

# LogiQA – A Logical Reasoning Dataset

CS 297

Presented To

Dr. Chris Pollett

Presented By

Adithya Harish Kumar

# Overview

- A Challenge Dataset for Machine Reading Comprehension for testing logical reasoning in natural language understanding
- 8,678 QA instances covering multiple types of deductive reasoning
- Sourced from expert-written questions for testing human logical reasoning
- Available in both Chinese and English

# Details

- Dataset Composition –
  - Total: 8,678 QA instances
  - Train: 7,376
  - Validation: 651
  - Test: 651
- Structure: Each problem consists of 8 lines:
  - Blank line
  - Correct answer choice
  - Context
  - Question
  - 5-8 contains four options

# Sample Question

- a
- In the planning of a new district in a township, it was decided to build a special community in the southeast, northwest, centered on the citizen park. These four communities are designated as cultural area, leisure area, commercial area and administrative service area. It is known that the administrative service area is southwest of the cultural area, and the cultural area is southeast of the leisure area.
- Based on the above statement, which of the following can be derived?
- A. Civic Park is north of the administrative service area
- B. The leisure area is southwest of the cultural area
- C. The cultural district is in the northeast of the business district
- D. The business district is southeast of the leisure area

# Applications

- Benchmark for reinvestigating logical AI in deep learning NLP
- Challenges models to improve logical reasoning capabilities
- Freely available for research: <https://github.com/lgw863/LogiQA-dataset>