분석기를 공격해서 강화할 수 있을까?

- AI를 이용해서 분석기를 개발하고 안전성soundness를 증명할 수 있다.

분석기를 공격해서 강화할 수 있을까?

- AI를 이용해서 분석기를 개발하고 안전성soundness를 증명할 수 있다.
- 당연히 정밀도completeness는 증명할 수 없다.

분석기를 공격해서 강화할 수 있을까?

- AI를 이용해서 분석기를 개발하고 안전성soundness를 증명할 수 있다.
- 당연히 정밀도completeness는 증명할 수 없다.
- 그런데도 억지로 증명하라고 잘 시켜보면? 반례를 찾아주지 않을까?

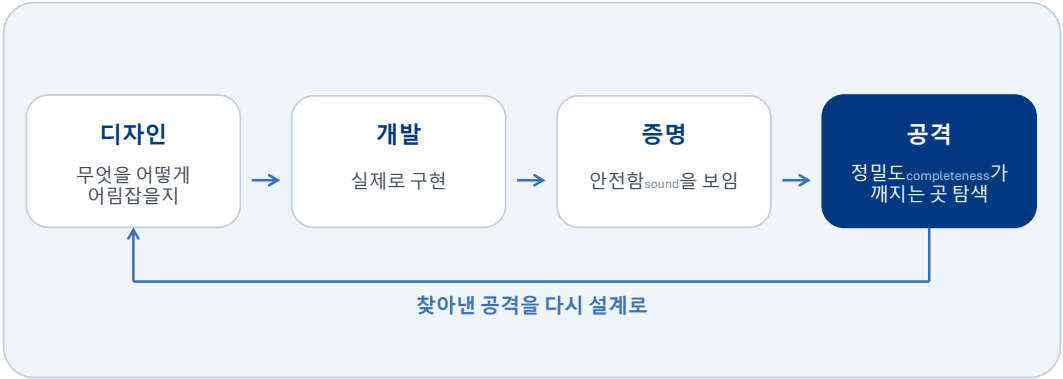
분석기를 공격해서 강화할 수 있을까?

- AI를 이용해서 분석기를 개발하고 안전성soundness를 증명할 수 있다.
- 당연히 정밀도completeness는 증명할 수 없다.
- 그런데도 억지로 증명하라고 잘 시켜보면? 반례를 찾아주지 않을까?
- $P \vee \neg P$를 증명하라고 하면?

분석기를 공격해서 강화할 수 있을까?

- AI를 이용해서 분석기를 개발하고 안전성soundness를 증명할 수 있다.
- 당연히 정밀도completeness는 증명할 수 없다.
- 그런데도 억지로 증명하라고 잘 시켜보면? 반례를 찾아주지 않을까?
- $P \vee \neg P$를 증명하라고 하면?
- 그러면 그 반례를 보고 분석기를 보완할 수 있지 않을까?

분석기 개발 사이클


디자인
무엇을 어떻게 어림잡을지
개발
실제로 구현
증명
안전함sound을 보임
공격
정밀도completeness가 깨지는 곳 탐색
찾아낸 공격을 다시 설계로

AI가 반례를 찾는 예시

```
Require Import Arith.

Example tempting :
  forall a b c : nat,
    c > 0 ->
    (a + b) / c = a / c + b / c.
Proof.
Admitted.
```

“이거 증명해줘”

AI가 반례를 찾는 예시

```
1  Require Import Arith.
2
3  Example tempting :
4    forall a b c : nat,
5      c > 0 ->
6      (a + b) / c = a / c + b / c.
7  Proof.
8  Admitted.
```

“이거 증명해줘”

• 현재 정리문은 거짓이라 증명을 완성할 수 없습니다.

반례:

a = 1, b = 1, c = 2

그러면 c > 0이지만:

(1 + 1) / 2 = 1
1 / 2 + 1 / 2 = 0

Coqtail에서도 두 계산을 확인했습니다. SigPL/260826/proof.v:3는 수정하지 않았습니다.

정리문을 고친다면 다음 중 하나가 필요합니다.

- 정확한 조건 추가: a mod c + b mod c < c
- 더 간단한 충분조건 추가: a mod c = 0
- 현재 명제가 거짓임을 증명하는 반례 정리로 변경

“여기 반례 있어”