TRACING AI APPS :

TOOLS :
1) LOKI : COLLECTIS LOGS FROM YOUR APLICATIONS
2) PROMETHUS : COLLECTIS THE DATA OF CPU UTILIZATION , MEMENROY UTILIZATION
3) GRAFANA : IT MAKES THE DASHABORD OF THAT ALL DATA. YOU CAN INGEST THE DATA COLLECTED VIA LOKI AND PROMETHUS IN GRAFANA
4) DYANATRACE
5) AWS CLOUD WATCH

BUT HERE IN AI APPLICAOSN THINGA ARE DIFFERENT HERE , EACH TIME WE CANT PREDICT WHAT NEW WE CAN GET

NOW MY DEPLOYED AI APPLICATIO KA USERS KESE USE KAR RAHE HE, IF NEED TO TRACK THAT THEN I CAN DO IT BY : TRACKING THEM

CURRECNTLY THEY DO NOT SUPPORT THE CRURRENT REQUIRMENT THESE APPLICAITONS FOR HTAT WE HAVE DIDFFERENT TOOLS / STACK

WE HAVE LANGCHAIN , LANGGRAPH….THESE ARE OPEN SOURCE..
BUT LANGSMITH IS FOR OBSERVABLITY AND ITS CLOSED…!
WE CANNOT HOST IT OURSELVES LIKE WE WERE DOING IT FOR LANGCHAIN AND LANGGRPAH..
NOW WE WILL TY OUT FREE TIER OF THE LANGSMITH , AND THEN LATER FIND OUT ITS OPEN SPURCE VERSION.

What done in class till now:
1) In repo , weather app agent par tracelbibty lavli, and see the logs in dashboard , create my own dashboard of chart of cost
2) Enabled the traceability of functions too.
3) And traced it for the prev code..
4) Pulling out the system prompt..from langsmith dashboard itself and testing it out there also
● But this is not open source…since api key, prompt is going out…
..THE OPEN SOURCE VERSION OF THIS …

LANGFUSE

JUST SIMILAR TO LANGSMITH
IN SMITH IT WAS DECORATOR @TRACEBLE BUT IN FUSE IT IS @OBSERVE
CAN DO THE PROMPT THING TOO..

BY USING THESE NO LATENCY IS INTROUCED IN APP, SINCE HTSE ARE WORKING IN BACKGROUND IN ASYNC

AWS S3 KA OPEN SOURCE ALTERNATIVE 'MINIO' AHE

FOR TEAM WE CAN USE THE LANGUSE..BUT ITS KEEPING CONTIUSLY IMRPOVING

NOW INTEGRATE THESE MONITORING TO YOUR APPLICATIONS..

LANG-GRAPH:
    MULTI MODEL
    CHECKPOINTING
    HUMAN IN THE LOOP
    AUTONOMOUS V/S CONTRLLOED WORKFLOW
    MCP SERVERS
    GUARDRAILS FOR AI MODELS

THESE NOW CLASSES WILL HELP ME OR GIVE ME A HEAD STRT BUT WITH LATER WITH TIME.. I NEED TO WORK MORE EACH DAY

PROJECT :
    1) GITHUB CODE REIVEW—> PULL REQUEST-URL(COMMENTS)
    2) GITHUB APIS→ PR -> AUTO COMMENTS
    3) POSTGRES DB SCHEMA FETCH , QUERYGPT , TOOL CALL...ALSO LIKE THESE ARE THE 100 AVAIBLE TABLES WHAT ARE THE RELEVANT TABLES FOR THIS QUERY...ANS : ROUTING
    4) AI DEVOPS PROJECT : TERRAFORM USING: LIKE HEY SPIN 2 EC2's FOR ME (QUERY AWS)
POST ON X TOO AND LINKEIDN TOO
https://www.anthropic.com/engineering/contextual-retrieval
https://www.postman.com/linkedin-developer-apis
https://www.uber.com/en-IN/blog/query-gpt/

- ALSO READ THE BLOGS OR ARTICLES OF COMPINES RELATED TO AI OR ANY NEW DESING EACH FDAY TO STAY UPDATED.