	58.44	37.61	54.70	61.32	47.59	49.51	34.47	42.64	32.02
5	21.17	25.68	19.25	26.97	31.61	39.75	37.74	35.60	54.15	53.49	35.06	32.35	34.40	30.15
6	23.46	22.69	21.85	34.00	29.06	58.83	35.16	41.33	57.01	48.58	29.45	29.96	35.95	30.13
7	30.07	30.14	26.47	38.92	28.45	68.56	37.21	39.07	54.73	42.91	31.54	36.39	38.70	36.05
8	18.10	17.56	18.89	35.48	28.83	66.47	33.94	29.80	53.06	37.32	19.09	29.95	32.37	29.81
9	13.92	20.38	26.75	33.70	34.31	50.88	49.94	30.73	49.82	32.48	15.66	45.14	33.64	38.41
10	4.53	12.74	9.13	12.46	19.25	25.66	31.54	35.52	42.62	30.50	11.14	21.79	21.41	22.35
11	1.44	6.32	7.11	7.57	14.02	10.42	32.92	17.51	29.67	23.16	5.72	20.38	14.69	20.37
12	1.33	5.99	6.81	7.84	14.55	14.52	33.03	19.87	31.47	22.59	6.39	20.42	15.40	20.38
Aligned Fine-tuning														
0	33.52	45.62	40.93	48.06	40.48	39.27	41.09	31.26	48.15	40.71	23.18	51.09	40.28	50.24
1	25.97	46.27	36.66	43.96	32.22	74.69	46.49	33.24	44.88	38.89	19.16	44.90	40.61	52.43
2	13.42	27.15	20.75	30.14	28.47	62.90	37.88	20.90	44.88	39.64	13.45	36.98	31.38	40.30
3	3.25	8.83	10.61	12.53	17.18	12.54	35.27	12.46	56.16	32.33	10.50	26.45	19.84	25.25
4	1.16	5.83	9.25	9.13	14.97	5.96	34.81	13.80	56.05	31.30	8.36	24.83	17.95	22.52
5	1.02	5.50	8.82	7.95	13.24	4.34	34.94	17.74	54.35	31.00	8.02	24.26	17.60	22.08
6	0.87	5.53	8.42	7.50	11.98	4.23	34.98	21.73	54.69	30.37	7.83	24.09	17.69	22.16
7	1.02	5.29	8.27	7.38	11.75	4.23	36.00	22.54	54.58	29.33	7.89	24.36	17.72	22.22
8	0.95	4.96	8.21	7.31	11.70	4.64	39.88	22.39	53.74	28.26	8.20	25.08	17.94	22.45
9	0.94	4.65	8.15	7.31	11.92	4.68	36.80	22.36	53.14	27.92	8.45	24.68	17.58	22.16
10	0.94	4.54	8.42	7.30	11.70	5.07	37.43	22.67	52.83	27.67	8.29	24.79	17.64	22.29
11	0.83	4.64	8.51	7.19	11.88	6.45	37.27	23.47	49.69	29.42	8.13	24.28	17.65	22.12
12	0.61	4.97	8.64	7.00	12.19	6.24	37.52	20.63	49.71	30.53	6.01	24.12	17.35	22.00

Table 11: *Wav2Vec2 Base Voxpopuli* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	36.03	39.09	36.05	38.58	70.41	49.37	49.45	66.59	65.15	51.81	18.74	45.95	47.27	48.77
1	26.46	27.53	29.18	39.04	56.79	37.63	43.86	54.82	60.11	45.31	13.75	42.79	39.77	43.50
2	33.54	36.19	33.67	42.61	53.96	54.99	46.30	50.20	57.04	43.89	26.75	42.56	43.47	44.19
3	26.99	29.49	31.23	41.00	52.72	57.47	47.07	32.74	55.96	39.10	16.68	43.12	39.46	44.71
4	29.22	30.79	27.92	37.75	48.67	56.93	45.20	42.49	59.60	35.38	19.05	42.42	39.62	43.45
5	20.45	23.64	22.53	28.97	42.29	52.96	42.74	38.01	52.98	44.60	12.42	33.07	34.55	39.99
6	14.54	20.67	25.65	33.02	43.85	28.89	47.67	27.94	48.39	37.01	7.43	41.54	31.38	41.08
7	9.10	12.65	17.32	22.11	23.82	27.51	36.85	37.36	47.55	35.48	8.13	25.36	25.27	31.47
8	6.32	10.31	11.24	18.89	21.77	21.64	36.21	39.50	48.96	32.86	8.15	23.80	23.31	25.22
9	12.49	12.49	13.43	20.08	34.71	18.19	35.88	51.94	52.31	37.15	18.35	29.92	28.08	28.67
10	4.95	4.56	10.26	12.68	19.77	13.38	34.69	48.07	36.31	23.60	22.34	24.26	21.24	23.21
11	0.51	4.31	9.77	12.27	19.72	7.77	29.41	24.63	27.03	19.61	5.11	24.13	15.36	22.59
12	0.36	3.37	9.09	13.59	11.36	12.82	30.10	22.11	27.22	18.11	4.37	22.07	14.55	19.74
Aligned Fine-tuning														
0	30.52	34.10	32.75	39.76	67.74	55.03	46.97	56.69	65.76	44.75	15.56	43.77	44.45	49.92
1	15.66	20.62	26.59	45.06	35.76	39.92	41.70	41.33	51.79	38.64	8.36	40.84	33.85	39.61
2	6.81	16.33	26.87	36.38	24.45	49.83	41.56	28.60	50.23	39.05	9.29	41.93	30.94	39.06
3	3.76	13.97	22.43	27.18	16.22	48.86	40.47	24.22	53.04	34.59	7.38	38.34	27.54	35.38
4	2.44	12.80	15.22	23.82	10.91	29.91	41.39	20.12	50.47	35.05	5.97	34.34	23.54	30.49
5	1.27	10.92	12.23	18.79	9.12	19.25	39.81	17.89	49.15	31.43	3.67	30.24	20.31	27.01
6	0.71	8.68	9.14	14.13	9.75	12.02	38.44	12.33	46.67	27.97	3.49	27.47	17.57	24.44
7	0.76	7.55	8.12	12.19	11.49	10.64	37.79	13.27	43.06	26.19	3.11	26.77	16.75	23.85
8	1.01	7.40	8.08	11.78	12.32	11.13	37.26	14.34	37.90	26.09	3.20	26.20	16.39	23.09
9	0.90	7.70	7.98	11.59	11.80	11.57	36.80	14.60	32.92	26.01	2.81	27.23	15.99	24.03
10	0.71	8.09	8.53	12.27	17.72	14.22	38.12	14.58	32.32	25.88	2.20	28.66	16.94	26.86
11	0.68	5.93	8.14	13.34	14.02	20.63	38.08	16.13	30.97	24.88	2.35	28.58	16.98	25.45
12	0.65	5.31	7.69	13.18	9.95	20.18	40.30	15.35	30.26	24.77	2.14	28.20	16.50	21.60

Table 12: *HuBert Base* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	44.68	40.77	46.79	44.15	36.17	38.74	46.97	48.74	41.56	52.24	52.67	43.77	44.77	57.28
1	45.56	43.68	51.48	47.74	51.15	73.20	41.70	56.30	59.93	43.93	39.57	40.84	49.59	57.52
2	66.01	58.92	62.01	52.56	48.71	35.02	41.56	49.27	59.59	50.41	52.18	41.93	51.51	56.72
3	66.32	65.27	59.53	46.99	35.62	53.63	40.47	35.47	55.54	53.50	50.79	38.34	50.12	56.95
4	81.20	79.00	72.16	55.36	40.74	47.35	41.39	53.24	58.00	54.42	61.17	34.34	56.53	56.44
5	68.09	63.25	57.39	41.02	35.61	40.10	39.81	50.60	49.19	50.55	71.39	30.24	49.77	52.04
6	68.99	66.18	61.65	50.40	44.28	54.10	38.44	67.54	55.63	39.95	71.11	27.47	53.81	50.28
7	60.18	54.09	45.93	44.30	45.80	60.34	37.79	64.25	51.13	44.81	76.87	26.77	51.02	48.78
8	23.55	23.54	25.10	38.36	40.62	57.16	37.26	32.06	54.41	32.59	39.96	26.20	35.90	51.46
9	18.59	21.08	26.33	38.81	48.02	55.96	36.80	23.31	53.08	32.46	34.16	27.23	34.65	54.07
10	4.04	11.92	12.52	25.29	31.88	38.87	38.12	13.22	44.52	29.98	12.17	28.66	24.27	54.66
11	1.71	6.58	8.49	11.06	19.60	21.51	38.08	11.93	36.40	23.23	10.14	28.58	18.11	53.75
12	1.23	5.96	7.56	10.80	15.55	4.14	40.30	11.10	26.61	27.98	8.18	28.20	15.63	29.16
Aligned Fine-tuning														
0	53.53	60.92	54.48	53.46	33.25	54.34	50.42	47.14	42.52	57.38	45.68	52.67	50.48	51.54
1	35.96	43.17	40.80	41.46	48.22	67.59	45.30	32.86	50.87	49.73	23.55	47.27	43.90	51.45
2	12.77	21.95	23.18	30.60	38.87	56.50	50.13	31.54	53.52	34.99	13.42	47.45	34.58	43.77
3	2.58	7.72	13.71	16.85	22.04	25.36	41.15	22.14	54.45	24.33	10.87	33.79	22.92	30.41
4	0.75	4.94	11.57	13.45	19.97	24.26	38.90	20.55	55.10	22.73	11.56	31.43	21.27	28.05
5	0.46	3.90	10.29	11.93	19.09	23.20	38.41	20.25	54.87	21.79	11.87	30.68	20.56	26.41
6	0.41	3.74	9.64	10.98	17.51	20.41	38.90	20.66	54.58	21.26	11.73	30.83	20.05	25.76
7	0.53	3.62	9.06	10.46	15.24	17.72	43.86	20.63	53.47	20.59	11.45	31.71	19.86	25.21
8	0.60	3.80	8.98	10.20	14.67	17.09	48.57	20.10	53.17	20.33	11.07	32.81	20.12	25.25
9	0.68	3.79	9.11	10.19	15.28	17.26	51.22	19.34	52.70	19.99	10.61	34.25	20.37	26.00
10	0.67	3.73	9.13	10.03	14.95	15.86	50.51	20.27	52.31	20.14	9.98	34.12	20.14	26.09
11	0.64	3.64	9.15	10.22	14.72	15.28	50.48	20.12	51.89	20.48	9.96	33.60	20.01	25.77
12	0.56	3.52	9.29	10.37	13.45	14.80	52.21	22.64	51.00	20.58	9.52	34.57	20.21	26.08

Table 13: *UniSpeechSAT Base* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	47.79	45.50	49.33	51.68	55.34	79.74	50.41	47.66	61.49	49.39	45.00	48.82	52.68	52.46
1	26.66	32.88	36.27	37.67	50.02	73.46	48.67	50.33	65.99	47.65	24.83	43.36	44.81	48.77
2	34.08	34.63	39.85	46.32	48.01	77.73	41.17	22.92	58.81	41.49	23.96	44.00	42.75	51.05
3	26.05	25.72	31.37	42.95	42.98	85.37	42.87	37.87	60.19	31.35	20.81	41.33	40.74	45.11
4	23.00	24.45	30.16	42.48	55.45	76.24	37.53	24.52	59.55	32.96	18.53	42.65	38.96	45.87
5	27.11	25.57	28.80	50.06	48.48	33.92	34.88	29.83	55.05	39.34	24.92	38.83	36.40	43.83
6	16.03	18.42	19.52	34.04	34.86	35.02	37.64	29.65	48.96	31.03	18.53	34.54	29.85	35.67
7	12.64	19.15	13.50	28.81	25.31	26.85	32.56	26.63	52.18	37.62	19.11	29.43	26.98	29.16
8	13.28	19.31	13.20	28.02	28.07	20.59	33.35	32.08	50.95	39.92	22.36	27.15	27.36	27.89
9	13.54	18.53	12.66	24.12	26.65	28.18	32.24	38.27	47.81	34.54	24.59	24.97	27.18	26.84
10	2.63	10.01	7.33	16.93	18.11	15.77	29.00	23.04	42.30	29.70	9.66	22.50	18.91	23.14
11	2.26	5.73	6.53	11.44	16.58	5.39	28.23	13.41	33.47	33.05	8.71	21.52	15.53	22.56
12	1.14	3.83	6.46	9.59	17.04	5.07	27.77	16.31	29.30	22.92	7.59	20.32	13.95	20.39
Aligned Fine-tuning														
0	66.65	58.15	58.23	55.70	48.31	24.26	57.98	61.06	58.02	44.38	68.49	46.50	53.98	51.23
1	46.53	46.66	40.64	46.42	31.14	29.78	48.49	63.84	57.98	43.54	39.02	37.72	44.31	44.03
2	13.09	17.16	18.47	35.10	26.95	22.40	43.60	25.17	56.25	31.48	13.66	38.12	28.45	34.35
3	8.15	16.19	13.46	20.03	25.97	14.85	37.27	21.86	60.44	30.73	11.68	32.25	24.41	32.97
4	4.70	15.84	11.34	13.26	21.61	7.21	34.71	16.03	60.72	30.43	11.36	28.68	21.32	29.97
5	1.47	12.73	9.30	7.99	20.07	6.93	32.89	20.00	58.52	26.13	10.50	26.66	19.43	26.25
6	1.27	11.99	8.90	6.02	20.98	5.31	32.66	20.12	57.92	26.79	9.98	26.13	19.00	24.99
7	1.46	9.98	8.10	5.75	21.01	3.50	32.48	20.27	55.01	23.92	10.66	25.72	18.15	23.85
8	1.74	8.83	7.86	5.94	20.29	2.83	32.25	20.13	53.03	24.73	11.43	25.29	17.86	23.24
9	1.74	7.76	7.73	6.10	18.25	2.44	32.24	19.60	52.09	25.54	11.26	25.22	17.50	22.88
10	1.54	7.26	7.84	6.06	16.94	2.40	32.28	18.82	51.62	26.13	11.12	25.45	17.29	23.02
11	1.43	6.30	8.27	6.17	14.15	1.40	32.20	18.01	50.70	26.51	11.07	25.52	16.81	23.07
12	1.05	4.76	9.17	6.85	12.01	1.55	32.88	18.04	47.57	24.14	10.61	23.57	16.02	22.38

Table 14: *UniSpeechSAT Base Plus* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	47.21	46.60	47.10	46.01	53.77	34.85	48.34	56.03	51.84	45.60	28.87	46.12	46.03	49.36
1	36.74	34.27	36.57	45.07	56.24	53.50	42.08	28.47	39.16	56.74	30.10	48.24	42.27	50.50
2	29.82	30.26	36.38	47.32	34.81	29.63	63.33	45.83	42.76	49.48	26.78	49.71	40.51	46.30
3	26.65	33.41	35.32	50.13	33.68	42.84	58.38	28.21	32.49	53.54	24.29	51.26	39.18	49.45
4	30.36	34.09	35.80	61.76	33.37	40.20	54.33	29.40	37.53	48.93	22.93	55.55	40.35	51.10
5	23.79	27.55	35.41	59.14	32.39	29.15	51.97	30.60	37.29	49.14	22.50	51.92	37.57	50.03
6	24.59	32.20	36.35	44.72	38.32	49.03	53.41	28.51	37.03	46.27	20.32	46.61	38.11	50.27
7	16.59	19.03	22.20	36.65	30.91	36.25	42.19	25.56	33.49	39.67	16.65	26.69	28.82	32.97
8	7.94	8.59	18.47	34.53	27.65	40.05	36.15	23.70	37.29	40.03	13.88	18.79	25.59	26.30
9	9.52	10.59	16.67	31.02	27.44	51.10	33.67	24.63	38.41	33.77	13.74	18.64	25.77	26.54
10	2.53	3.30	7.70	20.68	13.89	25.05	29.66	19.87	34.76	23.04	11.31	14.13	17.16	15.45
11	0.50	2.21	6.03	9.50	13.50	11.85	27.62	12.61	28.42	19.30	5.52	14.05	12.59	14.88
12	0.37	1.92	6.40	7.65	12.55	9.41	28.26	14.97	23.23	15.57	3.22	14.34	11.49	14.47
Aligned Fine-tuning														
0	38.74	41.98	43.18	51.95	28.38	25.98	43.60	27.57	45.88	44.18	37.77	43.69	39.41	46.70
1	34.65	42.30	42.49	48.35	46.13	43.55	44.71	33.27	49.00	55.57	29.19	41.04	42.52	48.03
2	18.06	20.19	25.04	46.54	36.71	56.58	45.76	29.14	49.79	31.51	17.42	38.76	34.62	38.28
3	10.10	12.69	17.67	26.16	24.54	45.90	39.08	36.69	42.15	31.25	14.21	27.99	27.37	27.88
4	3.76	5.99	10.87	15.11	17.27	7.16	33.74	17.87	31.27	28.92	12.70	20.45	17.09	21.74
5	3.22	4.06	10.01	15.30	16.96	4.10	31.30	17.51	28.02	28.36	12.29	18.85	15.83	20.63
6	2.54	3.09	8.92	11.10	15.64	3.50	30.69	19.45	28.31	24.23	12.43	17.10	14.75	19.39
7	2.17	2.27	8.55	11.52	15.17	5.27	30.70	22.52	30.12	23.71	11.06	16.24	14.94	18.16
8	2.20	2.09	8.30	11.67	14.51	8.44	30.54	23.85	33.76	21.40	10.59	15.76	15.26	17.44
9	2.38	2.06	7.81	11.06	14.19	6.93	30.54	23.17	34.24	20.37	9.96	15.56	14.86	17.06
10	2.39	1.99	7.40	10.64	14.11	6.37	30.47	21.20	33.11	18.97	9.38	15.48	14.29	16.91
11	1.78	2.11	7.04	8.90	14.06	5.91	30.41	18.54	29.56	20.53	7.43	15.59	13.49	17.03
12	1.20	2.59	6.99	7.69	14.17	5.78	30.00	19.07	27.74	24.56	5.64	15.84	13.44	17.36

Table 15: *WavLM Base* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	50.40	46.74	49.20	49.72	43.07	16.96	43.09	49.57	39.60	61.22	55.41	45.17	45.85	45.43
1	39.05	40.57	47.49	55.70	49.18	44.17	53.13	45.43	46.20	56.15	30.77	55.34	46.93	51.92
2	31.95	35.62	43.16	53.27	47.39	39.27	37.66	29.25	37.62	53.88	36.30	47.84	41.10	44.22
3	40.47	44.26	50.80	51.87	61.29	55.96	41.79	42.77	38.25	47.84	53.49	47.50	48.02	43.81
4	20.43	22.76	31.58	39.11	46.41	54.77	38.95	30.60	35.56	50.92	40.13	37.09	37.36	36.21
5	23.64	29.81	26.92	32.98	42.38	60.73	37.05	40.40	46.83	38.67	32.58	38.79	37.57	34.92
6	32.82	35.47	33.26	35.18	38.43	42.45	35.98	43.19	44.60	38.45	40.97	35.51	38.03	35.64
7	35.55	34.96	29.88	30.29	34.53	40.63	35.11	53.64	55.86	46.27	44.40	32.80	39.49	35.27
8	13.88	22.38	22.02	21.99	29.92	29.59	31.87	32.34	57.78	41.84	16.37	26.74	28.89	30.54
9	6.89	11.31	14.31	19.17	21.28	17.98	29.07	24.24	56.29	29.71	11.38	21.46	21.92	24.35
10	5.30	12.03	13.93	22.56	24.16	15.41	29.09	20.78	52.71	42.34	10.92	21.67	22.58	23.51
11	3.41	4.80	9.85	12.54	12.46	3.26	22.97	16.16	45.52	26.06	8.48	18.43	15.33	21.58
12	2.52	4.60	8.51	11.24	11.82	4.42	24.70	17.36	48.05	26.02	7.81	18.90	15.50	21.67
Aligned Fine-tuning														
0	71.36	59.70	63.10	54.68	61.66	16.16	50.26	62.64	37.60	65.14	76.18	54.12	56.05	48.53
1	44.23	45.79	48.49	47.10	37.87	13.38	47.48	17.61	28.94	62.80	63.54	47.65	42.07	41.68
2	24.32	36.68	29.67	43.12	44.60	25.01	37.00	17.38	26.49	62.45	48.56	41.51	36.40	37.75
3	10.67	17.36	17.49	24.88	47.72	11.01	34.36	19.73	36.22	44.35	13.56	23.35	25.06	28.91
4	5.51	11.97	13.27	19.47	27.26	22.74	30.48	21.08	41.20	44.20	14.03	22.75	22.83	31.56
5	3.06	10.07	8.99	9.80	25.31	29.78	26.75	18.39	33.25	33.03	11.33	20.48	19.19	29.33
6	1.82	7.61	6.60	5.68	18.71	13.64	24.82	15.50	35.21	31.56	9.33	18.35	15.74	27.83
7	1.14	7.94	5.99	5.79	18.08	8.03	23.83	14.97	32.59	35.32	7.13	18.22	14.92	27.26
8	1.12	7.38	5.93	5.68	17.65	5.11	23.46	14.17	39.29	36.62	6.28	18.11	15.07	25.23
9	1.04	6.32	6.01	5.49	16.47	3.88	23.42	13.90	41.39	37.73	5.57	18.33	14.96	24.03
10	0.94	5.78	6.04	5.53	15.85	3.41	25.15	13.41	40.58	37.49	5.32	19.07	14.88	22.95
11	0.84	5.12	6.22	5.50	15.15	3.54	24.63	11.51	44.80	40.18	4.36	19.29	15.10	24.33
12	0.84	4.81	6.69	5.45	14.23	3.80	25.11	13.64	43.86	40.70	3.41	19.74	15.19	22.50

Table 16: *WavLM Base Plus* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	51.79	50.70	59.05	46.09	70.15	51.14	46.39	54.15	42.66	66.30	72.71	44.56	54.64	50.50
1	42.09	45.60	39.51	47.30	50.75	61.26	45.86	45.80	47.77	38.48	34.89	41.80	45.09	45.61
2	25.87	27.25	33.30	37.07	53.03	53.75	44.32	53.01	47.27	40.66	26.20	42.21	40.33	45.68
3	34.45	39.68	34.85	37.37	45.79	59.67	51.46	36.69	51.46	34.15	32.88	44.00	41.87	43.51
4	24.15	27.03	25.21	29.69	46.86	57.34	49.26	39.07	44.81	39.30	28.22	37.47	37.37	38.97
5	28.52	30.03	23.57	21.47	33.88	63.60	37.56	48.60	50.27	29.62	55.58	27.54	37.52	30.15
6	44.12	42.93	27.32	22.15	31.46	55.61	38.82	56.96	53.52	36.20	51.18	30.04	40.86	36.73
7	16.66	16.48	14.03	10.45	20.89	30.04	26.23	62.09	47.70	29.13	40.87	19.02	27.80	26.11
8	8.57	7.28	10.34	15.61	18.04	37.10	25.75	32.44	51.89	30.30	14.86	18.86	22.59	22.95
9	1.36	4.14	6.07	7.69	12.95	33.53	23.97	23.42	53.71	24.86	7.75	17.64	18.09	22.89
10	0.94	3.35	5.49	4.77	11.82	16.57	23.31	23.44	47.46	20.17	6.34	17.26	15.08	20.88
11	0.58	2.43	4.85	3.11	11.45	8.70	23.56	15.63	49.17	18.33	5.52	16.82	13.35	19.43
12	0.52	2.04	5.21	3.10	11.62	10.73	23.72	14.15	45.68	19.32	4.65	16.57	13.11	20.32
Aligned Fine-tuning														
0	69.14	65.56	65.83	52.56	27.73	34.01	61.33	66.78	50.32	49.58	69.02	50.56	55.20	55.40
1	49.60	47.76	49.96	45.02	36.14	44.09	55.69	61.20	39.59	44.77	63.66	46.18	48.64	46.95
2	32.14	35.54	31.09	34.87	25.57	44.89	36.52	37.49	32.68	48.14	36.18	28.34	35.29	39.90
3	4.56	14.32	11.15	21.61	15.86	22.79	29.60	13.27	25.10	31.27	9.24	24.95	18.64	30.44
4	1.91	12.76	8.04	11.62	10.02	21.69	28.05	21.18	23.54	25.18	11.70	21.60	16.44	23.89
5	0.91	11.11	6.13	5.04	9.50	28.10	25.34	26.89	29.74	26.76	10.64	20.58	16.73	24.01
6	1.31	8.89	5.95	3.53	9.73	34.94	23.90	31.93	29.32	26.28	10.40	20.15	17.19	22.74
7	1.35	7.71	5.63	3.07	9.86	41.26	23.94	32.19	28.78	25.74	10.24	19.65	17.45	22.45
8	1.37	6.29	5.53	2.84	9.87	44.65	24.38	29.39	29.20	25.68	9.77	19.76	17.39	21.86
9	1.33	5.54	5.64	2.80	9.74	45.14	24.65	26.38	29.22	25.74	9.04	19.94	17.10	21.94
10	1.21	5.05	5.84	2.76	9.45	45.14	24.72	24.90	32.08	26.62	8.85	20.25	17.24	22.00
11	1.11	4.65	6.25	2.84	9.27	43.03	25.60	24.40	31.03	25.75	8.87	20.00	16.90	22.18
12	1.02	4.79	6.29	3.03	8.44	38.99	26.48	25.58	28.64	24.95	8.99	20.43	16.47	21.92

Table 17: *mHubert 147* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	52.07	53.85	44.67	50.81	46.71	71.69	47.20	45.02	43.77	40.70	42.34	51.75	49.21	50.67
1	57.50	52.44	47.78	50.06	49.58	40.29	48.55	65.30	40.23	41.78	46.08	48.80	49.03	50.75
2	66.28	64.09	53.30	52.93	31.82	21.99	40.21	55.50	46.68	53.72	42.82	50.40	48.31	57.98
3	54.41	57.86	48.89	53.92	43.70	53.80	39.61	51.26	46.13	46.01	38.67	42.52	48.06	52.39
4	40.52	50.36	39.16	49.08	49.66	59.02	46.54	40.13	46.92	39.96	25.87	44.25	44.29	52.31
5	31.97	41.48	29.75	32.07	37.57	55.83	43.03	75.48	54.67	33.72	29.08	36.76	41.78	41.42
6	28.95	35.68	28.51	45.78	42.94	33.04	47.95	51.93	56.17	42.16	24.15	42.59	39.99	45.50
7	23.29	33.98	29.27	43.13	39.35	57.01	53.60	39.60	46.27	42.53	19.55	52.44	40.00	49.24
8	4.70	10.44	7.42	12.60	14.02	21.77	23.02	29.27	35.71	33.63	8.20	17.00	18.15	20.25
9	2.65	7.22	6.22	8.59	12.60	16.62	21.12	21.33	25.56	23.08	5.95	16.54	13.95	18.44
10	0.86	5.35	5.33	4.54	13.02	4.19	21.84	18.14	22.75	18.14	3.08	16.58	11.15	18.67
11	0.54	2.17	3.90	2.50	13.50	1.90	25.46	14.30	19.53	18.20	3.97	17.54	10.29	18.15
12	0.44	1.41	4.34	2.61	10.84	2.20	20.70	13.77	19.67	18.38	3.69	15.53	9.46	16.82
Aligned Fine-tuning														
0	73.87	68.89	66.12	55.63	32.78	57.86	51.72	57.89	53.68	55.76	81.09	48.95	58.69	49.79
1	32.28	37.68	43.21	52.37	36.03	64.85	56.42	40.13	48.95	46.78	34.14	51.55	45.37	48.85
2	19.25	26.46	27.05	37.97	25.24	59.62	43.36	22.91	46.51	28.19	16.84	36.57	32.50	39.69
3	16.10	21.85	22.24	29.12	26.36	72.49	38.86	27.81	40.25	28.43	16.42	30.53	30.87	36.41
4	5.68	10.76	13.10	11.44	20.19	34.89	29.08	18.04	35.09	26.89	12.70	20.74	19.88	28.56
5	1.05	8.46	7.08	4.43	12.70	4.55	27.69	17.74	35.80	24.48	7.57	19.37	14.24	22.62
6	0.30	6.46	4.15	2.20	10.82	0.35	25.74	16.83	29.42	21.90	5.58	18.36	11.84	18.97
7	0.19	5.60	3.31	1.74	9.86	0.97	24.82	16.66	26.00	20.60	5.01	17.31	11.01	16.84
8	0.14	5.02	3.03	1.59	9.73	1.86	24.71	15.61	22.79	18.83	3.55	17.22	10.34	16.22
9	0.14	4.62	2.98	1.44	9.76	2.29	24.65	14.95	20.91	17.76	2.83	17.19	9.96	16.13
10	0.15	4.40	3.00	1.29	9.52	2.33	24.78	15.50	20.08	17.50	2.63	16.83	9.83	15.88
11	0.16	4.37	3.00	1.18	8.91	2.20	24.98	16.30	19.00	18.20	2.49	16.80	9.80	16.62
12	0.16	3.89	3.02	1.25	9.75	2.70	24.81	15.23	18.61	17.74	2.20	17.45	9.74	16.27

Table 18: *Wav2Vec2 Large* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	71.78	66.73	60.90	49.03	65.78	22.87	47.30	72.19	63.11	61.70	74.62	52.50	59.04	51.57
1	68.43	55.93	60.04	59.34	69.64	39.27	44.72	68.07	55.34	60.96	66.98	51.58	58.36	53.75
2	43.59	36.76	48.12	53.77	64.95	31.23	49.36	57.23	58.87	62.30	60.08	47.95	51.18	45.76
3	53.42	48.60	54.95	52.78	67.01	44.73	48.71	59.87	55.52	57.12	68.30	51.73	55.23	44.93
4	56.79	50.65	55.64	49.03	59.90	51.27	47.84	59.90	54.18	59.09	70.62	45.83	55.06	44.12
5	50.26	46.67	50.70	51.46	59.41	28.49	48.76	53.64	55.92	64.59	61.84	49.63	51.78	46.99
6	55.02	50.92	51.36	50.92	43.31	25.23	48.82	56.15	52.22	65.56	64.10	44.25	50.66	46.41
7	61.77	55.66	56.86	51.12	49.97	14.44	50.53	61.88	51.14	65.19	70.07	32.83	51.79	40.89
8	60.88	56.45	52.00	42.90	42.55	19.66	40.22	58.29	49.47	57.46	67.25	30.10	48.10	40.14
9	59.09	57.76	53.84	49.75	43.42	30.21	53.37	53.41	53.38	62.57	65.47	42.49	52.06	47.25
10	68.86	66.06	61.64	45.21	51.24	20.22	54.61	62.26	55.79	66.16	73.20	38.80	55.34	51.43
11	43.48	41.25	35.05	41.65	42.13	44.04	48.51	51.66	50.53	52.68	50.21	24.57	43.81	36.52
12	43.51	40.25	34.50	47.13	42.20	45.86	48.29	52.21	52.26	58.66	57.27	29.16	45.94	38.11
13	29.24	28.87	26.09	40.10	35.72	36.12	46.09	38.01	38.01	50.45	42.41	24.61	36.31	32.05
14	26.36	25.74	29.63	44.11	38.94	30.21	53.46	26.38	42.40	46.97	29.47	30.07	35.31	37.89
15	11.55	13.84	19.74	35.37	27.06	26.02	46.27	29.53	43.65	43.10	18.11	23.98	28.19	28.67
16	27.38	28.54	30.06	44.57	35.52	36.30	50.65	45.30	49.62	49.75	33.16	31.40	38.52	37.97
17	48.48	46.94	40.61	42.59	35.11	46.68	50.95	65.66	61.36	52.04	57.94	42.40	49.23	45.28
18	54.09	48.10	33.29	39.68	35.12	44.73	39.36	65.03	57.24	55.59	61.96	29.61	46.98	39.56
19	27.66	26.07	23.59	33.74	29.53	17.80	39.70	56.66	52.54	35.77	35.53	29.25	33.99	31.18
20	14.79	18.95	18.41	30.37	34.21	14.72	44.16	45.17	50.43	40.33	22.71	39.35	31.13	34.16
21	3.40	12.66	13.24	28.93	24.00	21.02	42.33	28.60	46.31	30.10	8.80	29.54	24.08	28.75
22	2.40	9.56	12.03	23.55	20.59	35.91	36.65	18.14	27.71	23.71	6.36	24.85	20.12	28.38
23	4.09	6.36	10.81	25.44	14.70	8.52	39.76	10.73	27.03	19.00	7.41	23.94	16.48	21.69
24	0.85	2.48	10.60	29.38	14.28	7.77	35.60	14.58	28.46	21.13	4.74	19.99	15.82	18.97
Aligned Fine-tuning														
0	29.71	32.71	37.82	48.93	43.03	57.60	42.94	29.14	37.34	50.41	25.15	46.07	40.07	46.65
1	17.80	17.18	24.51	44.07	36.93	34.46	44.13	20.91	54.84	35.38	12.93	38.74	31.83	35.02
2	13.28	14.30	18.95	38.58	37.98	34.05	42.76	27.03	58.44	28.06	11.61	35.54	30.05	32.51
3	10.45	13.17	16.13	30.29	27.99	20.67	41.13	20.80	57.10	27.82	9.96	31.24	25.56	31.27
4	6.50	11.46	14.22	20.03	21.56	15.19	39.14	19.35	56.10	24.80	9.69	28.53	22.21	28.46
5	3.00	8.27	11.68	14.69	16.37	11.18	38.28	18.67	55.04	22.56	9.17	27.49	19.70	26.35
6	1.21	5.01	9.48	9.28	14.39	3.37	36.80	19.62	53.60	20.01	8.24	25.86	17.24	24.36
7	0.77	3.47	7.38	6.86	12.79	2.11	35.96	19.85	51.70	19.08	7.52	24.11	15.97	22.04
8	0.75	2.80	6.45	6.13	12.42	1.68	35.80	19.73	45.54	18.69	6.66	22.97	14.97	20.70
9	0.83	2.40	5.97	5.57	12.35	1.47	35.60	19.47	42.00	17.89	6.01	22.28	14.32	19.94
10	0.94	2.25	5.80	5.49	12.35	1.81	35.42	18.52	39.46	18.09	5.72	22.10	14.00	19.48
11	0.91	2.22	5.78	5.53	12.22	2.20	35.36	18.01	36.89	17.58	5.16	21.68	13.63	18.87
12	0.93	2.14	5.80	5.34	12.03	2.33	35.22	17.87	35.21	16.76	5.11	21.37	13.34	18.50
13	0.93	2.10	5.88	5.19	12.04	2.29	35.28	17.74	35.50	16.11	5.11	21.27	13.29	18.45
14	0.98	2.04	5.94	5.30	12.04	2.16	35.31	17.61	35.37	15.54	5.37	21.18	13.24	18.40
15	0.92	2.06	6.01	5.23	12.01	2.24	35.31	17.61	33.95	15.51	5.16	20.94	13.08	18.31
16	0.95	2.07	5.99	5.38	11.92	2.40	35.81	17.36	34.95	15.50	5.25	20.98	13.21	18.27
17	0.96	2.05	6.09	5.45	11.98	2.29	36.01	17.36	34.34	15.32	5.34	20.94	13.18	18.24
18	0.95	2.07	6.06	5.19	12.01	2.40	36.20	17.24	33.70	15.55	5.29	20.96	13.14	18.23
19	1.01	2.10	6.11	5.23	11.86	2.24	36.63	16.83	32.65	15.95	5.38	21.11	13.09	18.27
20	1.03	2.16	6.08	5.08	12.08	2.03	36.31	16.83	31.97	15.87	5.48	21.09	13.00	18.41
21	0.91	2.31	5.97	4.93	12.23	2.16	36.46	16.45	31.94	16.79	5.55	21.15	13.07	18.47
22	0.82	2.37	5.92	4.81	12.30	2.40	36.26	16.53	30.79	17.59	5.62	21.20	13.05	18.52
23	0.79	2.44	5.79	4.73	12.26	2.33	36.09	23.32	30.69	18.20	6.24	21.32	13.68	18.57
24	0.79	2.43	5.64	4.54	12.37	2.70	36.01	16.30	30.61	18.45	5.46	21.33	13.05	18.61

Table 19: *Wav2Vec2 Large Voxpopuli* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	55.64	59.22	53.50	50.28	56.76	54.47	51.64	44.10	59.94	48.21	41.90	56.48	52.68	53.54
1	50.82	56.08	46.46	47.19	51.43	60.25	52.73	45.70	61.36	49.44	32.24	56.41	50.84	51.42
2	42.32	47.75	40.10	47.29	53.82	60.08	53.69	45.17	60.65	45.48	36.58	56.83	49.15	45.75
3	26.95	28.62	31.01	47.14	61.72	51.32	56.87	41.05	55.26	36.40	39.36	58.13	44.49	40.51
4	23.77	26.53	31.86	43.31	61.00	61.57	57.17	43.70	49.93	40.58	38.50	55.04	44.41	39.82
5	21.33	23.70	31.96	45.97	65.14	44.00	59.89	47.03	50.19	35.99	24.41	53.80	41.95	40.30
6	16.68	18.37	23.27	40.03	57.98	40.55	59.59	44.50	52.13	38.28	19.88	48.62	38.32	35.77
7	20.06	20.31	24.69	36.91	55.22	41.82	54.31	49.82	49.74	42.79	38.78	39.80	39.52	31.75
8	23.49	20.43	24.61	35.67	53.09	38.35	56.32	47.79	52.31	42.19	45.52	37.13	39.74	30.76
9	24.67	21.63	24.16	35.33	48.26	39.84	53.13	52.19	52.56	40.72	46.22	32.94	39.30	29.27
10	27.92	23.25	25.46	33.24	49.39	30.71	52.01	58.82	52.24	44.09	50.97	30.89	39.92	28.71
11	31.41	24.75	27.53	35.33	49.18	33.21	50.63	62.67	56.30	46.10	52.74	33.49	41.94	30.58
12	33.47	26.11	29.81	35.78	48.83	32.15	49.71	65.98	58.91	45.19	55.97	33.55	42.96	30.99
13	31.27	24.12	27.56	32.41	47.72	32.28	48.48	65.18	56.02	46.94	53.69	31.51	41.43	29.81
14	28.85	23.11	25.37	32.14	46.41	34.10	48.16	64.24	54.19	50.41	50.88	29.95	40.65	29.06
15	24.85	21.51	22.50	30.22	44.12	32.11	48.27	61.86	51.60	48.31	47.56	28.18	38.42	27.53
16	24.66	20.53	23.69	29.53	45.07	31.57	48.55	60.51	53.27	45.12	48.95	29.28	38.39	27.95
17	30.51	24.93	29.06	31.16	48.40	30.21	47.12	59.73	49.95	48.33	56.74	27.25	40.28	29.55
18	36.79	30.91	35.76	33.09	51.30	35.37	50.66	63.74	49.68	52.81	61.62	30.34	44.34	33.54
19	36.11	30.09	31.07	31.92	51.34	30.56	50.85	59.24	47.05	54.24	58.05	31.04	42.63	32.76
20	43.17	37.36	33.66	35.89	43.75	32.24	45.73	56.43	43.75	60.86	66.74	26.06	43.80	32.92
21	28.82	23.91	21.21	24.53	34.08	33.47	43.87	49.53	43.99	58.75	44.22	23.23	35.80	27.59
22	3.98	5.85	10.96	22.30	17.87	30.88	39.26	27.81	37.79	42.28	7.64	21.72	22.36	22.04
23	1.39	4.27	8.94	19.65	17.72	24.90	36.45	16.98	36.77	35.48	5.20	20.87	19.05	20.32
24	0.67	5.48	8.70	20.26	14.91	24.38	39.34	15.23	37.24	37.82	4.20	22.76	19.25	21.01
Aligned Fine-tuning														
0	43.28	42.27	38.06	46.12	48.40	44.43	50.87	45.70	48.07	35.32	26.87	47.01	43.03	43.11
1	49.09	51.45	43.52	44.00	36.56	54.94	48.74	47.29	45.70	44.79	19.85	42.91	44.07	48.46
2	36.11	38.09	31.90	48.43	36.83	38.91	50.49	54.44	48.00	48.61	19.05	46.79	41.47	41.25
3	18.10	22.05	22.42	35.82	25.20	38.52	48.70	48.32	37.06	47.70	9.12	37.01	32.50	36.16
4	10.42	13.17	21.40	34.64	26.08	31.48	47.46	38.31	41.48	42.51	7.92	35.80	29.22	33.02
5	3.97	15.39	18.45	25.98	25.68	22.31	46.36	46.64	46.36	44.00	5.87	32.31	27.78	32.05
6	1.82	9.73	11.04	18.55	23.70	17.72	42.64	28.21	42.42	52.85	8.34	28.66	23.81	26.60
7	1.02	7.36	9.56	16.73	19.83	9.84	39.27	21.45	51.73	52.76	7.45	26.92	21.99	24.50
8	0.66	5.78	8.55	14.09	18.80	8.61	36.80	16.54	48.54	46.48	6.06	25.32	19.69	24.39
9	0.72	4.95	8.11	12.68	17.06	7.68	36.48	13.11	47.64	46.71	7.08	24.43	18.89	22.63
10	0.63	4.14	8.04	12.61	16.83	6.97	36.33	13.37	47.86	46.60	5.83	24.35	18.63	22.23
11	0.65	4.02	8.43	12.49	17.21	9.37	37.20	13.11	49.08	52.97	6.57	24.38	19.62	21.78
12	0.53	4.23	8.14	12.27	15.45	12.90	38.19	14.45	47.58	45.70	5.23	24.00	19.06	22.16
13	0.44	4.13	7.99	12.76	15.24	13.92	38.77	14.70	50.30	53.16	5.04	23.73	20.01	22.31
14	0.35	4.02	7.74	12.91	15.59	14.01	38.66	13.37	49.14	53.91	5.02	23.62	19.86	22.19
15	0.38	3.75	7.81	12.95	16.63	14.61	38.78	13.52	49.36	46.13	5.15	23.39	19.37	22.46
16	0.39	3.54	7.85	12.75	16.89	14.26	40.24	12.19	49.66	46.86	4.79	23.11	19.38	22.08
17	0.38	3.66	7.50	13.11	16.48	16.12	40.22	11.79	50.66	54.67	5.25	23.57	20.28	22.00
18	0.36	3.55	7.87	12.94	17.53	19.21	40.02	11.68	50.99	53.34	4.83	24.37	20.56	21.73
19	0.37	3.72	7.38	13.29	16.68	21.86	39.19	10.73	49.05	46.25	5.34	23.43	19.77	21.54
20	0.38	3.11	6.67	14.01	16.23	23.54	40.18	10.86	48.03	54.13	6.92	23.42	20.62	20.27
21	0.39	3.31	7.19	14.35	17.16	25.70	39.05	10.85	49.39	52.60	6.71	22.78	20.79	20.57
22	0.35	3.04	7.59	14.88	17.48	20.63	42.17	12.71	48.63	46.91	6.52	24.04	20.41	21.64
23	0.42	3.97	8.04	17.00	17.37	24.30	43.26	11.93	49.63	53.29	3.18	24.57	21.41	23.49
24	0.39	5.04	8.31	15.29	13.72	24.13	44.22	18.97	51.89	47.52	7.24	27.51	22.02	23.20

Table 20: *HuBert Large* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	58.93	52.26	53.55	47.22	32.66	31.72	46.97	59.98	45.38	55.64	45.22	45.58	47.93	51.14
1	61.03	62.83	57.53	50.92	34.50	67.39	46.07	45.17	40.33	49.43	34.77	49.00	49.91	59.38
2	68.85	69.61	64.57	52.52	36.81	64.26	43.17	47.56	45.68	53.09	40.15	47.05	52.78	60.46
3	63.89	66.42	58.83	57.81	39.00	68.52	42.61	47.43	46.50	44.54	38.04	48.15	51.81	58.82
4	50.77	57.07	49.38	58.38	39.91	72.08	46.28	41.46	39.87	38.66	33.09	48.81	47.98	57.17
5	55.25	62.03	53.24	48.96	40.87	72.57	43.25	40.10	41.44	43.98	33.26	41.17	48.01	55.98
6	55.82	63.26	52.83	52.48	45.06	70.54	45.80	40.23	37.39	47.67	35.16	47.12	49.45	58.21
7	61.24	67.63	59.18	57.82	42.91	66.51	45.91	39.47	41.94	46.87	41.04	56.12	52.22	61.25
8	54.34	64.62	53.00	51.23	44.18	68.06	45.62	33.65	43.97	45.47	32.59	51.22	49.00	59.41
9	49.93	60.89	50.82	57.36	46.17	67.63	49.05	27.66	42.02	40.69	31.09	60.72	48.67	62.65
10	59.85	68.78	56.22	51.34	46.38	61.48	48.24	39.47	38.21	43.56	39.34	58.36	50.93	61.67
11	50.79	61.82	46.90	41.58	42.02	65.02	47.07	38.80	36.24	39.57	33.44	42.49	45.48	52.02
12	37.01	48.00	34.49	33.92	35.93	60.16	45.49	31.40	35.70	34.49	24.75	31.16	37.71	44.93
13	37.40	47.51	36.80	36.50	37.79	64.13	47.88	30.18	42.32	37.17	24.38	33.80	39.66	50.90
14	25.37	34.37	26.01	27.04	31.70	59.11	48.15	22.64	33.49	34.87	20.74	31.57	32.92	45.12
15	29.87	36.74	28.52	31.80	34.83	62.67	47.25	26.10	37.54	37.65	23.79	28.70	35.45	48.47
16	20.33	25.30	23.45	31.88	29.01	66.16	43.58	24.40	37.81	39.01	20.74	23.36	32.09	43.47
17	12.54	17.93	18.06	25.33	24.60	59.58	42.96	20.68	36.13	35.97	16.26	20.59	27.55	36.44
18	1.77	8.65	6.44	16.55	13.57	21.99	36.93	17.19	30.66	29.19	6.20	17.02	17.18	18.85
19	1.32	6.65	5.78	15.64	14.44	18.60	37.15	14.72	25.98	24.87	5.69	16.85	15.64	16.98
20	2.09	8.00	6.05	16.17	14.41	14.98	37.06	14.72	27.52	24.46	7.66	16.79	15.82	16.34
21	1.97	6.73	5.91	10.45	14.73	13.92	36.72	12.06	30.97	22.66	6.43	16.77	14.94	16.11
22	1.24	5.90	5.83	8.41	14.37	14.67	36.96	7.81	28.73	17.97	3.83	16.81	13.54	16.02
23	0.61	4.76	5.69	6.97	13.95	14.39	36.70	7.92	24.12	13.42	3.20	16.86	12.38	16.15
24	0.65	3.57	5.95	7.42	14.55	13.08	37.22	10.07	21.34	7.47	3.30	16.83	11.79	16.02
Aligned Fine-tuning														
0	28.73	30.94	40.21	56.05	49.00	34.85	55.62	43.70	44.28	36.63	51.56	51.48	43.59	43.31
1	22.88	23.87	32.15	52.11	46.57	50.84	58.09	34.45	50.20	28.80	19.23	51.56	39.23	43.37
2	14.19	15.91	21.89	36.54	27.52	41.39	53.45	24.24	56.18	28.69	15.95	46.56	31.88	38.97
3	6.89	11.95	19.84	27.83	19.88	26.33	48.38	29.17	56.35	32.00	11.65	37.63	27.33	37.53
4	6.96	13.43	18.46	26.16	21.29	28.70	47.57	20.40	54.90	27.50	12.75	36.23	26.20	36.06
5	4.72	11.33	15.33	18.14	27.37	27.97	45.51	24.78	54.48	24.72	12.44	32.31	24.93	32.64
6	2.39	7.76	12.53	13.66	24.50	14.48	43.29	20.27	36.99	24.42	11.61	27.37	19.94	28.72
7	1.29	6.49	9.17	6.13	23.15	11.52	41.39	19.34	29.47	24.04	9.52	24.07	17.13	25.46
8	0.95	6.53	8.37	4.47	21.84	13.03	41.70	19.07	26.79	26.82	7.67	22.91	16.68	24.72
9	0.85	6.34	7.38	3.52	21.12	11.09	41.20	19.63	25.13	26.51	6.38	20.98	15.85	22.80
10	0.73	6.08	6.95	3.07	19.72	9.75	40.77	17.49	24.85	27.16	6.06	20.53	15.26	22.57
11	0.79	6.05	6.57	2.39	21.04	11.31	41.93	15.35	24.80	26.49	4.74	19.67	15.09	21.25
12	0.57	5.26	5.93	2.08	19.17	9.41	42.31	12.96	24.87	25.30	4.37	19.04	14.27	20.26
13	0.53	5.22	5.88	2.20	16.49	9.45	41.74	13.24	24.62	25.38	3.92	18.65	13.94	19.50
14	0.54	5.06	6.01	2.12	14.16	10.03	39.37	13.64	23.86	24.55	3.53	18.40	13.44	18.87
15	0.50	4.97	5.83	2.46	13.98	12.90	38.39	11.54	24.84	25.61	3.41	18.27	13.56	18.52
16	0.37	4.50	5.70	2.88	13.80	15.54	37.92	10.08	23.48	23.85	3.06	18.20	13.28	18.22
17	0.58	4.48	5.95	2.84	13.92	14.44	38.79	9.53	25.06	23.23	3.81	18.14	13.40	17.89
18	0.69	4.65	6.00	3.82	13.94	15.28	37.86	9.67	24.39	24.21	3.60	18.22	13.53	17.96
19	1.11	4.19	6.17	4.20	13.51	14.98	37.32	10.73	26.42	23.76	5.27	18.13	13.82	17.49
20	1.24	4.38	6.22	4.48	13.75	13.51	38.34	12.86	22.11	23.77	7.11	18.11	13.82	17.60
21	0.80	3.85	6.23	3.82	13.16	15.32	38.11	9.70	22.65	21.01	3.97	18.02	13.06	17.61
22	0.50	3.88	5.75	2.65	12.41	13.03	38.19	9.27	23.35	20.47	3.44	18.15	12.59	17.57
23	0.53	3.24	5.90	2.87	12.42	13.83	38.60	8.90	22.34	17.95	3.18	18.02	12.32	17.49
24	0.60	3.39	6.16	3.41	12.94	15.06	39.09	13.77	20.94	7.94	5.14	18.13	12.21	17.52

Table 21: *UniSpeechSAT Large* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	74.00	72.19	65.69	45.47	47.58	44.93	43.05	65.96	49.99	69.74	81.37	44.76	58.73	48.74
1	34.40	31.57	41.94	46.95	62.88	44.04	58.58	58.31	55.67	70.40	60.61	47.98	51.11	39.76
2	29.59	25.45	40.44	47.60	69.46	56.58	60.77	56.28	68.81	41.94	51.53	50.94	49.95	41.21
3	15.50	15.31	31.52	46.84	71.09	56.54	58.86	43.97	46.39	55.34	34.37	54.03	44.15	39.91
4	16.47	18.35	31.78	42.82	63.91	52.40	59.73	49.68	50.78	34.28	40.96	45.59	42.23	35.69
5	16.10	19.73	30.48	38.99	56.08	62.95	60.51	41.30	52.47	36.00	36.00	41.24	40.99	34.09
6	25.34	23.86	32.03	32.00	52.21	70.37	59.85	50.56	53.59	36.26	45.13	42.36	43.63	36.08
7	32.74	31.41	32.75	33.43	54.88	70.18	55.57	55.76	48.47	58.07	26.89	42.98	45.26	40.47
8	23.47	24.30	23.24	16.40	43.17	67.72	50.49	54.17	44.46	54.55	33.52	34.60	39.17	32.38
9	36.42	34.80	27.60	16.32	44.12	45.77	49.59	55.10	48.22	60.70	43.38	29.65	40.97	31.35
10	15.84	20.08	12.45	7.50	24.84	43.68	37.53	38.92	51.25	52.78	20.41	22.33	28.97	23.73
11	33.90	32.17	18.80	8.79	29.06	23.09	34.66	45.70	47.17	65.35	35.93	21.80	33.03	24.90
12	17.73	23.31	12.91	7.31	30.15	20.67	32.58	25.17	46.93	64.83	22.18	20.63	27.03	22.00
13	6.36	12.13	9.57	10.15	23.26	34.31	32.48	17.36	47.11	65.10	10.66	20.18	24.06	20.08
14	6.49	10.81	10.41	11.40	22.87	42.14	31.54	15.63	47.16	65.07	12.07	19.07	24.55	19.51
15	1.02	4.01	7.51	9.05	16.91	62.37	29.57	15.37	47.03	65.09	6.87	17.58	23.53	18.80
16	1.00	5.07	6.54	6.40	15.04	43.63	29.25	12.16	47.29	66.95	7.15	15.52	21.33	16.84
17	0.84	3.60	6.01	5.57	13.02	48.19	29.94	11.66	47.32	70.17	7.90	15.40	21.63	16.67
18	0.84	3.62	5.86	6.17	12.18	48.45	30.42	10.35	53.29	35.39	8.18	14.93	19.14	16.79
19	0.68	3.27	6.01	6.93	12.84	55.65	32.22	9.40	57.12	40.82	6.92	15.65	20.63	17.30
20	1.22	2.81	5.97	6.94	12.85	65.69	33.33	10.05	53.05	35.59	9.22	15.12	20.99	17.24
21	1.37	2.73	5.46	4.88	11.78	72.08	30.89	11.00	54.41	36.99	9.89	14.86	21.36	17.21
22	1.28	2.58	4.32	4.77	11.59	71.02	31.83	11.11	59.66	47.46	10.43	16.11	22.68	18.09
23	0.92	2.61	3.51	5.34	10.82	61.13	33.18	12.57	63.25	68.09	9.54	16.87	23.99	18.02
24	0.85	2.91	3.58	6.33	10.80	55.07	35.71	13.27	45.22	64.91	8.96	17.07	22.06	17.95
Aligned Fine-tuning														
0	23.06	24.49	33.72	52.14	69.85	50.93	46.10	29.67	61.40	47.93	25.99	56.93	43.52	45.79
1	19.21	17.26	29.47	51.79	65.56	51.32	55.00	29.92	46.80	47.39	30.47	55.27	41.62	39.77
2	15.77	17.84	26.34	49.49	51.21	64.31	57.31	29.29	57.67	32.79	24.15	49.53	39.64	38.10
3	13.13	20.70	25.62	45.40	45.34	64.18	47.69	25.95	52.32	40.93	17.35	41.32	36.66	38.29
4	12.13	21.12	22.32	35.21	41.83	58.14	40.73	28.60	53.21	40.79	17.00	35.60	33.89	36.96
5	7.11	15.08	18.58	19.39	37.68	66.53	38.47	25.83	65.00	36.29	12.89	31.30	31.18	32.71
6	5.21	14.26	17.02	16.96	30.92	65.06	35.23	20.93	53.93	39.49	12.12	29.42	28.38	31.43
7	2.80	12.93	14.47	9.21	22.82	64.26	31.58	16.69	54.58	34.70	12.84	26.95	25.32	28.72
8	0.88	10.60	8.94	3.48	14.57	42.88	25.35	13.65	35.34	63.86	9.52	20.02	20.76	22.73
9	0.41	6.97	6.14	2.07	12.14	36.04	23.01	13.77	42.63	66.57	7.38	17.58	19.56	20.03
10	0.43	6.07	6.08	2.04	11.43	49.74	23.71	13.94	44.31	64.99	8.03	17.40	20.68	19.77
11	0.47	4.45	5.65	2.04	10.90	52.61	24.24	12.33	45.22	64.14	6.62	16.80	20.46	19.24
12	0.55	3.45	5.51	2.08	11.60	46.42	25.01	12.29	44.78	64.54	6.69	15.99	19.91	18.21
13	0.78	3.21	6.60	2.42	11.79	53.93	26.25	12.71	55.75	34.25	7.06	16.28	19.25	17.93
14	0.86	3.02	7.75	3.33	12.53	60.34	27.32	12.56	44.59	65.00	6.97	17.02	21.77	18.53
15	0.94	2.60	7.71	3.48	13.03	70.37	26.50	12.71	44.77	64.85	6.60	17.55	22.59	19.05
16	1.10	2.89	8.39	5.12	14.36	67.26	29.54	12.29	46.73	67.28	7.50	17.41	23.32	18.87
17	0.77	2.45	7.63	4.92	14.37	69.79	29.79	12.34	44.80	65.21	6.85	17.19	23.01	18.67
18	1.03	2.55	8.10	5.72	14.90	66.83	31.24	12.71	55.98	36.19	6.92	17.20	21.61	18.70
19	0.88	2.40	8.10	6.05	14.82	66.66	30.41	13.92	57.05	58.40	7.11	16.79	23.55	18.33
20	0.99	2.38	8.04	8.14	15.41	65.36	31.61	13.11	57.46	46.74	7.11	16.99	22.78	18.58
21	1.02	2.39	7.38	7.04	13.80	61.89	29.03	12.97	56.64	38.74	6.46	14.87	21.02	17.64
22	0.94	2.20	7.38	7.42	13.50	69.62	29.42	12.46	68.17	55.75	6.95	14.93	24.06	18.00
23	0.88	2.39	5.27	6.29	12.85	75.05	33.69	13.64	61.15	61.22	6.69	17.76	24.74	19.48
24	0.94	3.52	3.86	8.10	11.65	40.94	35.73	14.04	51.69	57.76	8.10	18.05	21.20	30.82

Table 22: *WavLM Large* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	22.16	30.72	37.04	47.44	68.75	30.60	55.80	29.27	47.01	30.68	26.27	48.20	39.50	58.82
1	24.94	29.22	38.14	44.91	56.33	60.34	53.82	28.51	55.01	65.56	27.73	49.80	44.52	61.25
2	38.19	40.58	41.18	40.44	51.55	53.93	55.40	47.31	59.66	46.95	51.60	48.65	47.95	59.83
3	31.54	33.46	34.29	40.21	58.37	41.65	55.77	36.03	53.36	34.46	37.76	48.96	42.15	51.26
4	29.03	31.65	30.19	47.11	67.11	41.97	52.01	33.37	48.25	31.98	38.02	49.75	41.70	59.28
5	21.02	25.96	25.75	40.41	53.84	48.53	42.75	32.04	51.03	70.06	30.82	49.59	40.98	49.76
6	22.82	27.81	23.59	31.74	51.11	28.98	46.06	45.30	46.64	62.29	42.99	51.09	40.03	40.55
7	22.56	28.32	27.05	37.98	52.66	26.20	46.35	40.68	47.28	58.21	40.52	51.60	39.95	50.69
8	35.17	40.31	41.26	58.31	63.30	24.30	46.30	47.66	47.11	61.25	45.19	51.23	46.78	57.61
9	40.00	49.20	51.42	55.47	61.40	24.13	43.97	39.35	47.07	68.22	42.03	49.03	47.61	54.74
10	53.25	61.41	56.04	53.47	61.55	20.50	57.29	48.74	50.94	34.80	56.69	49.42	50.34	54.93
11	55.59	61.51	49.40	59.61	64.96	34.23	53.93	54.47	53.34	38.09	45.96	48.85	51.66	53.85
12	51.99	61.76	58.38	67.43	71.46	32.82	57.69	46.63	51.68	31.94	40.55	50.81	51.93	57.91
13	60.64	67.60	56.74	63.00	59.65	39.88	54.45	50.07	53.67	37.31	33.70	49.22	52.16	58.28
14	52.17	61.22	54.26	64.60	57.13	54.06	55.33	39.22	52.40	39.52	32.00	49.34	50.94	58.65
15	29.63	40.98	28.66	46.64	39.61	29.07	47.94	38.80	47.30	65.57	23.36	53.07	40.89	60.68
16	9.14	13.47	10.34	22.64	23.37	13.25	53.97	24.92	55.91	40.68	12.89	50.86	27.62	35.29
17	3.85	6.13	6.21	16.58	16.43	11.57	47.45	19.47	49.32	62.71	7.90	51.75	24.95	36.53
18	3.83	6.42	4.87	14.72	12.93	10.08	54.42	18.67	53.53	39.21	8.19	49.09	23.00	31.48
19	2.93	5.46	4.09	12.00	10.88	7.42	47.28	15.90	49.98	63.52	7.20	51.19	23.16	35.20
20	1.93	4.16	4.55	13.18	11.40	8.65	55.12	13.77	56.78	41.33	5.57	49.52	22.16	32.21
21	1.02	3.46	4.05	9.17	10.66	9.95	83.81	12.33	87.31	83.06	4.04	86.54	32.95	49.36
22	0.45	4.54	3.58	7.81	10.37	2.33	56.07	12.06	51.16	46.63	2.16	68.74	22.16	40.47
23	0.36	4.27	3.62	9.31	10.94	4.42	51.50	11.26	78.20	80.91	2.14	75.85	27.73	51.22
24	0.36	4.24	3.65	9.06	10.94	5.78	54.33	12.06	52.29	50.41	2.11	73.01	23.19	39.37
Aligned Fine-tuning														
0	39.59	42.60	38.54	53.01	48.69	21.39	54.07	26.36	55.01	57.24	33.47	48.33	43.19	43.26
1	34.38	35.56	31.16	51.31	49.20	30.17	50.37	38.27	41.49	48.46	34.71	50.39	41.29	41.40
2	13.66	18.88	19.13	44.11	24.14	38.15	52.29	21.99	44.77	29.21	16.14	49.65	31.01	39.19
3	7.74	11.85	16.02	37.45	22.25	37.18	44.69	15.50	37.69	30.64	14.54	50.11	27.14	38.23
4	4.08	11.62	12.57	25.10	18.92	21.60	40.78	18.94	50.87	25.22	14.21	49.52	24.45	35.01
5	1.82	12.04	11.10	20.56	19.12	23.09	42.97	17.06	48.81	20.35	11.47	50.50	23.24	32.53
6	0.92	10.41	8.66	14.73	16.18	21.99	45.40	18.14	43.94	21.25	8.73	50.29	21.72	33.83
7	0.44	8.74	7.12	11.40	14.71	11.35	42.61	15.48	51.74	21.11	6.43	50.56	20.14	32.62
8	0.49	8.27	5.17	6.25	12.35	8.65	45.51	15.25	46.60	20.95	5.93	50.73	18.85	32.69
9	0.53	6.27	3.71	4.50	9.14	3.97	53.08	15.50	53.32	20.73	4.72	52.11	18.97	32.15
10	0.56	5.42	3.11	3.57	8.13	2.52	54.47	13.64	53.36	19.37	4.06	49.20	18.12	30.43
11	0.56	5.36	2.89	3.07	7.83	1.77	54.13	11.91	53.58	19.41	3.71	49.13	17.78	29.96
12	0.45	4.96	2.71	2.84	7.76	1.47	55.50	11.53	52.92	18.81	3.36	49.43	17.64	29.84
13	0.42	4.88	2.68	2.84	7.47	1.55	46.85	11.79	47.64	18.23	2.68	51.63	16.55	29.42
14	0.61	5.27	2.74	2.84	7.56	1.55	47.46	12.46	46.82	19.23	2.23	52.05	16.73	29.83
15	0.45	4.86	2.73	2.95	7.73	1.47	64.05	12.21	75.73	19.21	2.20	63.21	21.40	34.61
16	0.38	4.47	2.68	3.03	7.84	1.55	56.70	10.70	54.31	18.85	1.86	50.42	17.73	30.19
17	0.48	4.41	2.66	3.10	7.96	1.47	46.75	9.14	47.89	18.74	2.32	51.44	16.36	29.68
18	0.50	4.14	2.62	3.25	8.05	1.90	53.71	8.87	53.46	18.43	2.46	48.93	17.19	29.21
19	0.61	4.43	2.62	3.14	7.79	2.07	56.75	8.47	55.26	18.04	2.51	50.43	17.68	29.14
20	0.65	4.30	2.58	3.10	7.74	1.99	62.13	10.35	55.84	17.52	2.46	51.48	18.34	31.88
21	0.65	3.94	2.64	3.15	7.98	1.51	56.94	8.60	54.60	18.12	2.79	57.70	18.22	32.37
22	0.91	4.27	2.80	3.52	7.79	1.47	86.08	6.99	51.92	18.87	3.39	72.80	21.73	35.12
23	1.66	4.32	2.84	3.07	7.51	1.27	56.86	9.80	64.27	19.42	4.11	82.12	21.44	46.96
24	0.95	4.42	2.73	3.00	7.90	1.40	59.30	8.34	50.61	18.71	2.07	57.12	18.05	32.86

Table 23: *XLSR-53 300M* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	29.72	29.96	37.70	57.48	56.36	40.76	54.10	30.86	39.00	30.51	41.29	51.20	41.58	43.95
1	23.19	25.36	34.80	51.50	61.01	37.38	57.35	32.59	52.07	34.90	21.85	49.82	40.15	43.96
2	22.98	23.87	34.86	51.31	60.66	41.61	56.16	29.40	55.48	37.50	19.74	53.43	40.58	46.00
3	18.18	19.66	26.16	39.45	46.62	40.85	43.04	20.40	45.88	30.49	26.89	38.80	33.03	33.73
4	32.89	35.56	39.76	37.52	52.20	54.68	47.27	38.41	57.96	35.50	33.32	42.05	42.26	41.93
5	31.87	34.41	39.38	48.39	50.40	41.95	55.67	34.70	49.85	49.65	20.07	52.55	42.41	50.89
6	28.38	30.21	30.43	40.81	43.85	25.87	48.46	34.05	50.67	51.98	19.11	45.19	37.42	44.39
7	44.94	45.82	46.92	53.05	52.06	33.71	53.30	38.41	55.16	57.51	26.36	58.63	47.15	54.14
8	47.79	45.47	42.13	51.95	41.16	26.85	48.65	42.91	53.00	60.10	32.89	53.03	45.49	49.30
9	46.20	43.03	39.26	43.92	41.55	30.04	46.04	49.00	57.58	58.51	37.79	43.44	44.70	43.71
10	21.55	24.95	24.13	35.02	38.29	24.38	50.82	33.89	46.40	51.93	15.37	38.20	33.74	38.15
11	33.34	38.61	37.02	40.33	38.18	28.49	49.79	33.75	44.54	54.04	17.39	42.70	38.18	46.07
12	22.00	22.24	19.90	33.40	31.92	23.37	42.28	31.89	41.70	48.85	16.93	29.41	30.32	31.04
13	31.80	33.11	26.99	41.39	40.05	28.27	44.46	30.86	42.75	53.64	20.90	37.86	36.01	38.37
14	10.20	12.54	10.16	21.09	24.40	12.45	39.44	22.91	34.92	42.31	14.45	25.27	22.51	22.46
15	18.84	16.29	13.53	26.92	27.66	20.63	36.07	30.05	47.98	57.77	27.98	22.36	28.84	24.47
16	56.75	53.15	43.31	59.49	55.67	42.92	53.52	46.63	65.23	75.21	57.85	52.85	55.21	51.54
17	43.59	36.95	28.23	55.89	47.53	32.41	45.18	47.13	62.36	69.48	53.55	35.87	46.51	37.58
18	50.30	44.76	34.68	53.51	52.54	22.31	44.90	60.80	72.16	74.99	56.52	35.17	50.22	41.84
19	4.96	6.06	9.09	18.78	16.68	9.28	37.66	17.86	39.94	32.79	8.25	20.60	18.50	23.47
20	0.30	1.78	4.55	8.37	10.71	4.94	34.62	6.59	19.65	12.80	6.66	20.25	10.94	18.22
21	0.16	1.79	3.82	7.84	10.49	4.14	33.07	6.51	19.71	13.03	4.86	19.69	10.43	17.08
22	2.27	3.33	5.78	12.16	12.90	7.34	32.04	21.21	21.97	22.05	7.41	34.09	15.21	26.80
23	1.66	2.90	5.02	10.53	12.43	6.84	31.57	21.16	21.44	20.54	6.73	29.27	14.17	24.18
24	0.22	2.44	3.86	7.62	10.62	6.24	31.79	19.87	19.50	13.22	5.04	18.63	11.59	16.28
Aligned Fine-tuning														
0	35.76	37.08	48.03	55.81	44.90	58.05	53.79	36.03	42.88	54.56	41.57	53.35	46.82	51.71
1	14.41	16.86	27.03	45.21	39.19	53.84	52.28	24.48	46.74	30.78	19.09	45.87	34.65	43.37
2	8.01	18.64	19.11	29.58	30.05	66.92	46.86	34.44	48.68	31.43	13.12	38.36	32.10	40.33
3	4.04	15.08	14.39	21.92	27.26	35.07	40.66	23.06	55.83	25.32	12.61	31.85	25.59	35.19
4	1.59	10.82	11.01	14.34	25.19	23.72	39.70	20.42	53.14	18.71	11.03	28.68	21.53	32.24
5	0.49	8.39	8.54	9.62	22.11	18.15	37.24	17.77	45.35	18.35	8.18	25.63	18.32	28.86
6	0.27	5.46	6.12	5.91	22.56	12.02	35.97	16.79	39.03	17.51	6.78	22.67	15.92	24.88
7	0.18	4.84	5.14	4.51	16.04	6.24	32.59	12.33	31.82	20.68	5.34	19.09	13.23	21.61
8	0.16	5.01	4.87	3.56	14.43	3.22	31.27	10.61	28.03	22.77	4.83	18.02	12.23	20.43
9	0.12	3.81	3.99	3.07	13.92	2.91	30.19	9.83	25.46	21.29	4.74	17.59	11.41	19.23
10	0.11	3.25	3.68	2.76	12.97	2.74	29.65	8.74	24.87	19.30	4.43	17.03	10.79	18.24
11	0.10	3.21	3.71	2.76	13.30	2.52	29.88	8.21	24.71	18.91	4.51	17.00	10.73	17.77
12	0.10	2.97	3.65	2.91	13.89	2.48	30.79	8.07	25.05	17.36	4.01	17.19	10.71	17.70
13	0.10	3.12	3.73	2.83	13.76	2.33	30.76	8.49	25.15	16.99	3.74	17.26	10.69	17.65
14	0.10	2.99	3.57	2.84	13.31	2.74	30.93	8.24	23.71	16.63	3.99	17.46	10.54	17.59
15	0.08	2.64	3.63	3.07	13.20	2.70	31.11	7.96	23.14	14.38	3.53	17.37	10.23	17.58
16	0.12	1.96	3.85	3.52	13.02	3.04	31.24	6.63	22.01	12.39	3.41	18.01	9.93	17.57
17	0.12	2.03	3.91	3.79	12.83	3.26	31.53	5.96	21.37	12.67	3.30	18.14	9.91	17.48
18	0.12	1.94	3.95	3.78	12.50	2.74	30.60	5.56	22.27	11.89	3.35	17.88	9.71	17.38
19	0.12	1.81	3.86	4.21	11.91	2.48	31.48	6.50	20.91	11.48	3.16	18.03	9.66	17.25
20	0.12	1.87	3.91	4.66	12.76	2.61	32.16	8.46	21.75	11.06	3.35	17.88	10.05	17.35
21	0.12	1.94	3.93	4.47	11.76	2.48	31.31	7.54	21.45	11.72	2.88	17.74	9.78	17.68
22	0.14	2.27	4.81	7.50	13.68	2.44	31.27	11.66	21.51	12.38	3.48	29.36	11.71	26.32
23	0.15	2.48	4.46	5.41	12.27	2.33	31.05	12.87	20.22	13.29	3.46	23.92	10.99	23.00
24	0.16	2.90	3.77	3.86	10.77	1.55	32.14	17.87	19.77	13.81	4.76	18.52	10.83	17.38

Table 24: *XLSR 128 300M* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer. The Avg. column is the unweighted mean of EERs over 12 datasets.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	69.59	56.03	58.48	52.93	15.34	47.48	52.91	58.94	43.23	45.20	73.13	51.46	52.06	50.64
1	45.19	38.82	47.51	52.59	67.20	44.00	56.63	45.96	51.95	31.70	66.58	61.33	50.79	45.00
2	47.90	41.00	46.62	50.62	64.22	46.68	50.64	40.63	44.08	28.62	68.07	57.97	48.92	45.95
3	32.37	36.16	34.92	45.02	61.21	53.09	49.72	38.80	39.09	32.28	37.79	55.44	42.99	43.92
4	46.00	41.66	42.01	47.33	72.86	51.55	54.17	61.71	56.28	45.51	43.50	58.82	51.78	48.28
5	46.43	42.26	51.47	51.87	72.58	74.38	60.32	64.27	62.87	37.66	41.73	59.83	55.47	53.47
6	38.95	31.09	41.09	49.56	75.34	49.33	64.25	55.07	60.39	31.30	37.64	60.05	49.51	48.54
7	45.94	44.69	53.85	51.65	71.23	72.44	62.34	49.14	66.25	38.95	36.72	68.30	55.12	58.95
8	11.76	13.13	14.24	23.25	47.92	44.78	51.29	34.17	47.06	41.39	22.07	30.96	31.83	29.13
9	15.64	15.04	9.13	9.99	33.59	20.33	37.26	42.39	42.07	41.78	25.71	17.67	25.88	20.65
10	18.09	17.45	10.29	11.70	37.35	16.62	42.73	36.69	39.42	40.15	25.83	18.01	26.19	21.63
11	13.50	12.08	6.93	9.09	28.22	10.64	34.98	41.86	31.45	35.51	31.38	16.80	22.70	16.23
12	7.26	9.53	5.76	8.67	28.88	10.55	36.40	28.72	24.87	31.19	23.34	20.57	19.64	17.49
13	3.70	4.77	4.77	10.56	29.25	12.58	34.57	21.98	31.98	28.20	11.84	16.97	17.60	16.24
14	6.77	9.05	8.69	16.96	35.83	17.67	44.55	28.47	37.24	34.19	16.79	20.82	23.09	20.11
15	18.83	19.51	11.40	17.38	31.95	21.26	33.95	35.88	37.53	40.61	34.00	17.65	26.66	20.35
16	1.55	2.76	2.93	5.57	15.91	6.09	22.64	14.44	21.93	19.62	6.76	14.54	11.23	12.88
17	1.17	2.96	1.88	2.30	8.29	2.44	18.73	12.43	16.73	18.74	3.30	13.35	8.53	12.38
18	0.73	2.42	1.77	2.01	6.40	1.99	18.74	9.53	14.90	17.06	2.25	13.07	7.57	12.31
19	0.34	2.49	1.53	2.28	5.69	2.35	17.97	8.62	12.34	14.56	1.16	12.53	6.82	12.20
20	0.11	2.11	1.41	1.67	5.22	1.23	17.63	7.13	10.33	12.80	1.39	12.87	6.16	12.12
21	0.07	2.07	1.42	1.67	5.09	1.01	17.59	6.89	10.63	11.57	2.32	13.18	5.80	11.99
22	0.07	2.22	1.43	1.70	4.82	1.01	18.65	6.63	10.51	11.91	2.13	14.01	6.26	11.74
23	0.10	2.31	1.49	1.67	4.59	0.88	20.62	7.81	10.24	12.31	1.86	17.81	6.81	11.99
24	0.10	2.36	1.46	1.75	4.69	0.93	19.57	17.51	10.16	12.44	3.48	13.47	7.33	11.70
Aligned Fine-tuning														
0	21.98	22.08	35.17	53.23	49.12	82.24	49.20	23.57	54.53	34.74	23.37	50.51	41.65	43.52
1	31.84	31.85	34.56	55.51	45.51	80.25	44.75	22.28	47.42	46.41	27.38	52.08	43.32	51.35
2	24.16	27.40	28.89	48.88	34.11	83.92	42.20	17.62	39.30	41.14	20.06	49.29	38.08	45.60
3	20.16	24.84	25.02	45.77	26.48	73.20	40.90	20.00	40.16	40.68	16.84	42.10	34.68	41.65
4	19.17	19.20	19.92	32.97	24.41	50.00	43.29	18.39	35.50	36.08	16.79	42.73	29.87	37.96
5	7.08	9.82	10.73	11.66	15.39	22.23	35.25	13.37	39.15	24.54	7.94	46.32	20.29	33.46
6	1.17	6.03	3.52	3.33	8.40	7.90	23.89	11.76	18.97	20.42	3.72	41.16	12.52	27.81
7	0.29	3.96	1.94	1.82	4.87	3.22	19.95	7.94	17.96	14.71	2.53	35.58	9.56	23.80
8	0.23	3.91	1.86	1.14	4.12	2.52	19.33	6.38	14.07	16.10	1.33	21.63	7.72	16.07
9	0.11	3.62	1.76	0.99	4.07	2.33	18.69	6.91	13.80	15.28	1.35	22.24	7.60	16.36
10	0.10	3.38	1.67	1.02	4.00	2.07	18.58	7.03	13.89	15.00	1.55	21.23	7.46	15.76
11	0.07	3.02	1.60	0.95	3.97	2.11	18.57	7.13	13.51	14.09	1.51	24.56	7.59	18.23
12	0.07	2.98	1.52	0.99	4.05	2.24	18.47	7.29	13.60	14.08	1.40	25.31	7.67	18.40
13	0.07	2.78	1.55	0.92	4.01	2.07	18.33	7.54	13.39	13.48	1.53	19.56	7.10	14.61
14	0.08	2.73	1.55	0.91	4.20	2.07	18.39	7.67	13.70	13.37	1.69	20.09	7.20	15.07
15	0.08	2.67	1.61	1.02	4.17	1.90	18.58	7.79	14.00	12.80	1.92	19.18	7.14	14.41
16	0.07	2.44	1.56	1.06	4.09	1.77	18.38	7.43	12.94	12.66	1.93	17.00	6.78	12.87
17	0.09	2.42	1.62	1.14	4.36	1.68	18.38	7.67	13.89	12.79	1.74	13.83	6.63	12.48
18	0.07	2.00	1.58	1.36	4.37	1.64	18.24	6.89	13.96	11.53	1.88	13.81	6.44	12.55
19	0.08	1.92	1.54	1.40	4.45	1.14	18.38	6.63	13.90	10.89	1.67	14.29	6.36	13.42
20	0.05	1.59	1.54	1.55	4.32	1.36	18.37	5.93	12.90	10.64	2.49	15.21	6.33	13.81
21	0.05	1.59	1.45	1.63	4.42	1.94	19.27	6.50	12.52	10.01	2.92	16.31	6.55	13.86
22	0.07	1.59	1.44	1.59	4.44	2.74	20.33	7.43	12.20	9.51	2.62	16.77	6.73	13.69
23	0.08	1.71	1.49	1.59	3.98	2.44	18.60	9.29	12.28	9.57	2.72	15.76	6.63	13.21
24	0.10	1.96	1.51	1.63	3.62	2.57	18.52	11.40	12.28	9.50	3.83	14.97	6.82	12.58

Table 25: *MMS 300M* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	43.28	45.10	51.36	55.28	48.86	51.27	53.94	50.96	47.39	42.36	52.90	48.82	49.29	49.06
1	43.58	47.46	47.46	52.93	48.89	49.46	42.12	37.36	51.23	36.36	39.11	51.38	45.61	50.71
2	49.75	54.13	52.03	51.34	39.75	55.27	43.29	54.45	49.88	47.42	46.91	51.20	49.62	52.92
3	39.58	44.95	39.55	43.54	26.51	40.55	37.81	41.99	46.28	47.50	39.11	41.36	40.73	43.57
4	46.19	46.85	44.08	41.84	26.36	42.75	33.24	46.08	40.32	46.83	58.57	45.86	43.25	44.67
5	36.89	40.08	39.23	39.30	31.57	34.23	37.25	44.52	40.78	59.13	44.96	41.50	40.79	42.61
6	48.25	49.88	42.96	38.81	37.96	37.76	40.88	54.83	40.74	56.63	62.66	40.25	45.97	41.34
7	28.88	27.58	22.44	31.20	33.84	41.61	30.28	49.82	40.79	53.41	51.14	21.05	36.00	26.52
8	34.79	34.87	29.16	39.68	35.92	39.92	37.68	46.50	39.58	49.25	42.34	34.21	38.66	36.23
9	8.95	10.86	9.17	15.83	17.33	38.39	27.04	28.62	21.72	29.36	19.25	19.60	20.51	18.86
10	20.11	21.13	22.76	29.49	32.45	40.20	42.59	30.96	24.48	38.01	22.27	37.54	30.17	33.16
11	22.90	22.79	22.22	28.33	25.13	45.90	48.07	24.63	27.82	34.14	20.30	44.51	30.56	34.50
12	24.88	24.86	21.78	37.03	31.25	49.03	45.87	35.63	28.68	43.48	31.88	38.61	34.41	32.32
13	17.05	16.85	11.82	19.85	20.55	41.02	40.38	39.77	20.95	39.91	30.47	28.32	27.25	23.65
14	19.38	18.83	15.25	30.87	21.71	40.68	36.85	33.11	28.15	41.39	19.34	32.05	28.13	29.68
15	4.73	6.35	5.78	10.90	12.36	23.76	24.34	24.53	20.77	21.45	10.92	14.27	15.01	16.04
16	24.00	24.37	22.34	37.40	37.40	53.75	34.42	43.72	41.18	45.43	39.16	25.95	35.76	29.61
17	0.64	2.78	3.15	5.15	8.89	11.61	20.91	9.53	16.00	9.02	6.02	16.17	9.16	13.82
18	0.30	3.09	2.49	3.06	7.32	2.20	20.13	8.34	14.32	10.33	4.04	13.61	7.43	12.96
19	0.15	3.33	2.56	2.16	7.11	1.47	20.06	5.95	12.83	13.94	2.28	14.24	7.17	13.30
20	0.12	3.09	2.55	2.31	7.24	2.03	19.91	6.99	12.33	11.74	1.98	16.57	7.24	13.28
21	0.11	3.05	2.54	2.34	7.06	1.81	19.99	7.96	12.25	12.00	1.97	18.63	7.48	13.33
22	0.10	2.85	2.49	2.61	6.91	1.73	22.91	8.24	12.00	12.31	2.04	22.43	8.05	14.65
23	0.08	2.75	2.46	2.76	6.87	1.64	24.20	8.47	12.03	12.16	2.16	23.46	8.25	15.30
24	0.08	3.06	2.54	2.27	7.08	1.68	20.50	8.59	12.06	12.31	2.27	13.96	7.20	13.31
Aligned Fine-tuning														
0	42.58	41.67	36.82	51.12	37.94	44.69	50.89	53.75	62.73	34.74	36.76	50.20	47.53	42.80
1	19.39	25.14	28.03	46.34	37.14	76.91	49.51	28.87	42.86	46.41	16.51	47.28	44.88	43.97
2	10.01	20.32	23.61	35.89	33.48	80.71	49.54	27.03	46.81	41.14	15.54	44.81	39.10	39.19
3	2.87	15.07	16.91	22.26	29.69	56.09	48.82	22.39	47.18	40.68	11.52	35.49	35.56	29.38
4	0.76	14.32	9.66	9.51	17.70	24.51	44.57	17.36	48.90	36.08	7.24	27.73	30.08	24.53
5	0.26	9.05	7.91	7.27	15.67	21.23	42.09	20.13	48.01	24.54	5.88	26.22	22.70	21.78
6	0.14	5.79	5.62	5.00	13.41	8.98	32.26	17.46	35.24	20.42	2.93	19.91	15.47	18.00
7	0.14	4.45	4.61	3.86	10.45	4.29	28.19	15.60	21.68	14.71	2.77	16.46	11.65	15.99
8	0.20	3.99	3.98	3.03	8.96	3.67	26.66	15.48	18.27	16.10	2.30	15.11	10.57	14.52
9	0.15	3.47	3.75	2.84	8.73	3.45	26.48	14.57	16.02	15.28	2.37	14.64	9.72	13.78
10	0.14	3.37	3.63	2.88	8.56	3.93	25.73	13.79	14.84	15.00	2.16	14.34	9.62	13.61
11	0.15	3.34	3.55	2.84	8.54	3.88	25.49	13.90	14.62	14.09	2.23	14.27	9.25	13.45
12	0.16	3.37	3.50	2.84	8.46	3.50	25.05	14.04	14.08	14.08	2.12	14.21	9.07	13.43
13	0.12	3.02	3.54	2.92	8.59	2.83	25.11	14.19	13.84	13.48	2.23	13.95	8.90	13.27
14	0.15	2.90	3.62	2.88	8.62	2.44	25.51	13.92	13.44	13.37	2.11	13.95	8.83	13.07
15	0.16	2.73	3.52	3.11	8.74	3.09	25.23	13.62	12.68	12.80	2.04	13.97	8.78	12.86
16	0.16	2.66	3.56	3.18	8.72	2.91	25.65	13.50	12.95	12.66	2.28	13.84	8.73	12.80
17	0.16	2.62	3.58	3.28	8.67	2.87	25.60	13.50	13.65	12.79	2.37	13.98	8.87	12.63
18	0.15	2.58	3.49	3.10	8.77	3.22	25.88	12.46	13.27	11.53	2.44	13.94	8.66	12.39
19	0.15	2.64	3.52	3.10	9.29	3.58	24.95	12.99	12.33	10.89	2.32	13.65	8.58	12.60
20	0.14	2.64	3.46	3.07	9.05	3.32	24.78	12.72	11.84	10.64	2.46	13.63	8.54	12.69
21	0.15	2.46	3.45	3.10	8.85	3.80	25.35	13.52	11.55	10.01	2.25	13.76	8.49	12.21
22	0.15	2.40	3.46	3.07	8.77	3.67	25.43	13.62	11.48	9.51	2.34	13.71	8.27	12.00
23	0.16	2.43	3.43	3.07	8.62	3.75	25.25	13.64	11.38	9.57	2.42	13.76	8.26	12.10
24	0.14	2.42	3.30	2.96	7.96	4.01	37.07	14.30	11.56	9.50	2.83	14.09	8.80	13.08

Table 26: *Omnilingual 300M* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	22.31	22.58	37.10	49.15	67.91	39.79	55.91	27.41	58.43	41.06	24.04	49.80	41.29	43.30
1	32.69	36.82	45.63	53.96	64.37	65.65	57.20	44.37	54.07	35.48	35.62	55.38	48.44	50.38
2	28.08	34.03	41.95	52.03	65.21	85.07	55.63	37.23	56.93	37.78	23.78	56.14	47.82	54.20
3	29.85	34.75	42.26	53.73	59.44	71.82	53.52	30.47	55.63	38.40	28.10	56.74	46.22	52.08
4	26.81	29.91	37.12	49.65	48.37	66.51	55.79	47.43	56.49	34.99	21.20	51.01	43.77	51.52
5	44.23	53.15	49.74	54.04	46.04	55.27	57.92	54.58	56.22	42.43	32.98	58.14	50.39	58.95
6	19.63	26.19	30.96	44.57	34.60	52.11	56.05	35.61	52.48	27.86	23.62	51.75	37.95	46.61
7	24.93	32.59	26.34	29.09	36.68	31.57	55.14	49.02	49.85	31.40	30.49	37.04	36.18	37.52
8	29.03	37.69	29.35	31.09	28.82	32.07	47.79	49.92	52.56	26.54	32.54	34.33	35.98	37.98
9	39.98	50.70	38.44	39.91	36.88	25.05	44.35	55.38	49.35	43.92	41.13	34.31	41.62	43.73
10	16.98	20.35	13.80	20.60	16.37	7.29	36.53	50.07	38.68	39.58	30.10	15.21	25.46	21.98
11	15.99	19.45	12.75	21.24	13.74	11.01	29.19	46.89	32.08	40.24	29.91	13.09	23.80	19.64
12	14.55	17.36	10.99	18.82	10.13	6.15	27.10	44.22	29.99	38.77	28.01	12.39	21.54	17.52
13	27.03	30.94	20.68	25.90	19.13	12.45	34.23	53.50	35.66	45.02	43.40	14.40	30.19	23.82
14	13.20	14.72	10.24	15.07	10.82	3.67	24.94	48.34	26.12	33.67	28.98	12.38	20.18	15.71
15	9.46	10.78	6.51	9.58	8.40	1.55	24.31	43.97	18.23	25.08	25.57	11.54	16.25	13.23
16	2.88	3.62	3.78	7.30	5.94	0.17	20.82	21.06	18.40	21.76	9.13	11.37	10.52	12.20
17	6.01	5.83	4.34	9.43	6.87	0.67	22.96	44.24	20.44	19.33	16.86	11.22	14.02	12.16
18	5.63	4.69	3.51	6.21	5.32	0.39	20.44	45.95	15.58	24.12	28.22	11.38	14.29	11.59
19	0.60	1.83	3.12	3.97	5.94	0.63	19.99	28.64	15.44	18.14	3.60	11.76	9.47	11.62
20	0.42	1.82	3.11	4.17	6.15	0.43	20.00	26.86	16.79	15.88	3.72	12.44	9.32	11.83
21	0.38	1.86	3.01	4.09	6.06	0.43	20.16	20.66	16.81	15.57	3.11	14.68	8.90	11.88
22	0.36	1.83	3.02	4.20	6.09	0.43	20.23	19.37	16.77	15.73	3.07	14.37	8.79	11.90
23	0.35	1.82	3.03	4.17	6.09	0.43	20.33	18.82	16.73	15.59	3.06	13.92	8.70	11.94
24	0.16	1.31	3.13	6.11	5.96	0.48	20.22	16.54	16.80	15.34	2.81	12.13	8.42	11.89
Aligned Fine-tuning														
0	56.52	58.48	52.15	51.19	40.24	66.38	51.10	63.29	42.83	37.80	51.07	49.25	51.69	51.99
1	23.89	26.92	33.10	51.42	41.87	57.16	54.90	48.47	53.41	24.06	27.40	48.49	40.92	40.54
2	17.62	20.47	27.81	50.02	27.26	43.03	65.32	23.47	52.62	35.82	15.98	45.38	35.40	40.60
3	18.88	22.33	29.68	45.24	26.04	47.60	68.68	29.53	46.66	37.00	18.34	48.51	36.54	44.62
4	11.89	16.47	19.79	32.11	12.26	15.45	56.51	20.23	45.84	35.26	15.96	31.64	26.12	29.76
5	8.66	12.61	15.72	20.44	12.27	1.47	40.32	10.20	26.48	41.74	11.38	23.59	18.74	26.51
6	1.77	5.95	6.72	5.68	8.75	2.57	29.13	11.66	26.47	28.18	8.06	18.98	12.83	20.01
7	0.44	4.40	4.16	2.88	8.10	1.68	25.24	12.87	22.45	23.30	4.48	15.71	10.48	16.83
8	0.38	4.53	3.53	2.35	6.18	0.54	22.40	14.60	18.93	23.22	3.46	14.26	9.53	14.76
9	0.27	4.64	3.41	2.16	6.20	0.43	22.17	13.50	19.34	22.11	3.18	13.96	9.28	14.37
10	0.19	4.00	3.38	1.90	5.96	0.30	21.97	12.18	18.82	20.27	3.27	13.83	8.84	13.96
11	0.15	3.60	3.36	1.86	5.92	0.30	21.76	12.03	17.86	19.37	3.44	13.85	8.63	13.80
12	0.12	3.08	3.42	1.78	5.91	0.39	21.61	11.53	17.96	17.87	3.88	13.83	8.45	13.59
13	0.08	2.66	3.38	1.77	5.76	0.71	21.45	11.13	17.33	16.26	3.84	13.81	8.18	13.38
14	0.09	2.60	3.38	1.74	5.71	0.97	21.22	11.78	17.05	15.79	3.81	13.80	8.16	13.23
15	0.10	2.63	3.30	1.82	5.62	1.10	21.22	11.26	16.25	15.22	3.21	13.81	7.96	13.15
16	0.10	2.53	3.38	1.89	5.58	1.01	21.33	10.86	16.26	14.48	3.69	13.69	7.90	13.15
17	0.10	2.55	3.40	1.89	5.41	0.84	21.08	10.98	15.73	15.16	3.02	13.63	7.81	13.12
18	0.14	2.77	3.38	1.97	5.41	0.97	21.02	12.03	16.27	15.23	2.95	13.71	7.99	13.11
19	0.11	2.73	3.26	1.93	5.20	0.93	20.81	12.46	16.07	14.38	2.79	13.83	7.87	13.05
20	0.15	2.73	3.14	1.78	5.40	0.80	20.75	13.26	16.82	13.37	2.83	13.92	7.91	13.20
21	0.15	2.66	3.09	1.81	5.30	0.80	20.89	13.11	16.15	13.60	2.81	13.89	7.85	13.19
22	0.12	2.53	3.09	1.85	5.28	0.76	20.89	11.36	16.16	13.91	2.85	13.92	7.73	13.18
23	0.12	2.54	3.03	1.82	5.34	0.80	20.89	11.64	16.38	14.13	2.88	13.95	7.79	13.22
24	0.10	1.96	3.01	1.86	5.40	0.76	20.83	9.67	16.89	12.98	2.88	14.00	7.53	13.22

Table 27: *XLSR-128 1B* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	22.71	20.35	30.82	45.52	56.98	73.67	41.45	24.48	57.00	33.54	21.38	46.16	39.51	44.55
1	57.35	58.13	52.15	47.07	42.68	51.47	42.90	51.51	55.79	44.52	46.47	47.75	49.82	51.98
2	71.09	69.63	61.94	51.19	42.97	37.10	42.80	58.80	49.27	60.47	78.68	46.82	55.90	46.31
3	65.23	62.74	59.74	55.73	46.36	37.10	47.48	49.80	53.71	51.75	72.25	52.66	54.55	50.20
4	72.41	66.52	65.38	55.93	49.91	16.92	47.00	54.07	54.07	50.96	79.72	53.70	55.55	50.18
5	76.38	71.50	67.85	60.74	38.23	15.54	42.96	48.22	52.90	54.87	72.29	55.05	54.71	51.75
6	75.66	69.15	65.30	65.92	44.60	12.67	48.55	44.77	59.49	51.32	77.59	60.58	56.30	51.48
7	80.75	76.60	70.97	61.87	49.92	31.89	52.81	55.63	63.59	49.62	78.66	61.27	61.13	56.67
8	68.89	65.78	64.71	56.61	54.98	33.38	52.52	51.66	62.25	39.55	70.23	61.32	56.82	52.26
9	68.29	66.41	63.71	53.20	46.32	45.10	48.98	46.50	59.00	43.40	61.87	54.67	54.79	54.31
10	49.23	51.67	59.16	49.56	42.47	50.50	52.76	32.57	48.36	47.67	60.34	52.06	49.70	49.36
11	29.67	36.21	35.42	36.05	39.62	38.20	46.00	33.50	48.01	49.99	44.24	30.84	38.98	32.67
12	44.77	48.75	61.10	56.68	45.42	47.52	56.51	34.44	50.11	60.40	57.15	65.02	52.32	56.67
13	75.40	74.46	85.24	82.17	68.29	72.23	68.55	49.00	64.71	68.29	64.75	81.29	71.20	76.39
14	65.78	66.99	78.89	74.67	69.59	69.87	66.00	46.78	56.79	65.24	57.64	81.32	66.63	73.57
15	65.89	68.53	80.11	81.41	71.81	76.59	67.31	47.94	62.70	61.72	49.84	81.63	67.96	75.60
16	54.40	59.84	74.28	78.41	64.45	75.27	63.16	43.17	58.43	56.23	44.75	79.06	62.62	70.77
17	71.68	73.73	83.92	86.22	77.79	85.28	72.67	45.43	67.34	64.97	55.58	83.93	72.38	78.46
18	74.86	76.85	85.43	91.71	83.57	87.33	78.15	45.70	72.05	69.33	48.50	86.17	74.97	82.29
19	82.91	84.48	90.96	95.64	87.87	91.17	80.82	53.37	77.22	72.62	61.10	86.74	80.41	85.45
20	80.09	80.75	89.78	94.89	88.24	92.23	80.16	54.44	74.99	67.69	53.14	87.13	78.63	85.17
21	77.43	78.79	87.53	92.58	86.04	89.66	78.34	49.27	75.28	67.93	50.76	86.86	76.71	83.85
22	77.14	78.39	87.64	95.34	89.64	94.00	80.58	49.80	75.67	69.61	45.57	86.87	77.52	84.38
23	58.35	63.17	78.83	89.63	81.97	92.75	76.22	41.73	67.02	61.70	34.44	84.51	69.19	79.33
24	64.81	68.95	80.88	90.38	82.05	94.30	77.70	37.74	63.38	58.96	35.67	86.19	70.08	80.49
25	58.40	62.36	75.29	85.79	77.96	87.33	73.46	34.85	60.01	54.32	30.45	86.05	65.52	78.12
26	32.02	37.13	53.25	58.19	49.91	77.69	60.86	28.22	42.71	36.56	17.11	76.53	47.52	67.45
27	55.55	61.58	72.00	69.56	65.11	78.98	67.75	34.30	56.86	46.76	29.94	81.63	60.00	75.02
28	84.33	86.87	89.63	87.02	83.51	87.01	77.33	47.67	68.17	69.41	57.11	86.22	77.02	85.29
29	76.68	81.30	85.97	78.22	75.95	83.58	76.65	41.99	61.78	65.54	51.21	86.28	72.10	83.26
30	82.81	86.39	88.14	75.77	79.96	82.41	77.31	44.22	61.24	69.45	54.97	87.09	74.15	84.81
31	75.04	80.25	84.30	76.30	75.45	83.30	77.28	36.83	54.58	60.62	45.63	87.51	69.76	82.85
32	72.06	75.23	84.21	80.01	77.12	86.08	75.75	46.10	59.41	65.39	50.95	84.81	71.43	81.39
33	11.33	13.51	23.85	18.89	33.15	57.55	43.84	24.77	17.82	28.96	13.73	35.41	26.90	40.67
34	3.24	4.03	13.80	12.84	25.63	53.09	36.25	20.27	17.27	20.79	8.96	17.94	19.51	29.10
35	0.73	1.32	4.63	5.30	13.75	24.13	24.48	16.03	10.14	14.52	5.78	13.64	11.20	17.87
36	0.36	1.17	3.09	2.65	9.96	11.93	20.73	12.46	11.08	12.30	4.94	13.47	8.68	15.83
37	0.20	1.46	1.85	1.32	6.07	3.88	15.38	10.03	11.08	9.85	2.83	11.83	6.31	12.92
38	0.16	1.24	2.02	1.63	6.90	3.50	15.96	9.53	11.18	10.53	3.69	12.22	6.55	13.11
39	0.09	1.65	1.54	1.44	5.92	1.99	15.45	6.89	10.17	8.86	2.02	12.24	5.69	12.49
40	0.06	1.85	1.52	1.71	5.82	1.40	15.53	6.38	10.04	10.14	1.81	12.12	5.70	12.32
41	0.08	2.19	1.54	1.63	5.94	1.23	15.71	6.11	10.09	10.58	1.30	12.20	5.72	12.35
42	0.10	2.41	1.50	1.63	5.86	1.36	15.63	5.96	10.17	11.28	1.44	12.18	5.79	12.24
43	0.11	2.36	1.58	1.69	5.82	1.60	15.76	5.15	10.11	10.43	1.46	12.29	5.70	12.28
44	0.11	2.23	1.55	1.71	5.78	1.64	15.84	4.90	10.21	9.74	1.37	12.57	5.64	12.25
45	0.12	2.25	1.56	1.70	5.73	1.55	15.87	5.30	10.44	9.69	1.32	12.73	5.69	12.26
46	0.10	2.04	1.56	1.78	5.74	1.55	15.82	5.56	10.48	8.96	1.32	13.01	5.66	12.19
47	0.11	1.59	1.62	1.93	5.69	1.55	15.67	5.70	10.38	7.36	1.35	13.23	5.51	12.16
48	0.14	1.38	1.68	1.98	5.73	1.73	15.69	4.63	10.58	6.72	1.42	12.62	5.36	12.37
Aligned Fine-tuning														
0	32.51	34.49	35.79	49.11	50.98	60.55	41.29	28.21	38.82	45.39	29.61	45.67	41.04	46.64
1	33.41	41.02	37.56	52.40	53.49	35.99	48.56	30.30	37.28	36.49	21.48	52.22	40.02	49.78
2	36.03	38.51	35.85	52.48	37.96	35.61	48.18	37.48	49.71	30.40	24.89	49.43	39.71	46.46
3	22.82	27.08	29.97	46.31	47.56	40.33	48.24	29.53	48.13	32.48	16.44	45.16	36.17	45.89
4	16.39	20.30	24.47	47.67	43.12	61.05	47.84	27.56	40.48	24.47	14.33	40.03	33.98	39.11
5	11.51	14.23	22.77	39.57	45.86	70.26	51.23	25.30	46.41	24.24	13.75	36.88	33.50	39.52
6	6.76	10.46	17.60	27.34	42.75	57.60	45.13	21.73	50.48	20.84	13.56	29.54	28.65	33.96
7	3.85	8.44	10.45	12.96	28.14	50.26	36.97	16.96	49.26	19.49	13.09	25.71	22.97	29.34
8	1.47	5.17	8.14	6.78	19.53	38.91	33.65	13.50	36.65	17.12	11.10	21.12	17.76	24.89
9	0.60	3.80	5.95	3.79	15.08	32.61	28.94	10.88	32.11	15.99	9.01	19.53	14.86	22.81
10	0.24	2.97	4.55	2.72	11.34	24.34	26.03	10.60	28.34	13.98	7.10	18.28	12.54	20.88
11	0.08	2.07	2.76	1.44	7.55	10.51	21.81	9.40	16.58	11.61	6.23	15.09	8.76	15.63
12	0.07	2.03	2.15	1.33	6.84	6.84	19.72	9.93	13.92	10.44	5.34	13.94	7.71	13.79
13	0.07	2.01	1.78	1.22	6.15	3.13	18.20	9.68	13.03	9.43	4.23	13.26	6.85	12.85
14	0.05	2.09	1.59	1.17	5.70	3.32	17.68	8.60	12.85	9.23	3.72	13.13	6.60	12.59
15	0.05	2.22	1.43	1.10	5.42	2.48	17.36	8.49	12.27	9.00	3.67	13.01	6.37	12.31
16	0.05	2.15	1.39	1.09	5.25	2.40	17.10	8.59	11.69	8.42	3.60	12.92	6.22	12.19
17	0.05	2.10	1.40	1.06	5.15	2.24	16.99	8.60	11.63	8.13	3.64	12.94	6.16	12.12
18	0.05	2.08	1.42	1.10	4.95	2.11	16.74	8.59	10.88	8.02	3.55	12.84	6.03	11.92
19	0.04	2.02	1.41	1.10	4.83	2.33	16.56	8.22	10.25	7.56	3.67	12.89	5.91	11.84
20	0.04	1.98	1.52	1.17	4.71	2.48	16.33	8.22	9.39	6.93	3.53	12.94	5.77	11.77
21	0.05	1.98	1.58	1.21	4.65	2.65	16.22	7.94	8.77	6.45	3.30	13.07	5.66	11.72
22	0.05	1.85	1.63	1.24	4.64	3.04	16.14	7.54	8.18	6.02	3.23	13.17	5.56	11.70
23	0.05	1.83	1.63	1.29	4.66	3.00	16.01	7.29	8.07	5.78	3.16	13.40	5.51	11.66
24	0.05	1.80	1.69	1.33	4.59	3.26	15.89	7.03	7.84	5.51	3.11	13.37	5.46	11.59
25	0.06	1.82	1.72	1.37	4.65	3.17	15.87	6.86	7.68	5.38	2.95	13.44	5.41	11.58
26	0.07	1.82	1.75	1.44	4.64	3.26	15.82	6.89	7.53	5.30	2.90	13.60	5.42	11.55
27	0.07	1.82	1.73	1.44	4.61	3.32	15.89	6.89	7.38	5.30	2.79	13.76	5.42	11.51
28	0.07	1.78	1.78	1.44	4.64	3.45	15.84	6.91	7.42	5.37	2.78	13.66	5.43	11.44
29	0.07	1.78	1.76	1.48	4.71	3.41	15.80	6.36	7.33	5.33	2.74	13.80	5.38	11.46
30	0.07	1.82	1.75	1.52	4.71	3.37	15.94	6.11	7.33	5.31	2.58	13.59	5.34	11.46
31	0.07	1.84	1.79	1.59	4.71	3.58	15.92	5.70	7.36	5.20	2.46	13.52	5.31	11.48
32	0.07	1.85	1.79	1.60	4.70	3.93	15.91	5.55	7.19	5.32	2.35	13.31	5.30	11.46
33	0.07	1.78	1.84	1.71	4.69	3.63	15.94	5.68	7.36	5.09	2.25	13.44	5.29	11.39
34	0.08	1.70	1.83	1.78	4.62	3.71	16.06	5.15	7.46	5.09	2.27	13.67	5.28	11.34
35	0.07	1.73	1.85	1.79	4.59	4.06	16.00	4.50	7.50	5.19	2.11	13.97	5.28	11.31
36	0.07	1.67	1.93	1.97	4.68	3.88	16.12	4.39	7.36	5.17	1.98	13.55	5.23	11.25
37	0.08	1.70	2.01	2.01	4.67	4.14	16.09	4.00	7.46	5.51	1.98	13.34	5.25	11.18
38	0.07	1.69	2.02	2.08	4.61	4.60	16.11	3.97	7.31	5.31	1.93	13.25	5.24	11.15
39	0.07	1.65	1.96	2.20	4.76	4.90	16.70	4.10	7.36	5.09	2.09	14.35	5.44	11.27
40	0.07	1.62	1.99	2.27	4.76	5.11	16.90	3.60	7.43	5.16	1.58	14.97	5.46	11.36
41	0.07	1.86	1.81	1.93	4.73	4.47	16.36	3.44	7.61	5.20	1.44	14.88	5.32	11.46
42	0.08	1.98	1.69	1.93	4.72	4.47	16.46	3.57	7.56	5.65	1.35	15.47	5.41	11.52
43	0.08	2.04	1.66	1.89	4.80	4.90	17.10	4.50	7.52	5.76	1.28	16.97	5.71	11.88
44	0.10	2.07	1.67	1.93	4.70	4.90	16.62	4.50	7.52	5.93	1.30	16.27	5.63	11.65
45	0.10	2.15	1.66	1.89	4.77	4.73	16.56	4.50	7.53	6.03	1.33	16.87	5.68	11.63
46	0.10	2.08	1.66	1.94	4.68	4.73	16.48	4.37	7.67	6.03	1.35	18.11	5.76	11.85
47	0.09	2.00	1.66	1.97	4.67	4.81	16.64	4.10	7.72	6.00	1.37	19.46	5.88	12.39
48	0.09	2.65	1.59	1.58	4.66	4.77	16.08	3.84	7.66	5.97	1.49	14.47	5.40	11.56

Table 28: *MMS 1B* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	36.14	32.46	36.64	42.56	43.53	53.22	43.88	45.43	57.64	44.40	40.29	40.68	43.07	42.36
1	69.29	66.06	63.33	52.22	41.97	64.57	50.03	66.38	47.17	52.17	63.08	52.69	57.41	55.94
2	69.75	65.73	63.90	51.35	33.78	53.54	53.06	58.80	58.68	45.46	47.65	51.36	54.42	56.68
3	67.18	62.36	57.62	54.49	33.62	58.39	51.90	49.15	43.20	44.65	62.42	51.84	53.07	52.00
4	67.86	63.22	64.17	56.84	52.51	39.88	60.31	66.88	58.46	37.10	70.95	60.30	58.21	54.98
5	64.04	64.41	55.96	52.67	44.58	23.15	53.30	66.89	57.91	35.17	47.68	53.34	51.59	54.12
6	65.87	64.71	53.86	53.58	48.59	21.11	58.79	60.80	58.66	41.87	43.52	54.00	52.11	53.81
7	50.58	53.08	43.04	54.64	39.18	3.54	52.10	56.16	54.58	42.39	36.00	42.13	43.95	42.89
8	39.51	43.31	39.31	48.35	25.51	8.26	47.56	58.01	52.32	33.27	41.71	35.19	39.36	35.95
9	57.84	56.93	56.21	49.72	31.39	4.38	58.27	71.23	51.75	43.54	54.76	44.38	48.37	46.22
10	78.37	73.28	78.32	62.82	50.79	26.28	63.65	81.33	63.52	55.92	70.90	59.98	63.76	59.83
11	75.43	74.01	76.07	62.18	50.36	26.15	68.46	77.36	60.93	53.55	59.36	57.88	61.81	62.30
12	36.17	38.07	40.64	39.38	34.75	19.40	52.60	54.29	44.25	29.94	31.05	32.21	37.73	35.37
13	24.47	25.79	23.72	26.61	26.12	16.34	41.50	49.67	41.13	32.05	29.21	17.49	29.51	21.96
14	13.23	15.03	13.98	19.31	23.44	9.37	35.67	36.79	32.21	29.68	18.67	13.71	21.76	17.12
15	20.73	20.13	18.03	20.07	28.86	12.06	40.52	51.66	28.70	28.81	34.18	14.79	26.55	18.76
16	17.53	17.54	12.04	12.84	23.76	6.58	33.82	53.50	22.76	29.18	36.86	11.59	23.17	15.30
17	17.58	16.38	10.44	12.61	19.43	6.00	29.92	52.29	20.65	31.88	36.67	10.56	22.03	14.14
18	15.57	16.34	9.47	13.37	17.51	7.94	28.31	48.64	17.21	31.25	26.10	10.73	20.20	14.15
19	20.94	24.26	15.95	20.41	26.62	14.48	46.32	49.67	20.27	40.70	23.46	14.39	26.46	21.93
20	23.69	26.18	18.43	20.82	26.55	16.57	50.54	41.46	21.76	38.31	22.29	21.97	27.38	26.74
21	14.53	17.72	13.19	15.75	21.28	16.08	46.38	27.43	18.95	36.82	15.51	15.87	21.63	21.90
22	8.21	10.18	6.85	10.90	15.69	9.50	29.77	26.76	18.05	33.20	12.74	10.79	16.05	13.72
23	9.57	10.85	8.17	10.82	17.76	5.74	30.20	27.14	17.06	35.41	13.86	11.32	16.49	15.21
24	7.87	9.33	4.59	4.88	7.87	2.24	19.78	27.67	16.39	29.54	13.14	10.54	12.82	11.65
25	10.28	11.18	5.48	7.99	10.35	2.44	20.88	31.66	18.66	29.62	14.91	10.78	14.52	12.41
26	7.20	9.30	5.94	9.42	11.80	4.64	24.12	28.37	27.73	24.36	9.24	12.38	14.54	13.52
27	3.89	5.39	3.90	7.16	7.53	5.48	19.46	22.14	18.37	23.76	6.52	10.74	11.20	11.68
28	5.03	5.58	4.12	5.07	7.27	1.55	18.64	21.03	18.61	23.75	8.01	10.78	10.79	11.30
29	4.30	4.83	4.33	6.43	7.37	1.64	19.41	18.14	20.13	25.33	7.22	10.70	10.82	11.39
30	2.69	3.64	4.21	7.68	7.29	2.74	20.40	15.10	19.55	19.39	6.51	10.44	9.97	11.36
31	0.73	2.85	2.99	4.85	6.51	2.44	18.73	12.97	18.41	16.50	4.69	10.16	8.49	10.85
32	0.61	2.45	3.02	3.70	6.70	2.65	18.67	11.26	19.71	18.96	4.46	10.27	8.54	10.93
33	0.48	2.09	2.84	3.41	6.29	2.70	18.66	10.73	17.07	16.61	3.81	10.34	7.92	10.93
34	0.18	1.99	2.32	2.62	5.24	0.97	17.74	8.34	13.80	13.19	3.30	10.09	6.65	10.73
35	0.17	2.33	2.38	2.34	5.28	0.93	17.71	7.14	13.96	12.27	1.84	10.05	6.37	10.69
36	0.14	2.35	2.16	1.96	5.07	0.80	17.30	5.30	13.18	13.82	1.92	10.14	6.18	10.85
37	0.13	2.46	2.11	1.89	4.84	0.84	17.25	5.85	13.03	12.66	2.04	10.20	6.11	10.88
38	0.16	2.50	2.00	1.74	4.77	0.80	17.43	6.38	12.76	12.99	1.97	10.16	6.14	10.90
39	0.18	2.58	2.05	1.90	4.74	0.84	17.70	6.50	12.62	13.26	1.86	10.34	6.21	11.08
40	0.16	2.25	2.02	1.86	4.70	0.93	17.64	6.89	12.12	13.42	1.95	10.37	6.19	11.09
41	0.11	1.91	2.01	2.01	4.74	1.01	17.81	6.63	11.54	13.19	1.88	10.48	6.11	11.23
42	0.11	1.77	2.00	2.20	4.77	1.01	17.80	6.59	11.38	12.97	1.90	10.39	6.07	11.25
43	0.12	1.61	2.04	2.39	4.79	1.01	17.80	6.50	11.33	12.97	2.07	10.33	6.08	11.31
44	0.14	1.45	2.07	2.55	4.77	1.06	17.73	6.50	11.30	12.92	2.09	10.20	6.06	11.34
45	0.15	1.32	2.11	2.79	4.79	1.10	17.64	6.63	11.27	12.91	2.12	9.99	6.07	11.37
46	0.17	1.25	2.14	2.99	4.82	1.10	17.60	6.61	11.27	12.91	2.16	9.85	6.07	11.41
47	0.18	1.15	2.16	3.22	4.83	1.06	17.54	6.51	11.30	12.99	2.14	9.75	6.07	11.45
48	0.21	1.08	2.08	3.34	4.92	1.14	18.20	6.23	11.31	12.65	2.30	10.24	6.14	11.31
Aligned Fine-tuning														
0	35.30	34.18	39.22	49.34	50.22	77.88	49.37	41.46	47.02	64.05	32.89	40.65	46.80	48.14
1	34.59	35.10	37.20	54.71	57.80	61.22	56.29	34.30	46.02	59.42	52.69	46.60	48.00	42.79
2	23.70	25.11	32.74	56.53	49.80	48.94	60.76	21.33	33.85	52.97	33.09	48.59	40.62	40.51
3	23.02	25.95	30.22	48.13	48.08	51.77	51.60	21.59	39.36	57.92	22.64	35.77	38.00	39.22
4	13.84	15.75	22.64	37.21	43.04	52.20	48.28	17.61	45.27	46.89	16.68	31.15	32.55	36.45
5	5.09	8.11	10.30	17.46	25.14	43.96	36.46	13.24	35.27	33.34	12.98	23.38	22.06	25.13
6	1.44	5.17	5.49	7.72	14.82	34.94	29.64	9.52	37.00	21.55	10.80	19.63	16.48	21.59
7	0.21	3.56	3.26	3.94	9.90	13.64	27.50	7.84	28.23	21.46	5.50	17.46	11.88	18.82
8	0.10	3.00	2.20	2.61	9.20	3.26	25.47	7.56	25.70	20.39	3.42	16.08	9.91	16.69
9	0.07	2.54	1.88	1.93	8.37	2.48	23.77	5.95	22.01	20.44	2.63	14.87	8.91	15.13
10	0.10	2.56	1.75	1.66	7.72	2.33	23.24	7.29	18.81	20.00	2.81	14.24	8.54	14.36
11	0.05	2.00	1.73	1.63	7.57	1.40	23.15	8.60	17.18	18.84	2.93	13.90	8.25	13.83
12	0.07	1.82	1.72	1.63	7.29	1.73	22.95	8.60	16.67	18.76	2.79	13.66	8.14	13.57
13	0.06	1.58	1.79	1.74	7.17	1.77	22.80	8.99	15.34	18.14	2.44	13.49	7.94	13.37
14	0.07	1.44	1.86	1.93	7.23	1.86	22.76	8.36	14.88	17.90	2.51	13.40	7.85	13.24
15	0.07	1.38	1.90	1.97	7.18	1.99	22.65	9.00	14.75	17.82	2.63	13.35	7.89	13.19
16	0.07	1.28	1.93	1.93	7.29	2.29	22.49	9.00	14.18	17.09	2.35	13.24	7.76	13.09
17	0.08	1.30	1.91	1.89	7.19	2.35	22.47	8.85	13.96	17.00	2.19	13.20	7.70	13.02
18	0.08	1.17	1.98	2.09	7.18	2.78	22.35	8.72	13.09	16.17	2.30	13.00	7.58	12.84
19	0.07	1.11	1.98	2.27	7.19	3.04	22.30	8.34	12.92	15.57	2.51	12.98	7.52	12.81
20	0.08	1.07	2.02	2.38	7.19	3.09	22.29	8.74	12.56	15.14	2.44	12.91	7.49	12.71
21	0.08	1.03	2.07	2.50	7.13	3.04	22.24	8.74	12.41	14.50	2.36	12.84	7.41	12.58
22	0.08	1.03	2.13	2.54	7.05	3.04	22.27	8.49	12.56	14.29	2.42	12.84	7.39	12.58
23	0.08	1.01	2.13	2.57	7.05	3.13	22.26	8.74	12.66	14.35	2.44	12.79	7.44	12.54
24	0.10	0.98	2.20	2.69	7.03	3.13	22.23	8.74	12.51	14.33	2.44	12.75	7.43	12.46
25	0.10	0.97	2.20	2.73	7.07	3.26	22.42	8.77	12.43	14.35	2.51	12.76	7.46	12.46
26	0.10	0.98	2.21	2.69	7.04	3.13	22.35	8.62	12.34	14.37	2.55	12.74	7.43	12.43
27	0.10	0.99	2.21	2.84	7.06	3.13	22.48	9.15	12.42	14.24	2.53	12.77	7.49	12.36
28	0.10	0.98	2.18	2.80	7.07	3.17	22.37	9.17	12.40	14.18	2.49	12.76	7.47	12.34
29	0.10	0.96	2.10	2.73	7.03	3.13	22.77	9.00	12.44	13.84	2.67	13.04	7.48	12.33
30	0.10	0.95	2.09	2.76	7.04	2.91	22.67	9.68	12.58	13.52	2.72	12.85	7.49	12.35
31	0.11	0.94	2.15	2.80	6.88	3.09	22.98	9.55	12.59	13.41	2.62	13.17	7.52	12.31
32	0.10	0.95	2.15	2.76	6.90	2.74	23.27	9.15	12.60	12.88	2.67	13.19	7.45	12.31
33	0.11	0.94	2.16	2.84	6.89	2.48	24.59	9.53	12.54	13.01	2.62	13.66	7.61	12.22
34	0.11	0.94	2.21	2.95	6.81	2.61	24.55	8.59	12.61	13.13	2.36	13.67	7.55	12.12
35	0.11	0.94	2.19	3.03	6.65	2.40	24.06	7.54	12.41	12.83	2.41	13.52	7.34	12.09
36	0.11	0.88	2.19	3.06	6.54	1.99	24.15	7.04	12.18	11.92	2.37	13.48	7.16	12.10
37	0.11	0.89	2.21	3.07	6.62	1.77	24.87	7.03	12.48	11.66	2.49	13.68	7.24	12.08
38	0.11	0.90	2.18	2.96	6.59	2.03	24.57	7.54	12.15	11.66	2.32	13.40	7.20	11.94
39	0.12	0.91	2.23	3.07	6.62	1.86	24.62	7.52	12.23	11.51	2.32	13.34	7.20	11.93
40	0.12	0.88	2.21	3.14	6.62	1.86	25.41	7.43	12.06	11.20	2.28	13.62	7.24	11.99
41	0.11	0.89	2.18	3.10	6.57	1.73	25.06	8.72	12.09	11.29	2.19	13.37	7.28	11.98
42	0.12	0.84	2.21	3.10	6.56	1.73	25.04	8.60	11.78	11.18	2.20	13.37	7.23	11.98
43	0.12	0.86	2.21	3.29	6.66	1.90	25.08	7.29	11.85	11.69	2.18	13.36	7.21	12.00
44	0.12	0.87	2.19	3.23	6.65	1.81	25.47	7.14	11.72	11.67	2.18	13.55	7.22	12.05
45	0.12	0.84	2.19	3.29	6.64	1.86	25.34	7.03	11.76	11.77	2.18	13.51	7.21	12.04
46	0.12	0.86	2.21	3.41	6.63	1.94	25.16	6.89	11.83	11.78	2.20	13.43	7.21	12.04
47	0.13	0.85	2.23	3.53	6.67	1.99	25.35	6.63	11.91	11.84	2.32	13.48	7.24	12.03
48	0.12	0.88	2.30	3.60	6.72	2.07	24.17	5.95	11.84	11.27	2.81	13.05	7.06	12.02

Table 29: *Omnilingual 1B* under *standard* and *aligned* fine-tuning: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Standard Fine-tuning														
0	58.19	53.70	55.30	50.66	45.28	57.98	50.74	63.72	55.40	62.28	48.65	48.79	54.23	52.69
1	70.24	68.95	57.66	48.01	37.80	29.50	48.27	64.09	59.97	56.82	53.55	43.50	53.20	51.97
2	71.99	71.23	62.09	46.01	46.58	49.46	45.54	59.07	62.74	67.22	44.06	44.89	55.91	57.65
3	52.56	52.44	46.28	46.16	48.38	54.38	41.55	56.30	56.77	56.21	44.08	42.94	49.84	49.41
4	60.97	60.08	50.49	43.24	32.90	43.72	39.96	58.04	48.68	59.21	62.96	45.05	50.44	48.69
5	61.02	61.88	52.73	42.27	39.00	39.79	41.86	59.73	50.65	56.37	51.18	49.24	50.48	53.68
6	47.43	48.69	38.31	39.72	22.48	47.48	33.70	44.77	47.48	59.03	60.80	40.53	44.20	40.80
7	66.89	66.06	59.28	44.57	31.93	40.81	35.03	48.87	46.09	62.28	75.73	44.30	51.82	49.85
8	46.23	40.84	32.76	34.15	16.83	42.49	26.81	50.20	33.28	55.94	70.81	32.09	40.20	33.43
9	72.60	69.12	68.19	53.01	35.58	77.30	45.99	60.27	36.44	63.90	72.57	47.73	58.56	53.15
10	38.14	34.72	25.84	27.60	18.53	51.14	24.49	44.50	37.02	51.92	51.37	29.19	36.21	29.26
11	42.57	37.66	28.92	29.51	20.35	59.75	24.78	46.50	38.15	54.61	48.42	27.56	38.23	31.07
12	43.36	38.36	26.50	33.28	21.25	62.73	21.57	44.48	39.87	47.46	65.53	22.27	38.89	25.18
13	45.34	43.59	30.07	42.33	25.71	79.24	25.89	35.63	40.91	46.69	49.46	34.46	41.61	34.74
14	30.77	30.06	16.95	17.38	16.04	53.97	19.54	39.34	29.09	37.21	44.28	16.27	29.24	20.53
15	44.88	44.56	26.99	23.78	22.00	62.11	22.97	39.34	30.75	43.37	42.71	24.77	35.69	30.41
16	24.43	27.01	13.13	16.02	17.01	44.73	22.19	24.50	27.26	30.78	30.75	21.67	24.96	20.71
17	43.76	46.36	27.80	19.87	27.38	53.50	25.56	34.97	35.73	35.34	46.54	22.89	34.97	27.11
18	49.91	55.97	47.00	40.21	39.35	61.22	39.20	25.17	38.98	38.55	35.30	47.61	43.21	46.82
19	47.53	54.47	44.76	36.02	40.40	63.34	42.15	20.00	41.08	37.28	33.88	50.33	42.60	47.36
20	37.23	42.95	32.56	40.36	35.85	64.13	38.02	23.97	40.29	37.41	26.99	46.15	38.83	41.19
21	34.13	40.11	30.10	36.43	27.36	66.29	32.87	23.31	37.24	35.55	21.41	36.14	35.08	37.72
22	54.78	57.13	43.45	45.39	35.01	61.22	33.90	37.09	43.69	41.33	43.22	38.93	44.59	40.72
23	64.36	66.80	66.30	61.38	47.61	77.30	45.17	31.38	53.00	51.89	36.41	57.26	54.91	60.03
24	69.42	71.65	72.49	52.40	44.08	79.59	43.61	41.33	51.80	53.84	49.82	56.72	57.23	61.41
25	68.31	72.02	72.70	50.59	43.82	72.36	46.58	39.34	43.80	55.26	53.01	68.86	57.22	66.76
26	20.01	24.15	15.24	17.12	13.14	41.74	24.44	26.06	28.07	40.32	33.65	18.42	25.20	21.40
27	47.46	54.20	39.53	27.75	23.92	72.03	29.50	38.94	33.71	42.35	38.09	24.60	39.34	37.18
28	7.59	8.04	4.45	9.92	5.26	16.57	16.33	28.47	22.82	27.75	15.51	11.47	14.52	12.39
29	14.28	14.25	9.94	14.16	9.12	21.69	18.28	34.04	31.34	38.15	33.32	16.47	21.25	16.31
30	5.11	5.70	3.46	7.76	5.60	4.77	16.81	22.41	23.76	25.40	12.10	14.44	12.28	13.12
31	5.70	6.24	3.23	4.89	4.77	2.52	16.45	23.04	19.84	22.79	11.70	11.81	11.08	11.53
32	4.24	5.59	3.83	4.99	6.15	3.84	18.49	24.63	20.84	26.94	10.24	14.46	12.02	13.84
33	2.32	2.60	2.96	4.39	3.86	1.94	14.65	23.17	16.13	18.00	7.83	11.11	9.08	10.54
34	1.98	2.46	2.68	2.91	3.55	1.23	15.01	20.12	14.30	13.50	6.31	11.27	7.94	10.54
35	3.78	3.39	2.27	2.95	3.62	0.93	14.86	26.06	15.61	14.41	10.68	11.10	9.14	10.60
36	0.52	1.96	2.48	2.95	3.84	0.88	14.98	18.79	12.66	10.51	4.38	11.34	7.11	10.52
37	0.12	1.57	2.54	3.07	4.10	0.97	15.09	14.87	12.70	9.86	3.32	11.29	6.62	10.43
38	0.12	1.53	2.58	3.11	4.14	0.97	15.10	13.52	12.89	10.14	3.29	11.42	6.57	10.47
39	0.11	1.45	2.60	3.19	4.19	1.01	15.09	13.37	13.03	10.12	3.46	11.44	6.59	10.49
40	0.10	1.45	2.60	3.14	4.17	1.01	15.01	13.09	13.01	10.25	3.37	11.54	6.56	10.51
41	0.10	1.42	2.61	3.18	4.16	1.01	15.04	13.26	13.01	10.30	3.36	11.78	6.60	10.52
42	0.10	1.39	2.62	3.18	4.15	1.01	15.04	13.01	13.01	10.22	3.35	11.73	6.57	10.54
43	0.10	1.34	2.64	3.18	4.17	1.06	15.09	12.57	12.97	10.11	3.41	11.47	6.51	10.52
44	0.10	1.31	2.64	3.18	4.16	1.06	15.10	12.56	12.99	10.07	3.48	11.48	6.51	10.51
45	0.10	1.30	2.62	3.21	4.16	1.06	15.11	12.33	12.94	9.92	3.51	11.51	6.48	10.52
46	0.09	1.26	2.63	3.19	4.15	1.10	15.11	11.81	12.99	9.87	3.58	11.65	6.45	10.53
47	0.11	1.19	2.65	3.15	4.13	1.23	15.10	12.69	12.89	9.78	3.39	12.03	6.53	10.54
48	0.10	1.16	2.63	3.03	4.22	1.23	15.09	9.80	12.89	9.66	2.63	12.45	6.24	10.61
Aligned Fine-tuning														
0	61.66	63.01	53.82	56.91	48.78	26.46	52.02	61.43	47.72	52.34	72.80	48.97	53.83	46.11
1	27.00	26.76	36.48	56.15	46.62	47.93	51.58	31.79	39.76	36.81	48.91	56.34	42.18	42.22
2	24.32	25.74	31.63	55.09	46.24	60.81	56.33	21.99	45.38	30.48	22.78	52.73	39.46	41.11
3	13.75	14.86	24.89	47.22	40.59	63.12	51.46	23.57	43.75	27.52	14.46	48.07	34.44	38.46
4	9.73	11.46	19.31	40.74	38.97	52.20	49.69	21.20	47.01	25.82	13.84	42.76	31.06	35.00
5	6.65	9.45	14.26	26.73	24.96	25.66	40.70	16.16	43.20	27.82	12.61	26.62	22.90	27.86
6	2.31	6.65	8.55	13.32	11.60	6.84	30.94	17.49	32.50	30.18	10.17	20.68	15.94	23.06
7	1.24	6.30	5.60	9.66	9.23	4.86	26.87	14.57	29.33	29.18	9.52	19.43	13.82	20.20
8	0.68	4.32	4.02	4.32	6.56	1.36	23.27	12.87	24.78	25.82	6.79	15.20	10.83	15.60
9	0.72	4.70	3.95	2.99	5.52	0.35	21.48	10.60	22.37	27.30	5.41	14.19	9.97	14.12
10	0.59	4.82	3.75	2.96	4.97	0.35	20.49	9.27	20.90	25.61	4.53	13.79	9.34	13.51
11	0.54	4.87	3.60	2.80	4.89	0.17	19.69	8.72	20.47	23.97	4.46	13.66	8.99	13.27
12	0.45	4.64	3.43	3.07	4.93	0.17	19.42	9.02	20.74	22.51	4.00	13.58	8.83	13.02
13	0.38	4.37	3.26	2.76	4.76	0.26	18.95	9.17	18.95	21.89	3.85	13.50	8.51	12.72
14	0.30	4.12	3.00	2.69	4.95	0.30	18.61	9.42	17.29	21.41	3.67	13.34	8.26	12.48
15	0.30	4.04	2.90	2.72	4.98	0.30	18.41	9.17	17.12	20.99	3.69	13.30	8.16	12.39
16	0.23	3.69	2.81	2.58	5.03	0.35	18.26	9.39	16.92	20.24	3.95	13.24	8.06	12.32
17	0.22	3.46	2.75	2.54	5.16	0.39	18.11	9.29	16.50	20.21	4.13	13.18	7.99	12.25
18	0.22	3.35	2.70	2.54	5.34	0.58	18.02	9.27	16.53	20.38	4.00	13.15	8.01	12.24
19	0.23	3.47	2.73	2.54	5.72	0.63	17.94	9.12	17.06	21.12	4.01	13.15	8.14	12.27
20	0.23	3.40	2.68	2.57	5.59	0.50	17.84	8.89	16.71	21.25	3.97	13.10	8.06	12.19
21	0.23	3.15	2.76	2.80	5.83	0.58	17.83	8.87	16.78	21.31	4.06	13.08	8.11	12.18
22	0.23	3.22	2.77	2.80	5.85	0.71	17.77	9.27	16.89	21.32	4.02	13.08	8.16	12.17
23	0.24	3.20	2.82	2.73	5.95	0.71	17.80	8.47	17.69	21.74	3.86	13.03	8.19	12.15
24	0.24	3.01	2.83	2.81	5.97	0.67	17.69	8.07	17.65	21.43	3.86	13.00	8.10	12.10
25	0.26	3.04	2.91	2.89	6.09	0.67	17.75	8.07	17.87	21.60	3.69	13.02	8.16	12.14
26	0.27	2.97	2.83	2.87	5.89	0.50	17.63	7.71	17.38	21.01	3.51	13.02	7.96	12.04
27	0.26	2.86	2.81	2.88	5.74	0.35	17.61	8.37	16.92	20.20	3.30	13.04	7.86	12.06
28	0.24	2.77	2.75	2.92	5.45	0.35	17.47	9.65	16.76	18.95	2.92	13.08	7.78	12.05
29	0.24	2.58	2.76	3.03	5.14	0.39	17.36	11.66	16.15	18.41	3.27	13.07	7.84	11.92
30	0.23	2.39	2.62	2.66	4.88	0.35	17.18	11.53	15.40	17.46	2.78	13.08	7.55	11.82
31	0.27	2.34	2.60	2.72	4.92	0.35	17.11	12.71	15.69	17.58	2.23	12.92	7.62	11.74
32	0.37	2.37	2.57	2.84	4.79	0.43	17.13	14.68	15.53	18.08	2.18	12.97	7.83	11.68
33	0.37	2.17	2.55	2.92	4.92	0.54	17.14	14.87	14.85	17.63	2.18	13.07	7.77	11.74
34	0.33	1.93	2.58	2.95	4.93	0.84	17.13	14.60	14.65	16.71	2.11	13.18	7.66	11.80
35	0.30	1.89	2.54	2.91	4.63	0.93	17.25	12.99	14.00	16.48	2.07	13.22	7.43	11.84
36	0.26	1.85	2.54	2.95	4.58	1.10	17.14	12.33	13.82	16.62	2.07	13.21	7.37	11.85
37	0.26	1.91	2.55	2.99	4.41	0.88	17.29	12.71	13.88	16.51	2.07	13.36	7.40	11.96
38	0.24	1.81	2.52	2.92	4.47	0.88	17.27	12.71	13.73	16.48	2.04	13.40	7.37	11.97
39	0.23	1.70	2.51	2.84	4.48	0.93	17.20	12.33	13.63	16.14	2.09	13.40	7.29	11.96
40	0.24	1.70	2.50	2.95	4.50	0.88	17.21	12.48	13.68	16.31	2.14	13.36	7.33	11.94
41	0.23	1.59	2.50	2.96	4.56	0.88	17.23	12.43	13.58	15.97	2.18	13.35	7.29	11.94
42	0.22	1.57	2.47	2.92	4.59	0.88	17.21	12.19	13.64	15.83	2.16	13.32	7.25	11.90
43	0.22	1.53	2.44	2.99	4.57	0.84	17.13	12.06	13.55	15.57	2.18	13.29	7.20	11.88
44	0.22	1.50	2.43	2.99	4.62	0.80	17.12	12.46	13.74	15.72	2.16	13.27	7.25	11.88
45	0.22	1.55	2.43	2.95	4.54	0.84	17.11	12.57	13.61	15.74	2.14	13.33	7.25	11.91
46	0.22	1.51	2.42	2.92	4.52	0.84	17.07	12.21	13.70	15.89	2.16	13.35	7.23	11.90
47	0.21	1.48	2.41	2.87	4.55	0.84	17.06	11.89	13.69	15.73	2.20	13.39	7.20	11.89
48	0.23	1.28	2.46	2.96	4.43	0.93	16.94	11.53	13.37	15.13	2.20	13.59	7.09	11.93

Table 30: *XLS-R-128* 300M under *PLS* fine-tuning with k decay: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	2SDF	Avg.	Pool.
$M = 1.5, k = 0.01$														
0	76.82	75.70	61.77	50.06	45.34	16.53	37.09	71.13	47.86	60.85	72.83	49.78	55.48	52.26
1	65.40	69.86	56.02	49.35	39.04	50.09	40.15	45.96	53.14	51.30	42.68	50.96	51.16	57.78
2	58.03	64.38	57.05	44.07	39.72	44.04	41.06	51.64	49.05	56.31	56.90	43.13	50.45	53.03
3	42.83	46.80	42.93	49.57	30.70	34.27	42.98	42.39	38.15	53.56	45.19	41.36	42.56	45.58
4	29.56	39.94	29.60	43.09	27.07	36.88	42.90	34.04	37.56	62.85	42.62	40.74	38.90	40.10
5	12.12	19.59	15.64	22.46	19.52	23.76	36.50	29.40	35.41	43.82	27.43	27.33	26.08	26.59
6	1.66	8.17	3.51	4.32	8.67	11.70	25.12	11.26	16.89	26.17	6.41	17.92	11.82	17.27
7	0.38	5.62	1.76	1.25	5.59	1.55	21.04	9.14	12.36	19.20	4.42	15.56	8.16	15.00
8	0.24	5.72	1.60	1.02	4.04	1.40	18.71	7.67	11.99	18.28	2.70	14.65	7.34	14.16
9	0.13	5.19	1.53	1.03	4.01	1.60	18.51	7.43	11.95	17.28	1.77	14.48	7.07	13.88
10	0.10	4.54	1.57	1.10	4.00	1.47	18.44	8.34	12.30	17.51	1.49	14.23	7.09	13.63
11	0.10	3.83	1.51	1.06	4.06	1.47	18.08	7.66	12.23	16.17	1.55	14.31	6.83	13.58
12	0.08	3.37	1.54	1.10	4.30	1.81	17.71	8.19	12.60	16.14	1.55	14.29	6.89	13.48
13	0.08	3.28	1.50	1.10	4.32	1.42	17.39	7.96	12.42	15.84	1.69	14.09	6.76	13.33
14	0.08	3.08	1.55	1.14	4.31	1.81	17.53	9.14	12.93	15.33	1.74	14.02	6.89	13.32
15	0.08	2.92	1.52	1.06	4.30	1.81	17.18	8.62	12.42	14.96	1.81	14.06	6.73	13.28
16	0.08	2.73	1.50	1.10	4.30	2.03	17.16	8.74	11.23	14.25	1.77	13.94	6.57	13.06
17	0.08	2.61	1.39	1.10	4.32	1.73	17.01	8.47	10.87	13.46	1.81	13.78	6.39	12.90
18	0.10	2.79	1.39	1.10	4.33	1.81	17.38	8.87	10.74	13.74	1.70	13.96	6.49	12.85
19	0.10	2.68	1.32	1.28	4.41	1.77	17.46	8.75	10.67	13.04	1.37	13.72	6.38	12.83
20	0.10	3.06	1.44	1.25	4.38	2.16	17.87	8.74	11.48	13.74	0.79	14.12	6.59	13.04
21	0.07	2.13	1.56	1.25	4.52	2.61	18.03	8.47	9.67	11.24	1.91	13.93	6.28	12.91
22	0.08	1.73	1.59	1.45	4.61	3.26	18.36	10.45	10.06	9.88	2.09	14.01	6.46	13.14
23	0.10	1.58	1.68	1.48	4.55	2.91	17.86	11.23	11.48	9.17	1.39	13.95	6.45	12.85
24	0.08	2.09	1.66	1.44	4.19	3.09	19.11	11.41	10.20	9.29	2.00	14.19	6.56	12.31
$M = 1.5, k = 0.03$														
0	73.57	71.75	56.81	50.25	45.97	16.88	35.62	63.04	48.25	56.89	67.28	49.64	53.00	50.56
1	57.70	63.71	48.92	49.83	37.48	52.52	39.68	39.73	50.48	45.44	36.67	49.81	47.67	55.06
2	33.25	45.53	39.21	41.80	34.14	50.04	41.49	38.41	44.42	41.99	32.28	39.38	40.16	46.00
3	20.39	25.30	26.54	43.70	26.39	41.95	43.31	31.66	38.15	42.50	19.53	36.35	32.98	38.40
4	11.38	22.73	19.57	34.08	25.15	41.17	42.32	28.51	36.42	49.77	17.23	33.35	30.14	32.62
5	4.16	11.48	9.70	14.04	17.04	21.95	33.91	20.00	29.83	30.19	10.32	23.43	18.84	22.99
6	0.79	6.14	3.05	4.73	10.07	9.02	25.73	10.33	15.93	21.23	5.01	17.50	10.79	16.67
7	0.26	4.70	1.85	1.58	6.45	1.55	21.72	8.51	12.34	19.04	3.11	14.94	8.00	14.35
8	0.24	5.27	1.78	1.21	5.05	1.14	20.39	7.71	11.66	19.28	2.30	13.86	7.49	13.58
9	0.18	4.91	1.72	1.14	5.04	1.23	20.27	7.54	11.54	18.57	1.84	13.67	7.30	13.35
10	0.14	4.35	1.75	1.18	5.09	1.01	20.46	7.97	12.06	18.67	1.76	13.51	7.33	13.17
11	0.08	3.81	1.71	1.17	5.00	0.97	20.18	7.54	12.08	17.12	1.76	13.60	7.09	13.14
12	0.11	3.28	1.72	1.25	5.21	1.32	20.09	7.57	12.67	16.96	1.72	13.64	7.13	13.11
13	0.10	3.34	1.73	1.21	5.17	1.06	19.87	7.94	12.69	16.94	1.69	13.50	7.10	13.01
14	0.10	3.12	1.82	1.29	5.13	1.51	20.04	8.75	13.32	16.29	1.86	13.46	7.22	13.03
15	0.10	3.03	1.76	1.25	5.08	1.47	19.70	8.34	12.76	16.20	1.82	13.51	7.08	13.00
16	0.10	2.86	1.72	1.21	5.00	1.40	19.49	8.22	11.46	15.59	1.70	13.37	6.84	12.79
17	0.10	2.81	1.57	1.21	5.01	1.27	19.34	7.94	11.21	14.59	1.68	13.27	6.67	12.63
18	0.11	2.95	1.58	1.29	5.04	1.42	19.49	8.46	11.21	15.05	1.65	13.37	6.80	12.59
19	0.11	2.79	1.53	1.40	4.91	1.32	19.29	7.94	11.38	14.34	1.37	13.14	6.63	12.51
20	0.11	3.16	1.61	1.37	4.95	1.40	19.86	8.07	12.37	15.63	0.74	13.35	6.89	12.71
21	0.10	2.11	1.72	1.55	5.19	1.94	19.43	8.99	10.61	11.78	2.14	13.13	6.56	12.51
22	0.07	1.42	1.78	2.11	5.71	2.70	19.65	10.98	10.92	10.14	2.46	12.98	6.74	12.67
23	0.11	1.13	1.96	2.32	5.77	2.78	19.84	11.78	12.51	8.11	2.95	12.58	6.82	12.62
24	0.10	1.53	1.90	1.66	4.95	2.52	20.51	12.19	10.71	8.32	3.23	13.16	6.73	12.19
$M = 1.5, k = 0.05$														
0	50.48	49.39	50.62	50.89	40.74	51.06	47.09	59.07	36.46	53.00	76.06	49.12	51.16	46.50
1	54.88	61.57	49.11	47.06	43.94	53.50	41.52	55.23	44.15	49.65	63.21	43.88	50.64	47.58
2	49.98	57.82	43.82	51.95	42.56	46.46	43.98	59.73	47.57	66.30	37.37	47.25	49.57	51.83
3	35.15	49.74	35.46	52.74	52.93	70.33	44.47	42.26	37.83	55.79	25.71	45.15	45.63	47.20
4	21.68	27.52	29.42	51.53	60.82	77.77	55.63	25.05	40.42	25.47	20.27	44.50	40.01	41.02
5	7.19	10.14	12.88	27.98	39.38	61.93	35.78	22.39	41.45	23.20	14.21	24.25	26.73	29.44
6	2.86	6.13	5.77	13.02	15.64	31.81	29.49	18.01	23.00	25.44	11.22	17.11	16.62	19.34
7	0.24	3.27	2.11	2.61	6.57	8.87	22.29	9.93	14.54	20.99	4.43	15.09	9.25	15.14
8	0.16	3.13	1.40	1.33	4.76	3.71	19.53	8.74	11.35	18.21	2.99	14.08	7.45	13.48
9	0.14	3.36	1.44	1.25	4.59	3.09	18.86	9.27	12.37	18.24	2.20	13.82	7.39	13.07
10	0.10	3.14	1.35	1.17	4.41	2.44	18.60	8.46	11.71	17.19	2.28	13.69	7.04	12.84
11	0.11	2.96	1.32	1.06	4.35	2.16	18.26	8.87	11.10	17.10	2.32	13.32	6.91	12.64
12	0.10	3.25	1.37	1.10	4.37	1.81	18.00	8.87	10.44	16.85	1.88	13.29	6.78	12.65
13	0.10	2.93	1.39	1.10	4.44	2.29	17.94	9.93	11.32	16.94	2.39	13.30	7.01	12.57
14	0.11	2.83	1.37	1.14	4.32	2.61	17.91	10.08	11.25	16.50	2.09	13.24	6.95	12.53
15	0.08	2.50	1.35	1.14	4.27	3.00	17.94	9.93	11.81	15.71	2.51	13.26	6.96	12.56
16	0.08	2.46	1.30	1.21	4.15	2.29	17.75	8.90	10.49	14.78	2.15	13.07	6.55	12.41
17	0.10	2.65	1.28	1.18	4.16	2.07	17.68	8.22	10.21	15.27	2.06	12.99	6.49	12.25
18	0.08	2.58	1.33	1.21	4.27	2.29	17.93	9.17	11.07	15.14	2.16	13.18	6.70	12.30
19	0.10	2.74	1.35	1.21	4.13	2.61	18.35	7.97	10.39	14.06	1.70	12.91	6.46	12.20
20	0.11	3.16	1.35	1.11	4.01	2.65	17.93	8.21	10.58	15.74	1.09	13.07	6.58	12.22
21	0.10	2.67	1.30	1.18	4.55	3.45	18.35	8.51	9.68	13.31	1.70	13.47	6.52	12.41
22	0.07	2.24	1.21	1.33	5.48	4.99	19.38	8.47	11.09	10.72	2.11	14.74	6.82	12.99
23	0.10	2.68	1.32	1.33	4.88	3.22	18.52	8.47	11.21	11.18	1.44	14.87	6.60	12.68
24	0.11	2.88	1.39	1.29	4.20	3.04	19.73	7.96	10.74	11.07	2.13	15.31	6.65	12.03
$M = 1.5, k = 0.07$														
0	74.21	72.26	57.79	49.38	44.77	17.67	36.36	66.33	48.56	58.30	67.77	49.42	53.57	50.89
1	63.02	68.31	53.31	50.10	36.20	50.54	40.25	44.77	52.51	48.42	38.62	50.50	49.71	56.58
2	44.00	55.15	47.12	42.90	37.30	46.68	40.66	44.10	46.26	50.74	38.85	41.41	44.60	49.29
3	23.45	31.13	29.36	44.83	33.24	39.40	41.31	32.56	38.30	44.92	23.08	38.24	34.99	41.22
4	14.28	27.55	21.57	38.66	28.06	39.49	41.65	30.47	36.86	51.25	21.22	36.23	32.27	34.77
5	5.81	15.00	11.63	17.80	19.72	23.33	34.69	22.52	32.48	32.69	14.64	24.95	21.27	23.72
6	1.11	6.77	3.01	4.46	9.80	8.44	24.82	11.93	16.27	22.26	6.06	17.23	11.01	15.84
7	0.26	4.78	1.78	1.17	5.65	1.27	21.48	7.66	11.43	18.84	3.05	14.62	7.67	13.62
8	0.30	5.18	1.67	1.05	4.81	1.51	19.60	6.51	11.05	19.08	2.07	13.58	7.20	12.88
9	0.24	4.89	1.59	0.95	4.86	2.11	19.58	6.33	11.25	18.57	1.46	13.41	7.10	12.69
10	0.23	4.53	1.64	1.02	5.00	2.07	19.99	6.63	11.68	19.03	1.30	13.26	7.20	12.53
11	0.16	4.20	1.54	1.02	5.05	1.86	19.85	6.51	11.60	17.94	1.14	13.40	7.02	12.54
12	0.16	3.68	1.57	1.06	5.36	2.11	19.85	6.63	12.19	17.50	1.21	13.47	7.07	12.55
13	0.15	3.58	1.52	1.02	5.20	1.94	19.51	6.51	11.82	17.25	1.23	13.32	6.92	12.40
14	0.14	3.45	1.56	1.02	5.23	2.16	19.77	7.52	12.24	16.70	1.26	13.32	7.03	12.42
15	0.12	3.28	1.54	1.02	5.14	2.07	19.38	7.03	11.76	16.45	1.34	13.37	6.88	12.42
16	0.11	3.05	1.52	1.09	5.16	2.20	19.40	7.03	10.57	15.78	1.35	13.25	6.71	12.20
17	0.12	2.80	1.43	1.10	5.13	2.11	19.39	6.63	10.39	14.66	1.42	13.09	6.52	12.04
18	0.14	2.96	1.46	1.10	5.29	2.16	19.74	7.16	10.35	14.94	1.23	13.29	6.65	12.02
19	0.12	2.81	1.41	1.17	5.06	2.11	19.68	6.61	10.41	14.41	1.21	13.13	6.51	11.98
20	0.12	2.89	1.46	1.14	5.18	2.44	20.64	6.78	10.91	14.71	0.93	13.42	6.72	12.14
21	0.08	2.09	1.49	1.29	5.48	2.52	21.00	6.50	9.69	12.24	1.74	13.23	6.45	11.89
22	0.08	1.66	1.47	1.48	5.56	2.87	22.27	7.29	9.01	10.46	1.97	13.33	6.45	11.86
23	0.08	1.71	1.53	1.44	4.99	2.44	21.94	8.49	9.38	9.75	2.36	12.87	6.42	11.16
24	0.08	2.13	1.55	1.40	4.81	2.61	23.55	9.00	8.89	10.33	2.68	13.91	6.74	11.06
$M = 1.5, k = 0.1$														
0	64.11	62.77	45.88	49.22	41.97	20.67	36.08	55.12	48.53	54.91	53.90	48.20	48.45	48.03
1	39.28	44.98	37.51	49.38	35.94	64.18	41.51	32.84	47.99	41.06	26.48	48.46	42.47	52.57
2	15.89	26.88	26.47	35.82	31.30	64.13	41.29	30.60	42.77	35.14	14.89	35.33	33.38	44.67
3	10.97	21.42	19.83	36.84	25.45	58.74	41.60	28.60	44.07	33.19	11.75	33.33	30.48	39.50
4	5.44	20.19	14.35	27.22	22.84	47.04	38.64	27.43	41.34	38.38	10.80	30.56	27.02	32.39
5	1.58	13.09	7.70	10.79	15.99	21.60	32.00	17.09	35.25	25.22	8.50	24.15	17.75	24.05
6	0.34	6.38	2.72	3.67	9.96	9.88	25.75	12.16	19.85	21.23	4.93	18.19	11.26	17.13
7	0.15	4.74	2.00	1.67	6.81	2.78	22.92	10.43	14.89	19.93	3.60	15.79	8.81	14.75
8	0.15	4.79	1.89	1.33	5.53	2.20	20.91	9.80	13.40	20.03	3.08	14.70	8.15	13.86
9	0.14	4.24	1.83	1.21	5.52	2.16	20.55	9.39	12.54	19.38	2.53	14.43	7.83	13.51
10	0.08	3.66	1.88	1.25	5.48	2.20	20.74	9.53	12.67	19.19	2.51	14.22	7.79	13.30
11	0.07	3.15	1.85	1.18	5.55	2.24	20.42	9.25	12.66	17.63	2.74	14.24	7.58	13.23
12	0.08	2.67	1.88	1.21	5.79	2.35	19.98	9.27	12.77	16.99	2.74	14.21	7.49	13.17
13	0.07	2.54	1.84	1.21	5.68	1.99	19.71	9.12	12.64	16.48	3.18	14.06	7.38	13.05
14	0.07	2.27	1.89	1.25	5.73	2.48	19.82	10.08	13.25	15.37	3.44	14.03	7.47	13.09
15	0.08	2.09	1.86	1.21	5.68	2.20	19.54	9.80	12.93	14.98	3.60	14.04	7.33	13.06
16	0.07	1.94	1.76	1.29	5.57	2.57	19.28	9.92	11.30	14.43	3.46	13.87	7.12	12.83
17	0.08	1.85	1.65	1.44	5.56	2.11	19.08	9.68	11.03	13.23	3.58	13.67	6.91	12.63
18	0.08	1.90	1.60	1.44	5.56	2.11	19.09	9.95	11.06	13.61	3.18	13.72	6.94	12.55
19	0.08	1.85	1.51	1.52	5.47	2.03	19.19	8.87	11.24	12.86	2.68	13.48	6.73	12.52
20	0.10	2.04	1.59	1.55	5.37	2.65	19.36	9.55	12.01	13.87	1.95	13.75	6.98	12.75
21	0.10	1.44	1.65	1.78	5.54	2.52	19.44	10.20	10.55	10.17	3.78	13.50	6.72	12.54
22	0.14	1.09	1.68	2.08	5.34	2.96	19.31	11.25	10.68	8.61	4.48	13.22	6.74	12.52
23	0.20	0.97	1.74	2.35	5.12	2.91	19.28	12.46	12.14	7.19	5.04	12.89	6.86	12.31
24	0.20	1.07	1.72	1.74	4.94	3.63	20.12	12.84	11.94	7.00	5.55	13.70	7.04	12.15

Table 31: *XLS-R-128* 300M under *aligned* fine-tuning with variant ρ: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
$\rho = 1.5$														
0	75.03	73.50	58.52	48.81	44.33	17.31	36.65	68.21	48.51	58.63	69.10	49.22	53.98	51.10
1	61.66	67.29	52.78	49.16	37.31	52.74	40.55	43.04	51.95	47.93	38.44	50.80	49.47	56.53
2	41.67	51.89	45.57	43.35	37.58	49.37	41.73	41.86	44.08	49.12	35.51	42.15	43.66	50.19
3	28.78	36.88	33.65	47.56	28.05	36.79	43.29	37.49	37.00	46.75	27.94	37.93	36.84	42.04
4	17.32	28.24	23.45	38.39	24.85	39.58	42.87	30.73	38.36	58.52	23.31	36.83	33.54	36.91
5	7.44	16.13	12.35	18.30	17.79	21.15	34.93	25.96	35.51	37.46	15.05	25.84	22.33	24.98
6	1.06	7.09	2.91	4.06	9.12	6.97	23.71	12.82	15.80	23.46	6.01	17.63	10.89	16.24
7	0.27	4.85	1.62	1.17	6.07	1.55	20.39	8.64	12.31	19.61	3.50	15.20	7.93	14.03
8	0.29	4.88	1.57	1.09	4.54	1.68	18.49	7.69	11.63	19.61	2.55	14.23	7.36	13.25
9	0.22	4.62	1.53	1.03	4.65	2.20	18.35	7.43	11.67	19.00	2.05	14.05	7.23	13.03
10	0.16	4.13	1.56	1.17	4.72	2.11	18.50	7.94	11.94	19.24	1.90	13.84	7.27	12.85
11	0.12	3.70	1.48	1.10	4.73	1.94	18.23	7.16	11.89	17.92	2.41	13.89	7.05	12.79
12	0.11	3.25	1.51	1.10	5.08	2.24	18.06	7.67	12.23	17.64	2.00	13.92	7.07	12.76
13	0.11	3.06	1.48	1.09	4.98	1.94	17.90	7.81	12.07	17.09	2.11	13.73	6.95	12.61
14	0.10	2.86	1.51	1.17	4.98	2.40	18.01	8.21	12.33	16.43	2.28	13.71	7.00	12.63
15	0.10	2.64	1.47	1.10	4.89	2.11	17.79	8.07	11.71	16.14	2.23	13.76	6.83	12.61
16	0.10	2.45	1.48	1.21	4.87	2.29	17.57	8.07	10.56	15.56	2.16	13.57	6.66	12.34
17	0.11	2.29	1.38	1.29	4.94	2.29	17.56	7.97	10.35	14.50	2.23	13.41	6.53	12.20
18	0.11	2.44	1.37	1.40	4.94	2.11	17.72	8.21	10.30	14.79	2.14	13.52	6.59	12.15
19	0.09	2.29	1.33	1.44	4.88	2.20	17.83	7.41	10.26	14.06	1.79	13.27	6.40	12.07
20	0.11	2.50	1.37	1.40	4.97	2.61	18.13	7.94	10.88	14.53	1.35	13.59	6.61	12.26
21	0.08	1.65	1.43	1.63	5.70	2.83	18.51	8.74	9.55	11.84	2.49	13.41	6.49	12.17
22	0.08	1.32	1.47	1.93	6.23	3.50	18.76	9.67	9.60	10.46	2.62	13.22	6.57	12.24
23	0.08	1.32	1.53	1.90	5.74	3.32	18.15	11.41	10.66	10.03	3.18	12.73	6.67	11.71
24	0.08	1.78	1.48	1.55	4.76	3.22	18.64	11.68	9.73	10.09	3.92	13.42	6.70	11.15
$\rho = 2$														
0	76.65	75.03	61.03	48.85	44.86	16.70	36.80	70.88	47.97	59.42	71.99	49.47	54.97	51.68
1	70.36	75.09	59.90	49.52	37.31	48.53	40.41	49.95	53.99	52.83	45.75	51.06	52.89	58.82
2	59.88	66.03	58.58	45.25	38.38	44.17	41.96	56.69	48.38	55.74	54.51	44.70	51.19	53.50
3	48.13	55.83	49.24	51.95	30.63	32.41	44.73	52.19	38.83	58.36	50.51	43.18	46.33	47.79
4	29.75	41.90	30.49	44.57	29.42	37.59	44.37	35.23	37.28	65.06	43.99	43.00	40.22	41.08
5	15.15	24.77	17.53	25.11	22.86	21.95	37.36	34.83	38.01	48.14	31.93	28.17	28.82	27.47
6	2.11	7.34	3.69	6.89	11.87	13.64	24.91	15.63	17.19	27.79	8.69	17.35	13.09	16.24
7	0.42	4.54	1.65	1.71	7.57	1.86	20.71	9.80	12.41	21.35	5.13	14.99	8.51	14.24
8	0.34	4.51	1.59	1.33	5.37	1.40	18.39	7.92	12.20	20.92	3.09	14.08	7.60	13.53
9	0.18	3.91	1.59	1.29	5.37	1.64	18.35	7.81	12.32	19.82	2.53	13.86	7.39	13.28
10	0.14	3.29	1.67	1.36	5.51	1.55	18.48	8.36	12.82	20.48	2.30	13.71	7.47	13.08
11	0.10	2.90	1.62	1.21	5.48	1.23	18.03	8.07	12.55	18.78	2.58	13.74	7.19	13.03
12	0.11	2.52	1.68	1.21	5.77	1.73	17.93	8.37	13.42	18.61	2.37	13.78	7.29	13.00
13	0.09	2.31	1.68	1.21	5.61	1.55	17.73	8.60	13.13	18.06	2.69	13.61	7.19	12.88
14	0.10	2.20	1.69	1.29	5.65	1.94	17.86	9.40	13.61	17.30	2.90	13.59	7.29	12.89
15	0.08	2.04	1.65	1.14	5.69	1.81	17.43	8.75	12.80	16.99	2.95	13.60	7.08	12.85
16	0.07	1.94	1.61	1.26	5.49	1.99	17.28	8.59	11.19	16.25	2.77	13.43	6.82	12.61
17	0.08	1.79	1.51	1.40	5.43	1.73	17.05	8.24	10.57	14.84	2.90	13.29	6.57	12.42
18	0.11	1.96	1.48	1.44	5.38	1.81	17.09	8.74	10.50	15.34	2.57	13.42	6.65	12.39
19	0.10	1.84	1.43	1.51	5.41	1.73	16.99	8.21	10.58	14.48	2.16	13.19	6.47	12.31
20	0.12	2.12	1.49	1.49	5.22	2.24	17.50	8.47	11.38	15.22	1.49	13.44	6.68	12.51
21	0.08	1.38	1.58	1.82	5.70	1.94	17.65	9.55	9.86	12.25	3.30	13.33	6.54	12.40
22	0.07	1.13	1.59	1.97	5.78	2.33	17.46	10.73	9.77	10.63	3.79	13.21	6.54	12.44
23	0.10	1.03	1.69	1.97	5.41	2.03	17.09	11.93	10.87	9.35	3.97	13.12	6.55	11.97
24	0.11	1.13	1.68	1.70	4.75	2.07	18.17	12.74	10.38	9.06	5.06	14.41	6.77	11.70

Table 32: *XLS-R-128* 300M under *aligned* fine-tuning with equal weighting and aligned-only setup: EER (%, ↓) at each transformer.

Layer	In-Domain			Voc.	In-the-Wild			Diff/Flow		Other OOD			Summary	
	19LA	21LA	21DF	LSV	ITW	FoR	SC	DFADD	DIFSSD	MLAAD-EN	DECRO-EN	25DF	**Avg.**	**Pool.**
Equal-weighting Fine-tuning														
0	21.97	25.15	29.07	44.49	52.98	70.59	44.40	30.73	52.10	34.84	16.09	45.34	38.98	51.23
1	13.40	18.99	23.11	44.92	43.89	61.65	45.66	30.45	53.24	31.01	13.31	42.81	35.20	43.27
2	11.31	19.59	21.86	40.56	38.55	57.60	42.82	30.17	48.69	32.66	11.70	39.30	32.90	40.15
3	8.89	18.68	17.89	36.95	35.64	47.00	42.25	25.32	45.60	27.09	11.77	35.55	29.38	36.09
4	5.00	15.62	13.15	27.11	33.07	55.83	39.70	21.98	49.89	20.13	11.40	31.30	27.01	31.25
5	1.05	8.46	7.03	10.94	20.72	41.13	31.14	14.57	49.80	18.54	10.59	23.09	19.75	23.06
6	0.44	6.61	4.99	5.49	13.57	32.61	27.19	14.19	39.91	17.31	8.97	19.86	15.93	19.01
7	0.26	4.96	3.74	3.22	10.57	21.19	23.90	12.97	26.06	17.65	6.62	17.36	12.38	15.68
8	0.22	4.49	3.03	2.31	8.85	16.88	22.39	11.91	19.57	17.76	5.30	15.97	10.72	14.41
9	0.22	4.37	2.80	2.04	7.96	15.15	21.15	11.00	17.25	18.34	4.62	15.41	10.03	13.89
10	0.19	4.12	2.64	1.97	7.79	14.63	20.71	11.53	17.27	17.99	4.50	15.14	9.87	13.66
11	0.15	4.02	2.60	1.90	7.28	14.39	20.45	11.40	16.09	17.69	4.23	15.05	9.60	13.65
12	0.15	3.85	2.49	1.89	7.54	13.12	20.19	11.91	15.53	17.67	4.11	14.66	9.43	13.40
13	0.15	3.66	2.54	1.93	7.54	12.99	20.58	12.19	16.30	17.78	4.00	14.96	9.55	13.53
14	0.16	3.65	2.54	1.89	7.50	12.54	21.12	12.97	17.63	17.67	4.02	14.98	9.72	13.46
15	0.15	3.63	2.54	1.74	7.30	11.93	21.16	12.74	17.82	17.06	3.91	15.15	9.59	13.39
16	0.14	3.54	2.36	1.81	7.29	11.98	20.88	11.81	16.79	16.81	3.76	15.00	9.35	13.42
17	0.14	3.46	2.33	1.82	7.00	11.93	21.20	12.72	15.99	16.76	3.44	14.89	9.31	13.43
18	0.14	3.27	2.27	1.97	7.12	10.68	22.90	12.57	15.49	16.23	3.11	15.68	9.28	13.59
19	0.12	3.19	2.23	2.16	7.51	10.60	22.28	13.26	16.04	15.93	2.85	14.76	9.24	13.73
20	0.12	3.28	2.27	2.20	7.04	8.83	21.53	12.46	14.23	16.54	2.51	14.35	8.78	13.31
21	0.12	3.27	2.35	2.16	6.84	10.92	20.64	13.24	14.88	16.23	2.74	14.82	9.02	13.66
22	0.16	3.31	2.20	2.08	6.50	11.74	20.34	14.17	14.45	16.81	2.79	14.56	9.09	13.42
23	0.20	4.42	2.30	1.86	5.87	10.81	20.13	14.55	13.90	16.93	2.72	14.13	8.99	13.35
24	0.19	6.28	2.56	1.74	5.49	9.41	21.48	20.12	13.26	18.07	4.53	14.29	9.78	13.50
Aligned-only Fine-tuning														
0	48.20	47.66	37.54	49.15	42.23	25.36	37.45	47.43	48.40	51.62	32.56	46.72	42.86	46.17
1	19.05	26.70	26.36	45.44	32.14	69.57	40.59	29.27	49.97	33.61	15.37	42.97	35.92	46.70
2	5.79	21.75	21.16	31.62	28.03	65.77	40.01	25.30	46.03	27.74	9.68	35.22	29.84	40.33
3	5.29	20.02	17.57	30.74	24.69	59.62	40.92	24.62	47.22	24.38	10.68	33.73	28.29	36.66
4	2.65	19.06	13.69	22.68	22.61	48.64	39.66	24.92	46.79	24.76	10.31	31.11	25.57	31.60
5	0.84	13.19	8.36	11.85	15.81	31.53	32.97	18.14	42.39	20.86	9.52	24.41	19.16	24.16
6	0.27	7.69	4.34	5.11	10.76	15.82	28.54	14.20	31.08	19.36	6.61	18.82	13.55	17.90
7	0.15	5.43	3.11	2.57	7.70	5.74	25.81	12.44	20.46	18.94	5.09	16.48	10.33	15.45
8	0.12	5.18	2.56	1.93	6.57	3.97	24.14	10.83	15.63	19.20	4.30	15.40	9.15	14.48
9	0.11	4.64	2.33	1.82	6.50	3.97	23.76	9.80	14.10	18.74	3.46	15.21	8.70	14.16
10	0.10	4.06	2.35	1.82	6.41	3.58	24.00	9.78	14.00	18.31	3.19	14.98	8.55	13.91
11	0.10	3.57	2.43	1.74	6.24	3.97	23.70	9.52	13.62	16.86	2.97	15.12	8.32	13.86
12	0.11	3.17	2.46	1.78	6.31	4.68	23.47	9.65	13.80	16.16	2.51	15.16	8.27	13.84
13	0.11	3.16	2.41	1.78	6.20	4.42	23.12	10.05	13.75	15.92	2.51	14.96	8.20	13.75
14	0.11	3.13	2.53	1.82	6.15	5.03	23.36	11.23	14.54	15.39	2.55	14.92	8.40	13.83
15	0.15	3.21	2.55	1.71	6.12	5.11	22.87	12.08	14.99	15.19	2.55	15.02	8.46	13.89
16	0.13	3.04	2.49	1.75	6.09	4.86	22.97	11.38	13.54	14.89	2.37	14.90	8.20	13.74
17	0.12	2.98	2.33	1.74	6.18	4.68	22.98	11.00	12.62	14.21	2.16	14.74	7.98	13.54
18	0.15	3.10	2.29	1.74	6.28	4.68	23.25	12.97	12.85	14.50	2.23	14.83	8.24	13.46
19	0.15	3.07	2.15	1.78	6.09	4.60	23.13	14.04	12.89	13.78	1.88	14.56	8.18	13.49
20	0.14	3.25	2.21	1.75	5.98	4.51	23.49	14.04	12.99	14.70	1.53	14.61	8.27	13.56
21	0.12	2.86	2.27	1.93	6.31	4.55	23.62	11.26	12.43	12.72	2.33	14.56	7.91	13.37
22	0.12	2.54	2.20	2.04	6.26	4.77	23.74	10.96	12.12	12.56	2.44	14.30	7.84	13.25
23	0.12	2.36	2.15	1.98	5.98	4.73	23.32	13.12	12.51	11.61	3.85	13.90	7.97	13.14
24	0.13	2.62	2.04	1.97	5.57	5.31	23.87	13.64	12.58	10.99	4.90	14.16	8.15	12.96

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions and scope of our study are listed at the end of Section 1, and summarized in the abstract. The experimental results in Section 5 support our claims.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of our work in Section 6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We take an experimental approach to validate our hypothesis on redundancy in speech foundation model fine-tuned for deepfake detection. Details on the motivation and the experiments are presented in Section 3 including the training objectives of our framework.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Our proposed framework is described in detail in Section 3. We also provide details regarding its implementation in Section 4.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All datasets employed in this study are publicly available. We will provide the code for the final paper.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide experimental details in Section 4 and in Appendix B.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We did not report error bars or confidence intervals in the paper.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provided our configuration in Appendix B.3

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed and respect NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: As described in Section 1, our proposed framework is designed to detect deepfake audios, hence our work can be used for positive societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The employed datasets, code, and model backbones are all properly cited (Section 4).

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were only used for language editing and formatting and did not impact the core methodology.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.