

We pretrain on ~2.7k hours of EEG from ~5.5k people

17 datasets. 54% of the hours are child and adolescent EEG; neurodegeneration, psychiatry, injury and migraine cohorts are 16%.

2.7k hours

pretraining set (train + validation)
3.3k hours in the full corpus

5.5k people

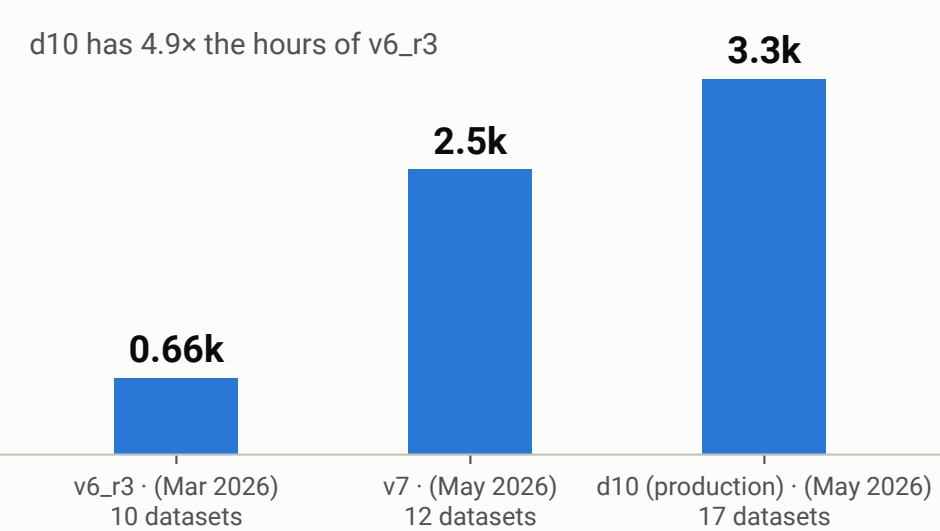
participants in the pretraining set
6.7k in the full corpus

17 datasets

child development, epilepsy, healthy adults,
neurodegeneration, psychiatry, injury

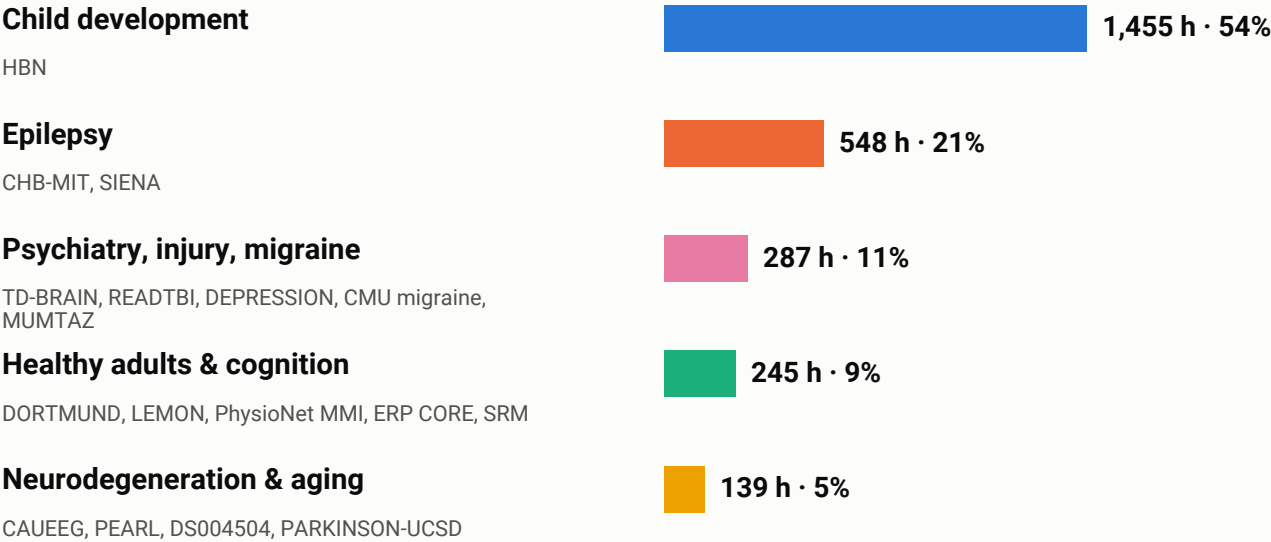
Corpus size by encoder generation

hours of EEG in the converted corpus



Pretraining hours by area

what the encoder has seen



Hours are unique EEG time from file headers. 14% of the pretraining hours (6 older datasets) are estimated from the May census because their processed files are no longer on disk. Evaluation cohorts are held out of the pretraining set. AD_EEG (a duplicate of DS004504) is excluded.