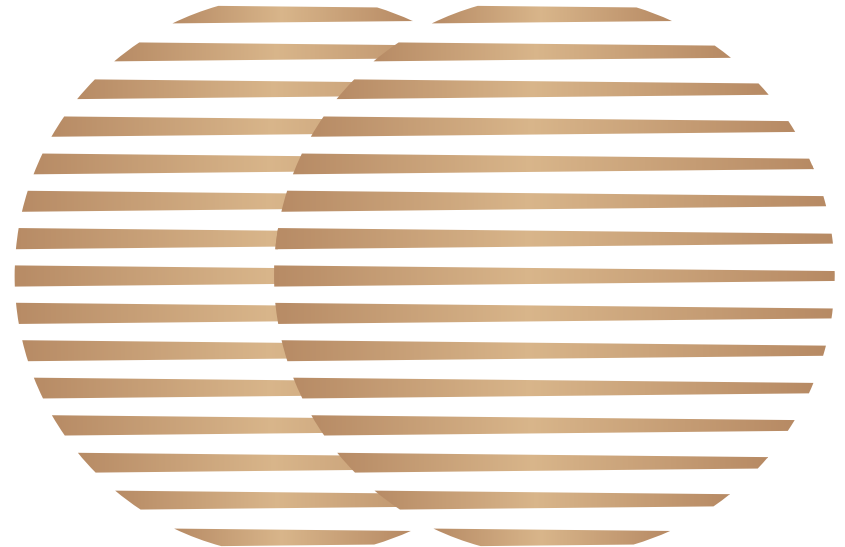


2024년 3월 5일 Study Meeting

# 논리 오류를 다시 물어보자

Let's ask about the logical error again

정지원  
성균관대학교 인공지능학과  
석사과정  
jwjw9603@g.skku.edu



## Contents

# 01

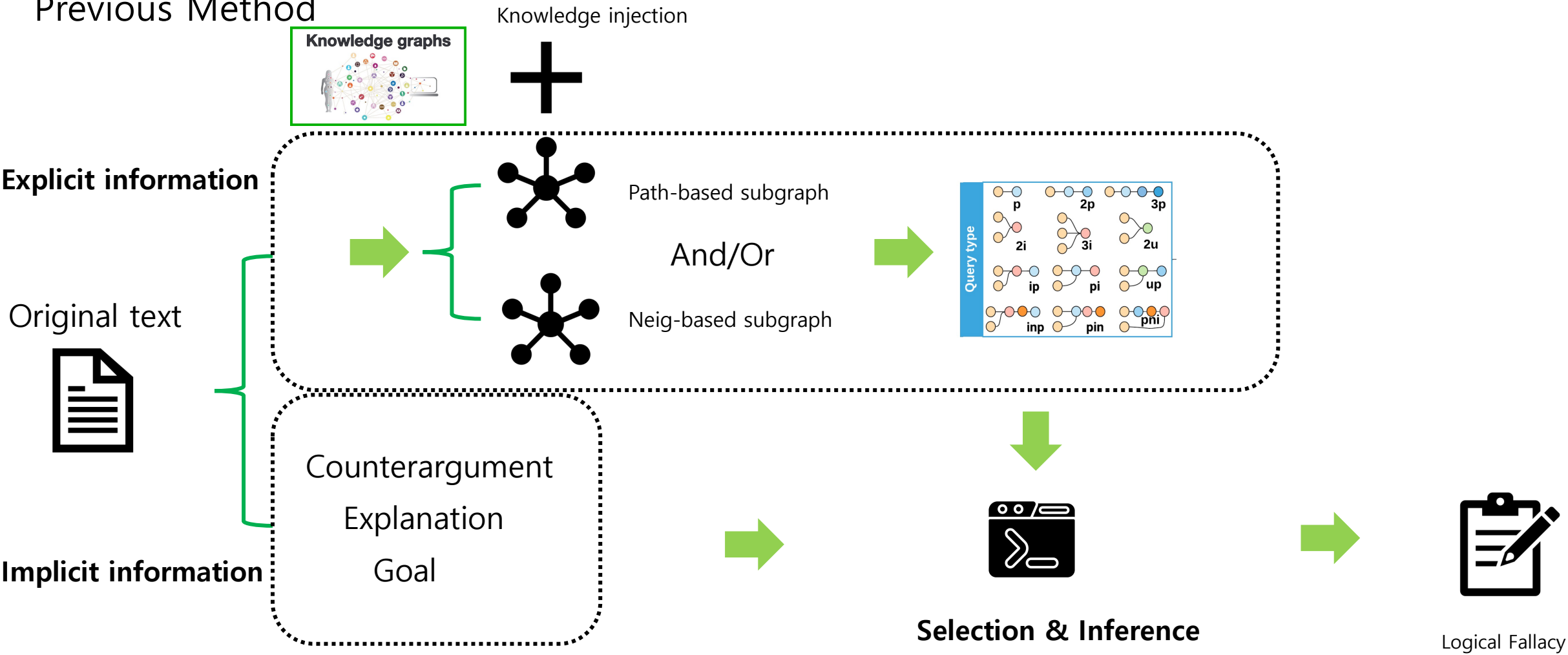
## Progress

- 진행 내용 Overview
- Previous Method
- Previous Method 문제점
- Faulty Generalization
- Experiment

## 진행 내용 Overview

- 이전 미팅 방법론에 대한 고찰
- 새로운 방법론 제시
- 실험

Previous Method



## Previous Method

1. Original text로부터 다양한 representation text(반문, 목표, 설명)에서 subgraph를 만드는 것이 아니라 original text로부터 subgraph를 생성한다. 생성된 subgraph의 정보를 **Explicit Information**이라 한다.
2. Original text로부터 다양한 representation text(반문, 목표, 설정)을 추출한다. 이 정보를 **Implicit Information**이라 한다.
3. **Explicit information**과 **Implicit information**을 가지고 Selection & Inference prompting을 진행한다.
  - 1) 여기서 selection은 Explicit information(subgraph정보)를 사용할 지, Implicit information을 선택할 지 고민하는 것이고, Inference는 선택된 정보를 기반으로 Logical fallacy를 감지하는 역할

## Previous Method 한계점 & 의문점

1. 논리 오류를 가지고 있는 문장을 기반으로 지식 그래프를 사용해서 subgraph를 만든다. 만들어진 subgraph가 어떻게 논리 오류를 해결할 수 있는데?
    - 1) Subgraph가 어떤 역할을 해줄 수 있는지에 대한 확신이 없음
    - 2) 논리 오류와 어떤 관계가 있는지 알 수 없음
  2. Explicit information과 Implicit information을 어떻게 선택하고, 추론할건데?
    - 1) 그래프의 정보와 텍스트의 정보를 어떻게 선택하고 추론할 지에 대한 구체화가 없음
  3. RoG방법론
    - 1) Subgraph가 아닌 RoG방법을 사용하더라도 RoG는 QA task이며, 이 방법을 그대로 사용하는 것은 novelty가 없음
    - 2) 논리 오류를 해결하기 위한 접근이 아님
- 논리 오류의 근본적인 문제점을 해결하기 위한 접근을 하지 못함. 단순히 지식 그래프를 어떻게 사용할까에 대해서만 고민했기 때문

## Faulty Generalization(Hasty Generalization)

"모든 백인 소녀들이 스타벅스를 좋아하기 때문에 애니도 스타벅스를 좋아할 것입니다."
올해 라스베거스는 작년에 비해 더 따뜻합니다. 그러므로 지구 온난화는 급속히 가속화되고 있습니다.
"내가 University of Florida의에서 수강한 두 과목은 그다지 흥미롭지 않았습니다. 좋은 대학이 아니라고 생각합니다."
뉴욕 번호판을 가진 운전자가 교통 정체를 방해합니다. 당신은 뉴욕의 모든 운전자가 형편없는 운전자라고 결정합니다.
"치과의사 5명 중 4명이 해피글로시 치약을 추천하니 정말 좋을 것 같아요."
가족 중에 서핑하는 사람이 없나요? 그런데 당신이 그 전에 캘리포니아에 살았다고 말한 것 같아요.
모든 여성은 나쁜 운전자입니다.
새로운 나라를 방문했는데 공항에서 처음 만난 사람은 무례했습니다. 당신은 이 새로운 나라의 모든 사람들이 무례하다는 메시지를 집에 있는 친구에게 보냅니다.
Kevin의 조부모님은 컴퓨터 사용법을 모릅니다. 케빈은 모든 노인들은 컴퓨터에 대해 잘 알지 못한다고 생각합니다.
남자들은 절대 길을 묻지 않는다고 하던데
좋은 식단을 섭취하면 심장 마비가 발생하지 않습니다.
십대들이 게으르다는 것은 누구나 알고 있다.
MeowMeowTweet은 미국산 스킨케어 제품입니다. 나는 그것을 사용했고 나에게 여드름이 생겼습니다. 미국산 스킨케어는 여드름을 유발합니다.
모든 선생님은 커피를 마시고 선생님은 사람입니다. 그러므로 모든 사람들은 커피를 마신다.
4살 아이들은 모두 말을 너무 많이 합니다.
나는 치즈 먹는 것을 좋아하는 키 큰 남자를 만났습니다. 이제 나는 키가 큰 사람들은 모두 치즈를 좋아한다고 믿습니다.
나는 성격이 아주 나쁜 빨간 머리 세 명을 알고 있는데, 애너벨도 빨간 머리를 가지고 있으니 그 사람도 성격이 아주 나쁜 것 같아요.
모든 소녀들이 쇼핑을 좋아하기 때문에 이 새로운 MallTime 게임 앱이 성공할 것이라고 생각합니다.
대도시 사람들은 작은 마을 사람들만큼 친절하지 않습니다.
내 동생 Humza는 치즈버거와 피자를 먹습니다. 그는 말했다. 치즈버거와 피자는 비만의 원인이 아닙니다.
"내 룸메이트인 파티마는 자기 철학 수업도 힘들고 내가 듣는 수업도 힘들다고 하더군요. 철학 수업은 다 힘들겠구나!" 바니가 주장했다.
방금 남부 사람과 통화 중이었는데 간신히 따라갈 수 있었어요! 남부 사람들은 모두 말을 빨리 합니다.
나에게는 여자 친구가 여러 명 있는데 그들 모두는 컨트리 음악을 좋아합니다. 모든 여성은 컨트리 음악을 좋아합니다.
그런 종류의 자동차는 제대로 만들어지지 않았습니다. 내 친구 하나가 있는데, 그것이 계속해서 그에게 문제를 안겨줍니다.
누구나 Post Malone을 좋아하며 그의 노래 중 적어도 하나의 이름을 말할 수 있습니다! (이것은 예입니다 ...)
"저 새 집 정말 멋져요! 사세요!"

## Faulty Generalization(Hasty Generalization)

1. 성급한 일반화(일반화의 오류)가 뭔지, 이 데이터셋은 어떻게 생겼는지를 다시 한 번 확인함.
  - 1) "A student argues that because they got an 'A' on a test without studying, studying is not necessary for success."
  - 2) 한 학생의 경험을 기반으로 성공의 필요조건을 결정짓는다.
2. 성급한 일반화의 구조는 다음과 같이 정리 할 수 있음.([https://en.wikipedia.org/wiki/Faulty\\_generalization](https://en.wikipedia.org/wiki/Faulty_generalization))
  - 1) The proportion **Q** of the sample has attribute **A**. Therefore, the proportion **Q** of the population has attribute **A**.
  - 2) Q : 샘플 데이터에서 특정 특성이나 속성을 가진 개체 또는 사례, 예시를 의미
  - 3) A : 주어진 문장, 문맥에서 관심 있는 특징이나 **주제**를 의미

➤ 즉, Q라는 일부분의 예시로 A라는 주장을 주장하는 형태가 일반화의 오류!! 이것은 Q나 A에 대한 검증이 필요하다는 것
3. 일반화 오류의 특징은 다음과 같음.
  - 1) 특정한 사례나 상황을 일반적인 규칙으로 확대(일반화)
  - 2) 충분한 증거나 데이터 없이 일반화 시도(부족한 증거)
  - 3) “모든”, “항상”, “절대” 등의 절대적인 양식(절대적인 양식)
  - 4) 특수한 상황이나 예외적인 사례를 고려하지 않음(예외 무시)
  - 5) 특정한 경우를 지나치게 과장(극단화), 그룹의 다양성을 고려하지 않고 단일한 패턴으로 일반화(고정관념)



## Faulty Generalization(Hasty Generalization)

### 4. 그렇다면 일반화의 오류를 어떻게 해결할 수 있을까?

- 1) **예외처리** : 일반화한 주장에 대한 예외를 식별하고 처리 -> 예외 처리를 할 수 있는 **질문 만들기**
- 2) **정보확장** : 지식 그래프를 통해 추가적인 정보를 얻어서 주장을 보다 적합하게 지원
- 3) **추론과 추정** : 지식 그래프를 사용하여 일반적인 패턴을 추론하고 예측 -> relation path 사용
- 4) **문맥 이해** : 해당 문장의 문맥을 이해하고 분석 -> various representation text 추출
- 5) **다양한 관점 고려** : Selection & Inference

### 5. 특정한 사례나 상황을 일반적인 규칙으로 확대하는 문장(일반화)에서 주장(A)이나 예시(Q)를 질문 형태로 만들고, 만들어진 질문 문장에 RoG 방법을 적용시키자.

- 1) **주어진 문장을 Question(query)로 만드는 것은 일종의 A(주장)및 예시(Q)를 확인하는 것, A와 Q의 관계를 확인하는 것**
- 2) RoG 방법론은 relation path로 계획(Plan)을 잡고 외부 정보(KG)를 활용해서 추론 과정(Reasoning path)을 선택하여 해결한다(Solve).

### 6. RoG방법론을 거치면 최종적으로 여러 Reasoning path가 나오는데, 이 Reasoning\_Path들은 다양한 관점, 특징에 따라 다양하게 나오기 때문에 해결하고자 하는 문장에서 발생하는 일반화의 오류 문제를 다양한 원인에 따라 해결하는데 도움을 줄 수 있는 가장 적합한 reasoning path를 선택하면 됨. 이것은 Explainability도 있음.

### 7. 이번에는 방법론을 먼저 결정짓고 구현하지 말고 단계별로 검증하고 지나가기

## Setting

1. 아이디어(방법론)을 fix하고 구현하기 전에 단계별로 검증을 진행한다.
  2. 논리 오류 문장을 질문(쿼리)형태로 만들어서 재질문하는 방식으로 성능을 확인해본다.
  3. 질문을 만들기 위해 데이터셋을 확인해서 패턴을 찾고 패턴을 기반으로 질문 형성의 규칙을 정한다.
  4. LOGIC 데이터셋에서 진행했으며, 총 480개의 Faulty generalization 텍스트 중 40개의 텍스트를 확인하고 패턴을 찾는다.
  5. 질문은 Faulty Generalization 문장에 대해서만 만든다.
  6. 13개의 레이블마다 40개씩 랜덤으로 뽑아서(520개), Faulty generalization와 나머지로 구분하는 이진 분류로 진행한다.
  7. Faulty Generalization을 잘못 예측할 시에만, text로부터 만든 Question으로 재 질문한다.
- ◆ 추가실험 : Text로부터 만든 Question의 영향성을 파악하기 위해, 정답(레이블)을 틀릴 시, Question을 사용하지 않고, 단순히 rethink하라는 prompt만 넣었을 때의 성능을 확인함 -> Rethink가 성능 향상을 줬을 수도 있어서, Rethink만 하는 실험을 추가함.

## Setting

### Logic [\[ edit \]](#)

---

A faulty generalization often follows the following format:

The proportion Q of the sample has attribute A.

Therefore, the proportion Q of the population has attribute A.

Such a generalization proceeds from a premise about a [sample](#) (often unrepresentative or biased), to a conclusion about the population itself.<sup>[3]</sup>

Faulty generalization is also a mode of thinking that takes the experiences of one person or one group, and incorrectly extends it to another.

- Q와 A의 대소 비교를 진행함. (e.g.  $A > Q$ ,  $A < Q$ )

## Setting(Example)

1. "Annie must like Starbucks because all white girls like Starbucks."

- 한국어 : "모든 백인 소녀들이 스타벅스를 좋아하기 때문에 애니도 스타벅스를 좋아할 것입니다."
- A : Annie must like starbucks
- Q : All white girls like starbucks
- A(애니) < Q(백인소녀)

2. It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.

- 한국어 : 올해 라스베이거스는 작년에 비해 더 따뜻합니다. 그러므로 지구 온난화는 급속히 가속화되고 있습니다.
- A : global warming is rapidly accelerating.
- Q : It is warmer this year in Las Vegas as compared to last year.
- A > Q
- Question : Does the fact that Las Vegas is warmer this year compared to last year necessarily imply that global warming is rapidly accelerating?
- Question2 : Is global warming rapidly decreasing?

3. "The two courses I took at UF were not very interesting. I don't think its a good university."

- 한국어 : "내가 University of Florida의에서 수강한 두 과목은 그다지 흥미롭지 않았습니니다. 좋은 대학이 아니라고 생각합니다."
- A : I don't think its a good university.
- Q : The two courses I took at UF were not very interesting.
- A > Q
- Question: Did the lack of interest in the two courses you took at UF lead you to believe it's not a good university?
- Question2: What was your overall experience like at UF besides these two courses? or Isn't the University of Florida a good university?

- 40개의 문장을 확인했으며 비율은 다음과 같음

- A > Q : 35개
- A < Q : 5개

- Q : 샘플 데이터에서 특정 특성이나 속성을 가진 개체 도는 사례, 예시를 의미
- A : 주어진 문장, 문맥에서 관심있는 특징이나 주제를 의미
- **A > Q문장은** 경험한, 주위 사람에게 들은, 본 내용(Q)을 기반으로 일반화적인 주장(A)을 하는 경우
- **A < Q문장은** 예외를 무시하고 대중적인 주장, 관념(Q)을 기반으로 소수의 경우, 주장(A)을 일반화 시키는 경우

## Experiment(Setting)

1.  $A \langle Q$  문장은 단순히 **Q가 사실인지를 되묻는 형태의 질문이면** 된다. 왜냐하면 이 문장은 소수(한 명, 친구, 가족)에서 주장하는 내용(A)을 대충적인 관념(Q)을 기반으로 사용하기 때문에, 근본적으로 Q가 사실인지를 직접적으로 물으면 논리 오류를 해결할 수 있음.
2.  $A \rangle Q$  문장은 경험한, 주위 사람에게 들은, 본 내용(Q)을 기반으로 일반화적인 주장(A)를 하는 경우로, 이런 문장 같은 경우에는 **A와 Q의 관계를 묻거나(Question), A를 되묻는 형태의 질문(Question2)**을 하면 된다.
  - 1) Question : A와 Q의 관계를 묻는 형태로 **일반화된 주장에 대한 논리적 결함을 직접적으로 다룸**
    - a. 주장의 근거와 일반화된 결론 간의 관계를 더 명확하게 이해할 수 있지만, 지식 그래프가 사용되기에 쉬운 질문 형태가 아님
    - b. 질문을 만들 때 고유명사(사람 이름, 회사, 앱 등)는 지식 그래프에 없을 경우가 있음
  - 2) Question2 : A를 되묻는 경우로 **주장된 결과를 의심하거나 부정할 수 있도록 유도하는 형태**
    - a. 일반화적인 주장을 직접적으로 묻는 형태로서, A에 대한 답을 추론하는 과정에 적절함
    - b. 주장(A)으로만 구성되어 있는 문장들에게 적합함.(e.g. All four year olds talk too much.)
3.  $A \rangle Q$  문장은 A를 되묻는 형태만으로 충분할 경우에는 Question2, 아닐 경우에는 Question으로 다루면 어떨까?

## Experiment(make Question)



User

I want to create a sentence with a hasty generalization error, along with Q and A, and create **Question** and **Question2**.

**Question** is intended to be in the form of a question based on the relationship between Q and A, While **Question2** aims to form a question with A in an interrogative form.

The proportion Q of the sample has attribute A. Therefore, the proportion Q of the population has attribute A.

Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.

A : global warming is rapidly accelerating.

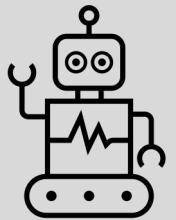
Q : It is warmer this year in Las Vegas as compared to last year.

A > Q

Question :

Question2 :

- Question : Does the fact that Las Vegas is warmer this year compared to last year necessarily imply that global warming is rapidly accelerating?
- Question2 : Is global warming rapidly decreasing?



Assistant

A < Q 문장 같은 경우는 Human Annotating

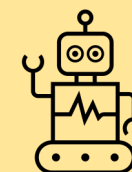
## Experiment(Method)



User

Your task is to detect a fallacy in the Text. The label can be "Faulty Generalization" and "The Other Fallacy".  
We refer to those logical errors but not falling into the "Faulty Generalization" as "The Other Fallacy".  
Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.  
Label :

The Other Fallacy



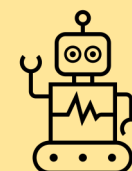
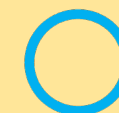
Assistant



User

I ask you to rethink. The Question is about the text you misclassified.  
Please detect a fallacy in the text based on the Question. The label can be "Faulty Generalization" and "The Other Fallacy".  
Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.  
Question : Does the fact that Las Vegas is warmer this year compared to last year necessarily imply that global warming is rapidly accelerating?  
Question2 : Is global warming rapidly decreasing?  
Label :

Faulty Generalization



Assistant

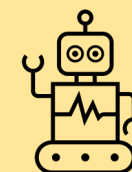
## Experiment(Method)



User

Your task is to detect a fallacy in the Text. The label can be "Faulty Generalization" and "The Other Fallacy".  
We refer to those logical errors but not falling into the "Faulty Generalization" as "The Other Fallacy".  
Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.  
Label :

The Other Fallacy



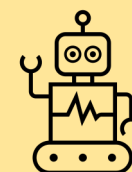
Assistant



User

**I ask you to rethink.**  
The label can be " Faulty Generalization" and "The Other Fallacy".  
Text : It is warmer this year in Las Vegas as compared to last year; therefore, global warming is rapidly accelerating.  
Label :

The Other Fallacy



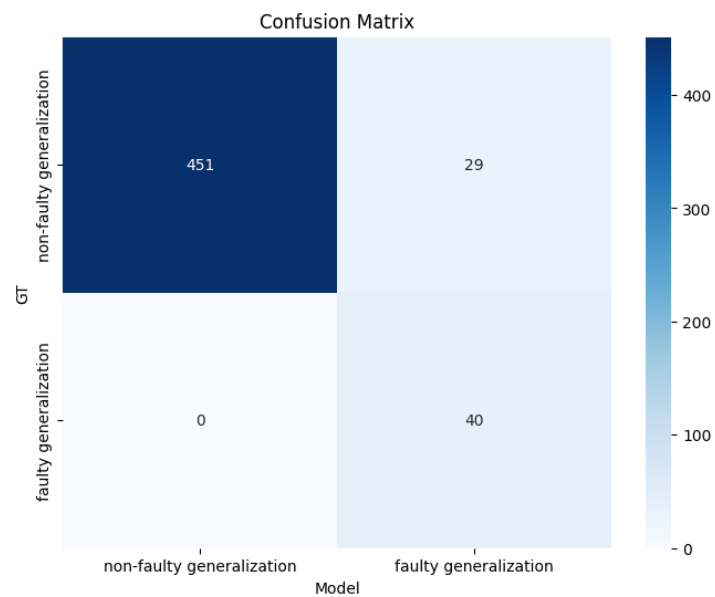
Assistant



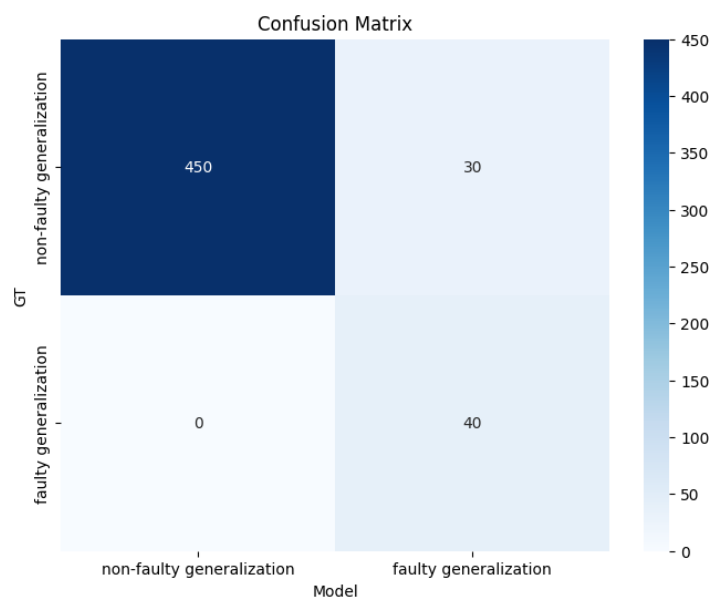
## Experiment(Result)

Type/Metric	Total Accuracy	Precision	Recall	F1
No Question	0.9( $\pm 0.01$ )	0.33( $\pm 0.03$ )	0.3( $\pm 0.01$ )	0.32( $\pm 0.02$ )
Question	0.95( $\pm 0.42$ )	0.61( $\pm 0.02$ )	0.99( $\pm 0.01$ )	0.75( $\pm 0.01$ )
Question2	0.95( $\pm 0.01$ )	0.61( $\pm 0.03$ )	0.98( $\pm 0.02$ )	0.75( $\pm 0.02$ )
Rethink	0.9( $\pm 0.01$ )	0.31( $\pm 0.04$ )	0.26( $\pm 0.03$ )	0.28( $\pm 0.03$ )

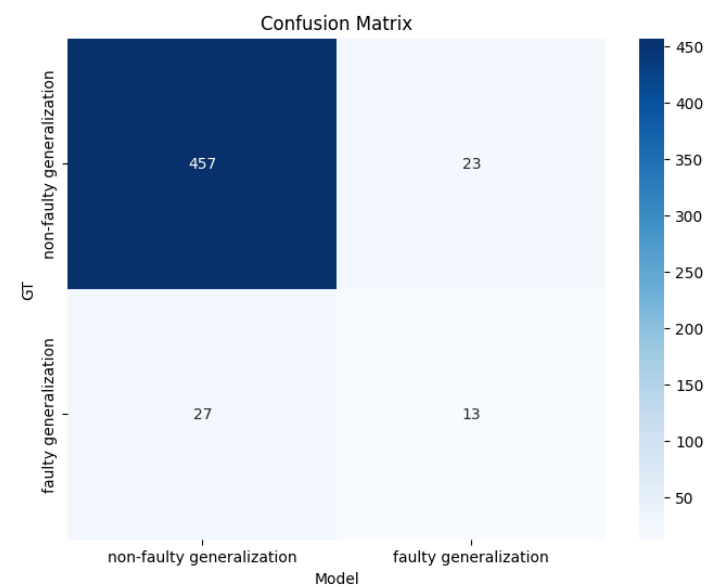
## Experiment(Confusion Matrix)



Question



Question2



No Question

## Experiment(Analysis)

1. 비록 Faulty Generalization 데이터 개수가 적더라도 상당한 성능 향상이 있었음
2. Question과 Question2의 성능 차이는 없었음, 다만 지식 그래프를 사용하기에는 Question2가 더 적합하지 않을까?(확실하지 않음)
3. Faulty Generalization 클래스 뿐만 아니라 다른 클래스에도 질문 형태를 만드는 방향이 맞을까?
  - 1) 만들더라도 일반화의 오류와는 다른 규칙이 있을 것임
  - 2) False Causality, Irrelevant Authority , Cherry picking, Post Hoc 클래스
4. 아니면 Faulty Generalization에 대해서만 진행할 것인가?
  - 1) 성급한 일반화 오류 데이터를 더 만들고(data augmentation), 오류가 없는 문장을 만들어서 이진 분류하는 task

# 감사합니다

발표 경청해 주셔서 감사합니다

정지원 성균관대학교 인공지능학과 석사 과정  
jwjw9603@g.skku.edu

