COT(Chain of Thoughts):→ Basically its a prompting technique, Kisi bhi conclusion par aane se pahle hamara model bahut baar sochta hai, phir kisi conclusion par pahuchta hai.

Imp. Topics:-

Architecture of LLM
fine Tuning
Training
quantisation → Kon sa technique use hota hai kaise humlog 32-bit se 16-bit par aate hain, 8-bit par aate hain, precision kaise change hota hai.

Hugging-face का bits & bytes library (gothrough)
Hugging-face का blog.

→ Quotient → marktechpost
→ Cohere → qdrant
→ pinecone

Langflow & flowwise → ye dono platform chatbot wade cheez, Rag, Advance Rag, ess sab ko drag & drop se kaam karwata hai.

Parameters of Training Arguments:

1) output_dir = 'Folder_name'

Jesme hota kya hai ki training ke time jo kuchh bhi hota hai unn sabhi cheezo ka snapshot leke uss directory me store hota chala jayega.

eg→ accuracy, score of evaluation, loss, value of weights and parameters., esn sabhi cheezo ke snapshot on a every particular steps.

Model checkpoints:→ ye ek neural network model ka snapshot hota hai, jo kisi bhi certain point par training ke time liya jata hai. aur wo certain point aap define karte ho, ki aap wo harek step par chahte ho ya ~~em~~ har epoch par. esike according aapko snapshot milega aur output_dir me chala jayega.

ess snapshot me rahta kya sab hai?

↳ esme model ka architecture rahta hai, layer kitna hai, perceptron kitna hai, weights ka value kya tha, hidden layers kitna tha. attention mechanism me kya-2 value use hua etc

→ num_train_epochs = $\underline{\underline{n}}$

↳ means hum apne training data par kitne baar apne model ko train karna chahte hain.

→ Per_device_train_batch_size = 8

↳ Agar aapne koi device assign kiya hai eg→ CPU (or) GPU.

let's suppose aap distibuted training kar rahe ho. means CPU par bhi train kar rahe ho aur GPU par bhi train kar rahe ho. ek hi saath me. means har device par ek saath 8 sample ka hi training hoga.

→ per_device_eval_batch_siz= 8

↳ esme training ke badle evaluation karte hain.

→ logging_dir = 'folder.name'

↳ ess folder me aapke jitne bhi logs hain wo jaake store honge. logs means, runtime kitna laga, error kya aaya, kitna slow tha, kitna fast tha, ye saare logs usme jaake store hua hai. (during training)

→ logging_steps = 100

means harek 100 steps ke baad wo log define honge.

→ save_steps = 500

jo aapne output_dir file banaya hai waha par jo aap store kar rahe ho, wo harek 500 steps ke baad usme save hona chahiye

→ evaluation_strategy = 'epoch'

↳ let's suppose agar save_steps nahi diya ho uss time par jo hamara evaluation hoga model ka wo ab har epoch ke baad hoga naa ki steps ke baad.

→ save_total_limit = 2 → aapke bahut saare check-points wo output_dir me save hote jaa rahe hain. usme lagbhag 50 checkpoints save ho gaye. yaha par jab hum save_total_limits = 2 karte hain tab wo last ke ya bahut hi recent ke 2 checkpoints store hua hai uss directory me sab usiko reflect karega.

→ load_best_model_at_end = True

↳ ye kya karta hai ki, jitne bhi snapshot store hue the hamare directory me, during training, usme jo best model hoga wahi uth ke aayega. aur wahi hum use karege.

→ metric_for_best_model = 'accuracy'

→ ye method hai ess baat ko check karne ka ki kaun sa ~~the~~ snapshot badhiya hai, ye cheez hum accuracy se measure karenge.

→ greater_is_better = True.

means jiska accuracy jitna jyada wahi badhiya.

Q Why We use accelerate library?

→ This accelerate library is made by hugging face. ye hamare training performance ko aur speed-up kar deta hai

Transfer learning:→ It's like using what you ~~know~~ already know to learn something new. Imagine you have learned a lot about animals in general. Now if you want to learn about a specific animal like a lion, you start with what you already know about animals and then focus on what makes lion special.

Fine tunning:→ Think of it like adjusting a reciepe to make it perfect for a special occasion. you start with a basic reciepe (pre-trained model) and then tweak it a bit to make it just right for what you need. so, if you have recipe for a cake (pre-trained model), you might adjust the ingredients or cooking time to make it perfect for a b.day party (Specific task or Domain)

[transfer learning is about existing knowledgle to help with a new task, while fine-tunning is about making small adjustments to improve performance on a specific task.]