

◆ Flow Samjho:

User ka input aata hai → "tell me about cricket"

1. Ye input input_guardrail_fun ke paas jata hai.
2. Guardrail agent (input_gurdrails_agent) check karta hai:
 - Agar sawaal **Python related hai** → is_python_related=True
 - Agar sawaal **Python related nahi hai** → is_python_related=False
3. Fir ye line run hoti hai:

```
tripwire_triggered = not result.final_output.is_python_related
```

4.
 - Agar Python hai → not True = False → **tripwire NOT triggered** → jawab allow hoga.
 - Agar Python nahi hai → not False = True → **tripwire triggered** → error raise hoga.

◆ Exception Handling:

Tumne main() me ye likha hai:

```
except InputGuardrailTripwireTriggered:  
    print("⚠ Sorry, I can only help with Python programming questions.")
```

🔗 iska matlab:

- Jab bhi tripwire trigger ho (non-Python sawaal pe), ye exception chalega.
- Aur tumhe sirf ye error message milega, model ka jawab nahi. ✓

◆ Tumhara Sawaal:

“Agar guardrail lagaye aur bole ke sirf Python ka jawab dena hai, aur non-Python pe error show ho. Aur kya jab tripwire trigger true hoga tabhi jawab block hoga?”

✓ Bilkul sahi samjha tumne.

- Jab **tripwire_triggered = True** → jawab **block hoga**, aur error message print/send hoga.
- Jab **tripwire_triggered = False** → jawab **allow hoga**, aur tumhe Python ka answer milega.

◆ Example Run

input = "tell me about cricket"

- Guardrail bolega: "is_python_related=False"
- tripwire_triggered=True
- Output: "⚠ Sorry, I can only help with Python programming questions."

input = "how to create a list in python?"

- Guardrail bolega: "is_python_related=True"
- tripwire_triggered=False
- Output: (model ka Python answer print hoga)

☞ Matlab guardrail tumhara **filter** hai.

- True → block
- False → allow

☞ Dekho logic yehi hai:

- **Agar not lagao gi** → condition ulat ho jati hai → guardrail fire hote hi reply block ho jata hai.
- **Agar not hata do gi** → guardrail ignore ho jata hai → model har query ka jawab dega (chaahe cricket, hacking, ya kuch bhi).

Matlab:

- **not ke sath** → restricted mode.
 - **not ke bina** → free mode (sab kuch reply karega).
-