



**UNIVERSITÀ
DI PARMA**

**Valutazione sperimentale
sull'individuazione
automatica di errori di
programmazione nel
codice generato da LLM**

Corso di laurea triennale in
Informatica

Candidato:

Manuel Di Agostino

Matricola:

332233

Relatore:

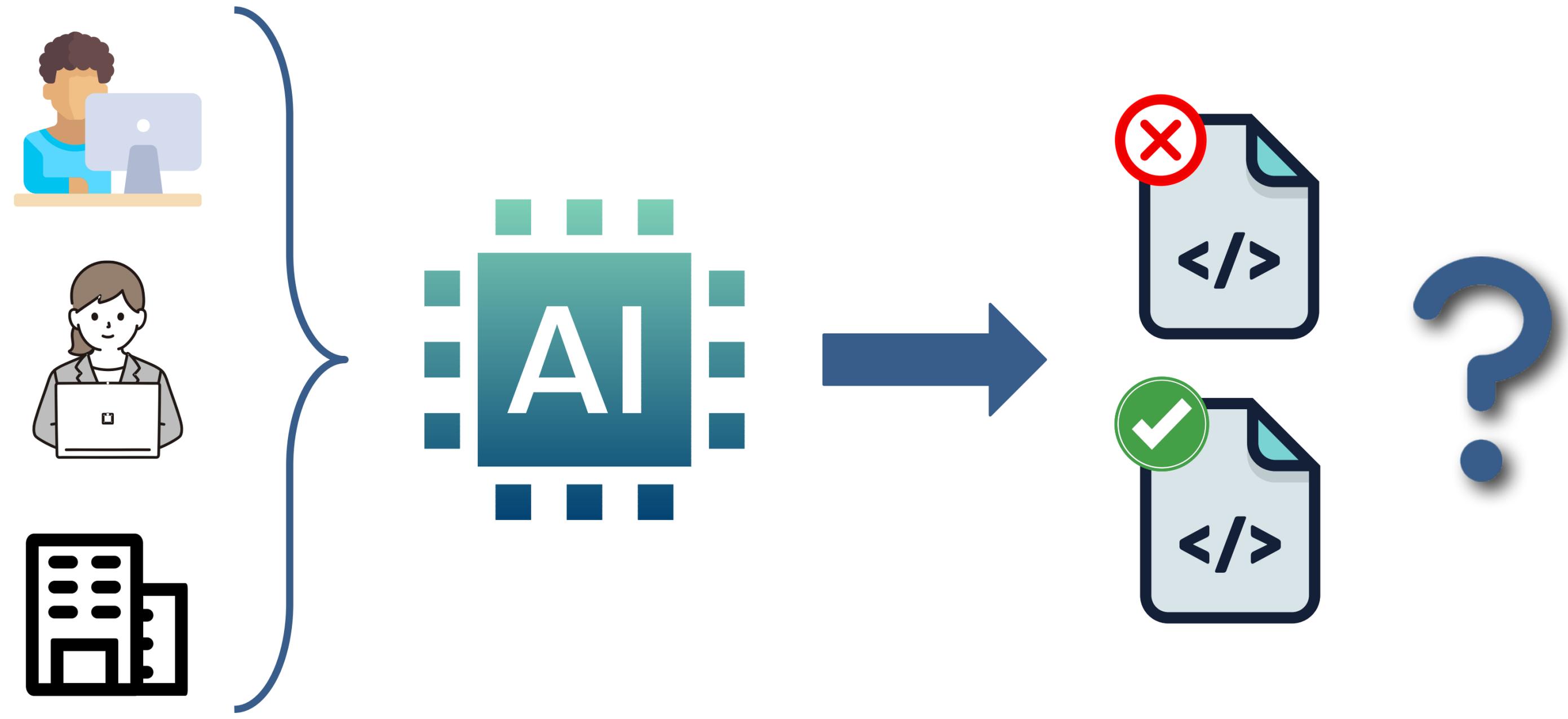
Prof. Enea Zaffanella

Correlatore:

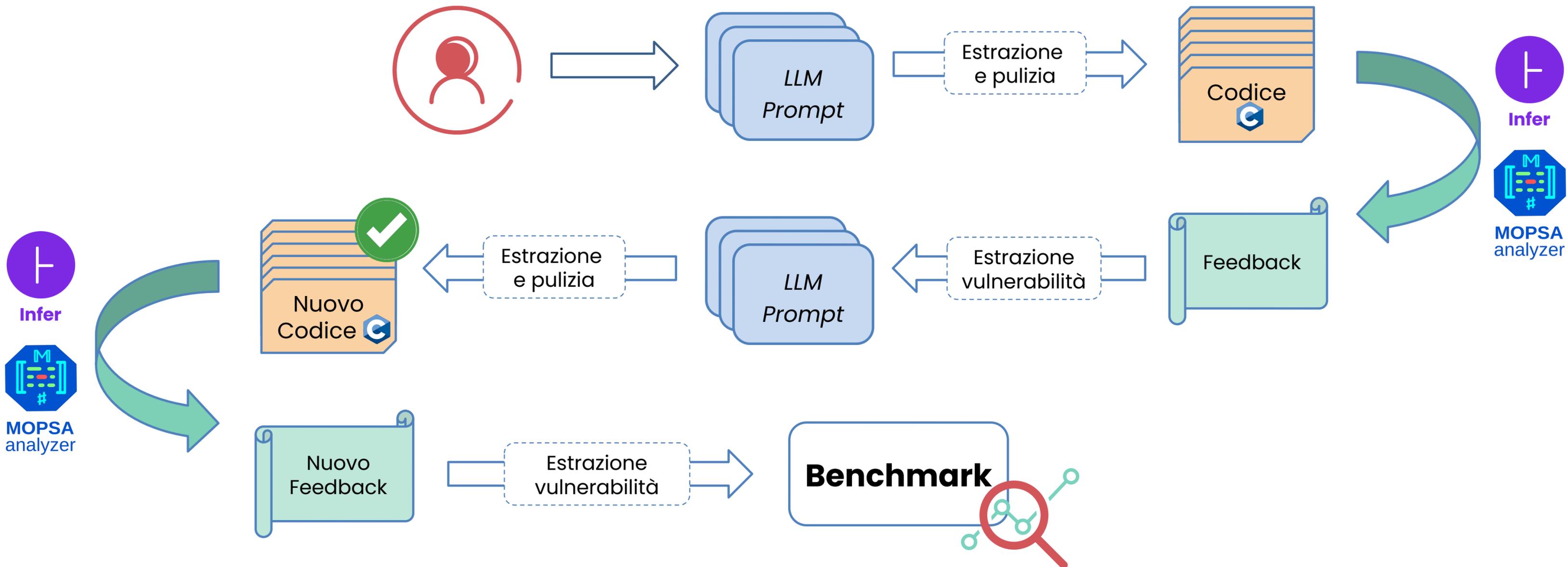
Prof. Vincenzo Arceri

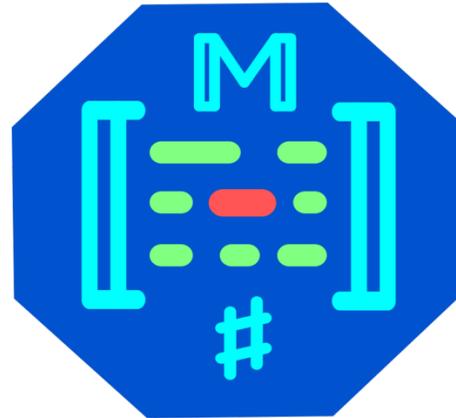
- Obiettivi e struttura dell'esperimento
- Tool coinvolti
- Verifica della qualità del codice prodotto dagli LLM
- Verifica delle capacità correttive degli LLM
- Risultati
- Conclusione e sviluppi futuri

Perché questo esperimento?



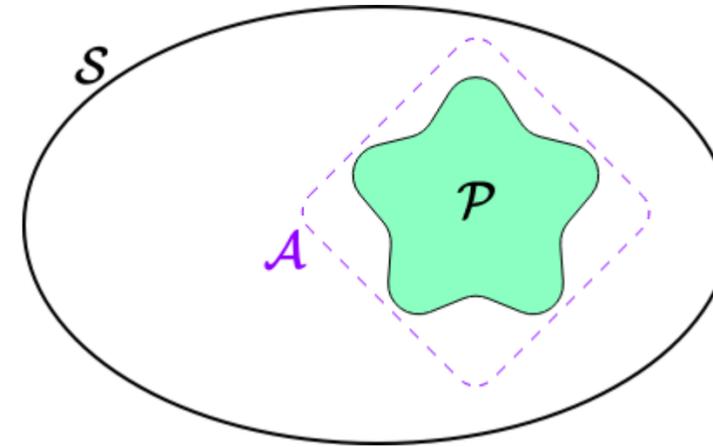
Struttura dell'esperimento



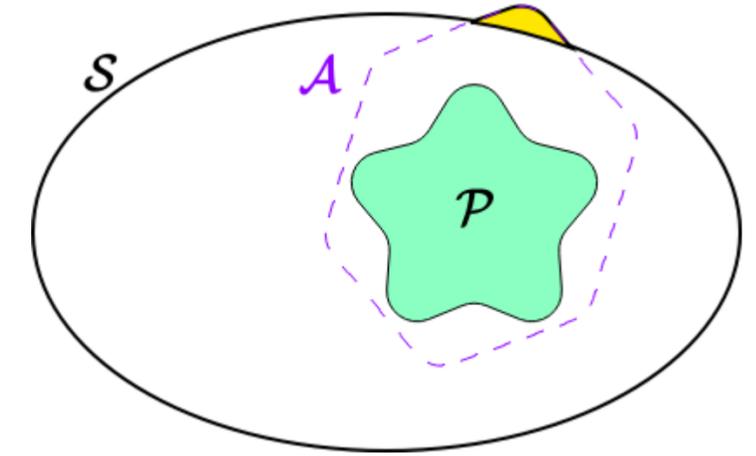


MOPSA analyzer

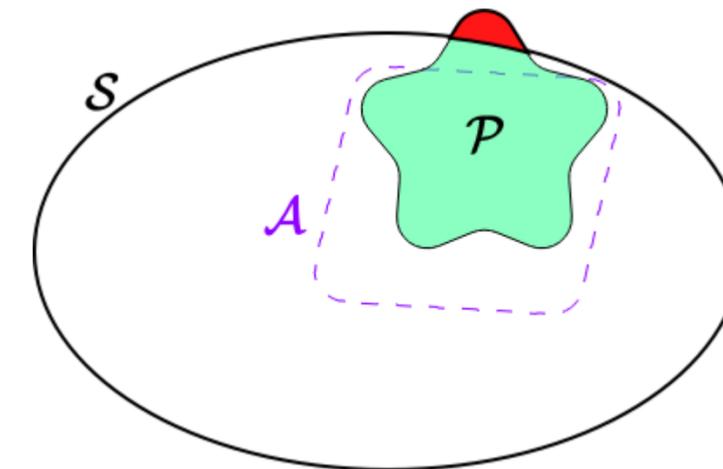
- Progetto open-source, finanziato dallo *European Research Council*
- Multilinguaggio, supporta sottoinsiemi di C e Python
- Fortemente basato sull'*Interpretazione Astratta*
- **Soundness**



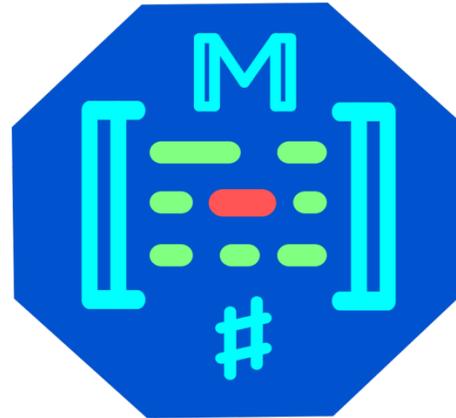
Sound, vero negativo



Sound, falso positivo

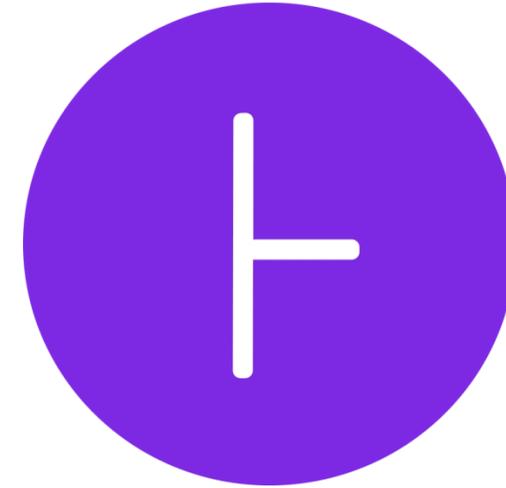


Unsound, falso negativo



MOPSA analyzer

- Progetto open-source, finanziato dallo *European Research Council*
- Multilinguaggio, supporta sottoinsiemi di C e Python
- Fortemente basato sull'*Interpretazione Astratta*
- **Soundness**



Infer

- Framework professionale di analisi statica
- Multilinguaggio, supporta i linguaggi C, C++, Java e Objective-C
- Integrato da *Meta* nella produzione del software
- Analisi *intra-procedurali* e di programmi incompleti

Sul cluster LEONARDO

```
SysPrompt >>> You are a chatbot whose purpose is to provide
the code implementation in the C programming language for
a task that is given to you as input.

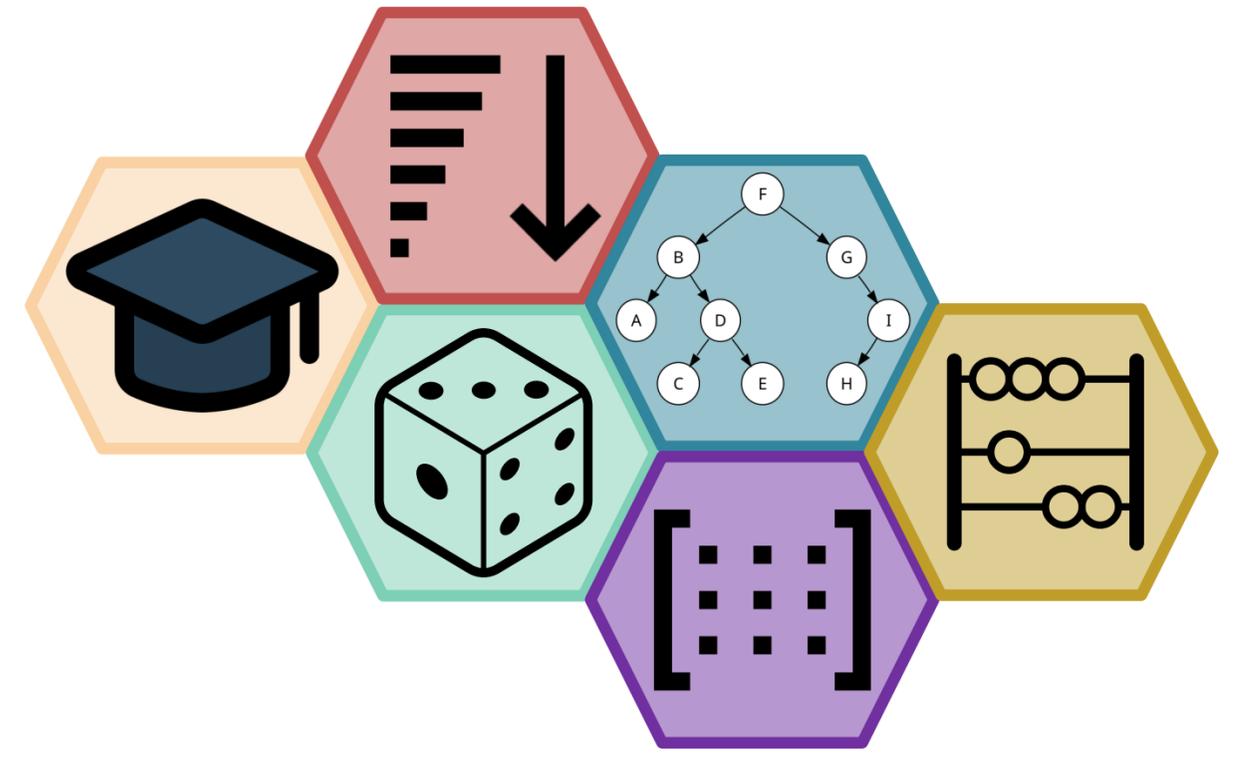
You always provide only the C code, without comments in the
code.

Your coding style is {gen_style} and the target for your code
is {target_context}.

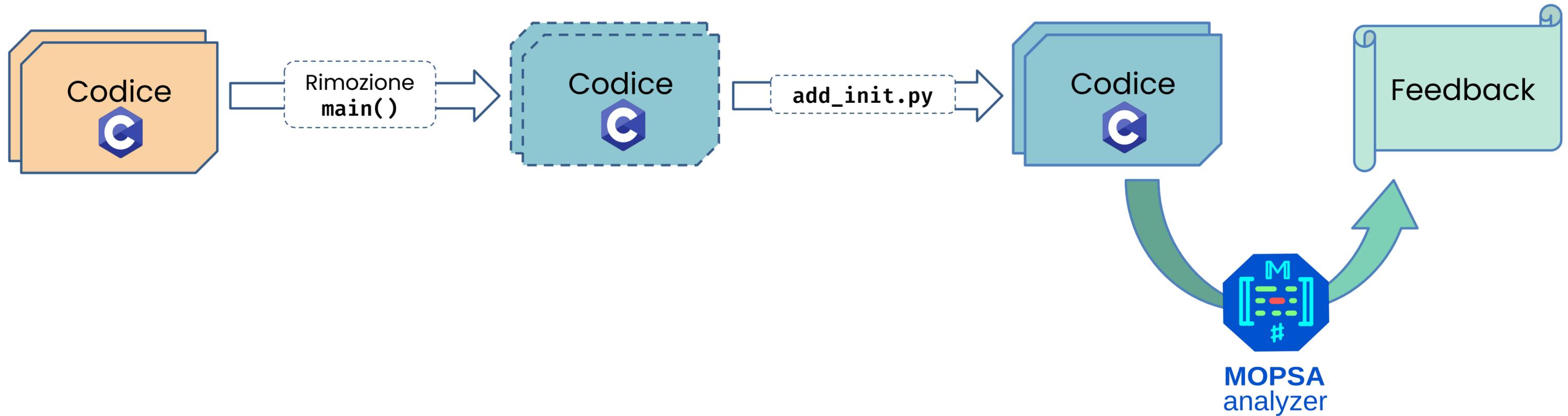
The code must solve the task correctly and compile, so you
must include the required libraries.
```

- Coinvolti **8 modelli** open-source
- Uno **stile** di programmazione e un **contesto** di utilizzo
- **10 macro-categorie** distinte, per un totale di 525 problemi
- Un job per ogni coppia <LLM, categoria>

 **Transformers**



Prerequisito: un *entry point*.



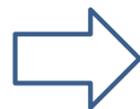
Obiettivi

- Inizializzazione *non deterministica*, tramite l'uso di **C Built-Ins** resi disponibili dal framework.
- **Invocazione** coerente delle funzioni target, considerando *tipo* e *numero* degli argomenti.

```
#include <stdio.h>
#include <stdlib.h>
#include <mopsa.h>

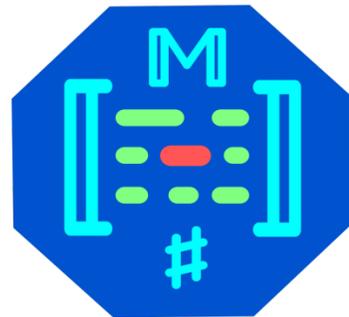
typedef struct {
    int value;
    struct __node *left, *right;
} __node;

__node add(__node n, int i) {
    n.value += i;
    return n;
}
```



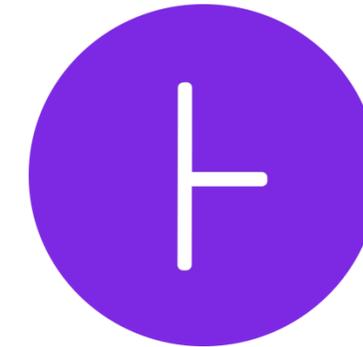
```
int main() {
    __node gen_var_name0 = {
        _mopsa_rand_s16(),
        (struct __node*)malloc(_mopsa_rand_u8() * sizeof(struct
        __node)),
        (struct __node*)malloc(_mopsa_rand_u8() * sizeof(struct
        __node))
    };
    int gen_var_name1 = _mopsa_rand_s16();
    add(gen_var_name0, gen_var_name1);
}
```

Sul cluster dell'Università di Parma



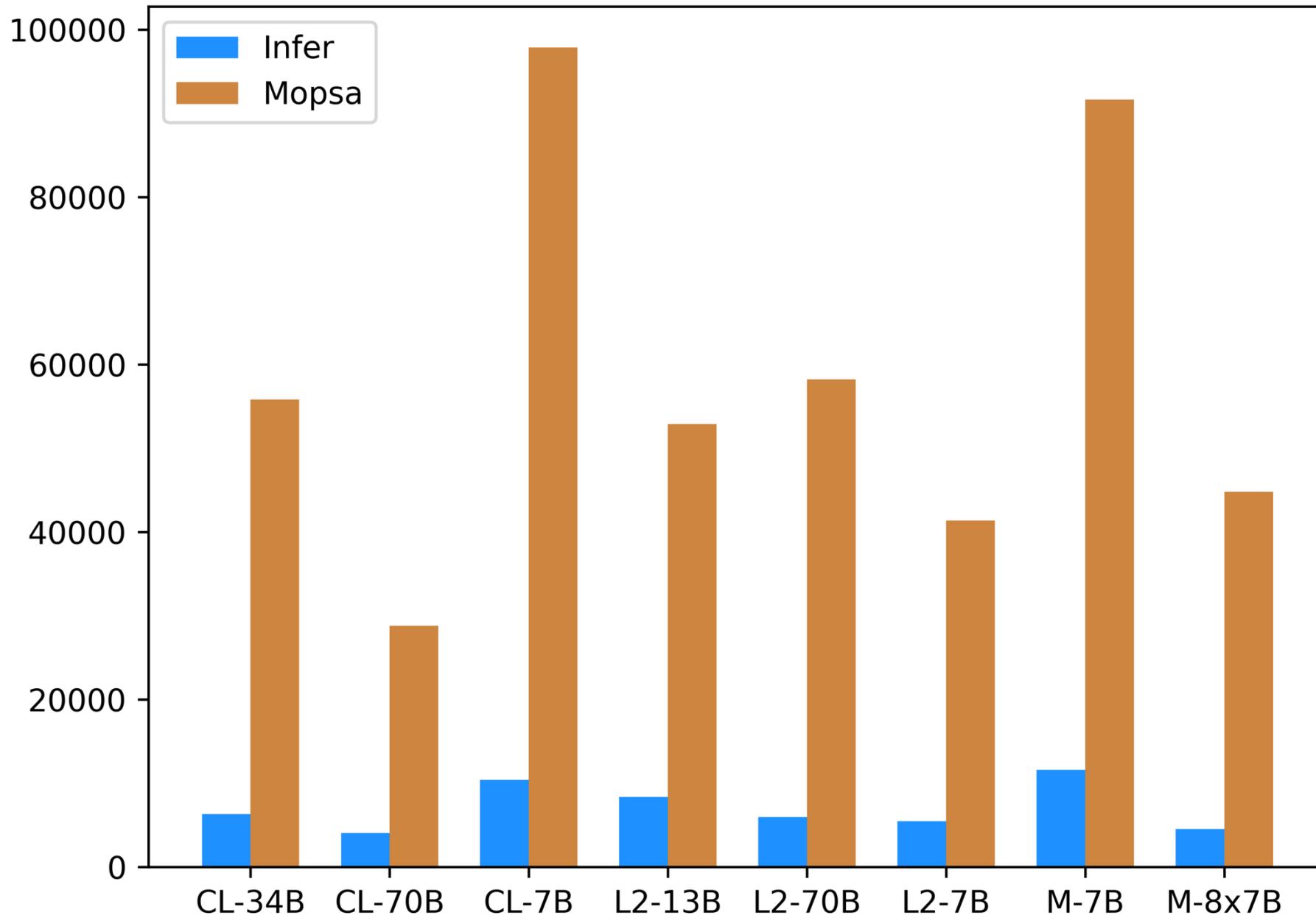
MOPSA
analyzer

- Configurazione `c/cell-itv-powerset-zero`, con domini **non relazionali**
- Controlli sulle operazioni aritmetiche
- Timeout 180s



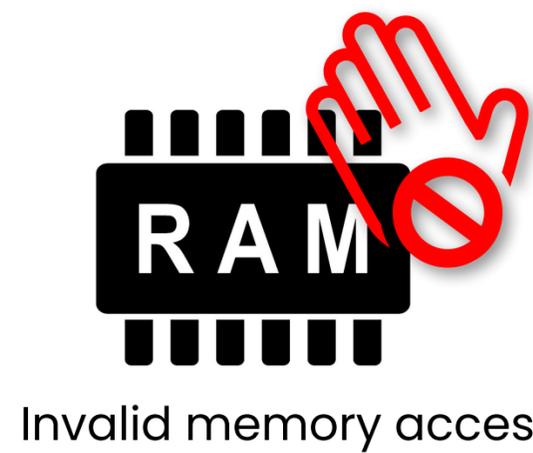
Infer

- InferBO
- Infer:Pulse
- Timeout 180s

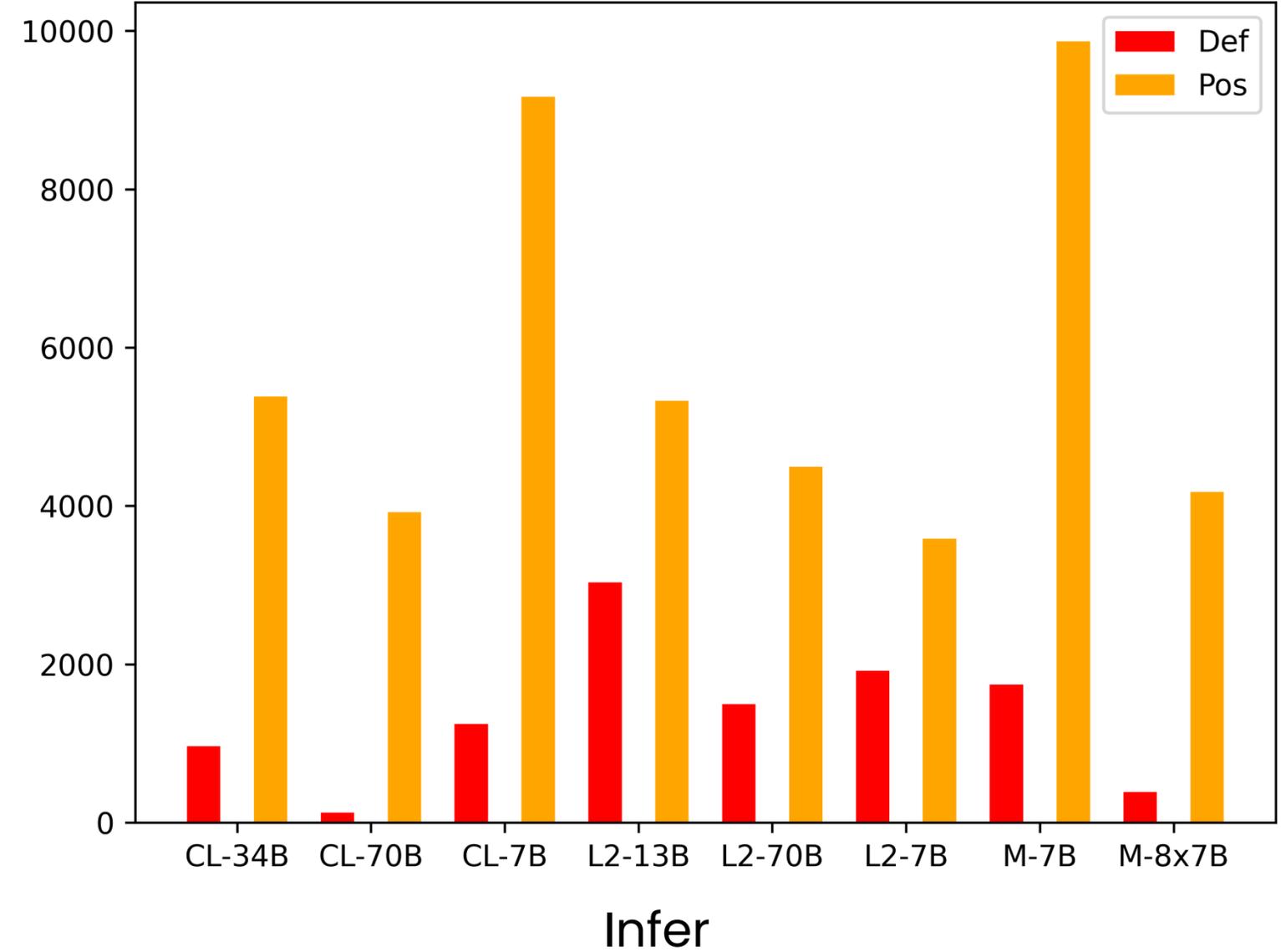
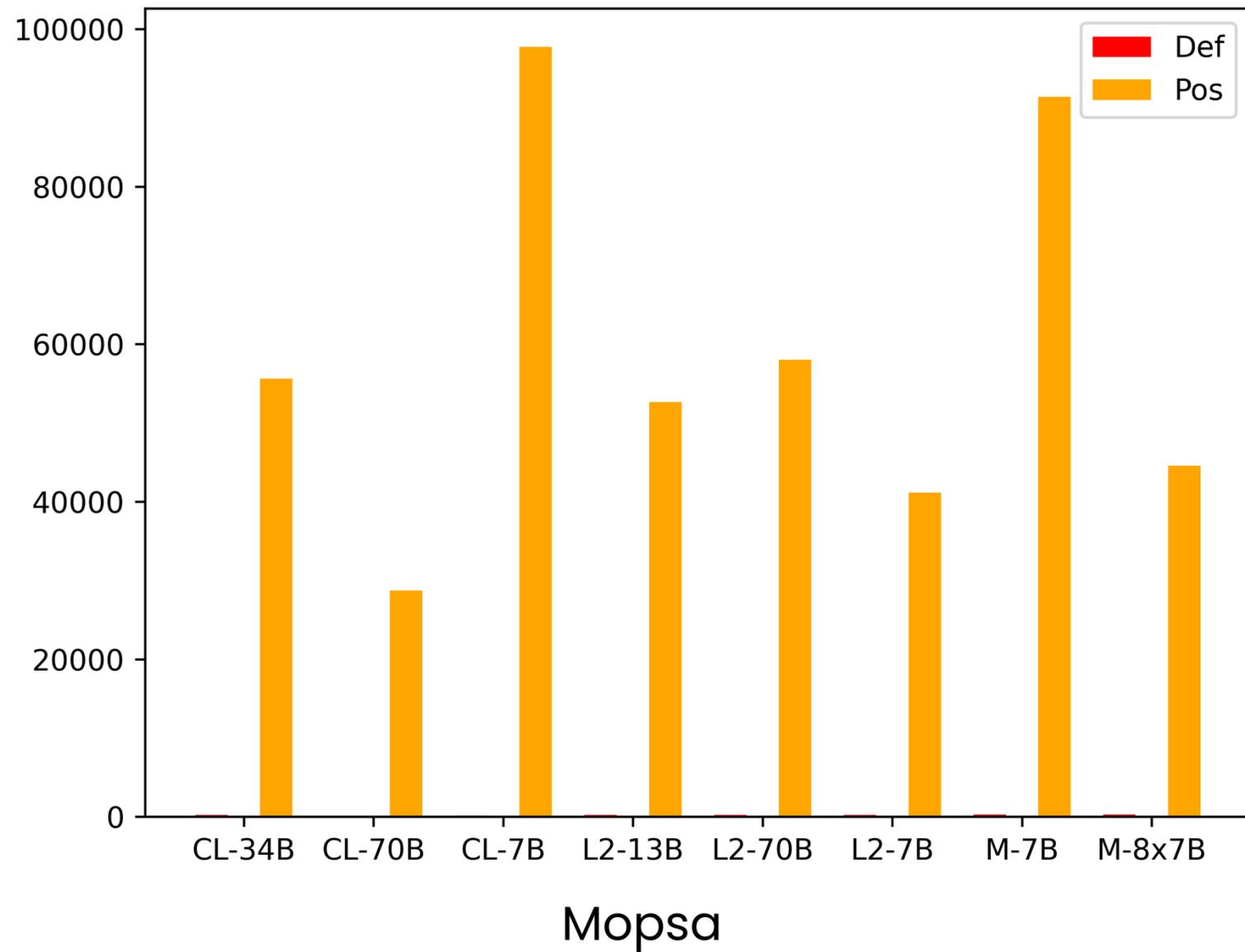


Numero di vulnerabilità segnalate da INFER e MOPSA per modello.

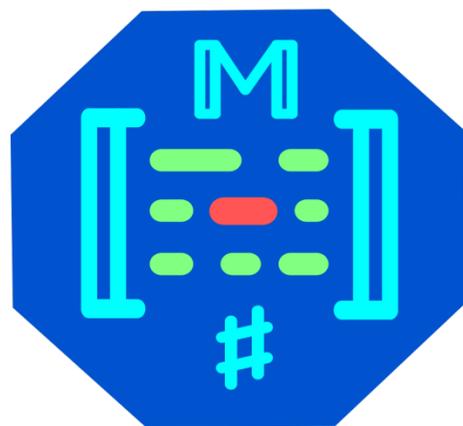
Vulnerabilità più diffuse



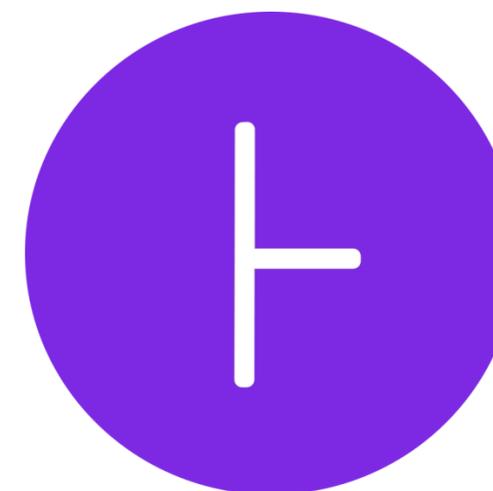
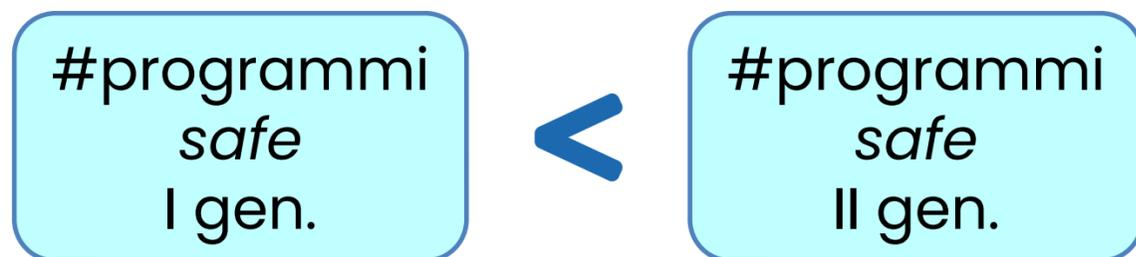
Vulnerabilità *definite* & *possible*



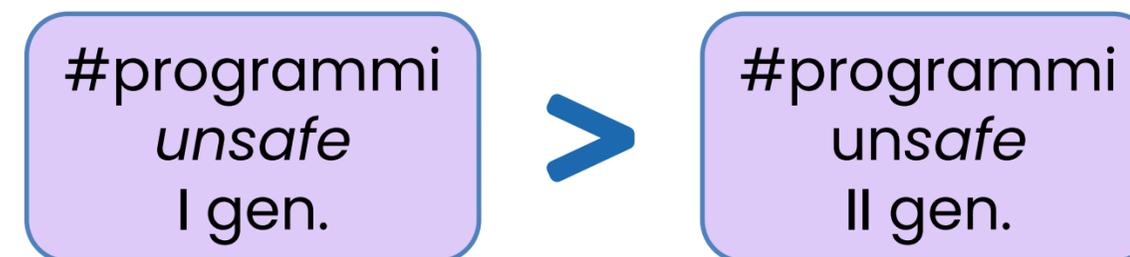
Due obiettivi



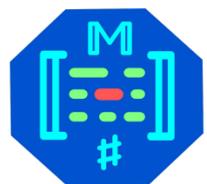
MOPSA
analyzer



Infer

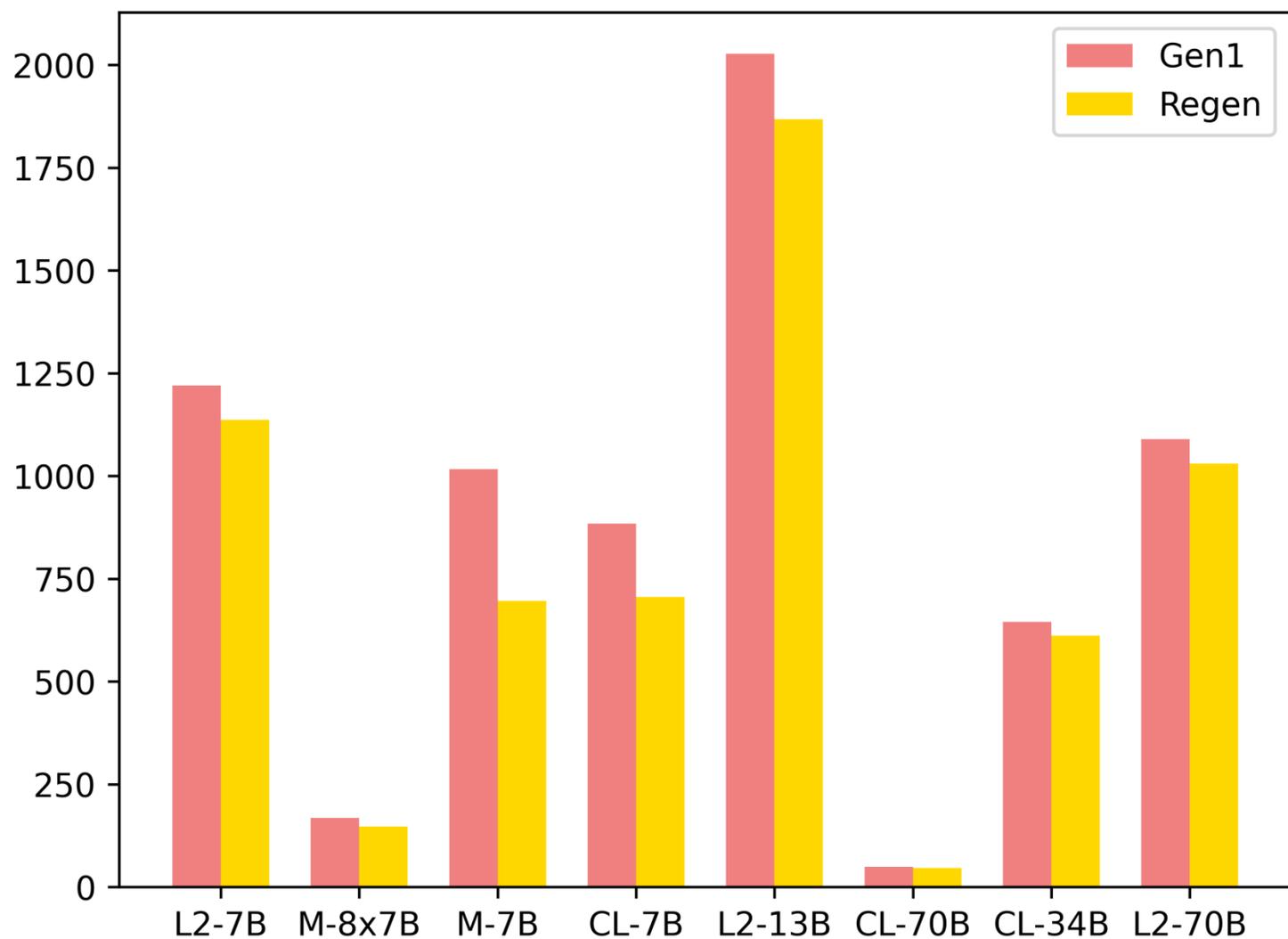


Benchmark



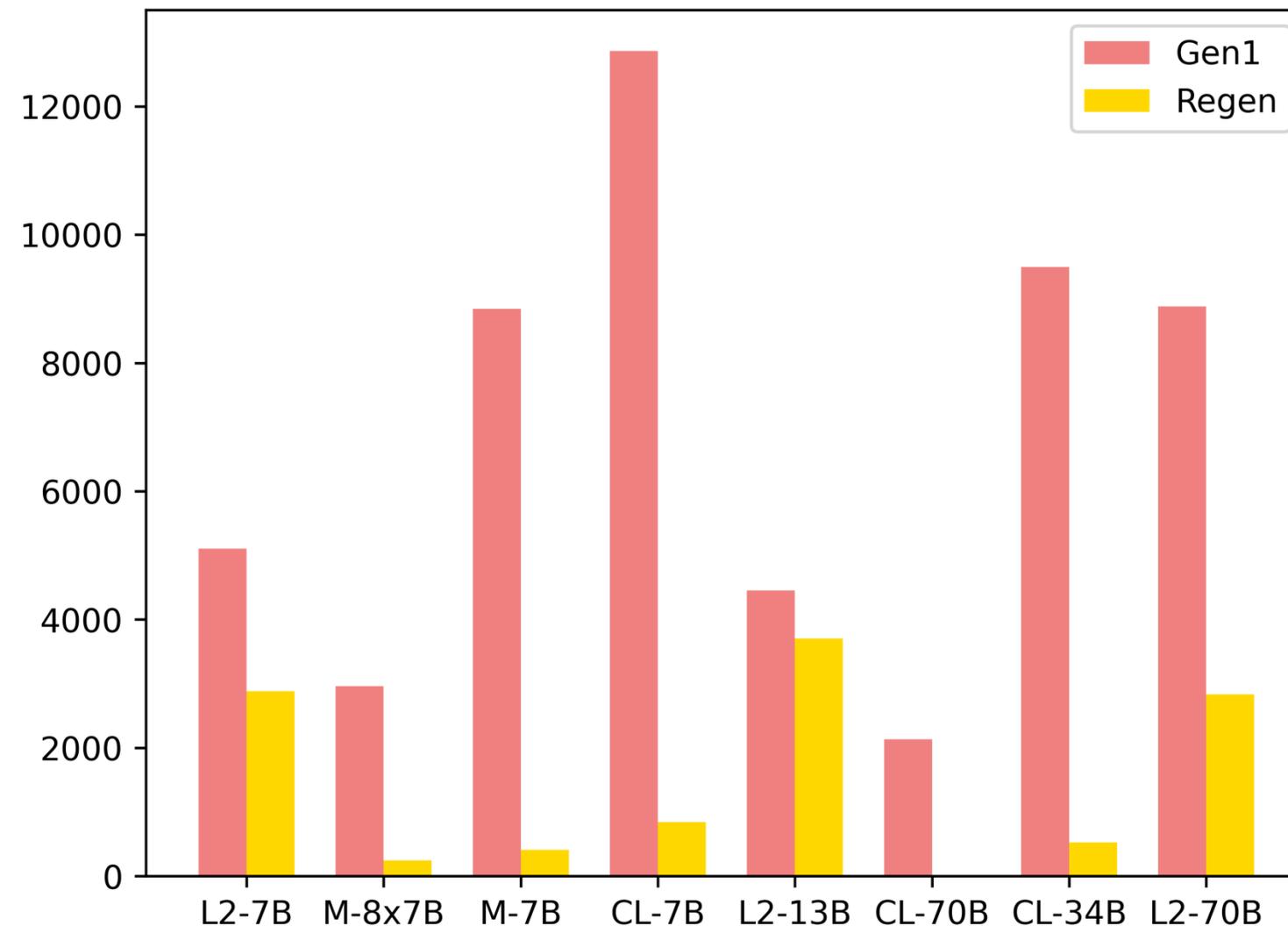
MOPSA
analyzer

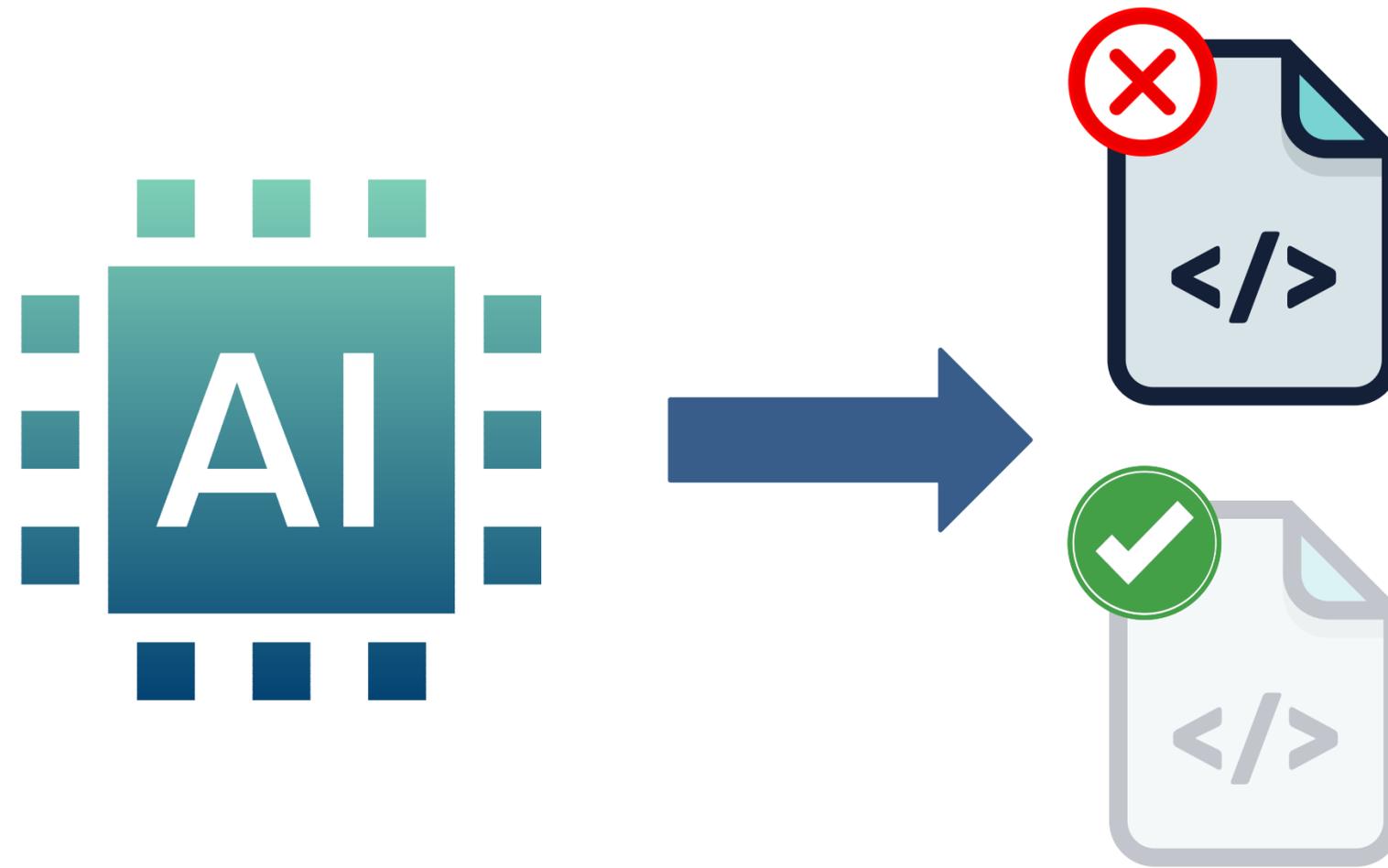
- Diminuzione degli errori *definite* con conseguente aumento dei programmi *safe*: da 785 a 3782, **incremento** di un **fattore 4,82x**

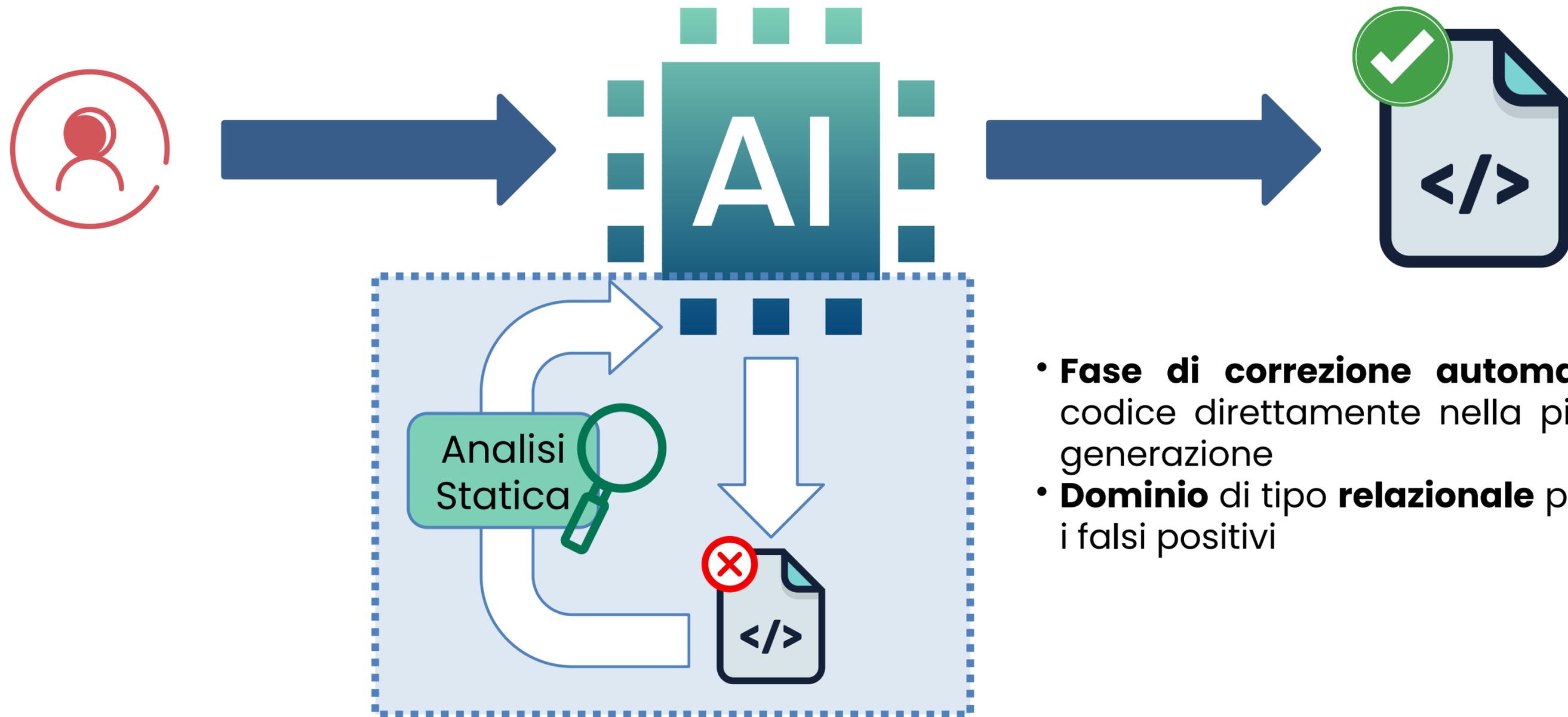


Infer

- Riduzione del numero dei sorgenti con vulnerabilità: da 7100 a 6242, **decremento** di un **fattore 1,14x**







- **Fase di correzione automatica** del codice direttamente nella pipeline di generazione
- **Dominio** di tipo **relazionale** per ridurre i falsi positivi

Grazie a tutti