CADT IDRI Institute of Digital Research & Innovation

Transforming Accessibility in Cambodia: Text-to-Speech Solutions for Education and Public Services

Presented by: POV PHANNET
Lecturer-Researcher, CADT

Date: 02/04/2025

Content

CADT IDRI Institute of Digital Research & Innovation

Introduction

As momentum on the SDGs stalls, AI's promise of exponential growth could offer much-needed rapid acceleration across the **2030** Agenda. To harness AI effectively, we must ensure it serves those most in need, and that all countries – not just those in the Global North – can tap into its development benefits.

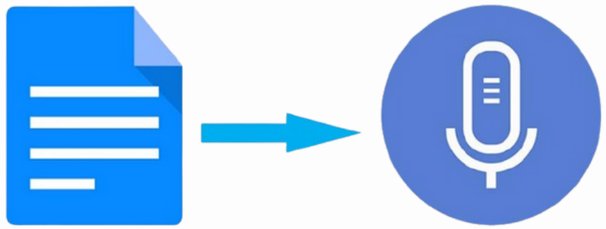
In Cambodia, AI-driven solutions, such as text-to-speech (TTS) technology, can significantly improve accessibility in **education and public services**, making information more inclusive and easier to access for individuals with disabilities, rural populations, and those with limited literacy .

https://sdg-action.org/can-ai-help-us-achieve-the-sdgs/

What is Text-To-Speech?

CADT IDRI Institute of Digital Research & Innovation

Text-to-Speech (TTS) is a technology that converts written text into natural-sounding spoken audio using **Artificial Intelligence (AI)**. It enables computers, smartphones, and other digital devices to "read aloud" text, making information more accessible to a wide range of users, including individuals with disabilities, those with low literacy, and speakers of different languages.

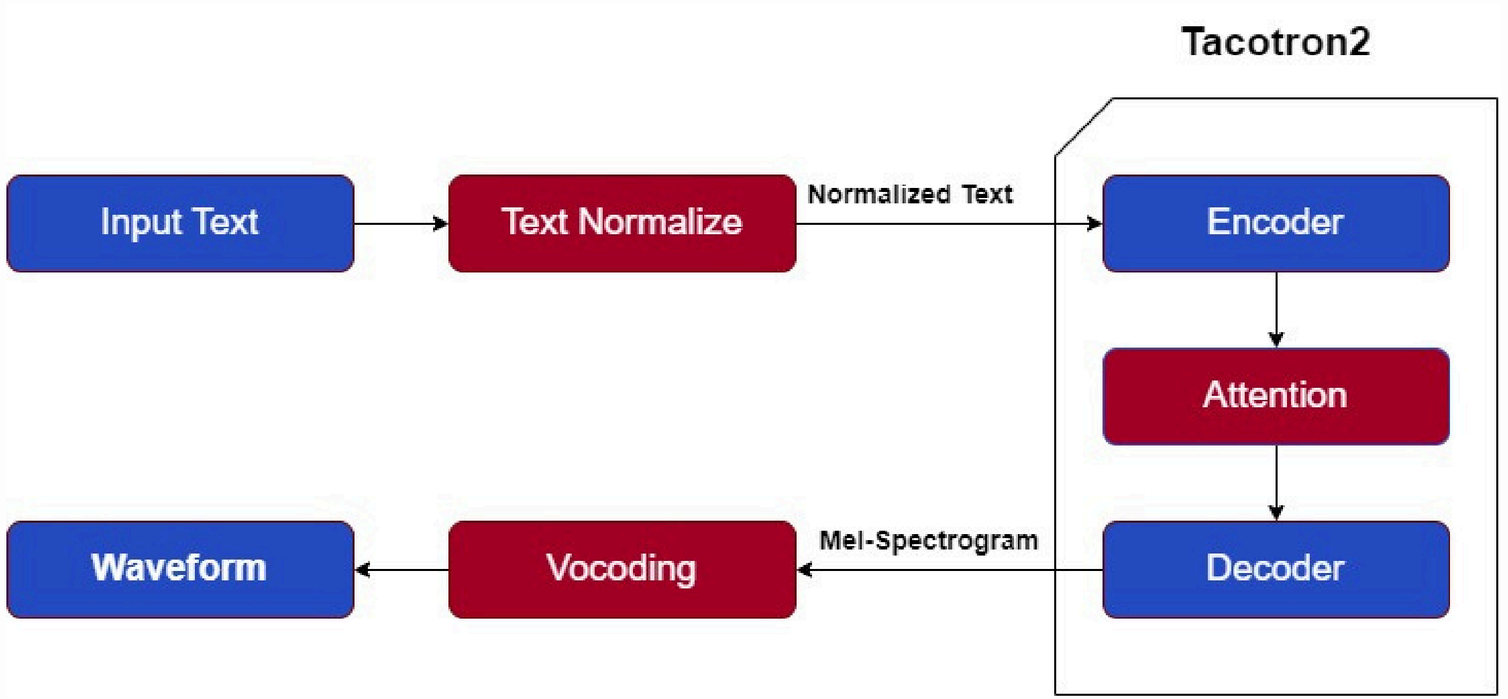


Text Speech

Current Model

CADT IDRI Institute of Digital Research & Innovation

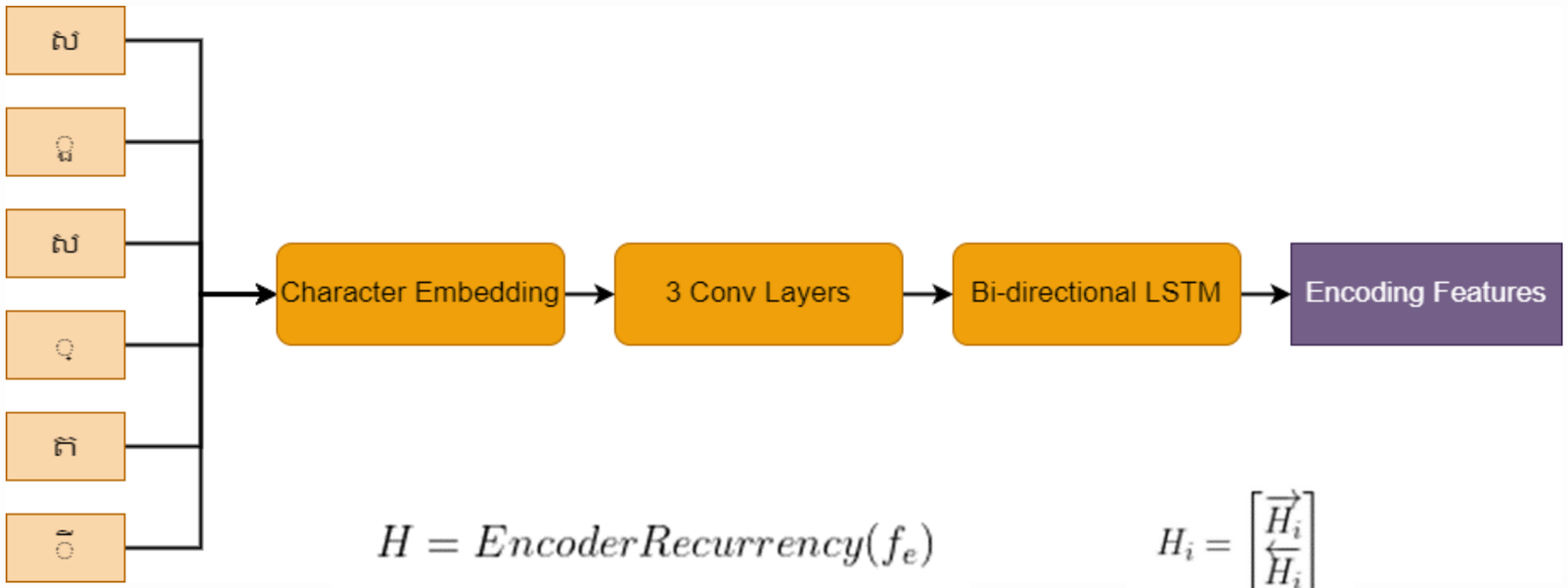
The **Tacotron2** model have 3 main component such as **Encoder**, **Location sensitive attention**, and **Encoder**.



Current Model

CADT IDRI Institute of Digital Research & Innovation

Encoder



$$H = EncoderRecurrency(f_e)$$

$$H_i = \begin{bmatrix} \overrightarrow{H_i} \\ \overleftarrow{H_i} \end{bmatrix}$$

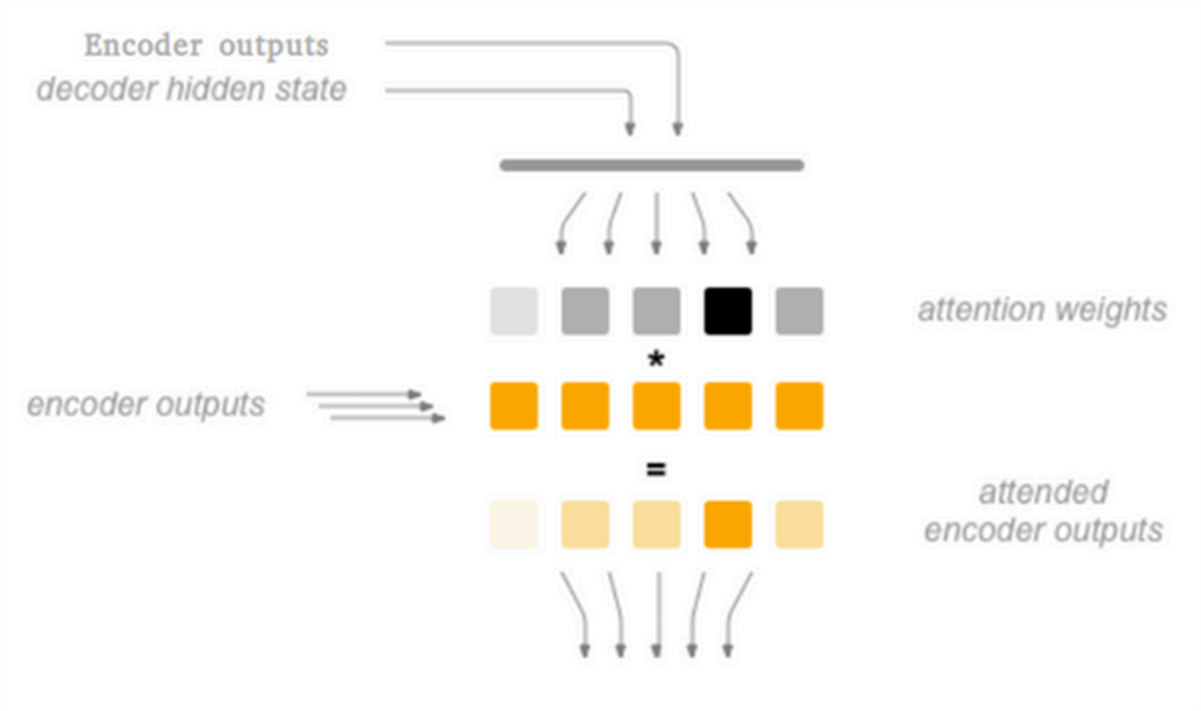
H: are the concatenation of the forward and backward hidden states

fe: is the encoder convolutional features

Current Model

CADT IDRI Institute of Digital Research & Innovation

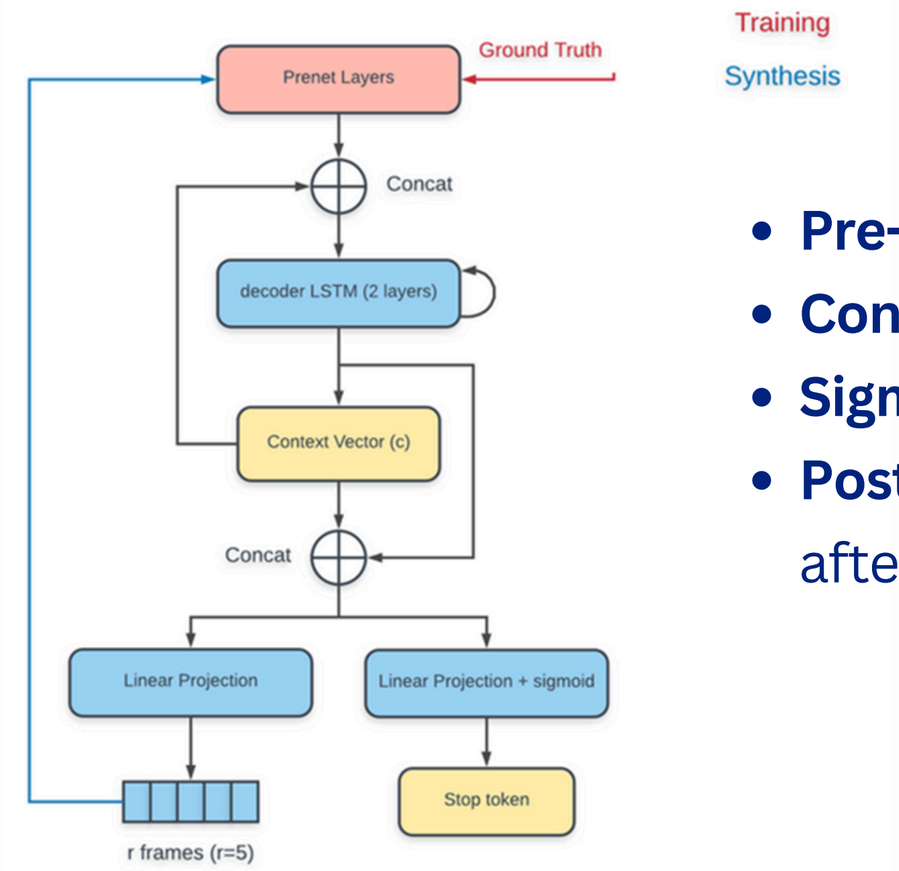
- **Learn alignment**
- Prevent some subsequences are replaced or ignored by the **decoder**



Current Model

CADT IDRI Institute of Digital Research & Innovation

The **decoder** is an autoregressive recurrent neural network(RNN).



- **Pre-net:** for learning attention
- **Context vector:** store the location of decoded word
- **Sigmoid function:** predict if the decoding process ended
- **Post-net:** checking and adding detail to Mel-spectrogram after decoding process is finished

$$py_i = Prenet(y_{i-1}) = Relu(W_2\ Relu(W_1 y_{i-1} + b_1) + b_2)$$

$$s_i = LSTM(LSTM(s_{i-1}, py_i, c_{i-1}))\ (*)$$

$$y_i = Linear([s_i; c_i]) = W_p[s_i; c_i] + b_p$$

$$y_{s,i} = SigmLinear([s_i; c_i]) = \sigma(W_s[s_i; c_i] + b_s)$$

Current Model

CADT IDRI Institute of Digital Research & Innovation

Tacotron2 Architecture

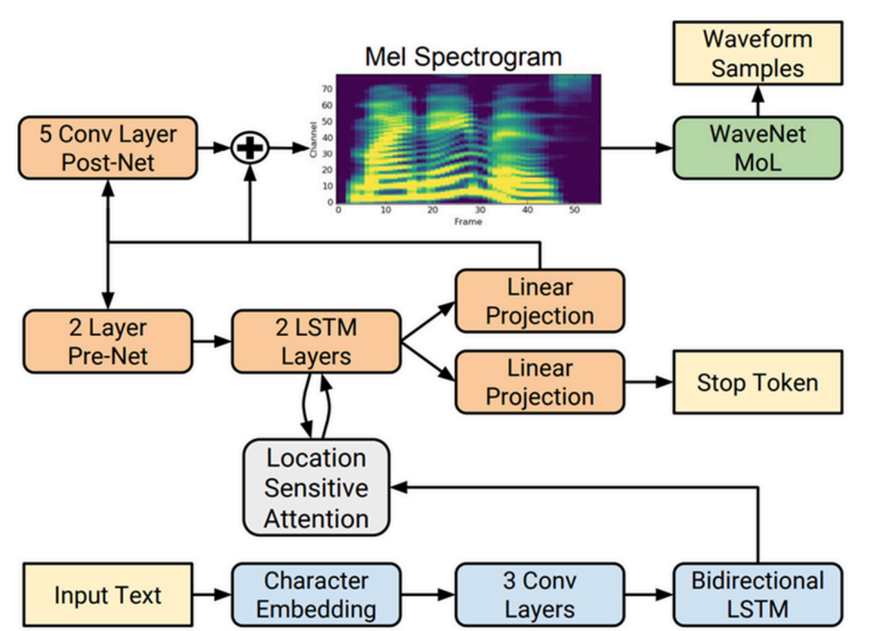


Fig. 1. Block diagram of the Tacotron 2 system architecture.

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4779-4783). IEEE.

Current Model

CADT IDRI Institute of Digital Research & Innovation

Khmer Corupus

Dataset	Male	Female
Total Duration	~ 3 hours	~ 4 hours
Vocabulary size	~ 3000 words	~ 4000 words

Current Model

CADT IDRI Institute of Digital Research & Innovation

Text normalization

Normalization	Raw	Normalized
Word segmentation	ខ្ញុំទៅសាលារៀនជាមួយមិត្តភក្តិ	ខ្ញុំ ទៅ សាលារៀន ជាមួយ មិត្តភក្តិ
Number to Khmer text	២០២៥	ពីរ ពាន់ ម្ភៃ ប្រាំ
Currency to Khmer text	៛	រៀល

```
kh_atr_m001_a3205.wav|ខ្ញុំ ចង់ នៅ ម្នាក់ឯង នៅ យប់ នេះ ។
kh_atr_m001_a3206.wav|ដៃគូ របស់ ខ្ញុំ បាន ភ្លាត់ និង គ្រែច ក ជើង ។
kh_atr_m001_a3207.wav|ខ្ញុំ បាន ហូប សាច់ គោ បី ថ្ងៃ ហើយ ។ ខ្ញុំ ចង់ ហូប គុយ ទារ ជម្ងឺតា ។
kh_atr_m001_a3208.wav|ខ្ញុំ ចង់បាន កាហ្វេ មួយ ដែរ ។
kh_atr_m001_a3209.wav|ការ ដែល សំខាន់ គឺ អ្នក ត្រូវ សម្រាក ឱ្យ បាន ល្អ ។
kh_atr_m001_a3210.wav|ខ្ញុំ យល់ បាន ខ្លះ ពី អ្វី ដែល អ្នក និយាយ ។
```

Current Model

CADT IDRI Institute of Digital Research & Innovation

Khmer Consonants

ក [k]	ខ [kh]	គ [k]	ឃ [kh]	ង [ng]
ច [ch]	ឆ [chh]	ជ [ch]	ឈ [chh]	ញ [nh]
ដ [d]	ឋ [th]	ឌ [d]	ឍ [th]	ណ [n]
ត [t]	ថ [th]	ទ [t]	ធ [th]	ន [n]
ប [b]	ផ [ph]	ព [p]	ភ [ph]	ម [m]
យ [y]	រ [r]	ល [l]	វ [v]	ស [s]
ហ [h]	ឡ [l]	អ [or]		

Khmer Dependent Vowels

◌ា [a] / [ea]	◌ិ [e] / [i]	◌ី [ei] / [y]	◌ឹ [oe] / [oe]	◌ឺ [eu] / [eu]	◌ុ [o] / [u]	◌ូ [ou] / [ou]	◌ួ [our] / [our]
◌ើ [oeu] / [oeu]	◌ឿ [oeur] / [oeur]	◌ៀ [ie] / [ie]	◌េ [e] / [e]	◌ែ [e] / [e]	◌ៃ [ai] / [ey]	◌ោ [ao] / [o]	◌ៅ [ao] / [ov]
◌ុំ [om] / [um]	◌ំ [am] / [oum]	◌ាំ [am] / [oam]	◌ះ [as] / [eah]	◌ុះ [os] / [ous]	◌ៈ [es] / [es]	◌ោះ [as] / [uos]	

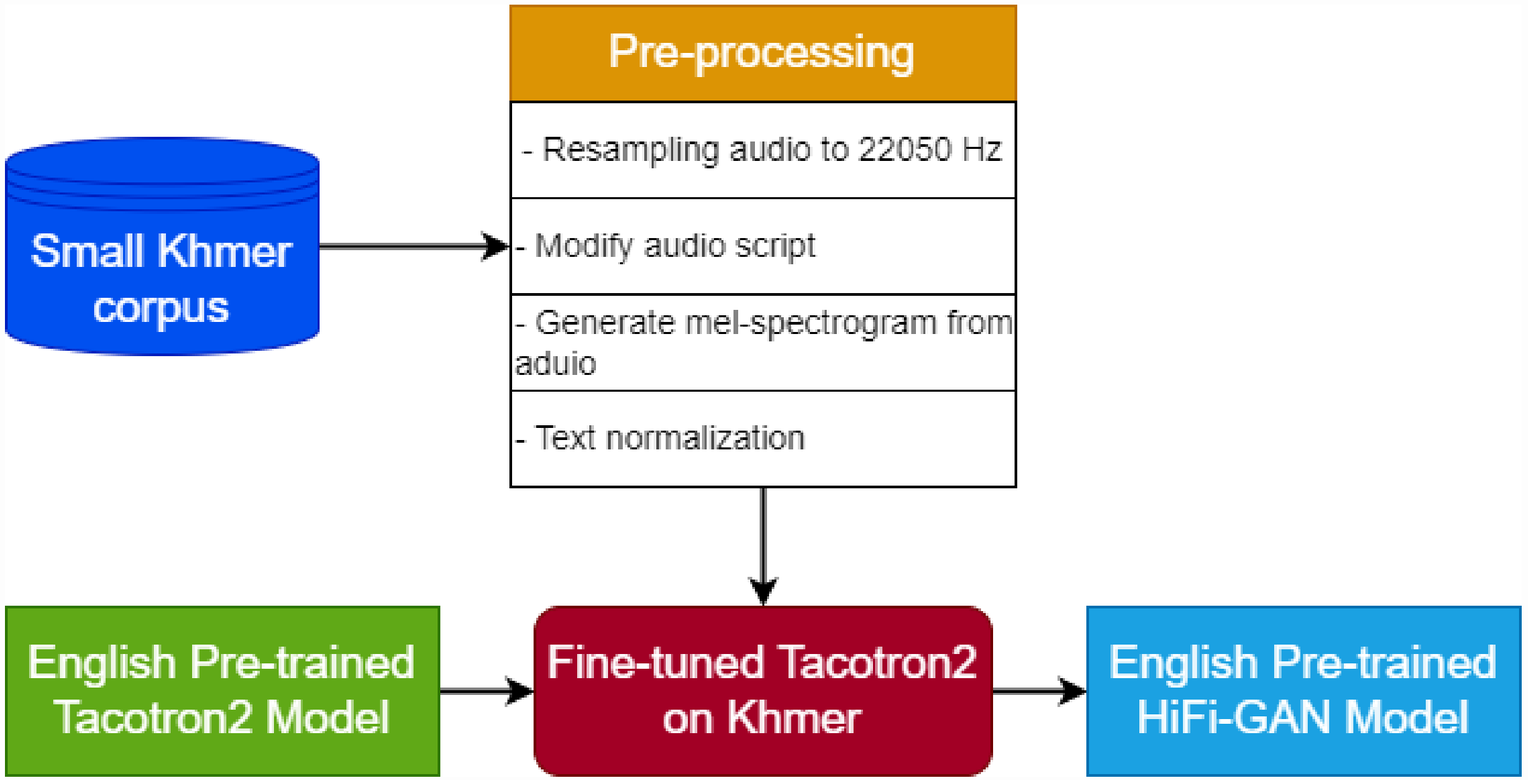
Khmer Independent Vowels

ឥ [ori]	ឦ [orey]	ឧ [oro]	ឩ [orou]	ឪ [orouv]	ឫ [roe]	ឬ [reu]
ឭ [leu]	ឮ [loe]	ឯ [ore]	ឰ [orai]	ឱ [oro]	ឳ [oro]	឴ [orai]

Current Model

CADT IDRI Institute of Digital Research & Innovation

Fine-Tuned Tacotron2



Application in Education

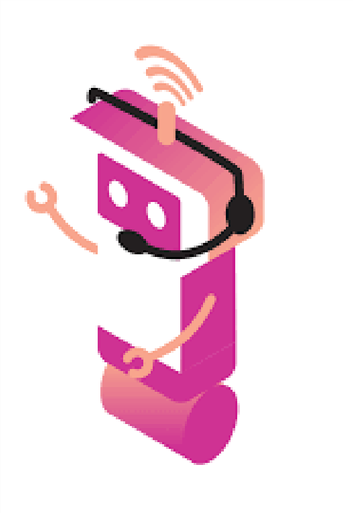
CADT IDRI Institute of Digital Research & Innovation

- **TTS in Language Learning:** it enhances pronunciation and listening skills, helping students practice languages.
- **Assistive Learning:** it supports students with dyslexia and visual impairments by reading text aloud.
- **Reading Comprehension:** it tools like Microsoft Immersive Reader aid in improving reading comprehension for all learners.

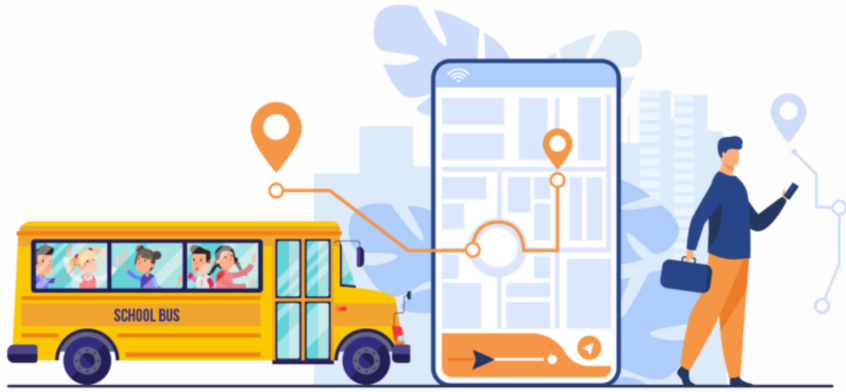
Education for People with Disabilities

- **Enhanced Communication:** effective communication between teachers and students using TTS.
- **Improved Learning & Education:** supports students with learning disabilities by reading text aloud, improving comprehension and engagement.
- **Enhanced Interaction with Technology:** enables hands-free interaction with smart devices and improves accessibility for visually impaired users.

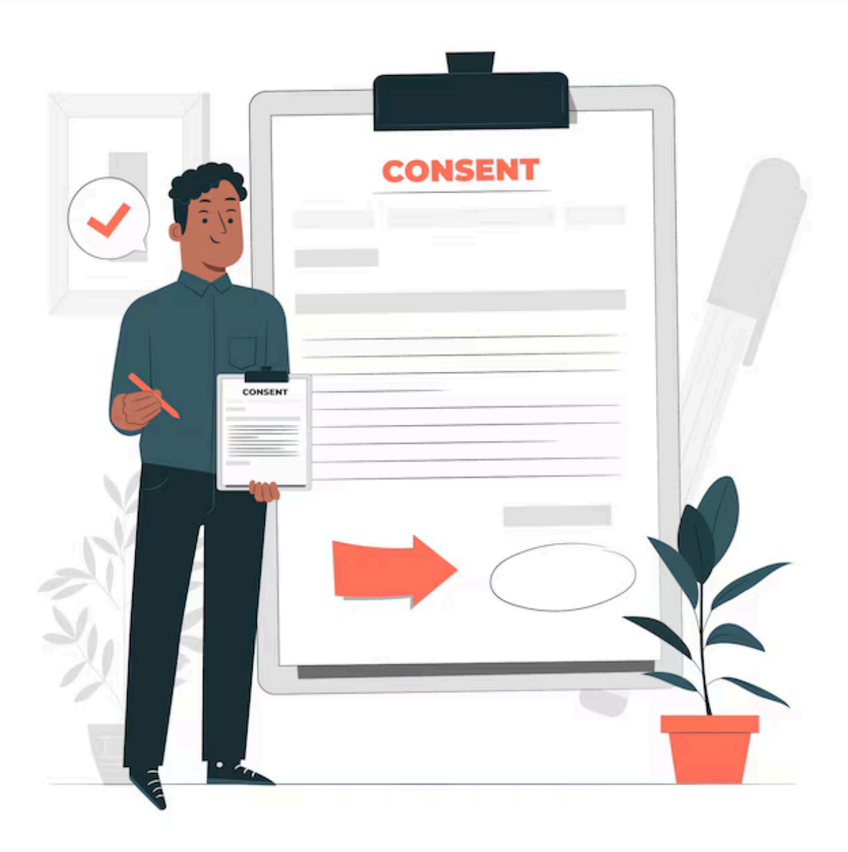
Application in Public Service



- **Public Transportation Systems: it** uses location tracking to trigger automated stop announcements and location guidance.
- **Automated Helplines & Customer Support:** it can power automated customer support systems
- **Healthcare Services:** it can be used in healthcare portals to read aloud medical guidelines



Conclusion



- **Education: TTS can be used to support inclusive education, making learning materials accessible to all students.**
- **Public Services: role of TTS in improving access to government services, healthcare, transportation and public information for citizens with disabilities.**

THANK YOU!

ANY QUESTION?