****[DLHLP 2023 Fall HW1 TTS](https://docs.google.com/presentation/d/1-hsO7qofYiTwJGefBq25fOW_HpHbvy7SokDSt4WBOE8/edit?fbclid=IwAR1JGbc75ajlyd2e9kRGwKMAYJVppbTeG1EayDCcNOrayWBy5Dzp8qrOmlY#slide=id.g117e4871a4b_0_5)

**基本問題**

1. 文字如何被處理成模型輸入?
2. 下載的原始音檔經過了甚麼前處理？
3. Input (text, feature…) “長什麼樣子” (e.g. alignment, spker embed…)
4. Output “長什麼樣子” (e.g. mel visualize, pitch contour)
5. 不同的TTS模型分別算了哪些 loss?
6. 訓練過程 (training curve, 大概要訓練多久)
7. TTS模型生成聲音的速度如何(生成 x 秒的音檔需要 y 秒)
8. Demo (好的例子/生壞的例子/缺點是甚麼)

↓這我老婆