

AI를 활용한 안전한 정적 분석기 생성

신채환

2026. 8. 26.

서울대학교 프로그래밍 연구실(ROPAS)

배경

정적 분석기 제작에서 인간이 해야할 일.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계.
- 분석기 구현.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계.
- 분석기 구현.
- 안전성 증명.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계. ← 인간이 해야할 일.
- 분석기 구현.
- 안전성 증명.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계. ← 인간이 해야할 일.
- 분석기 구현. ← 요약 방법에 따라 거의 자동으로 정해지는 것이 많음.
- 안전성 증명.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계. ← 인간이 해야할 일.
- 분석기 구현. ← 요약 방법에 따라 거의 자동으로 정해지는 것이 많음.
- 안전성 증명. ← 구현에 따라 거의 자동으로 정해지는 것이 많음.

배경

정적 분석기 제작에서 인간이 해야할 일.

- 요약 방법 설계. ← 인간이 해야할 일.
- 분석기 구현. ← 요약 방법에 따라 거의 자동으로 정해지는 것이 많음.
- 안전성 증명. ← 구현에 따라 거의 자동으로 정해지는 것이 많음.

인간이 설계만 하고 나머지는 최대한 AI를 활용해 생성할 수는 없을까?

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

입력: 정적 분석기를 구현하는 데 필요한 것.

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

입력: 정적 분석기를 구현하는 데 필요한 것.

- 정적 분석 대상 언어.

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

입력: 정적 분석기를 구현하는 데 필요한 것.

- 정적 분석 대상 언어.
- 요약의 방법.

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

입력: 정적 분석기를 구현하는 데 필요한 것.

- 정적 분석 대상 언어.
- 요약의 방법.

출력: 안전한 정적 분석기.

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

입력: 정적 분석기를 구현하는 데 필요한 것.

- 정적 분석 대상 언어.
- 요약의 방법.

출력: 안전한 정적 분석기.

- 정적 분석기.

목표

인간이 필요한 것만 입력하면 AI를 활용해 자동으로 안전한 정적 분석기를 생성.

입력: 정적 분석기를 구현하는 데 필요한 것.

- 정적 분석 대상 언어.
- 요약의 방법.

출력: 안전한 정적 분석기.

- 정적 분석기.
- 분석기의 안전성.

방법

Q. AI를 믿고 구현을 맡길 수 있는 방법?

방법

Q. AI를 믿고 구현을 맡길 수 있는 방법?

A. 구현과 안전성 증명을 '동시에' 시키면 됨.

방법

Q. AI를 믿고 구현을 맡길 수 있는 방법?

A. 구현과 안전성 증명을 '동시에' 시키면 됨.

Q. 안전한 구현은 너무 많은데, 인간이 납득할 만한 수준의 구현을 강제하는 방법은?

방법

Q. AI를 믿고 구현을 맡길 수 있는 방법?

A. 구현과 안전성 증명을 '동시에' 시키면 됨.
(AI가 안전성 증명을 완료하려면 자기의 구현이 정말로 안전해야 하기 때문.)

Q. 안전한 구현은 너무 많은데, 인간이 납득할 만한 수준의 구현을 강제하는 방법은?

A. 인간이 납득할 만한 수준의 구현임도 증명하게 하면 됨.

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

인간

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

인간

- 요약 방법을 Lean 4로 옮길 때 필요한 개념 제공.

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

인간

- 요약 방법을 Lean 4로 옮길 때 필요한 개념 제공.
- 요약 방법 설계에 도움이 될만한 기능 제공.

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

인간

- 요약 방법을 Lean 4로 옮길 때 필요한 개념 제공.
- 요약 방법 설계에 도움이 될만한 기능 제공.
- 안전성 등 AI가 증명하게 시킬 명제 제공.

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

인간

- 요약 방법을 Lean 4로 옮길 때 필요한 개념 제공.
- 요약 방법 설계에 도움이 될만한 기능 제공.
- 안전성 등 AI가 증명하게 시킬 명제 제공.

AI

Saphira 모듈의 목표

SAPHIRA: Static Analysis Proved by Human Idea and Restrained AI

인간

- 요약 방법을 Lean 4로 옮길 때 필요한 개념 제공.
- 요약 방법 설계에 도움이 될만한 기능 제공.
- 안전성 등 AI가 증명하게 시킬 명제 제공.

AI

- 증명에서 쓰일만한 도움 정리 제공.

예시

```
@[expose,reducible]
public def SetMem : AbstractionStructure where
  C := Set Mem
  A := Var → Interval
  γ a := {m ¦ ∀ x, m x ∈ a x}
  γ_monotone := sorry

public instance SetMem.A.instPrimcodable : Primcodable SetMem.A :=
  sorry
```

예시(계속)

```
public def all : SetMem.C →o SetProdMemInt.C where
  toFun x := {(m, n) ¦ (m ∈ x) (n : Int)}
  monotone' := sorry
```

$$\begin{aligned} \mathrm{all} : \mathscr{P}(\mathbb{M}) &\to \mathscr{P}(\mathbb{M} \times \mathbb{Z}) \\ M &\mapsto \{\, (m,n) \mid m \in M, n \in \mathbb{Z} \,\} \end{aligned}$$

예시(계속)

```
public def all : AbstractFun SetMem SetProdMemInt Concrete.all := sorry

public theorem all.optimal_or_not_exists_optimal :
    all.OptimalOrNotExistsOptimal :=
  sorry
```

실제 공간		요약 공간
$\mathscr{P}(\mathbb{M})$	$\leftrightarrow$	$\mathbb{X} \to \mathscr{I}_{\mathbb{Z}}$
$\mathscr{P}(\mathbb{M} \times \mathbb{Z})$	$\leftrightarrow$	$(\mathbb{X} \to \mathscr{I}_{\mathbb{Z}}) \times \mathscr{I}_{\mathbb{Z}}$

감사합니다.