**AI&DS**

**Building the dataset properly**

Understanding the required dataset format that is needed as input in the training process of the linked Jupyter Notebook above is essential. The format can be easily understood by looking at the example I provided [here](https://drive.google.com/drive/folders/1XnzpHBdD7FLmPwDI5J_MvylnnLGl4bcY?usp=sharing).

## Make it more real with post-processing

If we want to use the model for Text-to-Speech without target audio prompt, we can use VoiceFixer [3] to refine the generated audio. VoiceFixer is not trained on singing voices so the result is not good on those.

At ML6, we believe in the ethical use of our technologies, and therefore also anticipating potential harmful applications and identifying risks is crucial for us. Related to the voice cloning technologies, if this technology were to be used in a real life use case, it is fundamental to adhere to the principles of consent and disclosure (non-deceptiveness), and always aiming to create trustworthy and beneficial AI solutions.

## References

[1]Stanislav Beliaev and Boris Ginsburg. 2021. TalkNet 2: Non-Autoregressive Depth-Wise Separable Convolutional Model for Speech Synthesis with Explicit Pitch and Duration Prediction

[2] Jungil Kong, et al. 2020. HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis

[3] [Haohe Liu](https://arxiv.org/search/cs?searchtype=author&query=Liu%2C+H" \t "_blank), et al. 2021. VoiceFixer: Toward General Speech Restoration with Neural Vocoder