The generalist tokenizers were trained on the datasets listed below. If you use these checkpoints to compare to other trained models it will **NOT** be a fair comparison if the datasets are not the same. The datasets are in [other folders](https://drive.google.com/drive/u/0/folders/1gI36rS8irRZ32ibzKBPGncDmMXQtEf1C) of this same drive. Re-training the tokenizers is not time consuming, so promote good research and be consistent!  
  
Imputation:

* Weather
* Electricity
* ETTh1
* ETTh2
* ETTm1
* ETTm2

Anomaly Detection:

* SMD
* MSL
* SMAP
* SWAT
* PSM

Forecasting:

* Traffic
* Weather
* Electricity
* ETTh1
* ETTh2
* ETTm1
* ETTm2

Paper: <https://arxiv.org/pdf/2402.16412>  
Code: <https://github.com/SaberaTalukder/TOTEM>   
Video Overview: <https://www.youtube.com/watch?v=OqrCpdb6MJk>