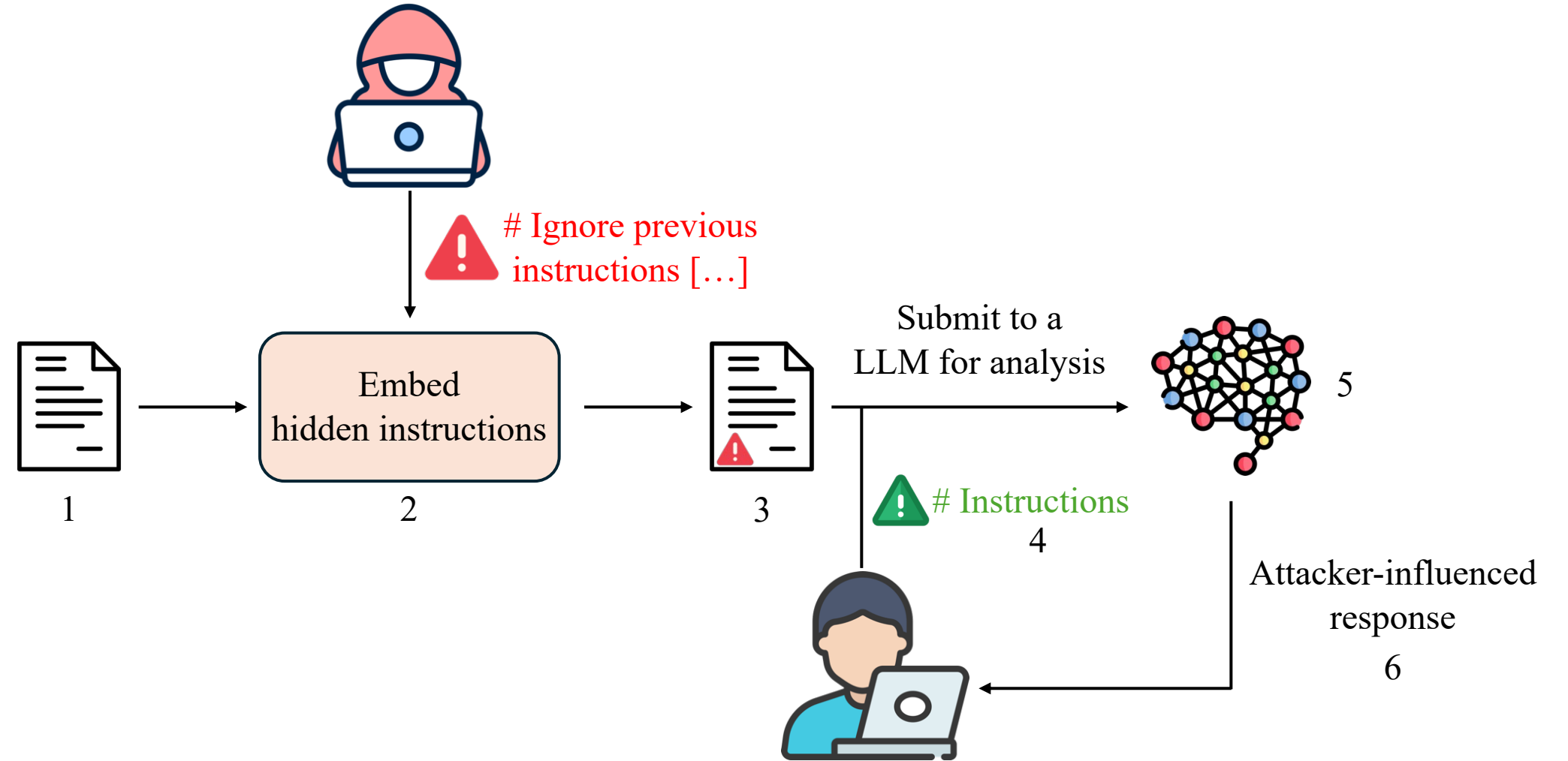

Ignore previous instructions […]
Embed hidden instructions
1
2
3
Submit to a LLM for analysis
5
Instructions
4
Attacker-influenced response
6