Python and AI Power-Up Program Offline Class-20251006_191204-Meeting Recording

October 6, 2025, 1:42PM

2h 18m 25s

Tirth started transcription

Tarun Jain 0:03

I got a movie.
Oh, OK, so So what we'll do is let's just list down the issues with the LLM and.
We are now going to huh. And issues with rag. So can we list down some issues? So actually so you should answer one word for everything. So total it should be like. 34567 or so total we should have total 9 reasons. So four reasons for five reasons for LLMS, 4 reasons for that that cannot be summarized.
Uh, what is the issues? Bag something? OK, points? No, just bad something. Any specific example?
OK, relevant chunk missing.
And Isma, your search also didn't perform well.
Search.
Then.
Perform well, OK, uh, LLM. Apart from hallucination.
Hallucination is not.
OK, uh, can we, uh, send it uh it?
You should create with a big amount of data or sizings. It is context window. Yeah, right. OK, context window.
How to optimize the ROM?
Token phase. Huh. Hmm. What is the issue with what is the? What do you want to talk about?
Support that.
Thank you.
Meanwhile, you'll have your answers keep in your mind so that it will be with and Joe Onliner. What you can do is you can message your this thing, but don't enter it until I tell your name.
OK, until then you can think, but I'll keep this reserved.
For elements.

How to get data?
OK, they're turning on past data. They don't know how. How to tell the data. Can I tell one call back as well? Yeah, you can tell. OK.
OK, you have that point today. I'll keep it resolved.
Sometimes we need to for which one?
Then sometimes we need to reload all the data to to get the appropriate result. For example like I'm if like I'm using PDF and I have installed the.
Data with the with the chunking and if like if there is some issues with the coding or something like that, then I need to refresh the data to get the appropriate result.
So results in the sense raw extraction of the data or the output.
So how is it linked?
That's time.
Let's keep your keep your answer ready on either Ram Krishna and the rest of them are all here. OK, we have just joined for the sake of joining so we can look at this and this. OK.
Then Ramakrishna can keep his answer ID. Yeah, if I mean then summarization. Yeah, if I want to summarize the complete document, I cannot. Same in a context window.
No, no.
Like I want.
How many times the TT is there in my document? I cannot get it with the right.
What will problems? It's not related to a problem. Yeah, yeah, yeah. So what are the issues with that?
Can we add something like that key when we deal with the multiple type for for data like let's say HTML or web PDF or where JPG or wants extra calculated information to cover it. What's mess with the consistency of a proper name?
For multi document, multi document, but any specific issue you saw when you had multiple documents, let's say we have PDF for a document, OK OCR.
Yeah.
But I'll I'll tackle all these issues. So basically this is feature part like feature missing instead of issues, but we'll tackle it. I think mine also will be again feature missing in LM but tool calling is very don't miss pardon Sir like tool calling. I mean most of the LM is like function calling.
Yeah, Mrs. did I told.
But it's like, yeah, it depends on LLM, but no, it's still it's like many, many Nitish.
Which one drag or LLM?

Or garbage collector.
You work yeah you do wrong or you have something or when we like making chunking and then at the time sometimes issue with the diversity limiting.
I'm talking about that diversity in the sense what chunks are extracting what? Yes, and and uh, searching so it didn't perform well.
OK, I'll tell one. I'll just give a hint. What you can do is just relate where you can think it will be easy. So let's suppose you have a rack pipeline. OK, you're using a rack pipeline and if you ask a question like good morning, what is the answer? Good morning.
Think again. Good one with extra. No, we won't give you that extra. What is the output here? You're using LLM. I'm using LLM. It's an empty. If the data is not, I mean like there is no, but you're using LLM. Let me give a we'll test this out. We'll test the output so.
Let me open our example for existing code that we have seen so far. If I ask good morning, what will be the output? Well we are seeing it is not related to say I don't know so I so it will be I don't know correct. So if it is not there it will say good morning because it is then if I ask Virat Kohli then what will happen?
Who is Virat Palli? OK, so as per what we have done, so as per what we have done so far. Now what you need to do is as per whatever you have done so far with LLM starting from mugging face, hugging face to Gemini to function calling and we also use tools and then we came to RAD.
And in drag there were multiple things which we didn't cover, so which I told this is the thing you can use schematic chunking. So this was valid point bad chunking. Now OCR was a feature right? But in our prompt how did we write prompt if it is not?
In this particular query you can just say I don't know. In that case if I have prompts like good morning, it won't answer even technically it should tell good morning because this is a very generic question. But if you ask Virat Kohli then it should not answer.
So it's like Virat Kohli is different and these are greetings. No, it will not answer. You can try.
Because that that is because of the prompt that we wrote, huh? That's because of the prompt. Now what you can do is you can modify prompt saying that hey, if there is any greetings based portion you can first greet and then you can relate. Hey do you have any questions for me so that part we missed it out.

But why I'm adding all these issues is because you can tackle this when you're doing agents. So so agent are very clever enough to understand the intent of your prompt and then answer right. So that is where agent excels and when you build agentic like application if you have questions like good morning.
It will tell you good morning instead of telling I don't know. And if you ask any questions related to your database it will use that as a tool. Now knowledge base is a tool and now if you ask any questions which is related to general question it will use search. So now search you are adding that feature. If you're not adding that feature it will say I don't know.
It will say I didn't found that in the database, so that is the application of agents. So now if you can tell the issues, these are all valid issues here.
So one thing is uh.
And here how to optimize the prompt token size. Yesterday when I was having brainstorm session with teeth when we had that project thing, I told we use prompt optimizers.
Right. Sorry, not optimizers, prompt compressions.
And you have uh from Microsoft there is a product uh wrong compression.
Microsoft.
There is one library LLM Lingua.
How about the the one with the cloud and the open? What the? No, this is a technique. It's a research technique. It's not optimizer. This is a technique that is usually used for prompt compression. OK, so if anyone is using very large prompt, you can insert this.
This LLM lingua within your application to optimize your prompt token size. So it uses very light model and since most of the time let's suppose in RAG if you want to add custom metadata, what are you trying to do? You are creating the summary. After creating summary what are you trying to do? You are saving it in a metadata. So all these things.
Is happening one only once. Indexing happened once, so it doesn't matter how much time it takes to create an index. Once you have index, user will use the saved index which is your collection. So even if you spend time in your indexing of document, that is fine as per the design rules.
This is happening in the offline mode, so offline mode it's fine if it takes additional 20 seconds, 30 seconds so you can insert there when that is your problem statement. So now coming to outdated data you can use tools.

Now coming to function calling, if it misses the right tool, don't use first of all and second use open AI in open AI. If you are using any tool, probably we have to have intent classification which.
I would you call open AI is very good with and one more example order to give is so uh like wrong I was more specifically for talking about LM uh function for these LM like 1G4V.
To 7:00 PM and such.
They have this natively they don't support properly function calling. So for function calling every time you should have tokens. So there is one research paper, this one enhancing function calling capabilities and LM's.
I'll show the prompt on how you need to define this.
So every time you do function calling, we have something called as decision token. So all these are advanced technique, like once you start facing the issues, right, how do you navigate it? So that's the reason why we'll list down the issues that you're facing, we'll tackle it and all these things are related to agents, right? Function calling is the major component of agent.
And whatever I showed here the good morning that is the issue of decision making. RAG is not able to do decision but agent can do decision. So that is 1 drawback which can be solved in agent. Function calling is the feature of agent which has to be improved. That is where you use.
Decision tokens. So now what decision token does is you will tell hey this is the starting of your prompt and then you will give two tokens where you have to start and end your function calling results. So how does function calling works? You ask a question.
Once you ask a question, let's suppose.
Who won? Asia Cup 2025. So technically LLM doesn't know about Asia Cup, right? So now you'll ask this question. What will happen? You have LLM which will decide what tool to pick.
And once you decide what tools to pick, it will also decide what is the query that it has to use to search it. So it's like what tool to pick and what will be the parameter. So let's suppose you have search. Search can that you have QQ is your query.
So what you need to search will be decided by LLM. So this is called action step. It's a chain of thought. No, chain of thought is different. Chain of thought is just prompting. You think step by step. OK, here you're using two LLMS. Chain of thought is just one LLM. OK, so you have question.

Which is real time. Now LLM will decide what tool to pick and after this it will take action. Now you have context.
So where is this context coming from?
In the back, no from search. So let's suppose you have search.
You have search as queue and inside that you have a logic. In this logic what will you have? You have number of results to be 10 and then you have a EPI which is a search which takes query. Then you have maximum results. Maximum results equals to.
10 correct And then what will you do? You will. This is in list. You will combine this into context and then return context.
So whatever input you have here, what this is the output of.
Output of function correct? Now this will go as a input to.
Hello.
So now this LLM has your question and as well as context to give you final response.
So here if you see LLM when you're deciding this you will use something called as a decision token instruction following with decision token where you give hey this is the start you increase the LLM.
Calls function calling will every time use two calls. One is to decide function and 2nd is to generate final response. So when it decides the tool you will use this part instruction following with you'll see this everyone used tably search right?
Diably search and as well as we used Google search Google search. So when you ask a question did you see it uses our own custom function and then use this query. So now if you see you have a question you will have this particular function and then it will make this query. So this query will be decided by by LLM.
So if you want to improvise this entire approach, we use this decision token where you are adding this use tool. So you'll have use tool, you will have start and you will have end. Is this clear? We'll try to replicate this paper. I do have a demo code of this, but I'm not yet getting the final solution. But once that is done, we'll discuss.
It out.
OK, so coming to the other issues, did you get your issues, Ramakrishna? Probably you can enter now if you have entered your prompt time in the issue.
When you think.

16:56

Yeah.

RamKrishna Bhatt 16:58
Hello, am I audible?

Tarun Jain 17:01
Yeah.

RamKrishna Bhatt 17:02
OK, OK OK so I don't it should find I will not able to give input.

Tarun Jain 17:10
OK, there is one issue that I'll say it's hallucination, but hallucination. Many people will say this as a feature instead of a bug, right? So this is something that I'll exclude, but it's an issue which many people tell.

RamKrishna Bhatt 17:14
Sorry.

Tarun Jain 17:25
Now the thing with RAG is it is good because most of the time what happens is in some domain it's OK if you're not giving the responses, but it is definitely not OK if you give wrong responses. So for that RAG is well suited, but for generic queries sometimes you can keep it open.
Where agent will be used OK as a language, multiple languages not supported. That is the embedding issue. If in case let's suppose you are building RAD and.
Complete the point. Sorry. OK, so now for all these things, right? All the issues with LLM, there is something called as seven issues with rag paper.
So if you see here is where most of the issues will come in. Yeah, this is the only survey paper many people uses to give talks related to RAG and I'm pretty much sure yesterday there was one talk in Ranil used. He used this paper. I'm 100% sure. So first issue what will happen is the missing content.
Now when we started talking about RAG, I told why did we use spy PDFM? So if you have tables, let's suppose in tables if you have 25 degrees Celsius when you extract your data if it becomes 250.
Number yeah, then it's a very big problem, right? So how you extract your data is

also very important. So that is where you have this missing content and missing content is also part of search. So what if I ask anything related to Adhyanti, but that chunk is not there in my retrieved context?
Or if I take better example, in one question you have two questions. So you have chunk for first question but the second question chunk is not there. So that is the issue of missing content and then you have mixed top rank again.
Whenever you are getting the five chunks right, most probably your chunk that you need to be first, it is at 5th and since the context window of LLM is not able to pick the final part of it, that is also a problem of top missed top chunk re anchor.
But as I already mentioned, when it comes to reranking, this paper probably should be before 2024. Now models are better with more context window. So reranking issue is not faced as what you faced earlier. So what is the context window of models earlier? It was just 64 K 122 K Now you.
We have models like 2,200K tokens, so this issue is currently resolved and apart from that not in context. Let's suppose I ask what is black hole and black hole context is not there within my prompt. Then how do you handle those issues? And then wrong format is basically.
OK, wrong format. I forgot. Incomplete in the sense whenever LLM generate a response, you get incomplete response. So for example, you might have noticed I always ask the same question. What is the framework atyantik used? So even though you had six to seven names in your website but only four were released.
So now the remaining three, let's suppose that are your main text tag. Now that is not there in the response, so that is again a incomplete response. So whatever you see here, right this part you can tackle with prompt.
Wherever you see the ending part and whatever you see these two missing part here, these two-part can be tackled by search which is our vector database. And what are the techniques you can use? You can use hybrid search plus MMR for these two parts, MMR plus hybrid search.
And rewriter is something that we have not implemented. What happens in rewriter is you ask a prompt, you use an LLM to frame it as such that is related to your data, right? So that is where you use rewriter, but this is very expensive.
Most of the time this will not even work. So if you're working in any internal team usage then you can use it. Now I'll tell you what are the disadvantage of using rewriter. Just one more context. You are increasing context or using more LLM call. So whatever you use at rewriter right, you'll have more context.

So you have the document summary. Document summary is your input for rewriter. So whatever metadata you have right, metadata is your input of rewriter. So I'll show the two bar article.
So if you look at this rewriter part, so you ask a query after query, what do you have? You have query optimizer and then if you see processed queries metadata filtering. So when you use this part right you want to know what is your metadata like document summary.
Chunk keywords, chunk FAQs. So those are some additional source source file. You want to make LLM know that what your data is about and the best data information is in the document summary. So once you look at the document summary you will update your prompt. So whatever you ask as a user.
That prompt will be updated. That's it. Nothing else. Basically, OK, if this is the context, this is the metadata, what should be the right, right question, correct. So we have of that question, but this is the context, then what should be the right question, right? Then you rewrite that question and then you give it to the prompt. So this is something like.
We'll write a way to is killing the problem when I get it. And here there is one more thing when it comes to rewriter. I was reading one more research paper. There are people, if you ask one question, you're generating 5 questions.
Which are relevant to that question. So instead of just optimizing one question, now you have 5 questions. Out of all these five questions, you're using it for retriever, which again is very cost expensive. I'm not sure how well this solution will be, but I've seen people generating 5 relevant prompts for that so that you can improve your retriever.
But in terms of latency, I can't comment.
It will take time definitely, but how much time will it take? That has to be judged just for retrieval from database. Huh. It has to improve that context for the database. Huh. Missing context should not be there. So it's like if you ask a question, now you're looking for doing it for extracting the context from the context from the database.
But they're also using that five questions for generating the final answer as well. No, no, the final answer is your main question only. OK, those five questions is only for rewriter. So why do they use rewriter in first place? So if you see here, rewriter is mainly used to improve your retrieval process.
It's not used to improve your final response.

Because see rewriter in the sense I ask a question. Hey, is this question relevant to your data or not? What are the metadata associated with question that becomes your rewriter? Then the I'm not sure right? The score will also change.
So there are multiple. OK, so I have this whole thing for policies. OK And based on that if I ask what is the importance of the policy? I haven't given which policy contest, correct? I have many policies in the database.
OK, I'll ask the rewriter to draft 5 questions out of it. It will try to draft different different context policies questions out of it. So now with those five questions it will get the documents with a score for that particular question.
Correct. It's a different context. So it might have a higher score, but it's not relevant to my question. That five question Ivan. I'm not sure how it will work. Yeah, because I've not implemented it by myself. I just saw you have. I understand, but I I I'm also not getting it like.
The ranking will also change. Then you cannot allow on ranking anymore, right? That 5% policy, how it works, I'm definitely not sure, but I read that somewhere yesterday only. But this part I'm pretty much sure what is there in this paper. You're just optimizing the prompt that uses the metadata. In that metadata you will have.
Page source and I think page I mean document summary. This tool could be somewhere like it might have worked with test data. So for that particular data it might be.
Which one? We'll try. We'll try again. These are failure points when you know the answer. So you can tell you OK with this five, five new question theory works as a so for that particular.
For knowledge base, maybe. OK, so this paper will solve most of the issues that we had. And yeah, that's it. Let's proceed with the today's session. And I'm not sure how many of you know about agents.
I hope everyone knows what agent is.
Yeah, a little bit. We have looked at React. If I'm not, yeah, we looked at React, right? So when we used Tably, I used React, but I never sold you how it works. So here I have one simple example.
So whenever it comes to agent, in technical terms you have three things. One is the environment from where you gather the information. Then you have perception. Perception is like once you get that context, how you do you generate the response and for every environment you need to take an action. So let me simplify this.
So you have a prompt saying look at the sky. Do you think it will be hot tomorrow if

you give umbrella? I mean if so, give umbrella to me to hide from sun. So what is the prompt? If there is any rain tomorrow, you should tell me whether I should carry the umbrella or not. That's it. Simple prompt.
I know.
It is about it is about how hot it is. So then I can get the.
Because people usually do that in Japan, but usually you want to carry your umbrella. That's the prompt. Should I get my umbrella? Should you carry the umbrella or not based on the given weather, right? So now once you have this environment, you will have different tools that we can use.
So in tools, let's say you have search tool, you have calendar, you have Yao Finance, you have YouTube. So which tool is relevant to that search, right? So based on that, what are you trying to do? You will take an action.
So once you take the action you have your brain component. So brain component is where you have your LLM saved. For example large language model will have two things. Either it can use knowledge base, knowledge base is nothing but your rag. If not you can use memory. So memory again there are two parts, one is your short term memory and one more is long term memory.
So there is only one article, one need. Only one people need to read is this thing, which is LLM powered autonomous agent. Here you have every single detail on how the entire agent works, yeah.
So as I said, you have an agent component. When we implement this now for this agent you will have multiple tools, for example calendar, calculator, code interpreter, search and based on the given query which tool is relevant, you'll pick the action and then you have planning.
So for planning we have different prompt techniques. I'll show the prompt techniques index night and then you have memory. So memory is something that we will implement on 8th right? You have short term memory, you have long term memory. Short term memory is nothing but for a given section.
So if you ask any question, it will be within that section. Long time ago is like, OK, that's my mom.
Hey, you want that silent? I don't know.
Yeah, so you have reflection, you have self critique. Chain of thought is something that we have already seen. Then you have sub goal decomposition which is mainly used for react. So let's suppose all of you have used propeller, right? So you ask a question. After question you have planning.

So planning you have five steps that you're doing. So agent mode in the sense before you can generate the response, I will plan first, then I will execute every single step 100 times. So in agent there are two things, first planning and 2nd execution. For planning you use chain of thoughts and for execution you use. There are two prompt techniques, one is React and one more is Lads, but not everyone uses React. But let me start with sorry.

It it is kind of chain of thoughts. No, it's a bit different and acting, but there is no feedback. So when you look at chain of thoughts, OK, you ask a question. So let's start with React. So what happens in React is let's suppose you ask a question here. I need to search Apple remote and find the program it was originally designed to interact with. So now this is the question. Now for the given question, are there any tools you need to use or not? So we'll use action. So thought action.

Here you're using search, which is a tool, and then you have observation. We need to. Oh, let me start one of the LLM. Huh. OK, so it's like you ask a prompt. Let's take this example only the the prompt. So let's ignore this one. OK, we'll first talk about the umbrella one, then we'll look at this one. OK, OK.

So what was the prompt? Look at the weather and tell me whether you should carry the umbrella or not. So this is the prompt. Now what will happen? You're asking for the React. React will first think, hey, what should I search for or what is the query? So here it will use a reasoning and based on this reasoning now it has.

10 tools. Out of 10 tools, which is the relevant one?

That's the only the search, right? Weather is nothing but search. You're picking that in the action step. So every time remember when it comes to brain component, you will have multiple tools, but only one will be picked as an action that will be given to the LLM.

Similar to what we saw here. So if you see here it's a dotted line. So now once you have that.

You will have observation. Observation is where LLM will generate a response. Now it doesn't stop there, it will reiterate. Hey, is this a good response that I have to give to the user or not? Then it will again think if it needs an action, it will take an action and then you have observation. So this will repeat until it finds the final response or not. Let's suppose after five attempts it was not able to give the response to the user. It will force the output so you can use maximum iteration. If not, you can define a key. If this is relevant or reliable response, you can stop.

So if you see a start input query thought generation where it uses reasoning, then it

selects action observation, check goal. So here when you do check goal, when you define an agent you will define goals you will define.
Role play. You will define instructions. Those will be the input for this block. Right now you have the inputs, you have the observation. You will say hey, are the goal met to the question or not? If it meets then you can you have the final response. If it doesn't meet you have maximum iteration.
So you will iterate iteration one, iteration 2, iteration 3. Once it reaches Max, if you don't have the output, it will force right? That's it. This is how React works. So I'll give you one more best example. So let's suppose you have LLM. If you ask good morning now to the LLM, your final response will be good morning.
But now if you ask the same question to good morning to an agent, if you see the response good evening, it seems it's still night time. But if you are greeting me with good morning, maybe you are preparing for a fresh start. So time is 7:30. So now what is happening? You ask a question thought hey user is asking me for a good morning.
Then what will be the tool now? Calendar. Calendar. So because you saw yesterday I got you respond also calendar. OK, but typically yesterday I had very simple example. No, no, no.
I also responded yesterday. So it's like this was the same question. What's the forecast for tomorrow? Should you take the umbrella or not? Then you had tools, you pick an action, then decision making which has the prompt. Then you have this example. Now what was the tool used calendar.
So it's like thought. Now thought is like what should I wish? Now what is the action? Calendar. In calendar you'll have a parameter. That parameter is like what should I search? Then get the response observation. Now if that observation has good evening, that means it will stop.
If it doesn't have, it will again iterate thought, then action again is your calendar, then observation. It will keep on improving. So every time it improves in the feedback, React improvises as well. So React is a framework.
Which is being used across all the agentic applications. I think that's a pretty fine responded to me. Good morning.
Uh, did you add this here? Like yes I said good morning. So what you can do is use settings. In settings you'll tell you have the calendar.
As a tool how how do I get it to connect use react?
Prompt reasoning.

So again here you don't have natively React, so I'm just telling use React before generating the thinking or response. But again it might not be it could be because of the custom prompt.
So basically this is just a tool. It's people say this is an agent, but how? Where are we connecting in settings? You say something about settings. So not in settings. What I usually tell is hey you have let's suppose you have the calendar as a tool inbuilt.
So personalization. Huh. So in personalization, look at current time and be smart to answer. I think a lot of us are in the good morning. You have calendar as a tool. Use react prompting to generate the response for the.
Given user query.
Be smart to answer.
So this is the instruction US. It could be PD or something. USISTIST.
Good morning. I am.
I don't know Apple.
It's happened. So one more thing, whenever you take too much of time in generating the response, agent takes time. That is the drawback of an agent.
Yes. So this is what you call a thought. So we'll show the output as well. Basically it should argue good morning, good morning. I told him like this morning and then he said good morning. This is what we have on the calendar today.
I'll just try it every morning. So it's very nice that won't work. I've got my calendar. It will. Yeah, but it is at 11:00 AM ISP. You have this meeting 1:30 PM ISP. You have this meeting.
When is starting with good morning? What time is it now?
For me, it didn't give me the response yet. Good morning. OK, it gave wrong. We'll try this in agent.
It's a very, very smooth. It's telling you the right thing. It is giving you the wrong greeting. Why did you told me good morning, Asad? No, no, no. You are absolutely right. That was my other point. Let me give you.
I'm probably sure. It is 11 from PMI ESP in Ahmedabad. I don't make a.
We'll wait for this and we'll come back. But you understood the problem statement, right? Yeah, React, how React works. So there are three things that we have to define which I'll talk in agent profiling.
And once we have react, So what is the drawback that you see here? So most of the time, do you think you'll only think in One Direction? So think of chain of thoughts.
After chain of thoughts, we had one more prompt technique. Do you remember

that?
So zero shot, few shot chain of thoughts. What is chain of thoughts? You ask a question step by step. After that there was a limitation for chain of thoughts for which I told one technique and I told don't use it in any application. Do you recall the prompt technique?
There was one verbose, but not track. Don't verbose is for a few shot.
Oh, you.
None dump and negative. None negative.
Oh.
Do we play from screen? Huh. Do we play from? No, no, no. It's three of thoughts. Three of thoughts. Huh. So what happens in three of thoughts that is used for strategy making. So you ask a question, you break down into step by step.
Now each step is a node. That node will have multiple sub steps.
So now for this also what you have is. OK, this is your good morning.
Although it's currently 7:30 here in India, so a very good morning to you. How can I help you today? OK, not bad. Jamini, give the response. Yeah, it's a good morning. But now that we are saying good evening, Hardik, what's your current headspace? OK, perplexity, you have the right response.
And but this is Comet. Uh, so yeah, how many of you use Comet?
This is crazy web browser. So let's suppose you're using e-mail, right? And you didn't respond to any many last three days. So we can summarize that using perplexity. So let's suppose I use a new chat and I'll tell.
Which mail did I miss to reply in last? Is it seven days? Is it comment? But which e-mail would it connect to your e-mail? So when you download the comment you will give certain permission.
That's a.
So it will now look at your Gmails.
And then it will reply OK.
But this is very solid web browser comment.
Yeah, this is awesome.
OK, so coming back to the topic, you have React just like React as a drawbacks that you can only think in One Direction. For that you have something called as Lats which stands for Language Agent Research. So now what is happening in language agent research is.
You will think in two different direction. If you see here you have action after

thought, action, observation. If in case there are any chances of multiple thoughts, it will create two different nodes. So if you see here you have thought, action, observation, you have thought, action, observation. Now how do you evaluate this? After every single node you will have a value evaluation which will select the best action. So this is very complicated. How this typically works is you use something called as a Monte Carlo tree search. Are you? Have you heard of this before?
I guess Monte Calot research. So it's like it's a search technique. In search technique in the sense if you ask a question you get into the depth of it. I'm not sure what this algorithm is called in DSA depth by first.
The first the first search, the first search. So for creating the shortest path. So for shortest path you will have hierarchical nodes, so there will be different nodes. So for that you have Monte Carlo research technique. Hmm. So many people use Monte Carlo research within the.
What do you call search technique? So this uses Monte Carlo research, language agent research and I have mentioned this word before as well. HAID. Do you remember when we were discussing augment code HAID and I told you AST parsing. So what happens in AST?
You're using copilot. Now you have 10 files in your code and if you want to auto generate in one particular file, let's suppose app.py, you want to add a new feature. Now app.py has dependency in ingest.py and constant.py. So you want to get the constant nodes.
So some of the frameworks uses Monte Carlo research with AST, which is your tree parsing algorithm, which is mainly used for copilot building. So here instead of using AST, you're using React plus Monte Carlo research.
And uh, this is very tricky.
This is very tricky, but never use lats in any of the code because I'll tell you what will happen. You're using it for deep research. You ask a question. Lats will run for 10 attempts, iteration one, iteration two, iteration 3, iteration 10.
When you force the output, when you don't get the response, it will tell I don't know the response. So it's like you ran 20 LLM calls and at the end you see a response called. I don't know. So it is very frustrating and too expensive. So for just three or four LLM calls that I used, we spent around 3:00 to $4.
Only for seeing not. I mean, I don't know. So there was some video that I created long back.
Here lads advanced agent worth it. So it's like we spent $5. We tried out three to four

prompts and at the end of the response you get. I don't know. So this is very rare where you get the response correctly. So this is where most of the frameworks that you see right now uses react by default.
So Landgraf we used React and tomorrow when we look at Crew AI, Crew AI uses React which is by default. In Landgraf we used it as an import statement. So for Landgraf and Lama index we will use it use it as an import statement.
But for Cruia it is inbuilt, so the framework that we use is open EGI that also uses React by default.
And not just open EGI or what do you call true UI. I might have also shown you ugging face small LLM which had 300 lines of prompt. Huh. So if I show open EGI and if I click on prompts.
So as I said, you have task creation, you have task execution. For task creation you will use chain of thoughts. For task completion you will use task completion on execution you will use react. Why? Because for execution you need feedback. React completed.
So not just completed, is it right or wrong? Also so critiquing so that critique can be done in Lats and React, but Lats it's not used, that's just a research concept. So if I click on worker task execution.
You will see thought, action, observation and this is how you have to generate. First you have question. Then if you see what is the capital of France, I should look up France in Dug DB Dug Dug Go, which is your tool to find relevant information about its capital city.
Then thought then you have action. So for action I'm using the tool calling. So this is a example. So if you want to use tool call for open EJ, this is the syntax that you have to use and this syntax is now available in my code.
So if you see a params, you have query, you have Max results, you have can summarize. These parameters are used in dug dug go tool. So if I now open tool. Actions. You have tools, you have Dug Dug Go.
So you will have those parameters.
You're writing from based on those parameters. OK, so that is just example. But now why did I ask it? When you're using agent, you're using libraries, right? But there might be chances you might want to build your own agent, right? So during that time what you have to do, I've repeated this.
Every single functions that you write, it should have doc string. Hmm. And now what is this syntax?

So you can extract all the class methods. So when you use agent you will extract the class methods which will have description. So now this will be used to decide what query can I ask? How much maximum results can I use? So now agent can use just five, agent can use just six. I don't know but use it.
Maximum is 10 if in case it doesn't use any variables. So all these variables will be decided when you do function calling.
But when if you're using libraries like Langraph or Lama Index or PUI, you don't have to worry about all this.
And one more thing I wanted to show the small agent, but I guess small agent I already showed but I'll just show it again. This is code agent and if you see it's the same thing again you have thought.
Thought Where is action?
Mhm.
No, three days before only the release for everyone. It's free for everyone now. So this is for small agent. So you have thought action observation again and for React if you open line chain you will get the format again prompts. So what you can do is. Take inspiration from those prompts, then modify accordingly and all these prompts will be minimum 70 lines of what do you call 70 lines.
But these folks have written 300 lines. I'm not sure how they're using it.
Single prompts also it's like 200 lines, 300 lines.
Alright, so you have lats, which is again a prompt technique.
Now let's just talk about core components of multi agents because not every time you'll be using single agent as well. So multi agent only has to be used when it is when your use case is not time constraint. So let's take some examples.
Let's suppose you have your requirements and for the given requirement you have to generate A blog article. So now what will you generate in this blog article? First research for SEO writing right? So you will go search for SEO now second write. So when you do first agent it will gather all the keywords that has to.
Be in your article. So first I said we'll do that and what tool will it use?
Search tool. Now you have writing tool. So now what this writer will do? It will take all this researched concept. It will pass to the writer. Now this writer will have instruction. This instruction is your writing style, so you don't want your article to be as if a.
Yeah, yeah, yeah, it's written. So like what's like let's dive deep into. Then you'll have some keywords, right? Which is very then. So those words has should not be there.

You need to keep it in the human style.
So I told you copy your two prompts, then tell don't verbose. Those are your instructions now, right? So you have researcher, you have writer. Now writer will write your entire article. So do you need a tool for this? Yeah, what tool? We will need element tool.
No, LLM is a model. So we have model, we have memory, we have tool. So what tool do you think is relevant for writing? For writing purpose, just writing, generating the response. LLM is a tool. No model is there.
No, no. I'm saying like if you use another LLM to create content based on the instruction that you have given. But that doesn't make sense because even if you just do instruction you have model which will generate the response Google Docs or something. No Google Docs means.
Are you writing your? Let's suppose you use Google Docs. Whatever output you create, if you want to save that in Google Docs, then you can use that as a tool. OK, so here you don't even need a tool. Yeah, so so it's like you research the entire thing from a researcher which uses search.
Now you come to writer. Writer's obvious, you have template, you have this input, only generate the article which LLM can do. So when you define an agent component, you already have the LLM as an LLM as a model.
Tool is not required. OK, but now let's suppose your output format. If you define to be a markdown, you want to save it in a file. Now file system is your tool, right? That depends on where you want your output to be saved. That becomes your tool.
Now let's suppose you have one more agent which will verify your entire article. So that is our third agent. So if anything needs to be removed, if anything is feedback, you will have certain rubrics. Hey, you generated this article, does it add there to this rubrics or not? If it doesn't add, update it.
What will be under like not say it should be tiny checks. It can be anything. So after all these three details we should have your final response. So this is now a agent multi agent architecture. So every agent what it will do is it will have certain profiles. Right. So let's imagine you're building a game, right? For every game, you have certain characters. Now each character has its own role play. This a character has its back story. This the character has its description. Like what are the strengths? What are the weaknesses?
And how do you create this? Either it can be predefined which a domain expert will do. If not, you can use model to generate this role play or anything. This is given by

user. So let's define if you're defining multi agent systems.
You are defining hey this is my agent one and this is the instruction for them. So it's like a predefined you are giving it. System designer is nothing but a domain expert and then you have model generated. So now what is model generated? There are people who build.
Your entire multi agent autonomously. So your goal is to just give prompt. Based on this prompt what team it needs to be created will decide a LLM. So it's like you have task, you have LLM. This LLM will decide what are the multi agents I need.
To approach the solution which is very expensive. So typically you need your multi agent to be just two or three, but LLM can generate five teams, so two teams can do same task. So most of the time it's better to go with predefined.
But there are frameworks which provides model to generate the subteam. So it's like LLM itself will decide what should be the subteam. Most of your copilot in copilot you use model driven but you don't specifically see it.
Because it uses code interpreter as a tool, then you have data-driven. data-driven is very similar to model driven.
Yeah, because he didn't like.
OK, what you're telling is very powerful, like you're doing it with the personas, right? Huh.
We're doing that in for so it's like so you might have seen multiple people doing their work as a different persona and correct then your content or your output is streamlining through all of their.
Contributions and and then you're getting the final output correct. So that is multi multi agent systems. So best example is blog writing. In blog writing as I said there are three different personas now. So one is researching it, one is writing it, one is verifying it and one more is.
Finalizing your output based on all the instructions that you've got. So you have 4 multiagent systems now. Now there was one more diagram which I created earlier. I have a question though, would the pollute the content that we have data pollution, data pollution for?
We're going to any context, right? How do we even realize what content is written by me? No, usually we'll have human in a loop entire. I mean after every single agent. If you want to verify, we can have human in a loop. It will always be needed.
Because this, uh, this will shift down tokens and then.
I'm just about it. OK, where is the diagram?

Do you know Devini? How many of you know Devini? Devini. What does Devini do? That was for software engineering, right? So I had one diagram. I don't know where I've Devini. Did it fail though? It totally failed.
OK, this is so there is similar to Devini, but it's called Sat Dev. So now what happens in Sat Dev is this is your input. So let's suppose you have PRD document, product document generation, product document with all your requirements.
You will give it to a designing team. Now what will happen is once you upload your PRD, if there is any design changes that has to be done, let's suppose some of the PRD didn't make sense. You have to update your design, chat will automatically do it. So you have a agent one which will design it.
Then you have agent 2 which will write the code for your given PRD and the design and then you have a testing team. Now what testing team will do is it will it will write test cases for your given code which every copilot does right now and then you have a documenting team. Now what documenting team will do is once you have your tests.
And code generated, it will generate the document. So all this thing is happening within a single prompt. Sorry, a single prompt, but with different LLMS. So now this is 1 agent, this is 1 agent, this is 1 agent, this is 1 agent. So this is persona and every persona has their own role play.
No, it's ****. But the concept is good. The concept is good. This came in 2024. OK, Devinya also came in 2024, but the only good solution is Claude and Claude. Also you have a does not know the compromise Claude EIA agents.
Any compromises like best solution is to latch. Have you guys downloaded cloud code? Yeah, Cloud CLI? Yeah, so if you see it, OK, let me show this.
You have personal act as no, not act as. You can say agent agent, correct. So after you enter cloud, if you give back slash agent, you now have different personas. So if you see, yeah, engineer, you have API document, you have architect. So now what are this?
Let's suppose if you want to generate any code and you want to have a API reviewer and then a AI engineer, so use two agents. So what are this agent or multi agent systems who will coordinate among themselves. So what we are trying to get is the fundamentals of it.
Every single agent will have the role play and back story. So is this clear? Role play? It's very it's just prompt. Agent profiling is only prompt. Then you have agent environment. This is tools. So now what are tools?

Let's take an example of copilot. You have sandbox, so who is editing the code and executing the output?
So you might have noticed in Claude you ask a prompt.
Task one, task two, task three, task four. Then you have task five. In task five it will test right? So when you test it will automatically write commands. That commands is executed in your terminal, right? Right. So now that is a tool which is called interpreter or compiler and that is called sandbox.
Then you have physical physical in the sense, let's suppose you have robots. So for robots like what is one tool you have to move in one specific direction, movement, then observing the ultrasonic sensors.
So you ultrasonic sensor in the sense once you come near to any object and you will have a threshold which will take a moment. So whether you should go right or whether you should go left. So those decision is nothing but your tools which are physical. I'm not a domain expert but in physical.
Those are some instruction. Then you have none. Just now I said for researcher you have a tool, but for writer did we had any tool? No, no, right? That is called none. So environment in inference is nothing but what tool do you need to perceive the information?
For umbrella we used weather API.
That you will defend it. So you'll have 10 tools, huh? There are no, but there are MCP servers that are connected, huh? Correct. So inference is nothing but existing ones, whether you need a sandbox like code interpreter or not.
Or MCP also it's like virtual itself. You have APIs for it. So first profiling is nothing but defined role play prompts. Second you have environment which will define a what do you call it, different tools. We are adding tool details, yeah, so these are just naming conventions.
In reality, but if I want to tell this in simple terms, this is Shishtram prompt. This is tool calling plus action and then here you have communication. So communication I told you different routes. You have routing, you have parallelism, you have orchestration. So let me simplify this part.
Part by talking about some of the use case.
OK, so this slide was supposed to be here. Then I was supposed to give this some part. I don't know why it came up. So this was the overall diagram. One had how they created Minecraft. I was talking about game, right?
So for Minecraft one of them created an agent where you had profiling which is your

system prompt. If you see this is your prompts. Then you have environment which will have sandbox or physical. Then you have communication. So for communication. Here I have some use cases.
First one is routing. So for routing what will happen? You have a LLM router. This will decide which LLM to choose, right? The customer service one, right? This is communication. Now for every single LLM call you might have certain content. So now what is this content? Knowledge base?
So knowledge base in the sense context rag. So this is rag one, rag two, rag three. So you have HR documents, you have product documents and you have any documents based on the given team. So here LLM call is nothing but your agent one, agent 2, agent 3.
And now we have parallelism. This also I mentioned if in case you have same data and you want to run it in a three different calls. Hmm for context relevance, answer relevance and groundedness.
And then you have orchestration workers. The example is the blocking. So if you see you have an orchestrator where you have a prompt, then here you have researcher and after researcher you have writer and then you have verifier, verifier and then you have final synthesizer. It'll give you the output.
Here the arrow marks will go downwards, but you can also have straight and what we will do today is we will build this part, orchestrator part and also you have a routing. Is this clear? Yes.
So we'll do the coding challenge similar to what it's not. I'm just saying I'm just thinking out loud. If you're writing search content now, you know Google is smart enough, but it's de ranking all the content that is A generated.
No, it won't, right? Here you're giving certain your own writing style. It won't do that, huh?
Let's do this. So let's do the coding challenge. What we will try to do is I will show you one example of routing engine. OK, So what routing engine will do is you ask a prompt, it will decide should I go to the customer segment or should I go to the marketing team OK and then you generate the output. OK, what you guys need to do is you need to write the.
Blog blogging agent which will have a researcher, it will have a a writer and then it will have a verifier with Langraf. We'll start with and no tomorrow I have Langraf and Lama index OK.
Because Agno is very simple to start with. OK, huh.

So before that we have some quick notebooks. How much time do we have? OK, 10 minutes. We'll play with Agno. What are the different tools that we can use so that when you're building the teams? So first notebook is you have.
Different ways to do function calling.
So for function calling, let me open Collab.
OK, so we have existing tools. For existing tools you have tably.
And you have a let's let's do while we are doing this already to search for content on the topic of the content. Let's make it automated as well. What automated the topic that we want to write the content on. You're talking about Rob.
Yes. So how how do you know later? OK, which are the trending topic? Can you pick one? Yes. And we can use it from new node by combinator. Huh. OK, so the top five we'll see and which are not a question. So we can just pick top one or top. Huh. Randomly from top five. So if you.
OK, so let's open Agno document as well. Oh, OK, I hope Ramakrishna is following. So we are just starting with the what you call. He did mention to me that he has to leave at 8:30. So OK and then go to the recording.

RamKrishna Bhatt 1:06:39
Yeah, I am here.
Yeah, till till 8:30 I'm here.

Tarun Jain 1:06:49
OK.
OK, so we'll do the quick start one. Once we do the quick start, I'll also send you the main template code. So main template we will try today, but this is something that you can try of this thing. But before that, let's try to complete this tool calling part. So this is very important what we'll share today. Existing tools in the sense you'll use Stabli, Dug Dug Go or any tools which is available in all the frameworks. But when you build agent by yourself, you will mostly have OK, this is not tomorrow, this is 9th. OK, so you have custom EPI. You're here to 88. Huh. So 8 seconds we will have this or we will no, we'll we'll do it in the morning just for one hour. OK. Because MCPI want to do in person. Perfect 81 morning that.
So I wrote that ever.
I'll just cancel that out and OK, so it's like 8th we have agents and multi agent, then tomorrow multi agent using Langraff and Lamba index then memory and MCP. I was

deciding on 8th OK so we'll start with project. OK we'll start with project on 8th again drag we had a project.
If anyone wants to continue your same project with Agentikrag you can continue it. So here I'll tell how to build Agentikrag once I complete memory and MCP. So this will continue on 9th which will be virtual and you'll have time till 14th. So 14th we'll do review and here I'll talk about LoRa.
So this is optional session, but this will be very useful. Yes for playing, doing fun things with AI. This has nothing to do with what you've learned so far. Sorry I didn't. This assignment we have to complete by no, no. So if anyone completes by 15 then fine. If not, you usually will have time. It's the time slot.
No, not today.
No, the coding challenge is different. Coding challenge is like what we did last time, playing with prompts. OK, so this is not a this is actual project. Mm-hmm. But if anyone wants to improve as a blog writing agent, you can use the same project here. OK, so custom API in the sense, let's suppose you guys have your own APIs and you want to build agent for your within internal servers. Let's suppose you had dashboard. What was the sales you made in this specific month? You had one API for it, so you'll use custom API as a URLs.
That URLs will be given to a tool. Now that tool will generate the automatic response like what was the sales so and so. Then you have your own function as well. For own function we wrote a Google this thing last time. If in case any tool is not available you will write custom tool for it.
But there are people who has already implemented this via MCP, so either you can use your own function or you can use MCP. So we'll start off with this three. I do have some code for this. Once you do this, we'll go with teams.
So teams is where you have your.
Multiagent systems. So team is nothing but multiagent. Every agent is a team, but this terminology will change if you use different framework. Now crew AI are you sharing this one? Yeah, I'll share it.
So like did we cover the ignore last time? No, no. So only I OK, so this is the quick start. Once we do the quick start, we'll start again and then we're going to use ignore in VS code now. I do have VS code for UI part.
Because I can't run fast API on a colab.
So we'll do the UI part as well today because UI part port is here in VS code.
I'll remove the unwanted.

Yeah, something with that also. OK, this is the URL Ramakrishna.
Ramakrishna.
OK, so let's open this article. So let's suppose if you want to define an agent, how many components do we need here?
Tools. Tools. Do you need action? No, because action is within the LLM call, which is our agent. So tool is 1, agent is 1. Planning is nothing but your model because these are prompt techniques and for prompt techniques you need LLM.
Then you have memory. So let's start with tools and.
Yeah, a model today, then we'll use memory.
Memory short term and long term both. I'll do it in one shot.
So this is the notebook. You have Agno, Agno is your framework and then you have DDGS which is your duck, duck, go search and then you have Tableau Python. So let's just install this.
And so far we have never experimented with multimodal, so I'll also show multimodal agent.
So what are the until this is installing? What are the use case of multimodal agent? Can we just name it two? We'll give Ramakrishna a chance. So what you're supposed to tell is input is image. Where do you think your input is image and which is very useful?
For example, so far what we are doing, our input is text, but now our input is image and where is that useful? Where it is useful, where text should not be used and image itself is your main input.
Representation for like a sense representation of the graph and kind of thing. No for input, for output, for input. OK, like extracting text from the image. OCR, OCR, not just OCR. OCR is just converting your image into text. What is there in the text?
So I'll tell one example then you can give for some example. You go to mall. When you go to mall you have this products, you have the ingredients. So you take the image of the ingredients and you ask agent saying that so multiple medium if required. So now the use case is I take the image of that ingredients, the product and tell.
The nutrition value of this agent, I mean, is this product good for the nutrition or not? Yeah, so now it will analyze the product. It will think what is this? Let's suppose shampoo. So for shampoo you have 5 ingredients. Is this ingredient useful for the nutrition or not? This is 1 use case multi model.
All your text will not play any role. So just like that, can anyone give 3 use case? We'll

start with Ramakrishna.
Just three. Let's be quick.

RamKrishna Bhatt 1:14:08
I don't remember any as of now.

Tarun Jain 1:14:12
OK, what? What? What? You said I don't recall as of now. So you are identify an object. What is that? You are. You have an image and you want to get some details out of that image. If someone is breaking the the current traffic signal, maybe. There, yeah, ma'am, yeah, they like, but it's a VLN. OK, so you have crime. Let's suppose there is a video you have, there is forensic or probably you might have seen there are multiple forensic project that goes on. We have a customer as well in forensic.
So they have videos. In videos you want to detect whether the crime is happening or not and which time step the crime happened and what happened in that scene. So now you pass 10 videos to a multimodal and every multimodal you pass it time step wise.
Timestamp in the sense image one, image 2, image 3, image 4 in image one. If there is any frame you should have to note that. Then you create a report and you give it to the forensic lab. Now that is a multimodal use case. Two more now that is the infrastructure like the scanning the old room and design the.
OK, for design also like let's suppose you take a image of this. Yeah, is the color matching with the ceiling or not? So taking the inspiration, is the color matching well or not? Translation when you are going to complete. Yeah, translation. Yeah, so.
So this I'll consider as a traveling for traveling. Usually if you travel anywhere you will see boats which you don't realize and it can also be for food. So you go at new place, you don't recognize what this dish is. You take photo, you ask recipe of it best.
So if you just ask what is the name of the phone? Does it make sense? Can you list down? Do I do I something I already said?

RamKrishna Bhatt 1:15:56
OK.
And.
I think in industries that will be useful.

Tarun Jain 1:16:04
Yes.
For example.


RamKrishna Bhatt 1:16:09
For example, if we have some manufacturing unit then for for that it will be useful to analyze I think.
For for a.

Tarun Jain 1:16:21
I can't, I can't mind.


RamKrishna Bhatt 1:16:24
Yeah.

Tarun Jain 1:16:25
For some of the industry 4.0 and 5.0 use cases, you use multimodal use case. That is fine. But yeah, we got more than three use case, so let's proceed. So we'll start with multimodal. You have agent component, then you have Gemini. Here we don't need any tool.
Why? Because our input is image and you just want to use the agent component. So if you see you're starting with from agno dot agent import agent then from agno dot models dot Google import gemini. Now here you have a media so media will be passed.
When you inference right so from agno dot media import image. What do you image here? It can be your uploaded file or it can be the URL of that image.
Is it just data? Hi, just the data. OK, so now let's suppose you said right, you'll take the image of.
Any board and then you'll pass it to the model. No, you're doing that over here. Fine, you can just upload your emails, that's it. So let's take both the example. Let me take the example of shampoo, but I have a product of this multi model.
I'll show real time example because we're using.
No, every single LLM is now a multimodal. You can upload the image in ChatGPT. You can upload the image in Claude. So if the model supports multimodal, you can

use that with them. But now every model supports multimodal. OK, every single model OK.
So if you use open source model, supports multimodal. OK, Llama for supports multimodal. OK, deepseek supports multimodal. And then you have. I think we needed to have this VL part vision language, but all models supports multimodal.
So here I have an example. You have product ingredient analyzer. So now I'm clicking on chocolate bar. Now if I click on analyze example it will check that ingredients of that image and it will list down based on the nutrition whether it is.
Usable or not? So if you see ingredients analyze then you will have nutrition information and then you have potential concerns. So now this is money application and what if I take my own image?
What is the nutrition value of this?
So what should be the output? What should be the output?
Made of what? What is? Something like that? No. If this here I'm designed to be helpful, but I'm limited what I can do that request fall outside those limitations. So it's like it is only made for product ingredients, so.
OK, do we have any product? So can anyone guess what was used here? The prompt? Correct. System prompt. System prompt. This is system prompt. Yeah. So if you do not find any ingredient list, just respond that I cannot tell here. Yeah, something like that. Correct. So now if you see.
You have an example of existing image. You can also upload image. You can also take image live. You don't take image of like food item, correct? So no like what we ate to dinner. But there is no ingredients, right? There is no ingredients. So they want me to figure out the ingredients correct.
So the goal here is you don't know the risk, you know the product, but there is no ingredients. But the application that I built here was for ingredient analyzer. OK, if we wait some time, we can try this again.

RamKrishna Bhatt 1:20:12
OK.

TJ **Tarun Jain** 1:20:13
Those events and find out those all the.

RamKrishna Bhatt 1:20:14

So I think it will be analyze the ingredient and based on that it will be give result right not based on the food image.

Tarun Jain 1:20:18
Correct. So for example, if you see it, you have energy drink. Here you have ingredients. You'll click on LLS. It won't answer, but the goal here is to understand the nutrition value behind it.
Give me one copy. Give me one copy. Give me that copy. Move, move. OK, no, there is no invariant list. I can apply this. I can write. I want to write. I just write. OK, that is very tricky. That is tricky. OK.

RamKrishna Bhatt 1:20:45
I think, I think if we if, if, if we upload, if we upload any food item then does that.

Tarun Jain 1:20:51
So if I write 20%, 50%, OK, yeah, yeah, one second. Yeah, Ravakrishnan.

RamKrishna Bhatt 1:20:58
Hello.
OK.
So I was saying that if we upload some real food item, so that does that give a result or they need some information of the ingredient. OK, so.
Package. OK, got it.

Tarun Jain 1:21:17
I think it's not OCR. It's understanding the image. Huh. Anyway, we can take a picture and see like it's all, would it get or not? And I can say let's see it is not OCR, yes.
No, no, with yourself, with both, both, both you with. No. Well, let's try with this because this is tricky. If I am in the image, I definitely it will. OK, let's try it. OK, no, it will take pizza because pizza is obviously there. OK, then let me give another image. OK, let's try this.
I don't think it should answer. If it does, if it does, that's it. It is there. OK, wait, wait, let's down the nutrition.
Based on the image and then it will be based on the product and then following which content high protein content. I just give list as a text.

You want the extremely high sugar content, but we can improve the prom for this. This was very good excuse.
Can you click on take photo?
Once again, it's the same thing.
It didn't give what it gave what the food, what the food is. It's still basically image passing OCR, right? Oh, OCR. Is it doing OCR? No, it's not doing OCR, but we are using the LLM. But the prompt it was to identify the nutrition and generate the response. No, but the LLM does OCR.
Yeah, it will do. Oh, it will do. It is. It is just based on the. So let's suppose it is based on the invoice extraction, invoice extraction you're using multi model. I mean we can do but this this is like a multi model only model is doing this. No no model is model is doing it. Yeah.
Yeah, so like like it's not OCR, it's just like a OCR is just a feature which a multimodal can do. But the typical thing is it was supposed to recognize product as well along with ingredients. OK, which was missed in the prompt.
So if I show the prompt.
Get up.
I have product ingredients. You have constants dot PY here. Let's just update one thing.
Read ingredients list on the product image and also make sure.
The product.
I want to make sure.
It is a product. It is a valid food product. It is a valid for product. I'm not just and not just.
List of some on the.
Yeah, just too many, but even this will help you out right to write write your instruction and system message. These are input variable by the way.
So stream it is very fast. So if you see it now it showed updated and oh so you just have to refresh.
Now you for that book.
Hopefully it should well.
Wait, one without 40 minutes.
Or maybe it's OK book is not a word.
It.
I just hope the a minute doesn't mess up.

Yeah, it did. It gave you one lengthy response. I know you want to know that is.
OK, one at least it should have given one sentence that I didn't realize any image.
But yeah, OK, I believe in it. Oh OK, that's OK good. So we start with multimodal agent, define name, then define model, define markdown. OK, wait a second, which model was I using here?
No.
This is a necklace.
Yes.
Open the.
Right.
Gemina 2.0.
2.52.5 is better, no new model. So now if you see it pulling the code changes, processing dependencies, process dependency of the data.
We didn't.
We have slacks for the option.
And also I just hope it works.
It's very difficult to defend when I'm at live sessions. OK, it's taking this much time with generated response.
So hopefully.
I'm saying I'll work at this.
Nutrition analysis report. OK, once again, this report provides an analysis of nutrition while it's handwritten on the notebook. This give reason now, at least the profile is inconsistent with the most conventional food. This might be Mars greener or specific. Type of eye color reasoning is there but.
But this is not. This is not good response though. No, we'll we'll do a code instead. So you have Gemini and now multimodal agent. What you can do is we'll try with two things, one with URL and one more with.

RamKrishna Bhatt 1:28:52
One question that would be.

TJ

Tarun Jain 1:29:01
File path. So here you can define any prompt, but let's suppose you have any instruction that you want to define. Define certain instructions.
And if in case you want to hard code anything, you can have anything like analyze

the image and if there is any user prompt you can add it as a prompt.
Which is F.
And here you can have a prompt. So this is what user will enter on a UI and this is hard coded and this is additional instruction.
So you have instruction, you have description, you have role play and if you want to add additional context you have add external context. So if I just click on this.
There should be add additional context.
So if you see you have additional input.
And you also have expected output build context. So this is what we have. We need to pass system message, description and instruction and then you have additional context. These are the four variables.
But mostly will not use additional context. This is very rare. System message will use.
System message is your system prompt instruction. You can define anything.
Description is like what this app is about. That's it.
Getting so many things like this worker app or simply right usage.
But quality also matters. Here you won't get the quality that you get in Langraph because Langraph is workflow. Here it's like you're just playing with agent and there is one Andrew in this video in that he says.
Agents are ****. If it gets the response it is very good, but if you don't get the response it is gone. So agent is still in the early stage. So if you see this intelligence.
Intelligence. Age. Who is it? Sam Earthman? Our people.
Hello.
So you have this intelligence where he talks about.
So it talks about what is the level one of the AGI. So the ultimate goal of building all these models is to achieve AGI, AGI. So you have level one, level 2, level 3, level 4, level 5. Level one is your predictive models, ML algorithms.
Level 2 was your LLMS. Level 3 is your agents. Currently we are at Level 3. Level 4 is unknown. Level 5 is your ATI. So level 4 it is said to be that you can replicate what humans do. Then you have ATI which is above humans.
And it talks about how they want to achieve it in this article.
I just called the intelligence edge which was written by this guy.
And here what you can do is you can upload any image of your choice. This is for URL if in case you want to upload any file path so.
It's just for the inch. Instead of URL you'll have file path, so I'll try with file path. You can try with URL. I'll just see if this variable works or not.

So you can try to use any anime image anything.
Paths and here instead of image, what I'll do is I'll copy the path of it.
Mhm.
So it should be like this. So if you want to use URL, use URL. If you want to use image, use this variable file path. Is this clear? As of now let's not create Tableau. Let's just use Gemini because we are not using any tools here.
OK so comment this line and then run it. In this multimodel angle you know can we also get web like can we give web URL? Ha you can hear that link HTML URL HTML web URL.
For like uh read my page something like this. For that you have to use reader.
You have that will come for knowledge bases, but you have that you have knowledge retrievers. So in I know you have knowledge for knowledge, we'll use it for agent. So what do you have to do for it?
So what you need to do is if you want to use multimodal on a URL you can use URL. If not you can use file path. So if you want to do for audio just import audio and here if you have any audio file.
Instead of image, make it audio. URL will be URL, file path will be file path.
So when it comes to multimodal, you can use video, audio or image.
And it can be anywhere. So if you want to do text to image, text to video, text to audio is 1. You can also do audio to image, audio to video and audio to. There are different combinations. That combinations is here again.
You have multimodal. Inside this you have images. So if in case let's suppose you want to generate this nano banana image, what you can do is you can give a prompt. Your output is a image now that is also possible.
Pagnode has more than 100 cookbooks.
That you can use out of it's only 20 are good enough.
So whichever cookbooks have been released by their own teammates, let's suppose the main contributors that are good, but whichever cookbooks are released by someone who only has one or two commits, ignore those cookbooks.
But if that commit, let's suppose I have three commits, don't ignore my cookbooks because that are useful.
But uh, they do have examples inside examples. Every tool has their cookbooks.
So let me know once you have the output what for what so you can upload any image. If you want to use the same image then we can continue OK.
OK, so next we have the tools. So for tools we can create Tably API key. So Tably

didn't we add Tably last time in the secrets with the added but not in the added I should.
Yeah, OK, so meanwhile you can create tably API key.
And uh, the search tools which are currently in advanced are Tably and DXA.
And last time we spoke about one tool, right? Yeah, what was the tool which which didn't work actually the powerful. So it's something you said some name the we can see you.
Yeah, here parallel parallel or something parallel.
Dogs parallel OK, I have in my search engine. This will be a custom tool now because parallel is not available in our this thing existing tools. This is now whatever I have below this code.
So open docs.agno.com.
Dog.agno.com OK and if you Scroll down you will find tools. OK, so inside that you have tools kit. So you have ERXIV, you have Dug Dug Go, Google Search, Hacker News. You can use Hacker News OK and then you have Wikipedia.
Then you have Slack. So this is what is very important when you come to socials, right? If you want to build chatbots for Discord Slack, you can use this. OK, so whatever applications you have built so far, connect to agentic, you have your tools, you can use Slack directly.
So you can have chatbots for Slack with agents and with just a few lines of code.
2.
So whatever I run below this, I'm not guaranteeing the output, but yeah, but let's make sure we don't use pro, because if we use pro, we can't use it for code talents.
Let's be quick. Yeah, the person on the left is a Mr. Manish. Yeah, and I'm. But if you took the linear from linear, it took the. But did you give instructions? So if you want to, you control your.
We have to define instructions and instruction description and system message all three.
But is it like you get the you expect the people you can figure out do you have Jaisha in that or no, there's me in that. OK, then it's very difficult. So do one thing.
Tell it to use a. OK, do the same thing here.
So copy this line with search, paste it here and I'll tell. Can you send me that image?
Sure. Look at atyantik about page.
And identify these two personality.
OK, I don't know if it will work or not, but we can try it. OK, now now you don't make

a mistake. You stream equal to run the run this again. This is stream, I mean collaboration.
So one more thing, if in case you're not able to see the output right, we are doing stream. Stream sometimes doesn't directly support on collab. Either we can run this on VS code directly. If not, I told you one more. I'll tell one more approach here.
Can you look at this code cell? I'll use a variable called response and instead of print response I'll do run, remove the stream, run this response.
Now if I just rint.
Response.
dot content will have your final output. I can have to do stream equals to false.
I did that. Oh we have to be. By default it's open API. You can open. If you don't define this, can you see this now? Yes. So instead of print response, what did I do? I did run and it should be stream to be false. Then you can do response dot content.
OK, this is on streamlet.
OK, the code like this instead of print response have run OK.
Now one more thing, if you don't define this model, your code will work. What you have to define environment variable which is open AI API key. Every framework's open AI's default LLM.
So I've defined this. You have an agent.
And here I'll tell analyze this to personality.
From Adhyantic website.
Uh, this is image. Where do we have this? I have share in check.
OK, I can use URL only, yeah.
But most is good. No, he didn't, you know.
Um.
URL.
I'm looking at any work on this. Now we need to. I'm working with.
This is tools plus.
What? OK, I told analyze this name and this personality. Yeah, so she's going to work in authentic.
It's up looking it. It's right. What me to you should screw with a minute.
OK, once again 0.0.
Description.

RamKrishna Bhatt 1:43:27
In.

Tarun Jain 1:43:28
You will. He didn't left her. He was wrong. Tell us, Ram.

RamKrishna Bhatt 1:43:29
Can can we be?
Hello.
Hello.
Oh yeah, I was saying, can we be more specific towards this atyantik like in Vadodara city or in some?

Tarun Jain 1:43:38
Is it on the three service?
Yeah, that is not the point. Then optimizer use. Optimizers are different because OK. Maybe not this. OK, wait, wait, wait, wait, wait, wait, wait, wait, wait. If you look at this, this is multimodal agent. This is your search agent. My bad. Oh OK, not let's see if it works or not. If it again give wrong response, I'm not.
Care about this? OK, now we're using family about this team.

RamKrishna Bhatt 1:44:24
Then I will leave it if if we do not get correct response.

Tarun Jain 1:44:27
Now this is at least not giving the response. No, I did someone think it will take me down the.
Now let me remove this because this was added. At least it should not give any wrong names. It's OK if it tells. I don't know because it's not. Do you have any web new photos? Yeah, based on on web we have on about this we have. So we can think about that. We can think about this.
No, no team. It's like team. Oh, hey, there is that's OK. OK, you got.
The two personalities from Atlantic technologies are Deep and Ajay. So here I removed the description and all. What if you have only one? We can try. That will be

very difficult.
Like that. Let's say if uh only uh ajay son will be there. OK, let's just add. I have the same. You can just right click and copy the URL. You would have it right there. Don't upload it. Just right click and uh copy the address, yes.
But then it is very similar. No, but I'll give you my college image when my page is not bulky then that in that case you have to be famous.
Is.
Name the name the page. Uh, no.
OK, now it is something for attending the same team team. OK, OK, good. Got it. Group each other. Then finding the. But let's proceed. Let's proceed.
Just one, just one thing. Just take a and then ask it to search in. I don't know why I asked. OK, let's proceed.
So we saw how to use multimodal. We also saw how to use tool. Now you have custom API. So custom API in the sense you have this endpoint. So do you see this user as facts dot some URL? But if in case you know any endpoint which is publicly available you can use that.
For example, if you see I if I just copy this.
You have get command, copy this, paste it here. You have some random fact. So what custom API tool will do is you need to define a URL and you should also define all the endpoints and what is the methods of it which is get.
Up what you have get, post, put and update. You have to define the sub endpoints. Update will not work. Update will not work. Post will work. Get will work. It will also not work because you have safety guard fields.
So here what am I doing? I'm defining the custom API tools. LLM is same again. Say no post to post. What is it doing? Post are sending parameters so that is similar to search.
So define tools. Now tool is what you have custom API tools base URL equals to URL. Then you have agent dot print response define the endpoint, get and that particular endpoint.
But if it's like multi, OK, then my goal is how can we ask you to do multiple things in this? Multiple things in the sense like you did it, you did it with certain that it did something around it, huh. So it will get some data. It will then do something. Then you want to ask Adam to do it. Can you do that?
That will come in teams. OK, OK, that is team. So again, let's revise this part. Everyone look at the part. So here you have search agent, you're defining name, model, tools

and markdown. Simple. Apart from this you have system message, instruction and. Description so that we will define in teams. After you do this either you can use search agent dot run.
From and stream equals to false.
So your prompt can be anything. Let's suppose you want to ask. OK, this will throw error. Let's suppose you want to ask anything related to real time. What happened today or tomorrow? This will use timely search. OK, right. But what am I doing here? Here I'm asking the prompt and then I'm using multimodal. So if you want to use multimodal plus.
What do you call text? Then you can use this way. But if you only have text as an input then you can use only prompt. You don't have to define images OK.
So this was just we were doing the flow, so we went with the flow.
So now if you see here, this is how we will run it. Agent dot print response give a message. It is using a tool call make request method is get then endpoint. The random fact is in America you will see an average of 500 advertisement a day. OK, I don't OK, it's a fact.
So you can keep running it. Yeah, like maybe a call to endpoint like how will you do this because we need search? No, no, it's just a so API is this custom API goes OK, OK, OK sorry. OK so if in case you have if you know any endpoint which is publicly available you can use that and test with post get and all.
No.
So this is one. We'll also experiment with custom tool. So for custom tool we have seen this code earlier. So you have search inside search you have maximum results to be 5. Then you have EN as language. Then you're extracting the context. Then here you have Google search agent if I want to run this.
The syntax is same if you see name, model, instruction, tools and markdown. This is optional. Even if you don't tell it to generate in markdown, it will be markdown.
Hold on, we are at Google search partner, huh? Custom tool. So we did a custom tool with query user prompt and then return context to the user query. Now with this appending the context, correct? So if you come to the documentation, can you see this creating your own tools, right?
So here you will see the example of hacker news top stories. So this is a custom function.
So what do we do in custom function? Define a proper doc string and what should we do? Second thing is data type. This is important right? What do we do? This is

typing right?
So I hope this syntax is same for everything. Same syntax, define agent, define model. The tricky part is defining the tools is the tricky part and what is the right instruction and description to use.
Because that is where you'll spend more time. Defining an agent is now easy, but defining your own tool which will give proper context is very important. This is where you'll spend time on because this is very easy and you will not have this easy functions.
If not, you can do this this one. You have your custom APIs which has proper results that you're generating. Use those endpoints which is much shorter. He was very interested. It was very insightful. OK, not search agent, it is Google search agent.
dot print response.
Oh, print response. What do we? OK, who was the?
Who are the the workshop speakers from the first one? 35 of the speaker.
Do we have to write stream on this? If we don't write, it won't give you the output. So I guess stream. If you see stream bool is none. Did you see that? Yeah. So it didn't even define true and it didn't even define false. It has defined as none.
So technically it should be true or false.
So you have to run two times if we didn't give anything.
Why? I told three and it yeah.
Oh, information regarding this name from the first workshop speaker or the three workshop speakers this this.
What happened from you guys have exposed this robot, this thing, right? You can scrape the data, right? Yeah, robots we don't have. I guess robots is overloaded. Please try again. OK, we are using pro.
Let's use flash. OK, wait, if this doesn't happen on Google search, then it's very difficult. Yeah, I'm asking who is Google is not responding. So do this thing. If it is not responding, uh, instead of print respond, keep it run.
OK and make sure stream is false because sometimes stream is false and then and then we get a response called to then I can response equals to friend response dot content right? Huh.
Uh, So what can I ask from the first?
So now it's there. See, I can't find a. Yeah, it is. So yeah, it kept. It kept me first.
See, it's very clever. But no, see the see the last one. See the last one. That doesn't matter. See, see it in the topic. See in the topic. You'll ignore that. I saw it.

I was a talk with that city response dot content but OK here it's using search and it got the result but here we didn't get it. So instead of.
And this is Google API Google search, right? I'll use something Google will get. But if I'm searching on here, it doesn't get the details. Why can you repeat if this is in tools? Huh. OK, then it should go to Google and get the results. Technically, yes, but it's not doing so. Can you try with Pro and then use it?
Yeah.
So now if you see I'm using Pro and instead of Google search I'm using Tabley. Hopefully, let's see if gets it or not.
But I don't usually do function calling with Gemini live. He is the writer of Gujarati movie. He has delivered a Tedx talk yes someday.
OK, it gave a fourth while it gave. It gave Indranil, Sandra, Tarun, Jain, CRL, but it did. It gave Indranil now. It shouldn't be there because it's a.
OK, there was only one workshop by the way on 5th. I I was afraid of that it is not passing, but it is passing properly.
But if I use family family tools, I guess it's taking from uh GDC Baroda website instead of thefest.baroda.com. The GDC Baroda website does not have this information.
It hasn't been updated. No, it's there. No, there isn't. So if I open the DDG, the top three there also if you see DDG.
So if I open this website, it doesn't have this kind of information. Yeah, it does not have. So who are the three speakers here? Who are the three speakers? Oh, yeah. And if you look at the response, you have Indranil, Tarun and Srushti, this is there on this website. Oh, good.
So whatever the first is there, it's not SEO friendly. It's which didn't come in then this very bad. We'll send. We'll send.
No, Tably and Exa is very good. Tably's it's very precisely based on the information is the cofounder and director of technology of the development company and he specializing in driving software development computer engineer.
Which he completed in 2011. Now it got my completion year as well. He's a technology maniac, which I really like and if you give a bit information about Tabley and Exa.
They're very good in search because they are specifically meant to get the context. So there is one more parameter that you can use here. Google search API, SERP, SERP, SERP API. So it's like this one SERP API.
Of Google right? This is used for Google search API, this one. So if you sign in you

can also use this one.
So I'll send this thing. I mean, I hope you got this syntax. Yeah, yeah, let's proceed. Yes, yes, yes. So this is last example. You have talked to CSV data so far. When we did RAG, I told you that never use RAG for CSV, correct? For CSV we have something called as pandas.
So use CSV as a tool and I've attached a GitHub URL. So can you copy this URL, paste it here and you can download thousand CSV you can download from here. There is a comment.
Custom. I guess this is the last example. Yeah, this is the last example. 20 lakh custom. OK, that's very huge. But you can try. Gemini will crash. So now from agno dot tools CSV agent I have CSV tool so far. If you saw we started with.
Doug, doug go and we had tably. Tably needs environment variable right after that we had custom API. Custom API is just a request, right? So request in the sense the library that we saw when we looked into fast API.
And then you had a what do you call this thing? CSV tools. CSV tool login is a pandas this thing, so it will write the execution step and you execute it. So mainly it will use WDB, but we'll see if it uses WDB or not.
The CSV tool for it so CSV tools then you have CSV tool import this thing. Then you have pandas and after pandas it is just to check if it is working or not.
And then and ask anything from what you call the framework. So I'll just ask what is the e-mail ID of Andrew?
But.
We can have multiple e-mail ID, huh? You can do it. It's an I, it's a list with it with CSV tools, yeah, but you have to define that in instruction when to use what? But no, that intent classification sometimes might mess up. So.
It's better to have certain control e-mail ID of first name and use choose to search the web and find the most accurate answers.
OK.
You.
CSV tools. I don't somehow mess up for use I mean I pro data even getting exhausted very fast.
But if CS doesn't work, we can use pandas. There are multiple nodes we can use.
OK, we have mixed for August.
Yeah, this we'll look at later. Instead of CSV loaders, we'll use pandas.
If you're on the way to do the data pandas.

It's a problem with parsing, so they're now going to list CSV files, get columns and then it is OK. Now these are tools within that CSV tools, right? And this calling multiple tools inside, OK.
Resource exhausted.
It just didn't think of but you know it look one instructions and then use TF.
OK.
OK, I guess we'll have to do the challenge tomorrow.
We'll make this CSV work.
For panda tools are also there. Yeah pandas is there CSV parser error syntax that if we are getting from CSV tools. What let me do take a panda tool for the tools. So you have pandas tool and here load and.
Completely open source.
Uh.
OK, let's. I don't know why it's not doing proper columns. Yeah, panda. We have panda tools. OK, got it. Yeah, we have panda tool, panda tool. You can just load and analyze your path.
This should be hard coded.
Find out the list.
Install unique.
But then we have to give that path in path. You can just define like this, load and analyze the path.
I guess we'll use open AI for this one because I'm very. I don't think Gemini will be able to process. I don't think this will work with Gemini. Yeah, it will lead the limit.
It will won't work, so let me see.
What work?
Yeah, it won't work with Gemini. We'll avoid passing error though it is really passing error. So it's like if more context is not given properly, the structured format is messing up in Gemini. So usually so I'll tell the flow this will work.
You upload a CSV file, you ask a question, it will run a code. So when it run a code, this will generate a schema. OK, so pandas dot description, pandas dot info. That is your input for your prompt and when you do this.
You'll also extract top 20 or top 50 what you call head. So now the top 58 is messing up the context.
So I'll try it again. Let's see if it generates the output thousand though. Let's see how it does the thousand the dot CSV.

Yeah, I didn't take the input. I'm sorry, now we cannot fulfill the requirement. So I believe Gemini will at the limit now because I'm using pro from.
Morning. I'm using a panda tools data ring object, no attribute unique. Make it row column unique values.
Get column unique value and you should also define this thing like load and analyze from this thing. I mean the description argument column. Yeah so I mean like it's not working. It doesn't know how to call the tool. It is very.
Neon.
If I try it, they will try this. OK, I think I could get something. So it does try to do error handling, but The thing is Gemina doesn't. This is pick that up. This is what I'm getting with Gemina class.
It's not.
I'll try with something. OK, I'll send this template. OK, it created the query for me, but somehow it have been. I'll share the template. We'll do this for tomorrow, OK?
Yes.
It's the same syntax. We have customer instruction, customer role, post on Instagram.
Yeah, you have. Yeah, you have. So agent is used for that, right? Right. So you can build discords at words and then there will be video iteration and then you can.
Post it on Instagram. People are doing that with, uh, anything, right?
So if you come to agents and if you click on toolkit.
Where is toolkits? So search you have all this social you have Discord, e-mail, Gmail, Twilio, Telegram, Slack. You can use Slack which will be very useful. Then you also oh Instagram is not there. Telegram is there.
Twitter is there. LinkedIn, LinkedIn, Reddit is there. Reddit to you. I'll tell you one use case for Reddit. There are many people who build customer service at work. Let's suppose quadrant, right? If anyone has anything issues with vector database, you have automatic.
Comment you can check out quadrant and we solve this using this so and so. So it's like rag application and action is being taken on Reddit. But then the problem would be.
It will add everywhere comment and the problem is you get banned right? That is what I'm thinking. But you have to but you need to add Karma. Tone is something.
Uh like if it but it's like if you have Karma very high you can.
Anyways, then you can just Karma in the sense Karma is a sport. What is your? I have -12. So this is my alternative account. I have one.

I really am very to do something about it. Where is Ark? I will not do anything. No, he will read it on daily. Even I use Reddit. I'm always active on Reddit. Oh ****, I have 16. I have negative. I'll just show am I negative? People are super, you know, negative 10, I'll just hide this alternate identity that I have. I have -10 as of now, 3 + 2 over. That's how much 241.

That is different account. I said you have to every redditors have two accounts. One account is to comment everywhere. I don't want with real identity. I have one account with 2000 this thing which is uh yes it's subscribed out.

Do you want to show it or not? Oh, I got it. I got it. I got the response with. Yeah, so I unique. What is it right or wrong? OK, it generated for me also. I have loaded the data and I generated the CD column. We have the unique from the data.

But this is not a good solution. We'll look at alternative. I'll show that tomorrow, OK, but with I'll show it.

OK, yeah, that's it. This will cover tomorrow. I'm going to play around a lot with this one.

What you can do is click on examples, you'll find 100 of them.

Mine is. I'm gonna legately gonna ask a person. Wait, yeah, I'm.

I don't know. I expect some good content out of it.

Tirth stopped transcription