

인간 시선 기반의 텍스트 하이라이팅이 의사결정에 미치는 영향에 대한 시범 연구

한양대학교 이창현, 차경진

Introduction

- 텍스트 하이라이팅은 글의 가독성을 높이고 작성자가 의도한 이해를 용이하게 만드는 단서를 남기는 것 [Ponce & Mayer, 2014]
- 최근에는 XAI (explainable AI) 등 AI를 활용하여 텍스트 하이라이팅을 적용하기도 함 [Ribeiro et al., 2016; Lundberg & Lee, 2017]
- 그런데 주관적 판단을 보조하기 위해 텍스트 하이라이팅을 하는 경우에, XAI를 활용하는 것은 오히려 편향된 판단을 초래할 수 있음
- 가령, 자기소개서를 평가한 인간의 판단이 편향된 경우, 그 판단을 학습한 AI도 편향될 수 있음 [Obermeyer & Lee, 2017; Ahsen et al., 2019]
- 이 때 AI의 높은 성능은 AI가 인간의 판단을 얼마나 잘 모방하는지에 대한 지표일 뿐, 판단의 건전성을 보증하지 않음 [Lee & Cha, 2023]
- 이 경우, AI의 설명가능성(explainability)은 판단의 건전성을 인간이 판단할 수 있는 근거가 됨 [Lee & Cha, 2023]
- 그런데 XAI는 레이블을 가장 잘 설명하는 요인을 찾기 때문에, 편향된 레이블이 주어진 경우에 신뢰할 수 없음

Introduction

- 텍스트 하이라이팅은 글의 가독성을 높이고 작성자가 의도한 이해를 용이하게 만드는 단서를 남기는 것 [Ponce & Mayer, 2014]
- 최근에는 XAI (explainable AI) 등 AI를 활용하여 텍스트 하이라이팅을 적용하기도 함 [Ribeiro et al., 2016; Lundberg & Lee, 2017]
- 그런데 주관적 판단을 보조하기 위해 텍스트 하이라이팅을 하는 경우에, XAI를 활용하는 것은 오히려 편향된 판단을 초래할 수 있음

RQ1. 인간의 시선은 텍스트 하이라이팅을 위한 신뢰할 수 있는 단서인가?

RQ2. 인간 시선에 기반한 텍스트 하이라이팅은 편향의 증폭 없이

- 가령, 자기소개서를 평가할 때 AI가 인간의 시선과 유사한 단서를 활용하면, 편향된 판단을 초래할 수 있음 [Lundberg & Lee, 2017; Ahsen et al., 2019]
- 이 때 AI의 높은 성능은 AI가 인간의 판단 속도를 뛰어넘는다는 것을 의미하지만, 인간의 판단의 건전성을 보장하지 않음 [Lee & Cha, 2023]
- 이 경우, AI의 설명가능성(explainability)은 판단의 건전성을 인간이 판단할 수 있는 근거가 됨 [Lee & Cha, 2023]
- 그런데 XAI는 레이블을 가장 잘 설명하는 요인을 찾기 때문에, 편향된 레이블이 주어진 경우에 신뢰할 수 없음

Literature Review

Text Highlighting

- 선행 연구에서 제시한 텍스트 하이라이팅에는 다양한 방법이 있음
 - Rule-based:의도에 맞는 단어 [Ponce & Mayer] 혹은 전문 용어 [Dudko et al., 2017]에 직접 하이라이팅하는 방법
 - 하이라이팅 되는 단어의 풀이 정해져 있음
 - Label-based: XAI [Ghassemi et al., 2021; Ozyegen et al., 2022]나 어텐션 레이어 [Lam et al., 2022]를 활용하여 레이블을 가장 잘 설명하는 텍스트를 추적하는 방법
 - 레이블을 참조하므로, 레이블을 신뢰할 수 없는 경우에 채택 불가
 - Encoder-decoder: 트랜스포머 구조 [Altundogan & Karakose, 2023] 등을 활용하여 글을 요약할 때 남길 텍스트를 추적하는 방법
 - 일반적인 상황에 fitting되어 있으므로, 작성자의 의도를 반영할 수 없음

Literature Review

The Paradox of Human-AI Interaction

- 인간은 정보가 너무 많아도 (complexity), 정보가 너무 적어도 (uncertainty) 합리적 판단을 내리지 못함 [Simon, 1957]
- 인간은 정보가 해석 불가능해도 (uncertainty), 해석 가능성이 너무 많아도 (equivocality) 합리적 판단을 내리지 못함 [Daft & Lengel, 1986]
- 이 때 human-AI interaction은 인간의 제한된 합리성을 보완할 수 있으며 [Jarrahi, 2018] 이를 위해서는 설명 가능해야 함 [Lee & Cha, 2023]
- 그러나 주관적인 판단의 맥락에서, 신뢰할 수 없는 레이블을 참조하는 label-based 텍스트 하이라이팅도, 판단의 의도를 반영할 수 없는 encoder-decoder 기반의 텍스트 하이라이팅도 만족스러운 human-AI interaction을 기대하기 어려움
- 이들은 오히려 편향을 증폭시키거나, 특색 없는 일반적인 의사결정만을 지원할 가능성이 높음 [Ahsen et al., 2019]

Literature Review

E-Z Reader Model

- 시각적 정보를 통한 인간의 인지는 Fixation과 Saccade를 통해 이해할 수 있음 [Henderson, 2003]
 - Fixation: 시선을 고정하고 오랫동안 바라보는 행위 (주로 집중과 관련이 있음)
 - Saccade: 시선의 빠른 이동 (주로 탐색과 관련이 있음)
- E-Z Reader Model: 읽을 때 중요하다고 생각되는 경우, 맥락을 추출할 때까지 그 줄에 시선을 고정하는 경향이 있음 [Reichle et al., 2003]

즉, 문해에 도움이 되는 부분에서는 Fixation이 발생하고, 그렇지 않은 부분에서는 Saccade가 발생한다고 기대해볼 수 있음

- 다만 글의 중요도에만 영향을 받는 것은 아님
 - 글의 난이도: 글이 어렵다고 인식되는 경우, 추가적인 정보를 획득하기 위해 Saccade하거나, 깊게 고민하기 위해

Fixation할 수 있음

Hypotheses Development

Human Gaze and Reading Literacy

- 문해력이란 주어진 글에서, 주어진 맥락에 필요한 정보를 추출/구축/통합/비판할 수 있는 능력 [Frankel et al., 2016]
 - 문해력을 측정하는 여러 지표가 있지만, 주어진 맥락에 적절한 정보 처리를 할 수 있는가가 결국 핵심 [Johnston & Costello, 2005]
 - 문해력이 높은 군: 주어진 맥락에서 중요한 정보가 무엇인지 일관적으로 파악하고, 거기에 Fixation을 보일 것으로 보임
 - 문해력이 낮은 군: 중요한 정보를 파악하는 능력이 떨어지므로, 주어진 맥락에서 중요한 정보가 무엇인지 파악하기 어려울 것으로 보임
- 정보를 파악하기 어렵다면, 거기에 집중하기 위한 Fixation, 혹은 추가 정보를 습득하기 위한 Saccade를 선택할 수 있음

Hypotheses Development

The Impact of Human Gaze-based Highlighting on Decision-making

- 텍스트 하이라이팅은 중요하지 않은 신호에 덜 집중하게 해 줌으로써 complexity를 낮출 수 있을 것으로 기대됨 [Jarrahi, 2018]
- 그러나 텍스트 하이라이팅은 사람들의 시선이 하이라이트된 텍스트에 더 많이 꽂히게 하는 효과도 있음 [Ponce & Mayer, 2014]
- 따라서 텍스트 하이라이팅은 의사결정의 속도 및 정확도를 높일 수 있을 것으로 보이나, 편향도 높일 것으로 보임 [Artelt et al., 2001]

H2a. 하이라이팅을 참조하여 의사결정을 내린 사람들은, 그렇지 않은 사람들에 비해 **의사결정의 속도**가 더 빠를 것이다.

H2b. 하이라이팅을 참조하여 의사결정을 내린 사람들은, 그렇지 않은 사람들에 비해 **의사결정의 정확도**가 더 높을 것이다.

H2c. 하이라이팅을 참조하여 의사결정을 내린 사람들은, 그렇지 않은 사람들에 비해 **서로 비슷한 유형의 오류**를 낼 것이다.

Methods

Highlighting Model Construction

- 시선 추적 기술을 가진 스타트업 기업 Visual Camp로부터 초등학생 시선 데이터를 받아 모델 구축
- 167 명의 학생들이 125-199 단어 분량의 글을 읽고 간단한 O/X 퀴즈에 답하는 과정을 시선 추적
- 10점 만점 중 9점 이상을 문해력이 높은 군, 6점 이하를 문해력이 낮은 군으로 구분
- 시선 데이터를 바탕으로 시선 분포(x축: 텍스트, y축: Fixation 시간)로 변환
- 3가지 하이퍼 파라미터를 중심으로 데이터를 처리
 - window: 사람들은 글을 읽을 때 집중하는 단어 외에 주변 단어들도 함께 고려함 [Schotter et al., 2012]
 - threshold: 얼마나 많은 단어들을 하이라이팅할 것인지에 대한 설정
 - time_limit: 읽기에 얼마나 많은 시간을 허용할 것인지에 대한 설정 [Ng, 2011]
- 각 하이퍼 파라미터 조합을 그리드 서치로 튜닝하고, 각 조합에서 문해력이 높은 군의 시선 분포 간 쿨백-라이블러 발산이, 문해력이 낮은 군의 시선 분포 간 쿨백-라이블러 발산에 비해 작다면, H1을 채택

Methods

Field Experiment

- 초등학생들의 시선 데이터를 기반으로 구축된 텍스트 하이라이팅 알고리즘으로 글을 하이라이팅
- 60명의 대학생을 대상으로 글을 읽고, 문제에 답하도록 함
- (1) 레이블에 의존하지 않음을 보이기 위해, (2) 실험 대상이 초등학생에서 대학생으로 격상되었음을 감안하기 위해, 문제의 난이도 상승

- $(X4 - X3) - (X2 - X1)$ 로 측정한 의사결정 속도 및 정확도가 양수라면, H2a, H2b를 채택
- $(X4 - X3) - (X2 - X1)$ 로 측정한 크리펜돌프의 알파가 양수라면, H2c를 채택

	Control Group	Treatment Group
Before Intervention	X1 하이라이팅 X	X2 하이라이팅 X
After Intervention	X3 하이라이팅 X	X4 하이라이팅 O

Results

Highlighting Model Construction

		Time limit						
		Window	0.75	0.8	0.85	0.9	0.95	1
Threshold	0.75	1	9.96	9.91	9.84	9.83	9.82	9.82
		2	9.91	9.78	9.76	9.77	9.77	9.78
		3	9.83	9.70	9.73	9.73	9.73	9.74
		5	10.08	9.90	9.91	9.92	9.93	9.94
		1	10.95	10.89	10.73	10.73	10.72	10.73
	0.8	2	10.75	10.66	10.64	10.65	10.65	10.66
		3	10.75	10.60	10.63	10.63	10.63	10.65
		5	10.95	10.73	10.72	10.71	10.71	10.74
		1	12.06	11.96	11.81	11.83	11.81	11.82
		2	11.78	11.63	11.55	11.55	11.55	11.57
	0.85	3	11.91	11.70	11.73	11.69	11.69	11.71
		5	11.87	11.66	11.65	11.63	11.63	11.66
		1	13.14	13.04	13.10	12.98	12.96	12.97
		2	13.01	12.95	12.85	12.80	12.79	12.82
		3	13.06	12.85	12.76	12.60	12.59	12.62
	0.9	5	12.93	12.62	12.55	12.52	12.51	12.54
		1	14.67	14.61	14.64	14.55	14.53	14.54
		2	14.75	14.69	14.69	14.42	14.41	14.45
		3	14.52	14.33	14.37	13.99	13.99	14.02
		5	14.49	14.18	14.11	13.82	13.82	13.84
	0.95	1	18.18	18.17	18.17	18.19	18.11	18.11
		2	17.74	17.59	17.59	17.59	17.59	17.59
		3	17.59	17.57	17.57	17.57	17.57	17.57
		5	17.19	17.12	17.01	17.01	17.01	16.97

← 문해력이 높은 군

문해력이 낮은 군 →

$t = 2.191, p < 0.05$

		Time limit						
		Window	0.75	0.8	0.85	0.9	0.95	1
Threshold	0.75	1	10.40	10.42	10.43	10.47	10.46	10.46
		2	10.35	10.37	10.38	10.42	10.44	10.45
		3	10.38	10.42	10.44	10.46	10.49	10.50
		5	10.45	10.51	10.53	10.54	10.55	10.57
		1	11.30	11.33	11.34	11.36	11.34	11.33
	0.8	2	11.31	11.32	11.34	11.37	11.39	11.40
		3	11.35	11.38	11.41	11.44	11.45	11.46
		5	11.34	11.38	11.39	11.40	11.42	11.44
		1	12.39	12.43	12.43	12.45	12.43	12.39
		2	12.38	12.42	12.44	12.46	12.46	12.43
	0.85	3	12.48	12.46	12.46	12.48	12.50	12.50
		5	12.33	12.40	12.40	12.41	12.43	12.43
		1	13.60	13.62	13.61	13.61	13.57	13.55
		2	13.52	13.53	13.53	13.53	13.51	13.52
		3	13.54	13.56	13.55	13.57	13.57	13.58
	0.9	5	13.55	13.52	13.53	13.52	13.55	13.57
		1	15.18	15.17	15.17	15.18	15.13	15.10
		2	15.07	15.05	15.01	15.04	15.02	15.01
		3	15.10	15.04	15.03	15.05	15.04	15.07
		5	15.02	15.01	15.01	15.03	15.02	15.02
	0.95	1	18.30	18.31	18.31	18.31	18.29	18.31
		2	18.19	18.26	18.26	18.26	18.26	18.26
		3	17.97	18.04	18.02	18.01	17.98	17.97
		5	18.03	18.09	18.09	18.09	18.09	18.09

Results

Field Experiment

Measurement	X1	X2	X3	X4	$(X4 - X2) - (X3 - X1)$	t (X2, X4)
Decision Speed	255.53	245.62	228.45	201.21	-17.33	3.40
Decision Accuracy	21.65	23.03	21.35	22.13	-0.61	1.22
Krippendorff's Alpha	0.503	0.575	0.562	0.628	-0.006	-

- Decision Speed: 초 단위
- Decision Accuracy: 30점 만점 점수
- 크리펜돌프 알파: 0 이상 1 이하 (일관성을 나타내는 threshold는 일반적으로 0.67 이상)

Discussion

Finding 1

- 문해력이 높은 군은 문해력이 낮은 군에 비해 모든 하이퍼 파라미터 조합에서 시선 분포의 쿨백-라이블러 발산이 작음
- 두 집단의 쿨백-라이블러 발산 간 차이는 통계적으로 유의했으므로, 이들에게는 유의미한 일관성 차이가 있는 것으로 보임
- 다만 시선 분포에는 글의 중요도만 영향을 미치지 않으므로, equivocality를 해소하기 위해서 기존 XAI를 활용한 보정 작업이 필요해 보임

Finding 2

- 시선 기반의 텍스트 하이라이팅은 의사결정 시간을 유의미하게 줄였지만, 의사결정 정확도 및 편향에는 유의한 효과가 없었음
- 시선 기반의 텍스트 하이라이팅은 사람이 글을 읽는 습관을 반영하기 때문에, 실제로 글을 읽는 프로세싱과 호환이 좋을 것으로 보임

다만, 이상은 같은 (유익한지, 악의적인지) 의사결정 정확도만 조금 개선했다는 점인데, 이는 시선 기반의 하이라이팅의 중요

Discussion

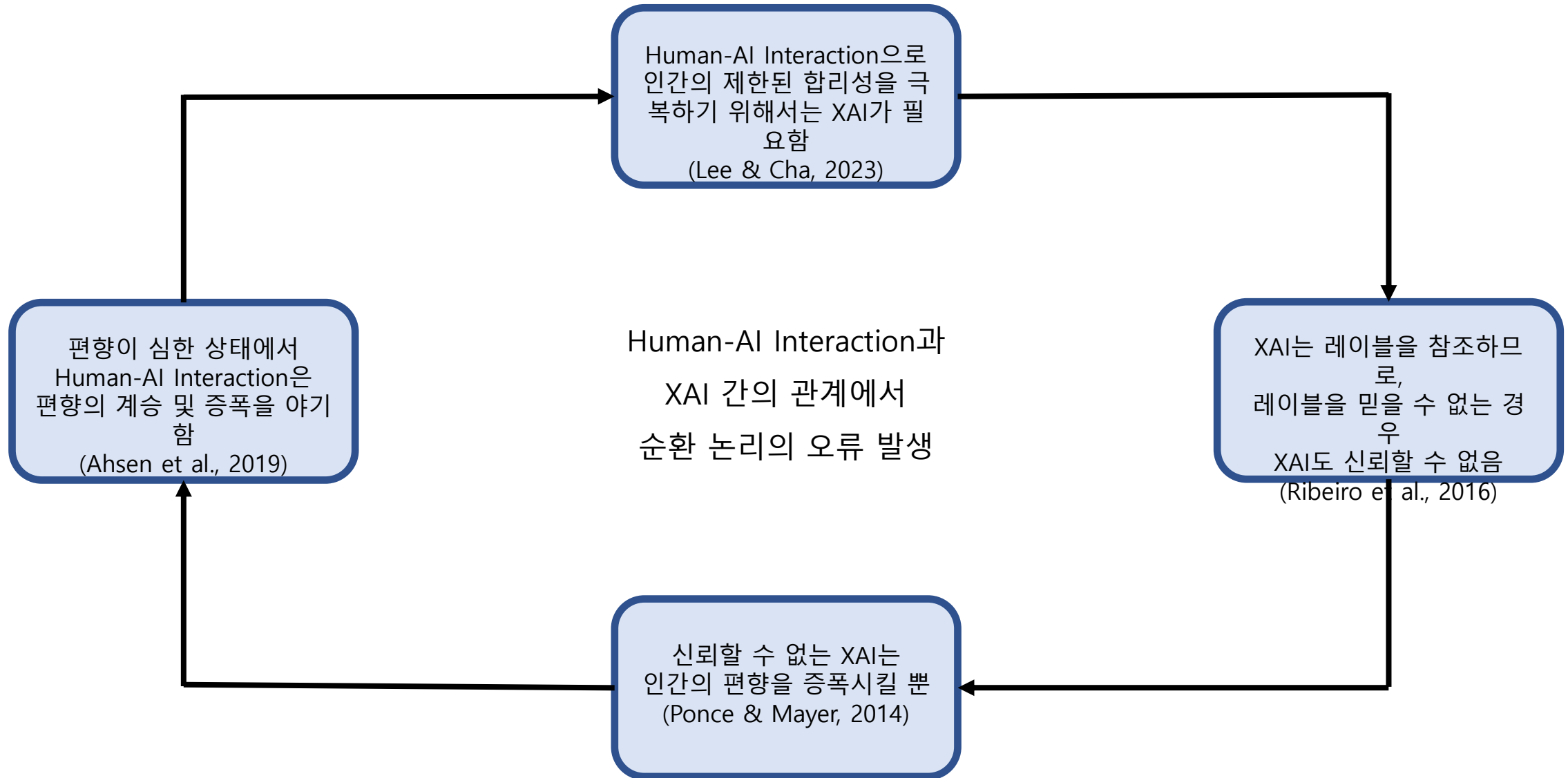
Limitations

- 본 연구에서는 기존 AI 기반 텍스트 하이라이팅이 맥락을 반영하는 것의 한계를 꼬집었으나, 본 연구에서 진행된 실험은 맥락에 특화되지 않음
- 텍스트 하이라이팅 모델링을 위한 데이터는 초등학생의 데이터를 사용했지만, 실험은 대학생들 대상으로 함

Propositions for Future Work

- 시선 기반의 텍스트 하이라이팅을 자기소개서 평가와 같은 특정 도메인에 적용할 필요가 있음
- 시선 기반의 텍스트 하이라이팅을 기반으로, 새로운 글이 주어졌을 때, 그 글을 하이라이팅하는 지도학습 알고리즘을 구축할 필요가 있음
- 시선 기반의 텍스트 하이라이팅 알고리즘과, 기존 XAI 알고리즘의 의사결정 패턴에 어떤 차이가 있는지 규명할 필요가 있음
- 시선 기반의 텍스트 하이라이팅이 궁극적으로 인간의 의사결정을 개선시키는지, 혹은 편향을 증폭시키는지 검증할 필요가 있음

Conclusion



Conclusion

