தளிர் - முதல் கூட்டம்

கலிபோர்னியா, அமெ. மார், 19, 2023

பெயர் காரணம்

- தமிழ் எந்திரவழி மொழிமாதிரிகள் - உத்திகள்
- Thamil AI/ML Models Resources

தேவைகள்

- அடுத்த கட்ட செயலிகளை எளிதில் உருவாக்க AI/ML உதவும்; Chat GPT, DALL-E

தமிழ் சார்ந்த

- மொழி மாதிரிகள்
- தகவல் தரவுகள்
- கலை
- பாரம்பரியம் சார்ந்த விடயங்கள்

பொருள் தேவைகள்

- எந்திர மொழி மாதிரிகளை தயாரிக்க பயிற்சிவிக்க நாள் ஒன்றிற்கு $80-$300 தேவையாகிறது

முதல் கட்ட ஆய்வு

➔ “Tools for constructing AI/ML solutions for Tamil,” (INFITT உத்தமம் 2022) PDF

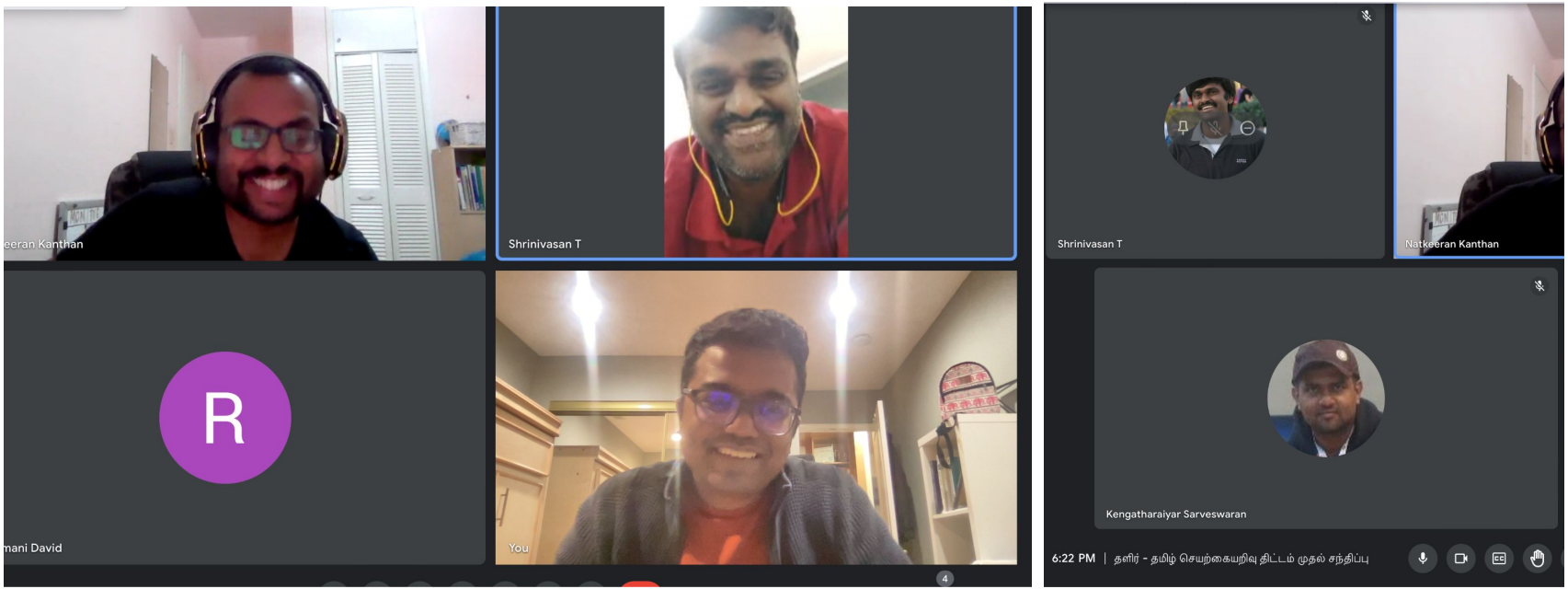
கட்டமைப்பு / குழுக்கள்

- **மொழி மாதிரிகள்** (Language Models)
- Vision Model
- ASR/TTS
- Generative AI

முதல் செயல்பாடுகள்

- உரிமம் தேர்ந்தெடுப்பு
- https://github.com/thamizha/thalir

சந்திப்பு


eeran Kanthan
Shrinivasan T
R
mani David
You
4
Shrinivasan T
Natkeeran Kanthan
Kengatharaiyar Sarveswaran
6:22 PM | தளிர் - தமிழ் செயற்கையறிவு திட்டம் முதல் சந்திப்பு

பின்னூட்டங்கள் மற்றும் அடுத்த கட்ட நடவடிக்கைகள்

Natkeeran

- LOC - English newspaper capture
  - Noolagam - newspaper
  - 300k - 500k pages
  - Computing resource by Canada University Consortium possible if aligned
  - Challenge in training
    - Common Voice Day

A Muthu

- We need socio-cultural review of AI models in Tamil
  a. What are inherent biases ?
  b. Are societal biases / discrimination carried into model ? (mostly yes - but how ?)
  c. Alignment problem
- Able to sponsor
- Working on TTS/ASR

Dr. Sarveswaran

- Creating awareness of AI technologies is important as so much can be accomplished using off-the-shelf tools.
- Building quality resources for Tamil is crucial. Despite the potential for LLMs to handle Tamil well, it may take some time. Therefore, focusing on fundamental building blocks is necessary because none of the existing models has been trained on high-quality data. We don't have enough resources even if we want to fine-tune or test. None of the available resources is representative and is biased towards certain genres, dialects, regions, or domains.
- We need to create templates to democratize the usage of LLMs and AI in Tamil. Additionally, creating a model zoo is a good idea.

பின்னூட்டங்கள் மற்றும் அடுத்த கட்ட நடவடிக்கைகள்

Rajamani David

- Working on Tamil OCR - interested to build a image recognition model
- Able to sponsor

T Srinivasan

- Training from scratch
- Finetuning
  - HW resource
- Inference - application development

TamilPesu.us Model - demo site

- Demo site of Inference Applications
- Application development

Summer of Code Engagement Model

- Raise money from organizations with matching
- Run a program annually / bi-annually / rolling basis
- Repeat till goals are met/funding exhausted

மேலும் பார்க்கவேண்டியவை - **Bias & Harms in AI**

https://docs.google.com/document/u/1/d/1bG0yIdawiUvwh7m0AnXV5W6JHkK9xwXemuVjSU5tbhQ/mobilebasic