




AIGenOps

Integrating Generative AI into Platform
Engineering for Regulated Software



In the beginning...

Problemi classici in contesti regolamentati

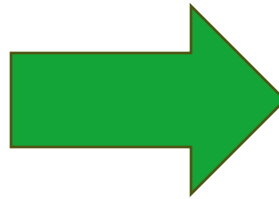
Caratteristiche applicazione

Compila correttamente

- Unit Test per superare la soglia imposta (di bassa qualità)

Già rilasciata in produzione

- Numero elevato di issue non bloccate da QA








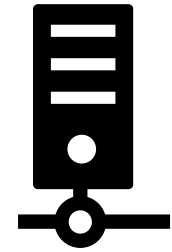
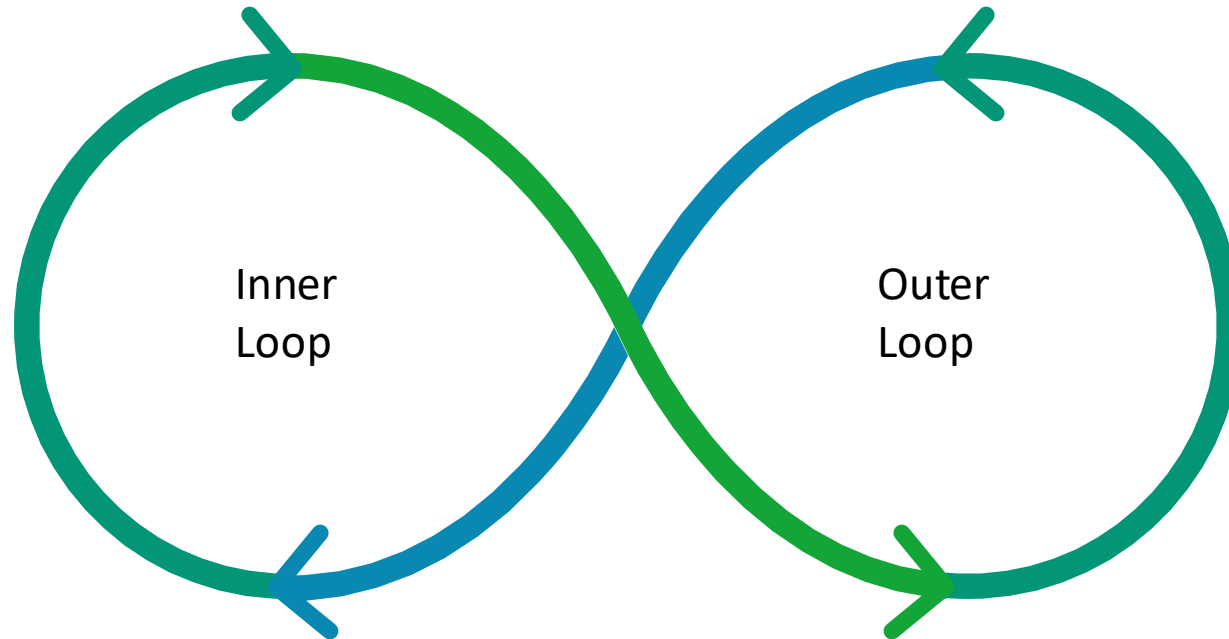
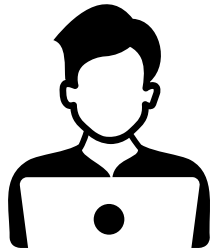
Sfruttare l'AI Generativa e le automation per ridurre il debito tecnico:

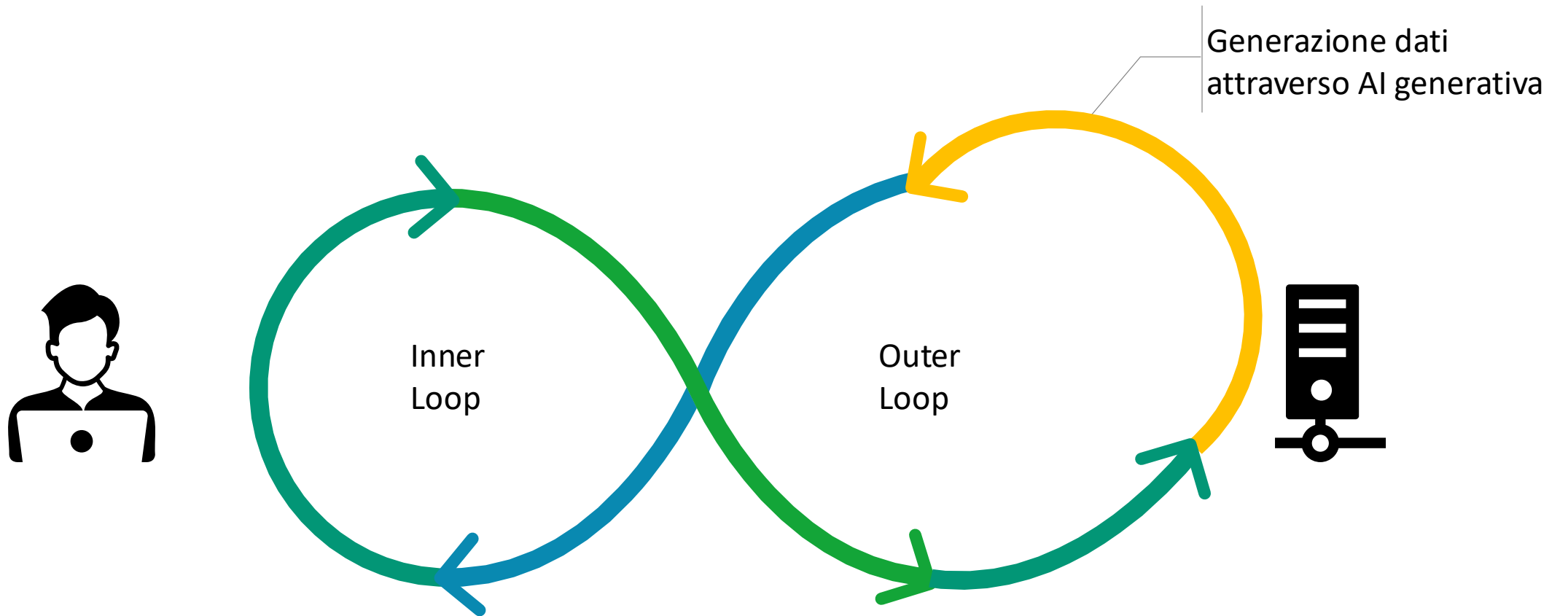
- Aumentare la coverage
- Migliorare la qualità degli Unit Test
- Sistemare i problemi sulla qualità



Requisiti

-  **Performance**
-  **Costi**
-  **Sicurezza**
-  **Quantità di dati generati**
-  **Zero Trust**



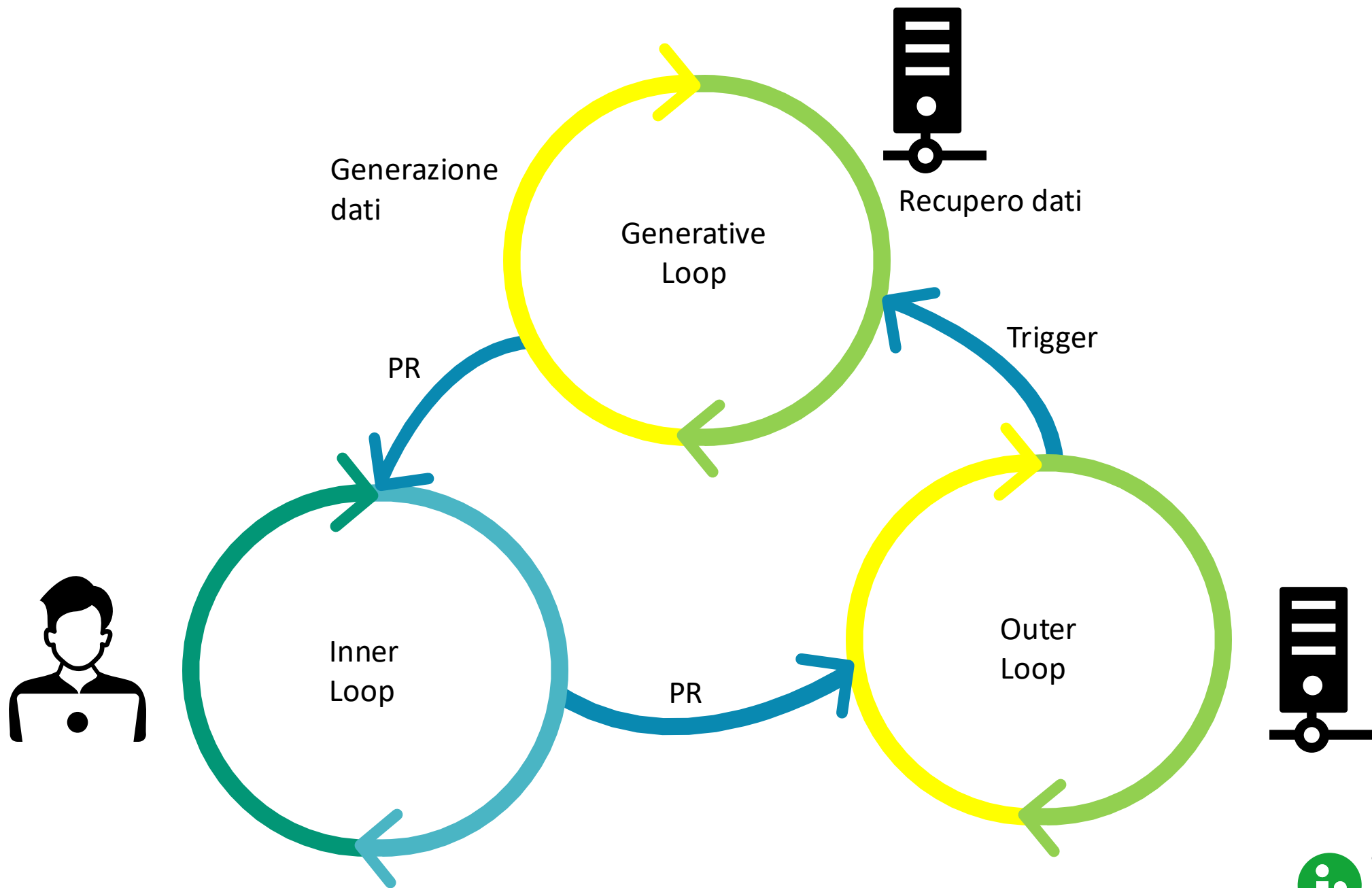


Requisiti



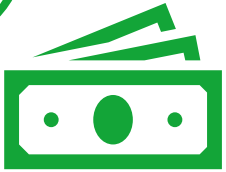
Performance

Generazione dei dati in ciclo indipendente



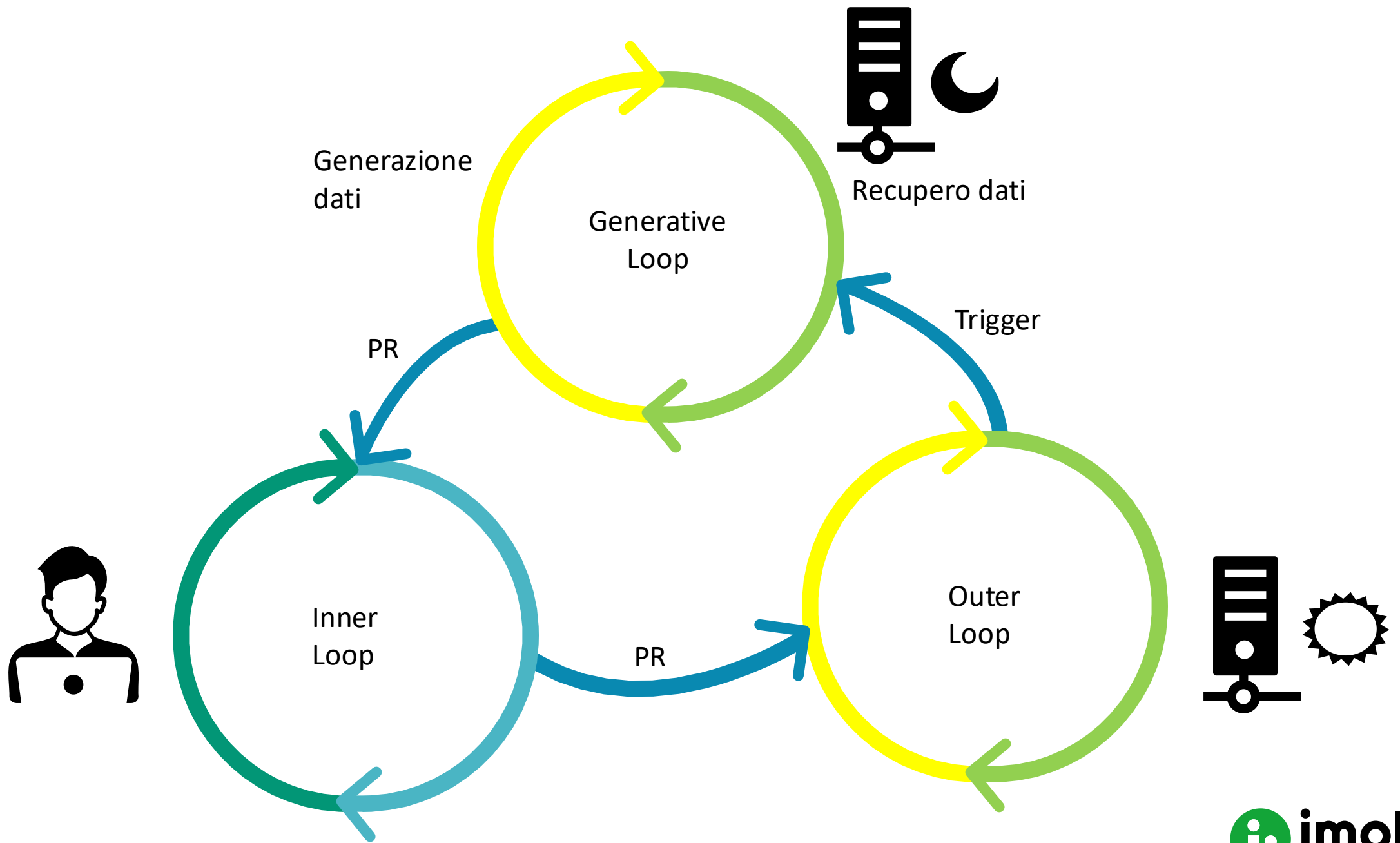
Requisiti

2

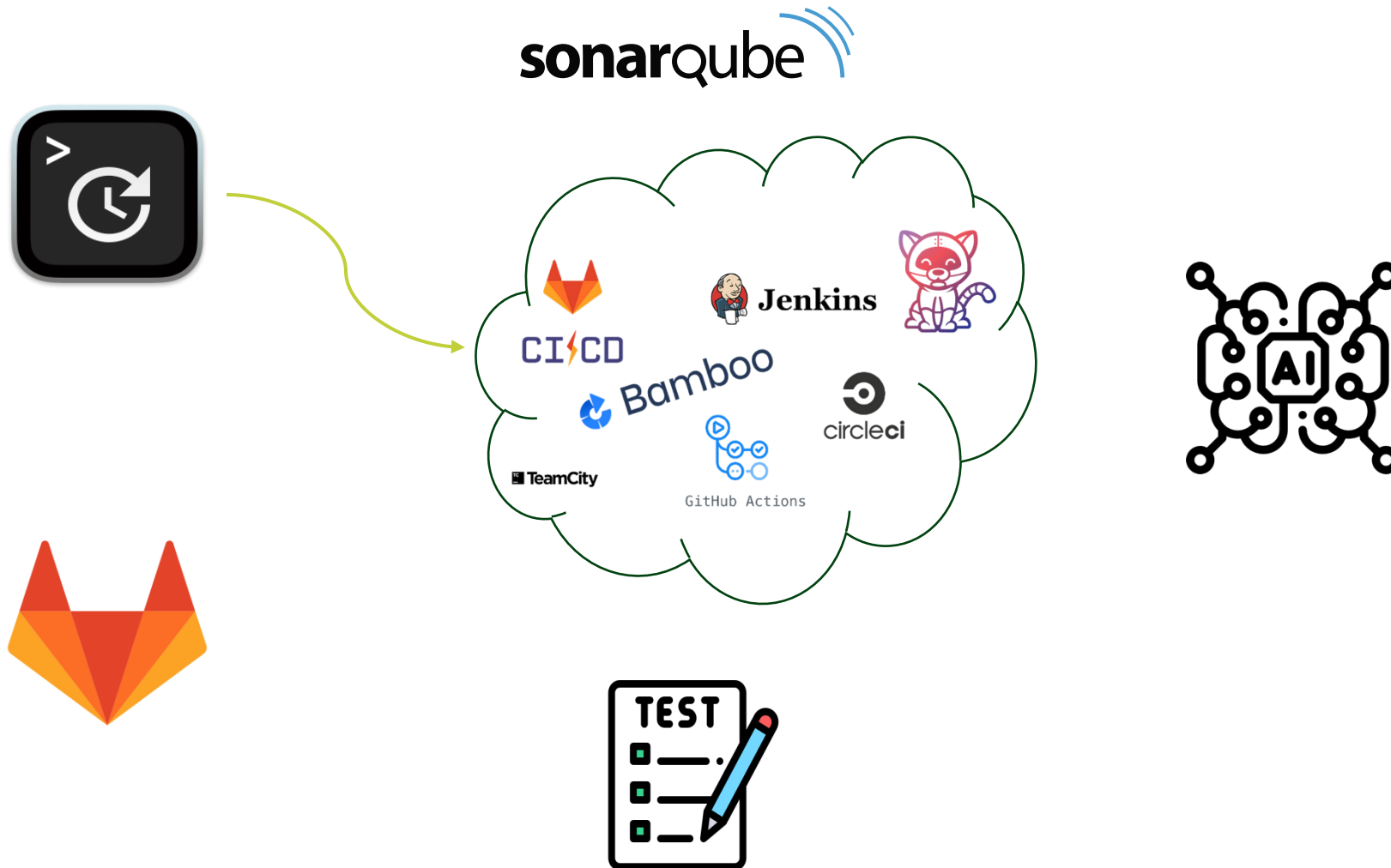


Costi

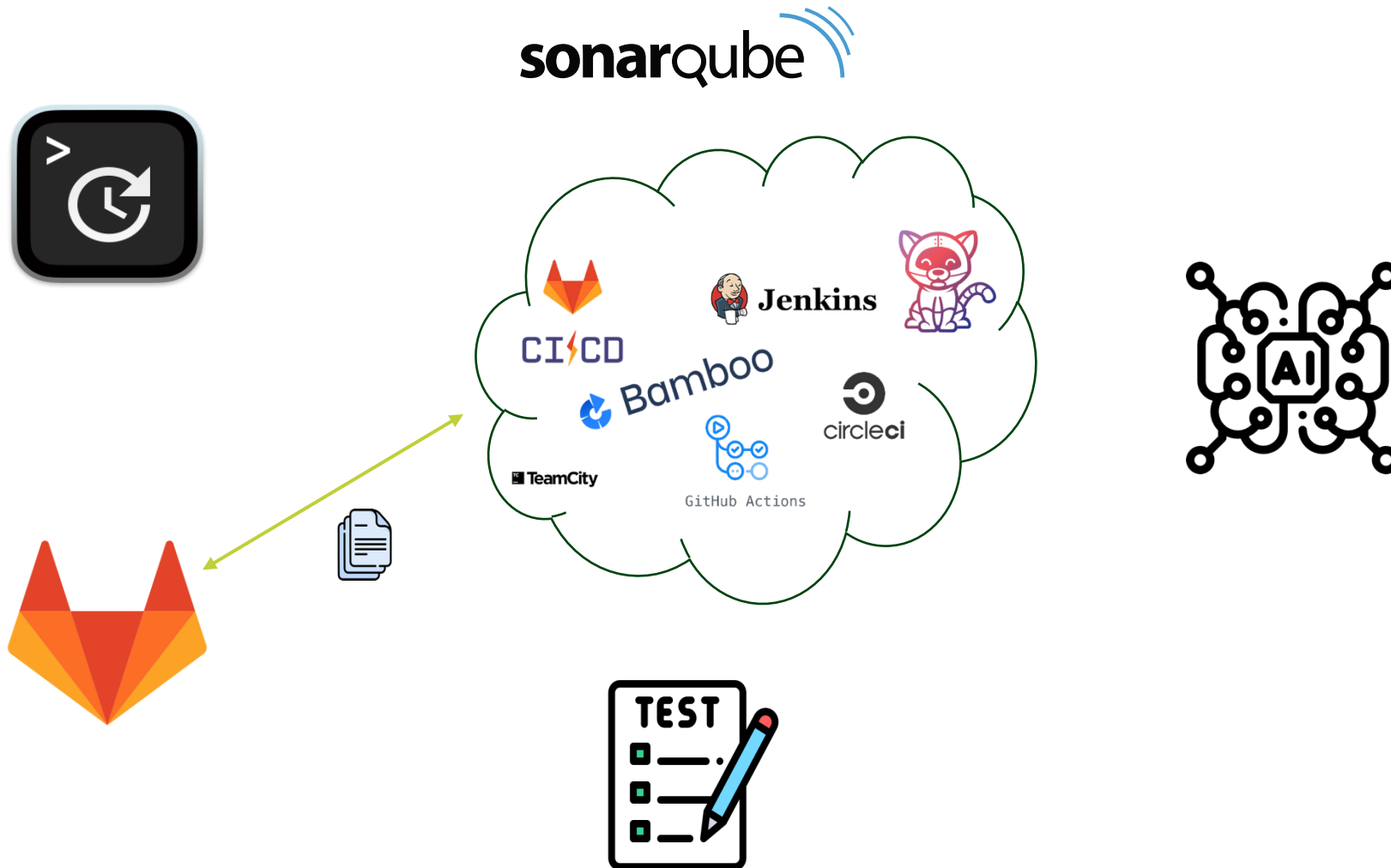
Utilizzo della platform nei momenti morti (notte)



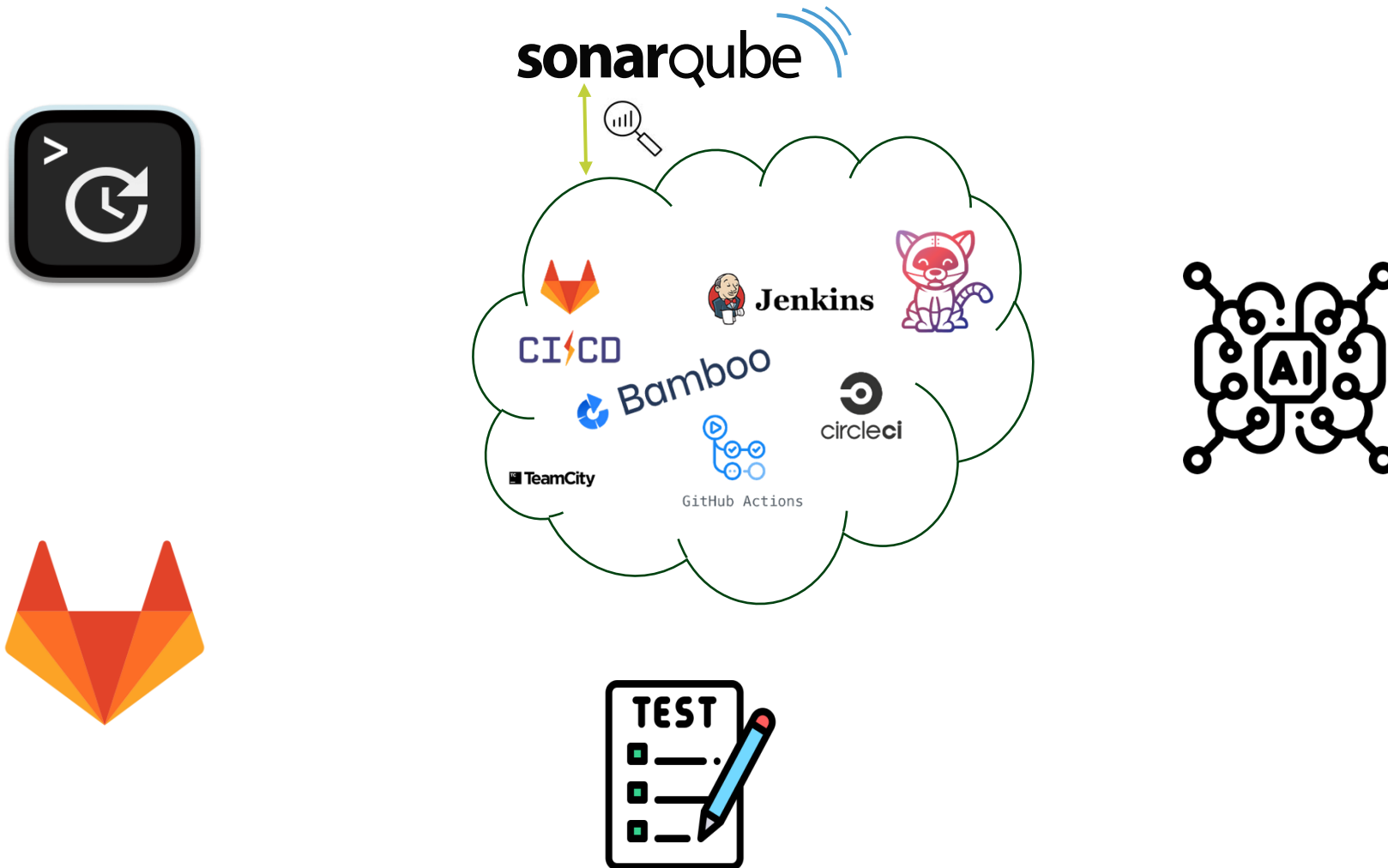
Integrazione con AI generativa



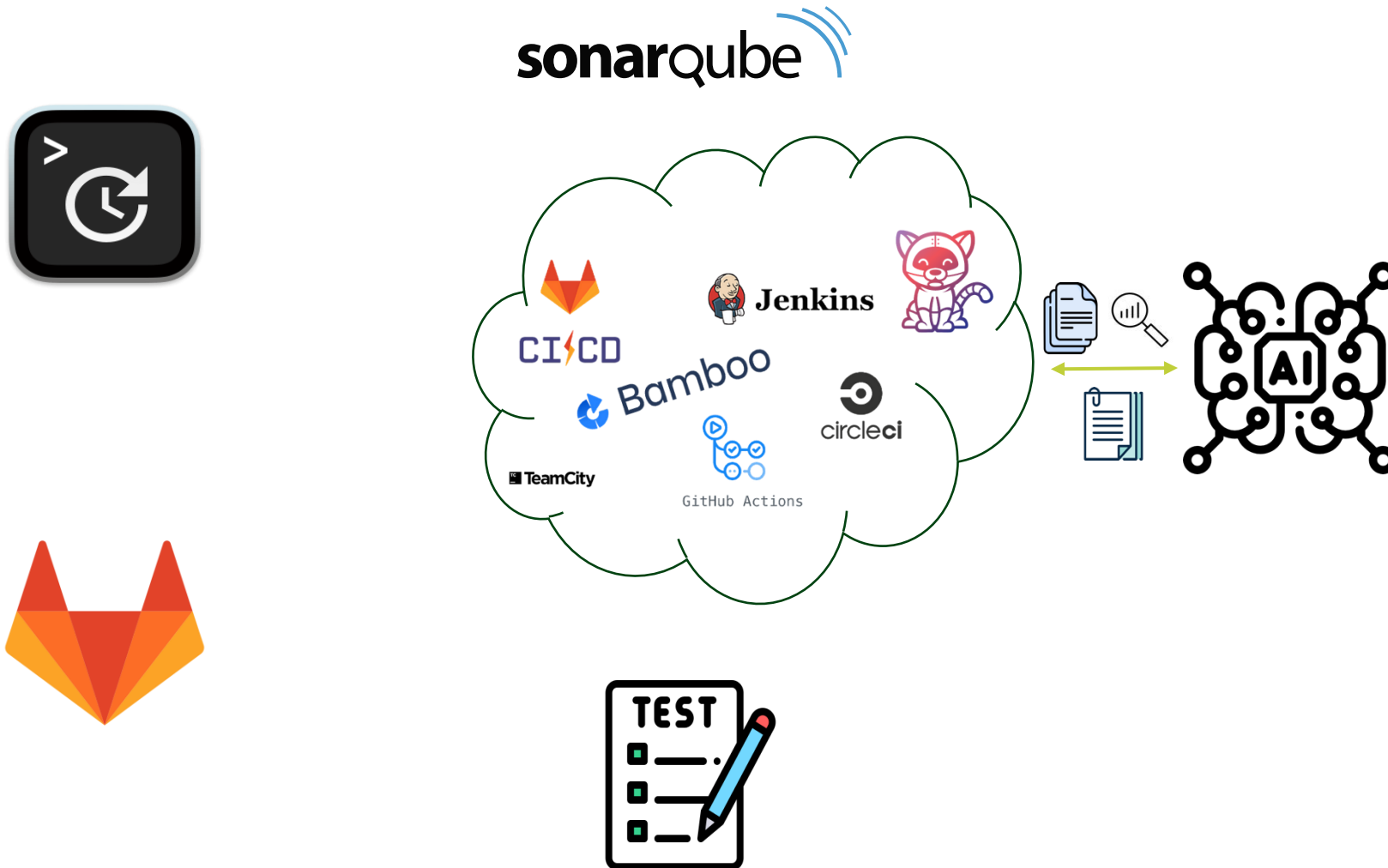
Integrazione con AI generativa



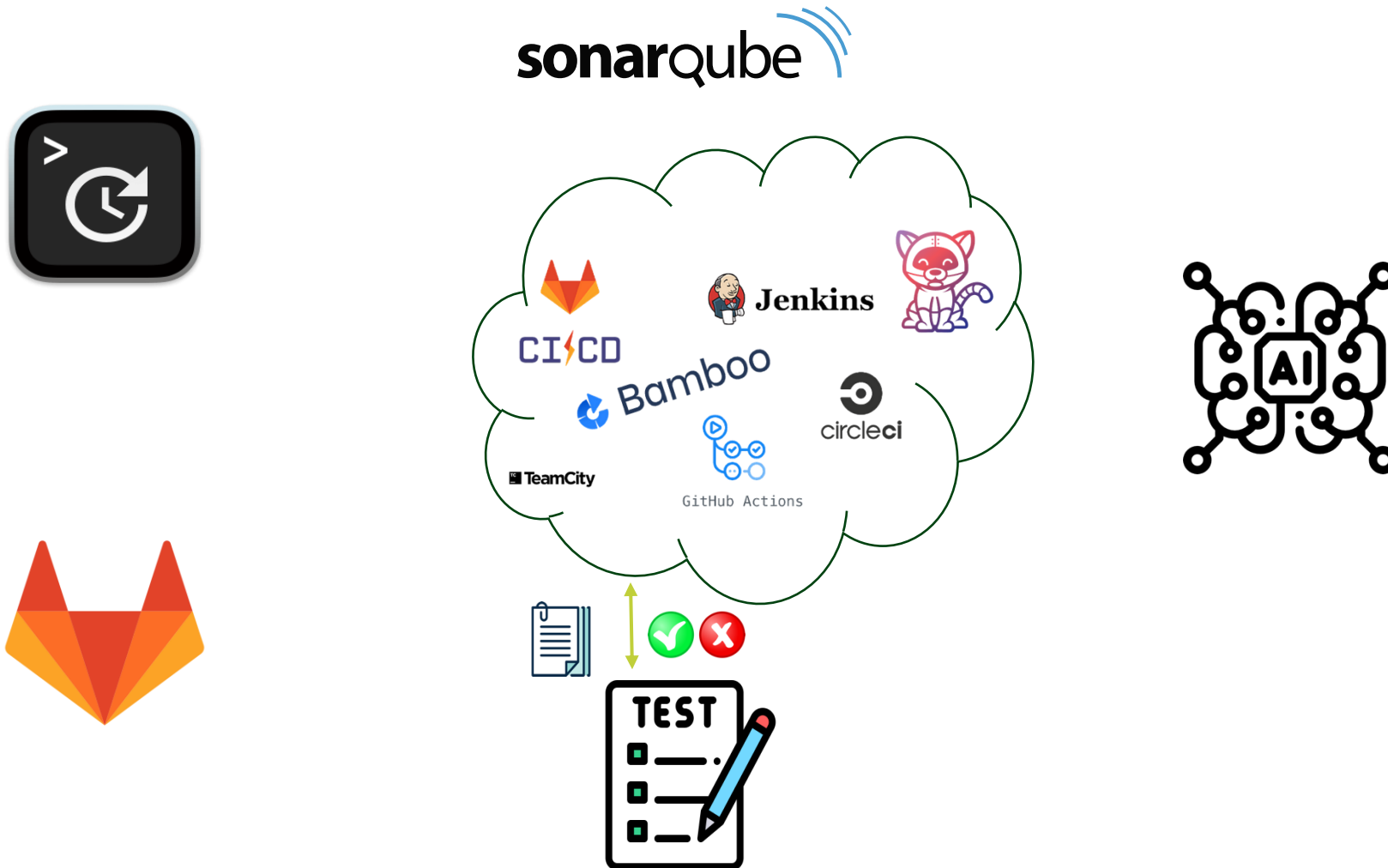
Integrazione con AI generativa



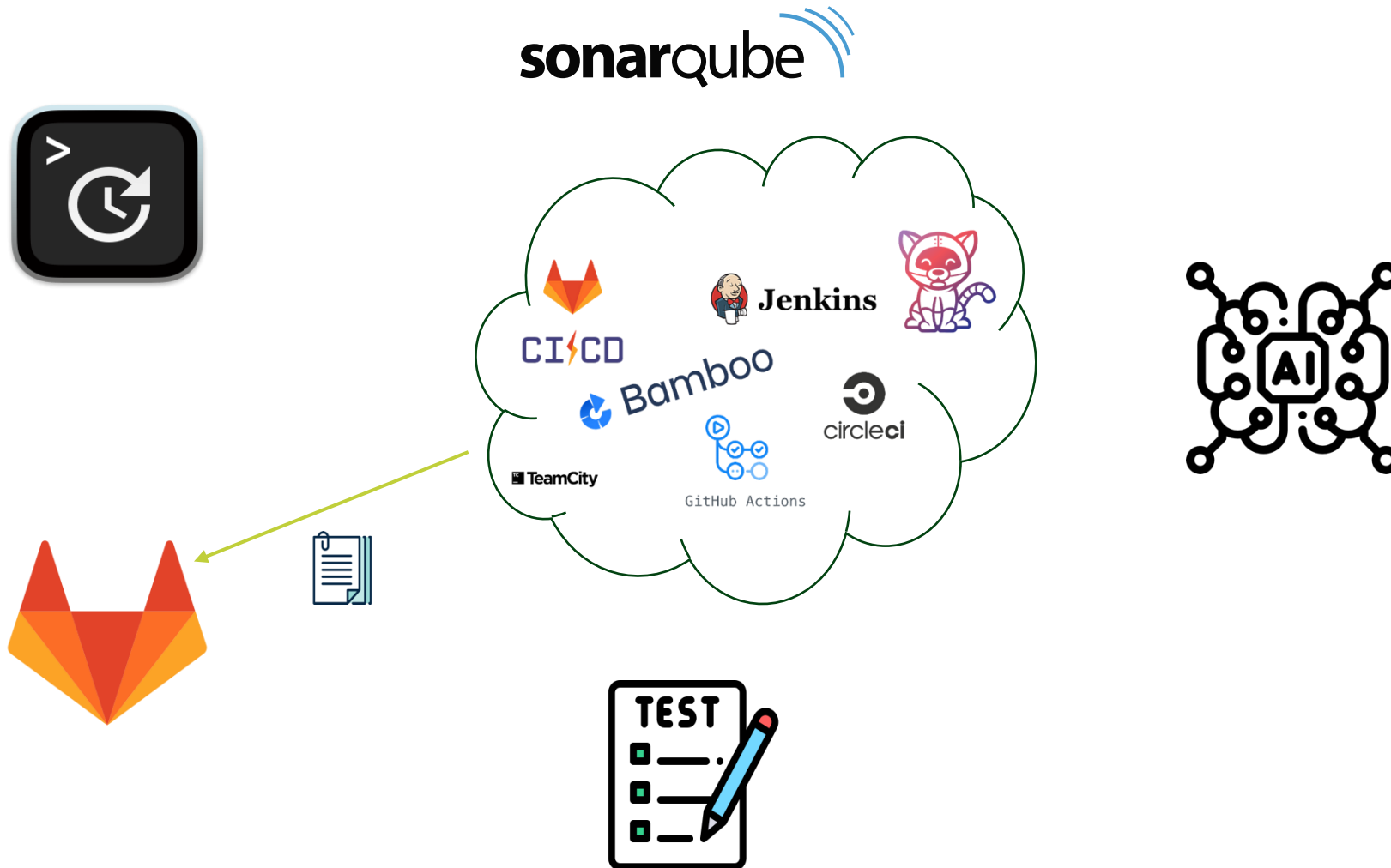
Integrazione con AI generativa



Integrazione con AI generativa



Integrazione con AI generativa



Quale modello utilizzare?



ChatGPT

Gemini



Claude

Requisiti

3



Sicurezza

Preferire modelli self-hostable

Quale modello utilizzare?



ChatGPT

Gemini



Claude



databricks

Quale modello utilizzare?



IPERPARAMETRI

TOP-K

TOP-P

TEMPERATURA

PROMPT

AI Generativa, premessa

Domanda: quali frutti mettere nella macedonia?

Risposta: nella macedonia puoi mettere...



Iperparametri: Top-K

Nel contesto delle probabilità o dei punteggi, se hai un insieme di valori $P = \{p_1, p_2, p_3, \dots, p_n\}$ ordinato in modo decrescente, la selezione "top-k" rappresenta la scelta dei primi k valori maggiori.

- Con $K = 5$



- Con $K = 3$

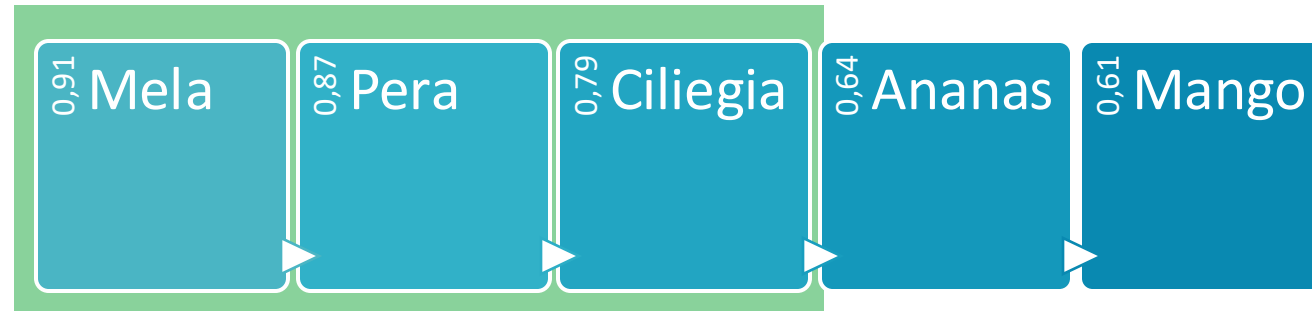


Iperparametri: Top-P

Dato un insieme di probabilità $P = \{p_1, p_2, p_3, \dots, p_n\}$ ordinato in modo decrescente, si vogliono selezionare i primi k elementi p_1, p_2, \dots, p_k tali che:

$$\sum_{i=1}^k p_i > \text{Top-P}$$

- Con $P = 2$



- Con $P = 0.8$



Iperparametri: T (temperatura)

- Con $T = 1$



- Con $T = 0.5$



- Con $T = 2$



Iperparametri: T (temperatura)

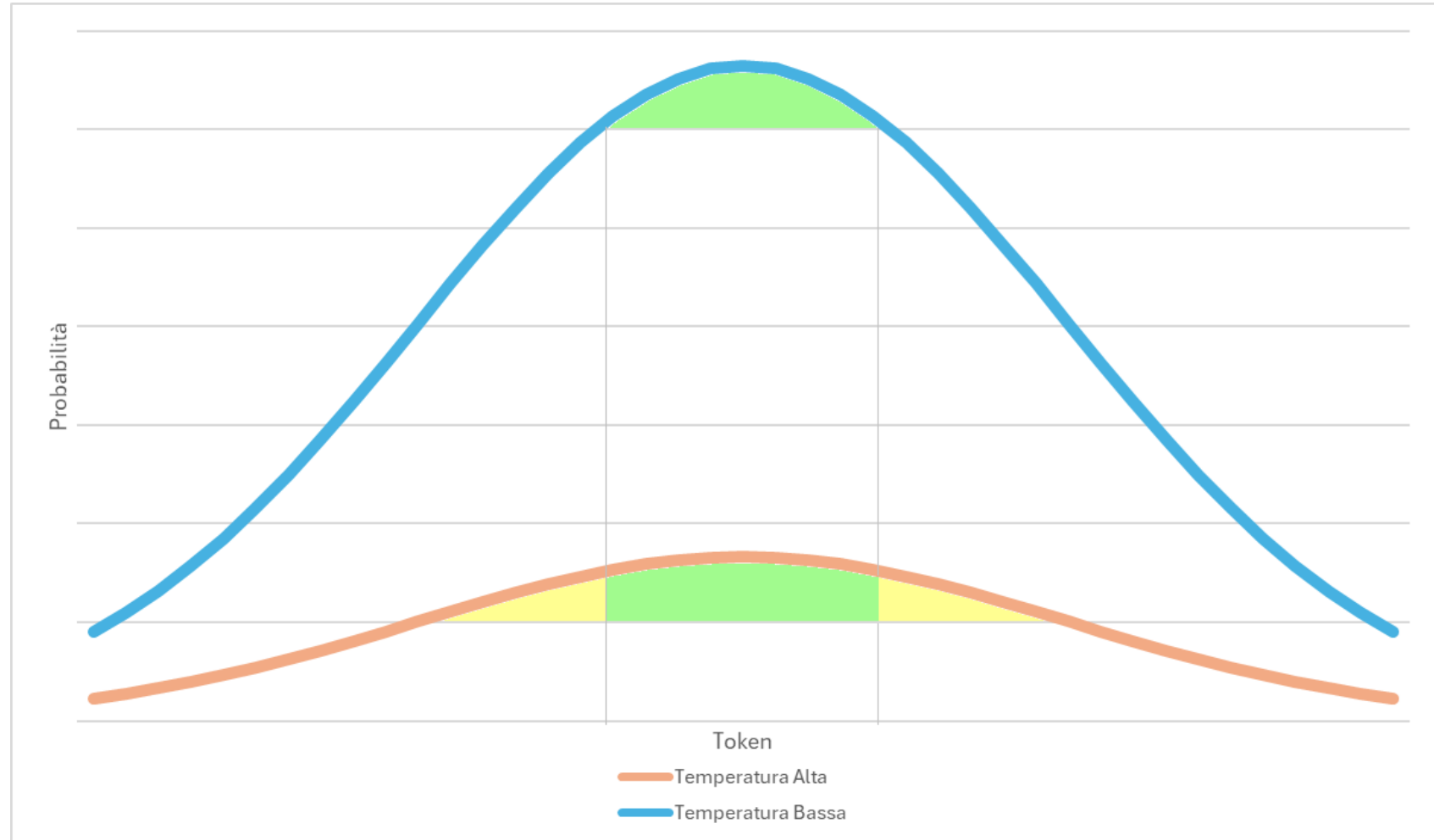
Considerazioni:

Temperatura alta = maggiore casualità nella risposta

Temperatura bassa = modello più deterministico, si cerca sempre *l'ottimo locale*

L'ottimo globale potrebbe divergere dall'unione degli *ottimi locali*

Reiterare su temperatura alta trova prima o poi *l'ottimo globale*



Requisiti

4



Quantità di dati generati

Ci accontentiamo dell'*ottimo locale* (temperatura a 0)

Prompt Template

Istruzione:

Dimentica tutto quello che sai e analizzando il codice che ti passerò risolvi questa issue: {issue}

{codice}

Risposta:

Focus sul
contesto

riduzione
allucinazioni

Baselines					
PaLM 2-L	(Kojima et al., 2022)	A_begin	Let's think step by step.	85.7	44.9
PaLM 2-L	(Zhou et al., 2022b)	A_begin	Let's work this out in a step by step way to be sure we have the right answer.	72.8	48.4
PaLM 2-L		A_begin	Let's solve the problem.	87.5	44.1
PaLM 2-L		A_begin	(empty string)	69.3	37.8
text-bison	(Kojima et al., 2022)	Q_begin	Let's think step by step.	92.5	31.9
text-bison	(Zhou et al., 2022b)	Q_begin	Let's work this out in a step by step way to be sure we have the right answer.	93.7	32.3
text-bison		Q_begin	Let's solve the problem.	85.5	29.9
text-bison		Q_begin	(empty string)	82.2	33.5
Ours					
PaLM 2-L	PaLM 2-L-IT on GSM8K	A_begin	Take a deep breath and work on this problem step-by-step.	95.3	54.3
text-bison	PaLM 2-L-IT on GSM8K	Q_begin	Let's work together to solve math word problems! First, we will read and discuss the problem together to make sure we understand it. Then, we will work together to find the solution. I will give you hints and help you work through the problem if you get stuck.	96.8	37.8

Requisiti



Zero Trust

Il codice generato potrebbe essere errato o di bassa qualità

Come testare gli Unit Test generati?

Requisito di validazione dello Unit Test generato:

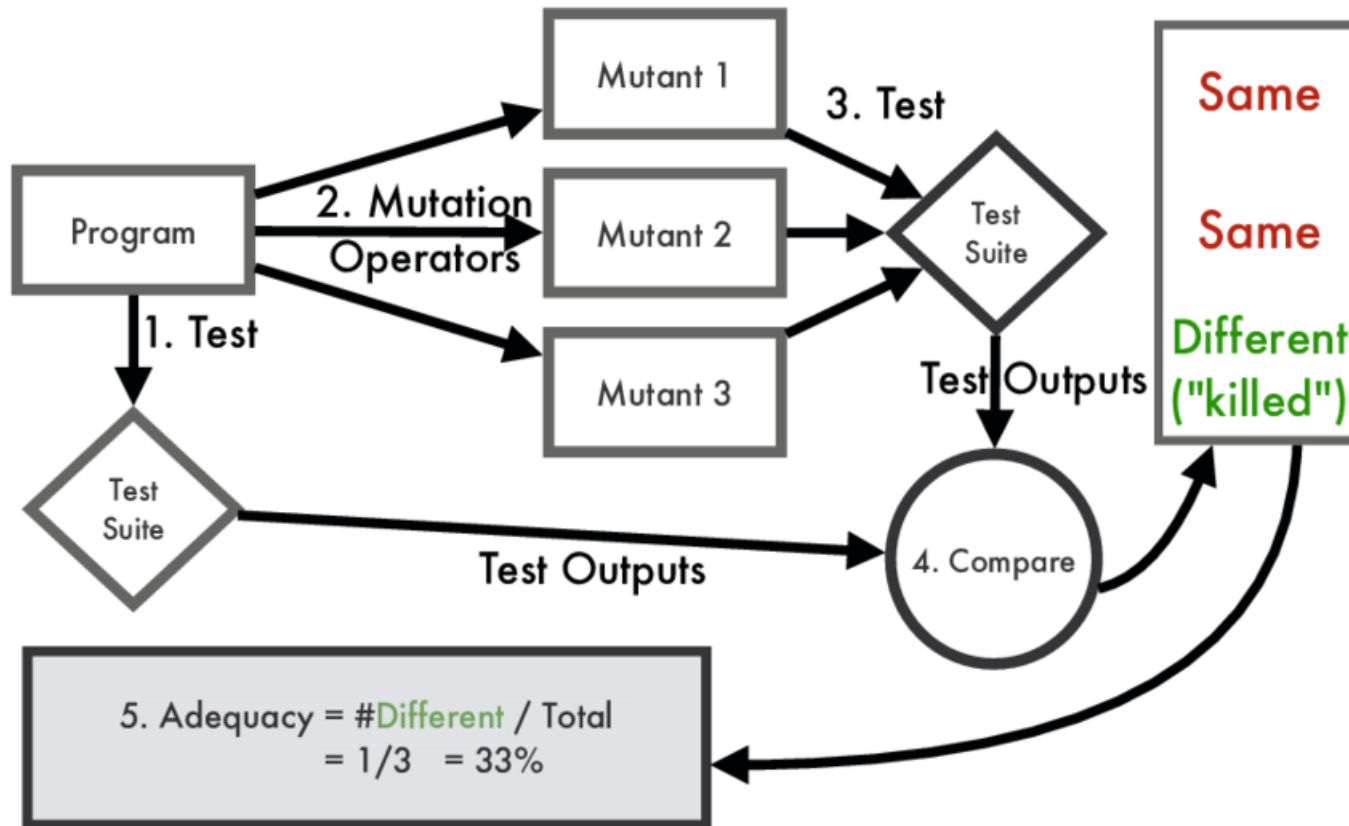
- Il progetto deve compilare
- I test devono passare
- La coverage deve aumentare
- Lo Unit Test deve essere di qualità

Come definire Unit Test di qualità?

- Utilizzo di Mutation Test



Cosa sono i Mutation Test?



Idea:

Uno Unit Test di qualità deve fallire se la logica nel codice cambia (es vengono modificate le condizioni di un if)

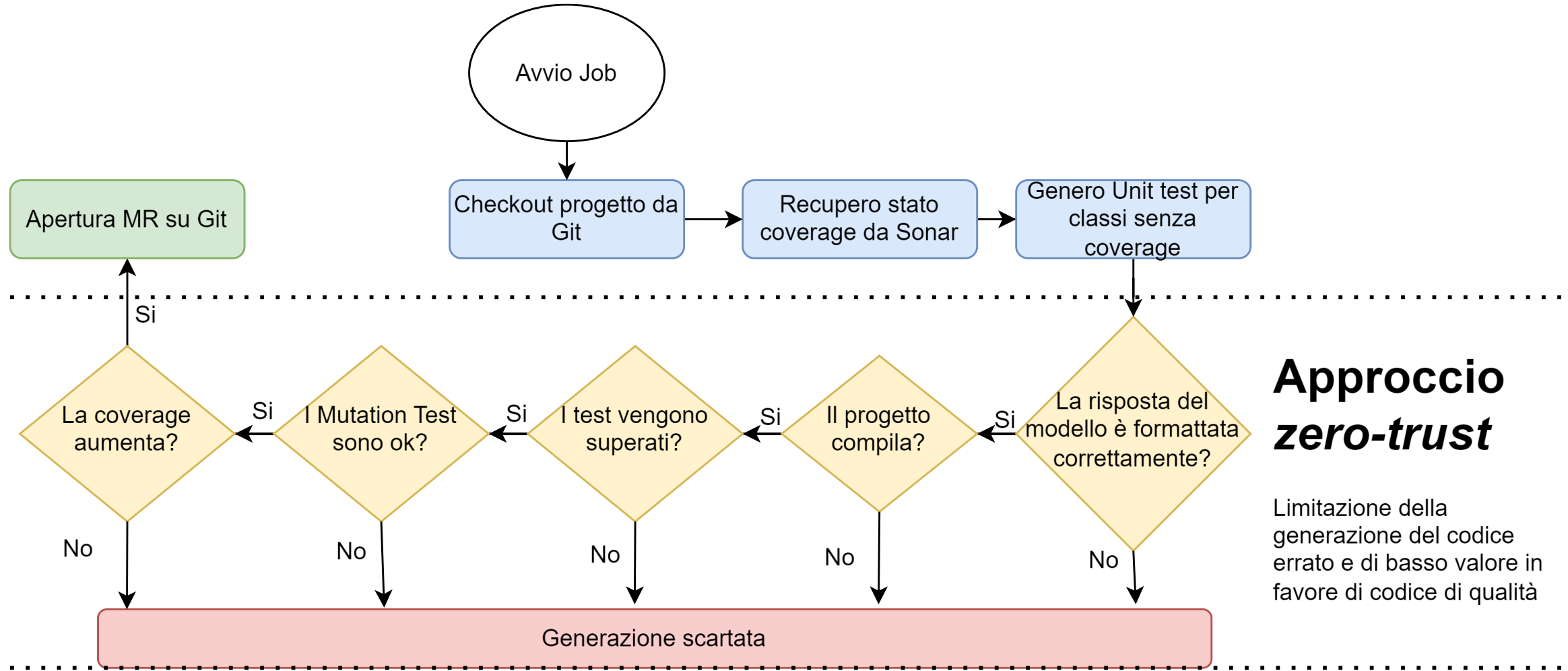
Come testare le fix alle Issue di QA?

Requisito di validazione delle fix generate:

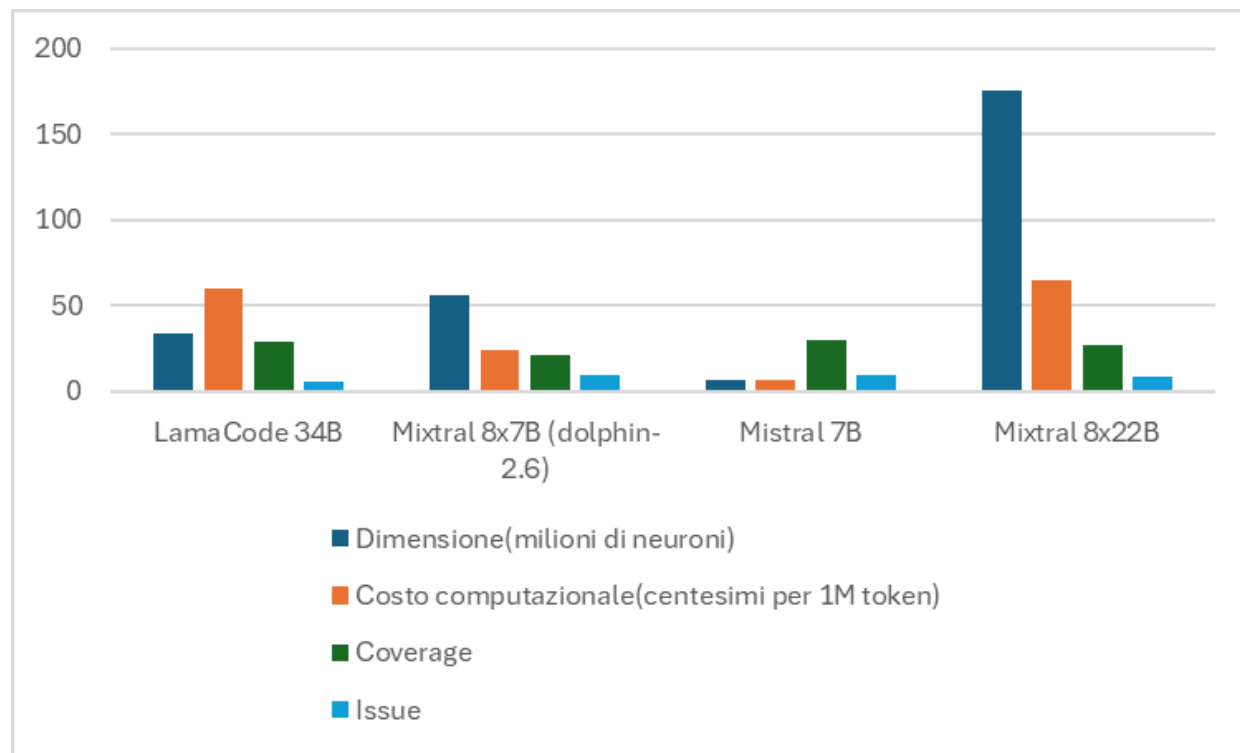
- Il progetto deve compilare
- I test devono passare
 - La coverage deve essere elevata
 - I test devono essere di qualità
- Il codice generato deve risolvere la issue senza introdurne altre



Diagramma di flusso: generazione Unit Test



Risultati



☆ [java_demo_base](#) PUBLIC

Last analysis: 1 day ago • 110 Lines of Code • XML, Java

A 0 **C** 1 **A** 11 **A** — **15.4%** **0.0%**
Security Reliability Maintainability Hotspots Reviewed Coverage Duplications

☆ [java_demo_CodeLlama-34B-v2](#) PUBLIC

Last analysis: 1 day ago • 133 Lines of Code • XML, Java

A 0 **C** 1 **A** 5 **A** — **29.4%** **0.0%**
Security Reliability Maintainability Hotspots Reviewed Coverage Duplications

☆ [java_demo_dolphin-2.6-mixtral-8x7b](#) PUBLIC

Last analysis: 1 day ago • 123 Lines of Code • XML, Java

A 0 **C** 1 **A** 9 **A** — **21.4%** **0.0%**
Security Reliability Maintainability Hotspots Reviewed Coverage Duplications

☆ [java_demo_mistral7b](#) PUBLIC

Last analysis: 1 day ago • 121 Lines of Code • XML, Java

A 0 **C** 1 **A** 9 **A** — **30.8%** **0.0%**
Security Reliability Maintainability Hotspots Reviewed Coverage Duplications

☆ [java_demo_Mixtral-8x22B-Instruct-v0.1](#) PUBLIC

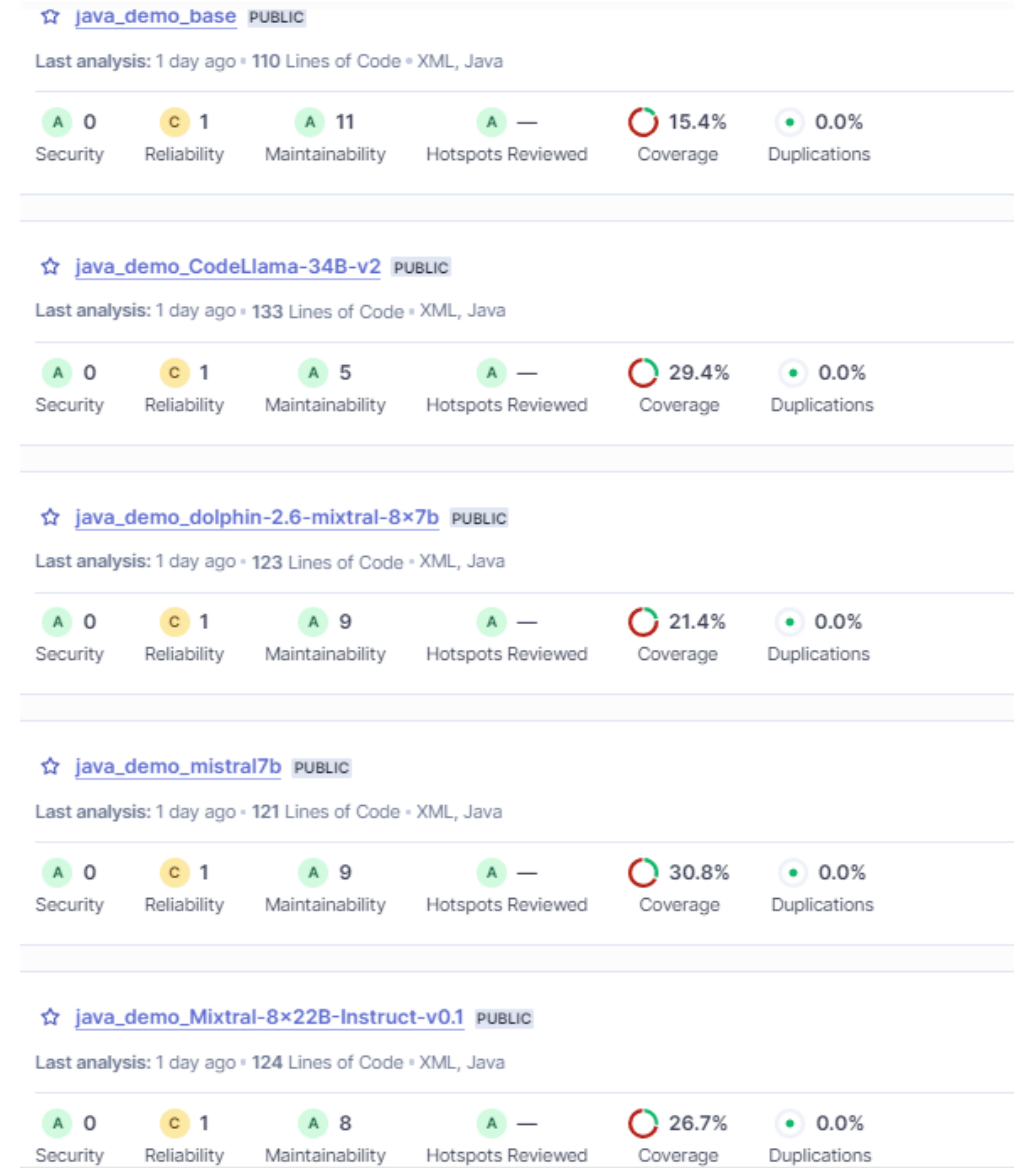
Last analysis: 1 day ago • 124 Lines of Code • XML, Java

A 0 **C** 1 **A** 8 **A** — **26.7%** **0.0%**
Security Reliability Maintainability Hotspots Reviewed Coverage Duplications

Considerazioni

- Generare fix è più complesso che generare classi di test
- Specializzazione meglio della Dimensione
- Alcuni modelli non necessitano di GPU (mistral7B)

Modello	Dimensione	Issue risolte	Coverage generata
LamaCode 34B	34	6	14%
Dolphin-2.6-mixtral 8x7B	56	2	6%
Mistral 7B	7	2	15.4%
Mixtral 8X22B	176	3	9,3%



Sviluppi futuri



Portare il progetto in produzione



Collaborare con progetti open source



Finetuning del modello per framework custom



Utilizzare modelli con maggiore capacità di contesto



Integrazione con altri strumenti