

**MAI 475– Large Language Model**  
**IV MSAIM**  
**22-07-2024**

**Regular lab Question**

**Lab Exercise 6: Logical Question Answering in English and Indic using LLM-based Models.**

**Objective**

To implement and analyze the performance of LLM Models on **Logical Question Answering in English and Indic Languages**. The core emphasis is to see **how transformer models understand and resolve logic-based questions**, such as those asked in UPSE/SSC/School Logic Tests in Indic Languages.

**Task Focus: Question Answering – Logical Reasoning**

**Note: You have to take any two types of inferential reasoning (See below) in both Indic and English.**

**1. What is a Logical Question Answering System?**

It involves **inferential reasoning**, such as:

- Syllogism
- Direction-based problems
- Blood relation
- Number series
- Puzzle-based logical questions
- Cause-effect or assumption-based MCQs

**Datasets for Logical QA**

**Public Datasets (Download Links)**

You may use these:

- **SQuAD v2** – General QA in English (for training/ Finetuning)
- <https://github.com/csitfun/LogiQA2.0>
- [iNLTK Hindi QA dataset](#) – For Hindi question answering
- **IndicQA Dataset** – Extractive QA in 11 Indic languages

- **AI4Bharat Logical Reasoning Dataset** – *Logical & reasoning-style QA questions in Indian languages*

### Sample Indic Logical QA:

#### Kannada

##### ಉದಾಹರಣೆ 1 (ರಕ್ತಸಂಬಂಧ)

Q: ರಾಮನ ತಂದೆ ಮೋಹನನ ಮಗನಾಗಿದ್ದಾನೆ. ಮೋಹನನಿಗೆ ರಾಮನೊಂದಿಗೆ ಏನು ಸಂಬಂಧವಿದೆ?

A: ತಾತ (ಅಜ್ಜ)

##### ಉದಾಹರಣೆ 2 (ಸಿಲೋಜಿಸಂ - ತಾರ್ಕಿಕ ಪ್ರಶ್ನೆ)

Q: ಕೆಲವು Aಗಳು Bಗಳು. ಎಲ್ಲಾ Bಗಳು Cಗಳು. ಎಲ್ಲಾ Aಗಳು Cಗಳೇ?

A: ನಿರ್ಧಾರಕ್ಕೆ ಬಾರದು (ತೀರ್ಮಾನಿಸಲು ಸಾಧ್ಯವಿಲ್ಲ)

##### ಉದಾಹರಣೆ 3 (ಭಿನ್ನವಾದದ್ದು ಆಯ್ಕೆಮಾಡಿ)

Q: ಪುಸ್ತಕ, ಮಾಸಪತ್ರಿಕೆ, ಪತ್ರಿಕೆ, ರೇಡಿಯೋ

A: ರೇಡಿಯೋ

#### Hindi

##### Example 1 (Blood Relation)

Q: राम का पिता मोहन का बेटा है। मोहन का राम से क्या संबंध है?

A: दादा (Grandfather)

##### Example 2 (Syllogism)

Q: कुछ A, B हैं। सभी B, C हैं। क्या सभी A, C हैं?

A: निष्कर्ष नहीं निकाला जा सकता (Cannot be determined)

##### Example 3 (Odd One Out)

Q: पुस्तक, पत्रिका, अखबार, रेडियो

A: रेडियो

## Malayalam

### Example 1 (Blood Relation)

Q: രാമിന്റെ അച്ഛൻ മോഹന്റെ മകനാണ്. മോഹൻ രാമിന്റെ എന്താണ്?

A: പതാമഹൻ

### Example 2 (Syllogism)

Q: ചില A-കൾ B-കളാണ്. എല്ലാ B-കളും C-കളാണ്. എല്ലാ A-കളും C-കളാണോ?

A: തീർച്ചയല്ലാതിരിക്കുന്നു (നിർണ്ണയിക്കാൻ കഴിയില്ല)

### Example 3 (Odd One Out)

Q: പുസ്തകം, മാസിക, പത്രം, റേഡിയോ

A: റേഡിയോ

## Tamil

### ഉതാരണം 1 (இரத்த உறவு)

Q: രാമിൻ തന്തെ മോകനിൻ മകൻ. മോകൻ രാമുக்கு என்ன உறவு?

A: தாத்தா

### ഉതാരണം 2 (தர்க்கம் – Syllogism)

Q: சில A-கள் B-கள். அனைத்து B-களும் C-கள். அனைத்து A-களும் C-களா?

A: தீர்மானிக்க முடியாது

### ഉതാരണം 3 (வித்தியாசமானதை கண்டறிதல்)

Q: புத்தகம், மாத இதழ், செய்தித்தாள், வானொலி

A: வானொலி

## Performance Evaluation

### Metric

### Explanation

EM (Exact Match) Checks if prediction matches label

F1 Score Overlap between predicted & gold span

Human Judgment Accuracy, logic clarity, and fluency

Compare the performance Evaluation Metrics of both the models using in English and Indic Languages. Prepare the chart as follows:

Model	Language	Task	Accuracy / EM	F1-Score	Precision	Recall	Notes
<b>BERT-Base (uncased)</b>	English	QA (SQuAD)	80.8% (EM)	88.5%	-	-	Extractive QA
<b>mBERT (cased)</b>	Hindi	QA (IndicQA)	70.4% (EM)	78.2%	-	-	Logic & factual
<b>IndicBERT</b>	Kannada	QA (Logic - IndicQA)	66.2% (EM)	74.4%	-	-	Logical reasoning
<b>GPT-2 (fine-tuned)</b>	English	QA (generative logic)	-	85.0%	-	-	Evaluated manually
<b>mT5-Base</b>	Tamil	QA (Translation + Logic)	60.0% (EM)	72.5%	-	-	mT5 + fine-tuned

NB: No marks will be credited for the pipeline implementation.

### Program Evaluation Rubrics

Model Selection and Implementation	6 Marks
Timely Submission	2 Marks
Viva	2 Marks

### General Instructions

- The file you have to save with your name, last 3 digits of register number and program number "Anto\_501\_Lab1".
- The implemented code you have to download and upload in the Google Class room in the given scheduled time.