

# Python and AI Power-Up Program Offline Class-20251007\_191000-Meeting Recording

October 7, 2025, 1:40PM

2h 9m 40s

● **Ajay Patel** started transcription

**TJ** **Tarun Jain** 0:14

So either you can use projects up or let me use VS code.

Mhm.

We start.

After the morning.

 **RamKrishna Bhatt** 1:35

Oh yeah, I I I'm able to hear you, but your voice is not clear.

**TJ** **Tarun Jain** 1:43

Um.

Your phone is mine.

You you might want to. Should I rejoin or now it is fine. Now it's fine, right?

Ramakrishna.

Uh, hello.

OK, I'll share this color notebook again. Uh, so we'll just continue with what we left yesterday.

So this is the template notebook. Yesterday what we did so far was we define agent component, then you have tools and then you have LLM. But now we will just add this new component called team.

My voice is correct, right? Margi is also there on only Ramakrishna. Hello. Yeah, OK, it's audible.

OK, OK, so these are the two tools and then we have teams. So basically when you have teams right, you can do three things. One is coordination, second is collaboration and 3rd is routing. So whenever you do collaboration, coordinate and routing, earlier you had one parameter called mode.

So it's like when you define team, you have a parameter called mode where you could define collaboration.

Then you have coordination.

And then you had something called as routing. But now in their latest release, they've merged it into a single prompt. So whatever instructions you want to give, you just have to define a prompt. That's it. So you don't have to define mode and agent will automatically take care of itself. So we'll define this team component.

But before that we have to install package.

And you can create the copy of this. And if anyone wants to run this on VS code, we can run it on VS code because after you run it on collab we will be doing this part. So this is where you will create the UI for an agent.

So if you're running on VS code, you have Agno, then VDJS which is for dug dug go, tably, Python and then you need to install Google Gen. AI. If it is on VS code then you by default you have open AI installed.

Will it take Azure? Open AI is Azure. Azure Open AI requires open AI library. That's it.

OK, this is the library required for what do you call Azure Open AI and Agno by default as agent component which uses Open AI so you don't have to install it.

But if anyone others probably will have to use Google's NEI.

If you're running it on VS code for collab, it is fine.

Everyone has to run on VS code because of the UI, but the only thing is first we will see if it is routing properly or not. When I say customer service query, it should go to customer service. If I ask anything related to marketing service, it should route to marketing service.

OK, but once the app is done, we have to check this on UI. So UI is very well designed when it comes to add no that is a probably the plus point in the latest release.

OK, now the installation is done and I'm just importing the statements.

And when we when it comes to teams, right, we have two agents. Now if you notice the agent component, you have a parameter called model. So that model can take any elements. So now let's suppose you think customer service at bot, Gemini is doing better, you will use Gemini.

And marketing agent, if you think open VI is using better, you can use open VI. So every agent can have separate LLM. So if you want to use the same LLM, you can use same LLM, but if you think this particular.

Use it for example multimodal agent. Which one is better? Open. No multimodal is a minute writing. OK, OK, whatever writing task is there is general knowledge. You have open AI and if in case you have anything related to code you just plot. So based

on what use case you're building and.

What agent you need, you can pick the other now, but the code is same, but the only thing is we have our own instruction and role OK.

And you have new 2 parameters, but when you define the team, team also needs one LLM, right? So now what is the purpose of this team? It is just to route, right? So even if you use small LLM, it should be fine, but that small LLM should be from Openyard. Hello.

Is it clear? Yes, we are just installing the package. That's it.

Do we need a? Yeah, we need Tableau. So Tableau also what we'll do is for customer service, we will use Tableau because customer service you can go check the web browsers, get the information right. Dubdubgo, it's very simple, so we'll use it for marketing agent.

No, doesn't need API.

Doug doug on Google says doesn't need API. Brave needs API, Exa needs API, and Tali needs API. So Brave doesn't make sense.

For plain Google Python you don't need it, but if you're using SERP you need a API.

Will you be what? SERP. SERP is Google API. OK, but if you're just using a Google API you don't need it. That is Python SDK. That's a very native. We wrote one custom function right? For that we'd never use the API. OK, but if you're using SERP.

Probably what we'll do is let's use SERP today because we have not used SERP so far. So this is the URL. Instead of dug dug go, we'll use SERP OK.

Ramakrishna, we have to create API key. This is the URL, so I'll just sign in. No, this is we are not using. We'll be using OK, this idiot. We have to verify phone number. I'll give my API key.

It will take time, no, but it's not that much of a difference. We can use Tabuli. We have to. So yeah, we can try it now when we are using Tabuli, one more tool will one more agent will use different tool.

I do have a.

Meanwhile, everyone can start writing the code. I hope the syntax is very similar. So here what you can do is just define customer agent equals to agent.

And here you will define name. So name is.

Customer service assistant.

And then we can define model. Model is nothing but copy Gemini pasted ID equals to Gemini 2.5 flash.

And then what else did we add? We had tools, so your tools will use tably.

And markdown equals to true till here we wrote yesterday. So you had one name, you had model, you had tools and you have markdown. Today we are just adding rule.

Because when you're routing, you want to know for which member you're routing. So member is nothing but your role. So these are some common parameters within some of the frameworks, role instructions, system prompt and back story.

So backstory is for pro UI but tag no it users add add additional context which we saw yesterday here.

Role is customer role.

Do we are doing OK for customer agent where we huh.

Because customer agent, you need a better search tool.

So I'm waiting for this. What you can do is repeat it for marketer agent. Until then I will send you the API key and where where should you check for documentation? docs.diagnore.com. So you need to replace dug dug go.

OK, this is the API queue.

Yeah, this is the API key.

And I'll go to Tools, Toolskit, then search.

And inside search we have SERP API.

So here right instead of I mean below tably create one more OS dot environment SERP API key equals to whatever I shared in the chat role is for.

You are an expert at what we say. We did this yesterday at home. No, no, we didn't do it. So it's like, how do you define system prompt? You have role play, then you have inspection, and then you have structure output parses. You are separating it, huh?

All the frameworks separates it and many people they repeat. It just works.

Typically what we did was when we started open EGI, even we gave only two parameters. We give just add your system from, add your user from. But then we started giving new parameters. We saw better routing most of the time what happens in.

Agent is how well are the parameters assigned when you do tool calling because that is the major part. And when you give more parameters, more context and one thing I didn't understand like most of the people repeat some of the statements. So for example if you see.

Description and back story doesn't make sense. Like why do you need a back story example if you can just give that in instruction, but still you have that parameter.

So this was the first agentic framework. Whichever framework that you're seeing, right? This was the first one that was developed. If I open the documentation, you have documentation.

Which starts. So if you see here you have role, role. It's like simple thing what to define in system from. Then you have goal. Goal is nothing but your instruction and then you have back story. So if you see same things are repeating here. OK so it doesn't make sense having.

Different parameters. At the end of the day it's been upending in your system prompt, right? So why do we use role and instruction? These two are going in your system prompt. But The thing is people don't know what system prompt is. You might have seen whenever as system prompting events.

You'll hardly see two people raising the runs. It's very rare you will see at least more than 50%. So system prompt is unknown because no one even adds this in the documentation, right? If you just search for system prompt tradition in Gemini, if you get that in first attempt in documentation, I will leave GD.

So it's that worst. So same goes for open AI. But open AI I will not say because open AI you can directly add role. So if you see you have role, you have a system, then you have content that is open CI, but Gemini it's not direct. Claude again it's not direct, but whichever is.

Is open AI alike for example Fireworks, Grok, Samba, Nova, Uggings Face. There you can define role system, then content, your prompt.

So this is just your system prompt but separated. OK, so just add this line.

And we need to import this line.

What we create in our customer agent and marketing agent. So remote the dug dug go. Instead of dug dug go we are adding surge API tools. So this was one line which we added new and this was the second line and we also have to install some package.

What's that again? Google search results. Yeah, Google search. Google web to install Google search results.

Pulwani. So now we'll just copy this customer agent, paste it here and instead of customer I'll just add it as.

Marketer Agent and this is Marketer Marketer Service Assistant. Model is same, but if anyone wants to switch models we can switch it. And for tools I'll be using SERP.

And role is this marketer role?

And marketer instruction is marketer instruction.

OK.

OK, marketer hold in the comma at the end. No Sir, API key provided. OK, I didn't run this.

Oh, what comma in that? OK, this will become tuple tuple.

Uh, now it's correct. So I didn't run that the OS dot environment. We can run it.

Model any other endnote of models dot H dot dot open El.

Hello.

Yeah, is this done? These two lines? No. Serif Pi tools and OS dot environment serif API key and I've shared the key in the chat.

OK, so till there is it done, we have defined both the agents. Now we have main instruction. So if you check at the main instruction, if the query involves marketing product promotion or social media strategy, route it to marketer agent. So whatever you want this.

This particular agent to be very specific to, you can define it. Then if a query involves customer support, product troubleshooting or service assistant, route it to customer agent. Earlier this was very simplified. You just had one parameter called mode.

Route. But now this parameter is not there. So we define team of agents or it can be anything, any parameter. I'll just call it as worker equals to team.

And if you see you have members, members is a list. So if you see members and you have list and this list is nothing but your agent. So we have.

Customer agent. I will copy this customer agent.

And I'll copy this marketer agent. And what else do we have? Memory will cover memory. I've not added.

So, so there are two things you have to notice here. When we talk about context engineering, there is one entire concept that we have to cover. There is a slide. So here if you see so far whatever we have covered is we just covered these two parts. Context engineering and we have never touched the memory and state so far, right? So the difference between these two is memory is like long term. So you have to save your preferences. State and state is nothing but if you're chatting with an agent, what you asked earlier, those kind of question you have session.

So here also when it comes to agent you have to define that and for that what we will be using is we'll use SQL Lite.

But we'll cover that in this team we also have to set model. Team also needs a model if you don't define for routing. So if you see that you're defining this instruction, so this instruction is going to the LLM that you define inside the team.

So here I'll define model equals to Gemini and I'll keep very light model.

We need any meaning or anything.

So basically if you're using Azure, it makes sense to use mini here, huh? Not GPT 5.

You can use GPT 4 or mini also that works.

GPT A4 mini is good for function calling also.

No, it's a bit expensive. I mean, like Mini and Rhino are cheap, is not Devina 2.5 Flash.

I'll just take a bit and I'll add light. Let's see if it routes or not.

But not cheap, I link for.

Then instructions equal to main instructions.

And here you have one parameter called show.

It should be show members. Can you see this parameter show members response to be true. So now what is happening here? When team makes a decision I want to see to which team it is calling it. OK so true and then markdown to be true again.

And whenever we work with invoice extraction, what you're supposed to do is yeah, tomorrow instead of doing this, no. When we looked at multimodal agent, we should have taken invoice extraction example instead of playing what we'll.

Because there what we could have done is we could have used a schema like I want my output parser to be in Jason and then instead of using markdown to be true, we could have used use Jason as true.

But they cannot work together both. No, obviously no. So we either will use markdown or Jason. But when you define what you call, when you define use Jason mode to be true here for your agent, we have to define our schema.

So we have to define schema which is from pidantic, but let's try to do this today only. This is very it's like you just have to define parameters so if you just over on agent.

You'll find minimum 502030 parameters. You have knowledge. This is that.

Knowledge. Knowledge itself is right. It's like 2 lines of code, only two lines of code.

You give a URL or something. Uh huh. That's it. And if if you're using a separate embedding, you just give the embeddings or even use your and then you just say hybrid too. That's it. It will do both. And one more thing it uses called agent.

Chunking by default. So basically we have recursive chunking, token chunking, then you have semantic chunking. Agno has something called as agent chunking which is done by default. So that is their feature again Agno's it's so abstracted like.

If we never learned the previous thing, we would never understand what it is going on over here. Huh. So in abstraction, we know what is happening. So frameworks, it's

very important. If in case you want to contribute to this, right, it becomes very easy now because you know what are the parameters if some parameter doesn't make sense.

You will write your own custom code so but having understanding of how to write custom code is important.

So we have name, we have number, we have model instruction, show numbers. Why is it showing the red mark? OK, my instruction.

So we have name, we have members model instruction and notice model model. I'm using light and I'm not using flash. Let's do one more thing here. I'll make it pro pro. This is one of and now what you can do is same worker dot print response.

Every tools does need API from, huh?

No, but it is already getting on this environment. It's not picking up. OK, I'll see the spelling mistake.

Did you make spelling mistakes fine?

No, we didn't. It's the same. Well then you might need to run it. So for print response you can ask any strategy based questions. I'll just copy paste from.

So I'm just asking what are the strategy should we take to get good audience for events? Stream equals to true.

So which member should it call now? Marketing agent?

It's searching for Google. Google is used for marketer agent.

So firstly, this is the Uh search parameter that it is using efficient even marketing strategies 2024.

So it has delegated to. It has delegated to marketer service assistant and this is the output.

Query that we added. What is the query? OK, generate a very lengthy report on audience. Getting the best audience for any event. What are the what are the strategy for audience should we take to get best audience for e-mail?

Mm-hmm. We can add any marketing query and I change it to. So now I'll again copy worker.

Worker dot print response.

I have delivery issues.

With look got.

What do I need to do or do I need to do?

Stream equals to true and what is the alternative if I don't want to stream worker dot run and then round how much stream to be false?



What content? What if I give this as true?

What is true? If this is true then it will response the response will be a stream then it should be like for but for response for collaborate.

That there will be some this thing, but it will still work.

There are a few people working on VS code, so once this is done we'll move with the UI part.

We can have any questions, right? I take them from. We can have any questions. No, you can ask any question. Just see if it is routing properly or not.

If you are getting limits, switch to Flash. You've got got the details. Yeah, I'm saying if the customer support related query involving Flipkart delivery issues and how to do the customer service assistant.

Let's proceed here a step by step by to resolve this. Now how do we know this is the response? That's all just because we wrote it. Do we not have a something around it? For which one? So they said I'll route it to our customer service assistant.

But how do we know that tool? So only I use different tools. So if you see here what is it using here? Can you notice the screen? You have web search using tabbly for customer service earlier if you see it is.

Search Google. So this made clear that tools was used based on the given agent that you defined. OK, so if you don't define tool, it was very difficult to identify because without tool both them both of them will do proper things.

So if you see here you just verify if the number is correct or wrong because here it also gave the number.

But OK, if you're doing this in VS code, if you're doing print response with screen to we just have to do it print but we can't do it. No, you can do on VS code. It will show one word at a time.

OK. OK. Bye.

OK, this is working. How everyone of you got the response? Now do one thing. Copy this file, then download Python file.

So that we don't have to copy paste this entirely on VS code.

Once you copy paste this create a folder. In that folder what you can do is you can first step is create environment variable then install libraries.

So I'll repeat the steps here.

Download the.

Collab as Python dot PY.

No, the phone number it gave to me. It is for national toll free helpline numbers for

consumer.

The customer is so with all the number for for anything which one new temperature at 0.

OK, wait, you're using what? Which model? Azure 5 mini 55 for this. OK and do you have 4? Just try with four. I personally believe 4.1 is much better than five.

I don't know why they release GPT 5. OK, even when you run evals right, it's same.

I'm I'm not seeing any difference. Last time when Aardeep said I didn't run any evals, but last week I did run evals. Scores are same.

With 4.1 and five we haven't changed. One of our product is using GPT 4.1. That deployment is still there. It's same same. We didn't delete the instance. You need to create a new one and use so download the collab as PY.

Then create a folder. That folder can be any name. You can just keep it as customer service.

And inside that add that file and then just create a what you call environment variable virtual environment Python 3.

VNV OK ENV 4.1 is much better response. 4.1 is good.

Then you have Python 3M. This is the for virtual environment, then PIP install agno. Agno Tabli, Python, Google Genii.

I can do this.

And then what was one more library Google search results search results?

So yeah, this we need to do and so far just run this and then new 2 libraries are pip install fast API and which is the library that we need to install with fast API. Does anyone recall? No with fast API.

So let's suppose I want to do the UI and I want to use fast API. So if I want to run fast API, there was one command.

To that library has to be installed with fast API every time.

So I hope Ramakrishna got the response.



**RamKrishna Bhatt** 32:28

Uh, yes, in Colab I am.



**Tarun Jain** 32:30

OK, OK.

Should I scroll above anywhere or?



**RamKrishna Bhatt** 32:36

No, no its fine.



**Tarun Jain** 32:38

OK, so I'll just keep this screen. We don't have to install anything for Azure. So once you define worker, you just have to define worker dot print response, then give a prompt and then stream to withdraw. That's it.

So what you need to notice is you notice if the tool was used properly and also here you'll have member.

If you scroll below you will see this member delegate task to member.

Which has to be marketer for first prompt and for second prompt it should be customer service.

I will remove this.

Let's oh, is it done? No. Anyone can tell this? I don't have, but I'll make sure I'll deliver it.

Fast API end.

OK.

So how did we run the code? Let's suppose I have a fast API code. How did we run it?

But how did we get that URL?

You should run the web server. Huh. Stream you so you have UV con.

dot run and what are we supposed to do? You can't run app app then post 0.0.0.

0.0 then you have port, but we won't need for this. We won't need fast API for this no, but this agent OS for this part it's using fast API.

And for serve it is using UV con. So whenever you see any syntax like this right what you call multi agent then app undoubtedly that's UV con. This is the syntax that you use for UV con.

And then you have this.

Reload option which is true.

is different and is different.

Huh. So worker is a team.

So if you see an agent OS equals to agent OS.

You're defining the ID. This is nothing but you're defining your.

Fast API. OK, so for a fast API you define an ID. You are giving a description, then you're defining agents. So how many agents we have? We have two agents for. Let's

suppose you only want to use customer service. You can chat with the customer service separately.

So if I don't define this, let's suppose I don't define customer service, I can only chat with marketing service. But if you want to chat with the team then you can use teams. So in the UI you will have all the three what you call all the three agents.

OK, sorry, I used. Oh, OK, OK, OK. I sorry, I I don't know what to do because I was using OK.

This is worker and this is worker. OK, let me change the variables. Then what else do we have? This was.

Marketer agent.

And this was customer agent.

That's it. Remaining everything is same. So I'm we are still using doug douggo, but that's fine.

The.

Mark, do you want to hear that as well? No, not the can you scroll down to the Uh voice part so we can copy it ID.

OK so till worker your code is same right? OK whatever you have downloaded that will remain same. You first have to import this line which is from agno dot OS import agent OS.

So at this time, or else I'll just copy this and add it below. You can have multiple teams as well, but that won't make sense, right? So it's better to have multiple agents, then give it to a team, but having multiple teams?

But why do we need if we are giving agents already? Why do we need agents? So it's if you want to chat with that separately. Let's suppose you have 10 agents right now and now you have worker. You will only chat with worker for separate this thing for routing.

But if you want to test only one single agent, then you can directly go chat with it. So it's not necessary that you have to use it. OK, so when you build UI that will be there, OK.

So from agnode OS import agent OS then just ID description. This is similar to what you call fast API and these are two new features. You just add agents and teams and then a app equals to agent OS dot get app. This will trigger fast API.

And then if you want to run it, you just have to use agent OS dot serve. So basically it is UV con dot run, but since we are using external library they will have their own naming convention and they have kept it as serve.

And now why is this multi agent code? Because my file name is multi agent code. And why is this app? Because this variable is app. So if this is app 2 but I'm using it with poetry then how do I do it?

Like to.

Oh, what word?

So it goes with the command. So it comes differently. I'll I'll I'll experiment something and see what it does. No this this won't make any issue right? This is nothing but your file.

OK, this is the same file which you downloaded. OK huh. I named it as multi agent code. OK, I'm getting this.

Because did you install all this package? Yeah, so if you see it's not in environment variable.

So always notice this whenever you create environment variable you should have.

Source.

PPNV win activate. You should see this line PPNV.

And after this you have to do pip install.

Mhm.

Read this. So let me copy this line.

Let me add this here. So before you can do this, we have to activate first.

Great. Is it done? Uh, you didn't create environment variable do LS.

So it's not PPN. You're also is ENV. OK, so I I made it as PPNV. What I did for poetry, I removed that part. I gave app equal to app directly. How you kept it as app dot PY then? No, no, no.

No, not be late. I just gave your F equal to have to.

OK, OK, then the reload functionality won't work because it depends on it. But if you give it app equal to F2 without reload, yeah, I got it, yeah.

OK, that's just a second thing. This part is done right everyone. So here PPE and me because when I created virtual environment.

I was using PPNV. OK, yes. So you you you have to use ENV. OK, so here it will be NV, here it should be NV, right? So this is anything. So this variable you can keep it anything.

Then the requirement then pip install Agno, tably, Python, Google, Genai, Google search results.

Yeah, first you have to activate, then run it. Now it's installing in that interpreter and once it is done, also install fast API and UV con. What is the URL? I've not. You'll have

os.agnode.com.

What?

No, the only problem would be like what we think about this. You have to add your URL. I'll show that part local host, but everyone would be running on local host and the AP would be only once.

Means I'm local, huh? You're local. It does not expose or something like that. OK, it does not expose the code. OK, right. Everyone completed this. We'll run it at the same time.

Yeah, just a second. Still, it's not going. So that's fine. Just do Python.

It's base. No, you don't have to do Python V Python connect the region.

The ha, this remove Google collab. Instead of Google collab you have to use these two things and what are you supposed to do dot ENV. So dot ENV will have tab the API key, Google API key and it will have SERP API key.

And then copy this under thing the key.

What is the code for agent device to connect to agent device? So I'll just copy this. Paste it below so that it will be easy, but make sure this is above. This three line should be above.

Is it done, everyone? Yeah.

And DNV, just take a look. I'm not adding any colons. Yeah, Table API key equals to whatever key is there. Yeah, Google API key, Serv API key.

I'll copy this and I will push this above. Yeah, I'll just send this in chat so that Ramakrishna can add.

OK even I'll run the code now Python multi agent dot code. So when you run it you should see something like this. Are you able to see these two things? Yes and what you need to do is this is your end point so copy this.

And if I paste this URL, this is my end point. Let's just see if it has Swagger UI docs. You have the Swagger UI. So what is the URL to add Swagger UI? Just copy your endpoint slash docs and then you will see all these details. So if you see you have core agents, teams, workflows, health, home and all.

So this is just endpoint but the UI is currently at.

Oh.agnode.com so copy this oh.agnode.com.

Uh, I don't know if it will ask you for sign up or not.

So after sign up it will look something like this.

I'll just look at some other screen and not see. I'm getting what error what error?

What we are what? So do you call? No is requesting to HTTP local code 777 right?

It does not allow. OK, let's just we are not in environment variable.

We need to install it. Um, Google.

What is it? Open. Do you see? I mean why are you opening file?

It is Python. No, I said the names to be right.

So it will, uh, OK, but then why did not?

So it will do. I don't know why. I mean.

Still run it from this thing. VS codes have been so can't have any Python code locally.

I need something that's more is. I don't know why it is just and they should be listed.

Oh no, now this is not in environment variable. Switch to environment. OK, click on this interpreter the site and.

And to select recommend this one, this recommended OK, where's the one on today do LS we need to.

The agent, let's do a list. Did you get this? How many of you are on the screen? You watch that? Nothing. You don't have to do that. Just remove all those things.

Remove all the three lines. OK, so how many of you are in this screen?

Only one. I'll keep keep the code if in case anyone is still writing code. Python multi code. So here I've kept app two so it should be app two. If you're keeping app it should be app.

And they should be up, yeah.

No, you didn't add this code only. Oh sorry. Oh the the IT should be the file name.

Multiagent code. Oh OK.

So this is fast API syntax.

OK, for sure this nobody is getting.

Now I won't. I've copied it in the chat as well. I want to see someone who is using it.

OK, I'm getting something.

I'm doing agent OS API 777 so just copy this one.

Copy this URL. OK, OK, so this is the endpoint which has swagger UI and all those things. But we need the UI which is a os.tagno.com. So we'll have to log in, no log in.

OK, no, no. OK something else.

OK, So what are you able to see here at the top where you see executive agent, what are you able to see?

Hello.

The router hyper agent with a red mark. OK, can you click on this agent OS? Agent OS where? Yes. Can you see the screen? Yes, guys, everyone follow this.

You will have your name. Below that you'll have something. Just click on that. Then

you will have agent OS. OK then click on plus.

OK, how many of you are following?

And here you can keep it local. Mm-hmm. Here give the endpoint. Endpoint is your 777 right? Local host right name we can give customer service.

Tag it can be workshop or anything. It can be anything which is optional and then click on connect.

OK.

Now it will take some time to connect and then it will show teams and everything that we have. Yes, I didn't possible.

I have a, yeah.

OK, it didn't reload. Let me reload. What I did was I changed the app to to app just to show the code.

So can you see now you have a market agent, you have customer agent. If you want to chat to that specific thing, you can chat. But I don't want to chat with customer service, I only directly want to chat with the main orchestrator. OK.

So understand the difference. So there might be chances you you want to test some individual agent in your UI so you can test them. But what will I do? I'll chat with the main orchestrator, yes, and I'll just do what team members do you have.

And you got the response.

So this was first question, second question, I'll just copy this.

I have delivery issues with Flipkart. What do I need to do?

I'll come back here, paste it and if you see you have something called as behind the scenes. Now what is it trying to do? Delegate task to member and this is tool calls.

Tool calls.

And if you see member response will be there and what member was used customer agent.

And if you see here it will tell what is the time, what was the result from the I mean tool call. This is very solid UI.

And then you have the final response.

Yeah.

So I'll repeat again here. After your name, click on this. You don't have to give name.

Click on this plus icon. Local paste your URL name can be anything tags. What tags you want that particular tag to be known as which is optional. We can keep it.

It anything that is relevant. Once you do that, this is how your screen will look like.

You'll have both your agent, you will have your teams, and once you use your



knowledge base you can also configure your knowledge base here. Is this clear?

That's very good.

We don't talk on that in this.

So you are like this here, right? How many of you got the response?

I I got mine. You got the response. No, I haven't tried. Do we use main orchestrator?

Main orchestrator? OK.

What is the question? What is? Are you running it? Can you come back to your code?

Come back. What did you connect? Click on that.

Add added. It should be bad customer service. There should be no space.

Copy or this thing? How many of you are getting errors?

Uh, Ramakrishna, did you get the response?



**RamKrishna Bhatt** 53:55

Actually I will not able to run in VS code. I joined from the company laptop so I can't get that I will not able to run in VS code. I joined from the company laptop so if it's Python is not the.



**Tarun Jain** 53:57

Oh, what happened? I can't hear his voice. Hello.



**RamKrishna Bhatt** 54:10

Are you able to hear me now?



**Tarun Jain** 54:11

Yeah, now I can hear you.



**RamKrishna Bhatt** 54:13

OK, so I joined from the company laptop so in which there is no Python installed so.



**Tarun Jain** 54:20

Oh, OK.



**RamKrishna Bhatt** 54:21

I will not able to run that in VS code.

**TJ** **Tarun Jain** 54:22

Oh, there's this part. OK, Mitesh, copy this URL. Yeah, I did.

Open in a new screen, paste it. You should see this thing. So whatever ID you are given, what is the ID? Our ID is router agent. Then you have description and then you have name. Are you able to see this? Yeah, yeah. So now this is working. OK.

Come back here, click on this. Now if you click on this, edit local page that you are a local restaurant service without space and then you will see this green one. It's not working for me. No, it's good.

Uh, we have to change this. I just added this in the middle. I did this. Can you change this?

Uh, where is agent or is different? Agent or is different? That's why I'm all that work. OK, where you always speaks for me here. Oh, because this is in. OK, so I'm doing it inside me and with poetry. I'm doing it with uh teams one.

Let me go to 18 documents and 18 documents, right? So if we have to do it at the background, active is not greatly defined. So if we do it like this, it won't. It won't have a of scope. I don't need any of it. I will go for that because of poetry's qualification.

See. Yeah. And then I do. I do it for the one like version.

Yes, I got the answer.

In in a man modified.

OK, I'll do it. I put anything. I you're not paying anything. Should I install those back in there? OK, there, there.

Yeah, yeah, yeah. But usually we don't install.

Do you work it?

We also want to be a sponsor. Thank you. We're coming to agent too, but when the opportunity with you type in the.

You can't pop of this.

Yes, but you get to my.

I thought she's often a man. Why is it happening? Like you have a like getting it out. She's at. She's at. She's at. No, no, no. OK, this is working.

So I get a. OK, that is it. Yeah. I need you. I'll see with this.

OK, now that's a long before it. What's this?

I want you to know. I'll have you. I'll have a.

It's a.

That's a.

What's big?

No exit. I remember I was presented the response.

Maybe it's because OK, let's take this.

Uh, if you have anything with.

I mean the way you have defined it. Oh, or OS dot agent OS. That's right. OK.

This. Yes, that's what the available added. So what the response the delegation tool just verify the delegation part.

I'll be. We'll get close to those who already got the response. Start with the blog writing this thing. So what you're supposed to do is define a research line. Now the what is the task of researcher to gather all the SEO keywords?

Now this researcher needs to give the information to writer. Then you need to have verifier to write the final article. So how many agents you need to create? Three agents, researcher, writer, verifier and verifier should give your final output.

When you define our orchestrator, you should be like very specific. First use research, then use writer, then use verifier column. So for verifier, if you have any rubrics, you can define it. If you don't want, you can. So you can start with that part.

Thanks. Can you come back to the second option?

And what? Agency team.

It should use one of the one of the agent.

And I can when I got the I install.



**RamKrishna Bhatt** 1:02:24

I don't.



**Tarun Jain** 1:02:26

No. So that's it, buddy. We can sorry.



**RamKrishna Bhatt** 1:02:37

Tarun, are you?

Are you able to hear me?



**Tarun Jain** 1:02:45

Hi.

Hello, Adam.



**RamKrishna Bhatt** 1:02:49

Oh, OK.

Tarun, are you able to hear me?

No, I can share my screen so I have switched the laptop so.



**Tarun Jain** 1:03:15

What? Hello.



**RamKrishna Bhatt** 1:03:16

If you are free, I can share my screen and I just have some doubt. I switch my laptop.



**Tarun Jain** 1:03:21

Yeah, yeah, yeah.

Yeah, I'll stop sharing. You can share.



**RamKrishna Bhatt** 1:03:26

OK.

So like I already have ENV created and I have downloaded that Python file and just paste it over here.

I have created ENV file as well. So what else do I have to change to run this in my locals?



**Tarun Jain** 1:03:59

No, the ENV is wrong. So can you come back to ENV? So you should never give this thing E what you call codes. So Google API key there should be no space. Remove the space as well.



**RamKrishna Bhatt** 1:04:10

OK.

Sure.



**Tarun Jain** 1:04:16

Same do it for all the things.

Uh, now save it. Save controllers.



**RamKrishna Bhatt** 1:04:28

OK.  
Yeah.



**Tarun Jain** 1:04:32

So dot you know this is when can you Scroll down. You have to import agent OS where have you done that?



**RamKrishna Bhatt** 1:04:39

Uh.



**Tarun Jain** 1:04:40

So this is the code I've sent it in the so can you come to the chat?



**RamKrishna Bhatt** 1:04:48

Sure.



**Tarun Jain** 1:04:51

So you have this code, uh, the second code, huh?  
Agent was from agent was, yeah, from there only from agent was.



**RamKrishna Bhatt** 1:04:57

OK.



**Tarun Jain** 1:05:02

Copy this at bottom.  
So what you can do is you can remove this worker dot print response 60 to 69 you can remove 60 to 70.



**RamKrishna Bhatt** 1:05:17

OK, this 60 to 70 with this also. OK, got it.



**Tarun Jain** 1:05:19

Oh.

Huh.

Scroll up, scroll up. Now save it.

Now try to run this.



**RamKrishna Bhatt** 1:05:32

These imports are fine, right? From this they already. OK, I have installed that everything.



**TJ Tarun Jain** 1:05:35

We'll try to fix this.

Once again, can you come back to your code?



**RamKrishna Bhatt** 1:05:43

Good.



**TJ Tarun Jain** 1:05:44

Scroll down full bottom.

So here multi agent code is my file name. Your file name is template LLMMMA Agno.

You got it. Yeah. This is your final.



**RamKrishna Bhatt** 1:06:00

Yeah, yeah, yeah, this one, right.

OK.



**TJ Tarun Jain** 1:06:05

And.

Yeah, now London. So guys, if anyone is done, start with the blog.



**RamKrishna Bhatt** 1:06:09

OK.



**TJ Tarun Jain** 1:06:20

Oh, you have to do PIP install fast API. What happened?



**RamKrishna Bhatt** 1:06:20

Acknowledging what?



**TJ Tarun Jain** 1:06:35

And also install UV con, UV con.  
Space.  
2.



**RamKrishna Bhatt** 1:06:45

You, is that correct?



**TJ Tarun Jain** 1:06:47

M.  
RNRNRN for technology, CORN, but it should no branding. Yes, branding not include anything related to development. We are now run it.



**RamKrishna Bhatt** 1:06:55

CORN OK UVICURN OK.



**TJ Tarun Jain** 1:07:05

No, but not anything will do. Happy pinch all Google.



**RamKrishna Bhatt** 1:07:08

Google Gemini not found.  
Sure.



**TJ Tarun Jain** 1:07:21

OK.



**RamKrishna Bhatt** 1:07:24

OK.



**TJ Tarun Jain** 1:07:31

Now run it again.

OK, I guess you have not added this. I'm I'm sending you a line. Can you copy that?



**RamKrishna Bhatt** 1:07:38

It.

Sure.



**Tarun Jain** 1:07:49

Oh.

Yeah, I just copy this line. I'm sending that in chat.



**RamKrishna Bhatt** 1:07:58

OK.



**Tarun Jain** 1:08:01

After this you can turn it.



**RamKrishna Bhatt** 1:08:03

Yeah, I think I have done this. You have sent this, right?



**Tarun Jain** 1:08:08

No, it's probably trying to do it again.



**RamKrishna Bhatt** 1:08:11

OK.



**Tarun Jain** 1:08:18

Uh, now try to run it.

It's.



**RamKrishna Bhatt** 1:08:26

OK.



**Tarun Jain** 1:08:27

Cool. So copy this [os.agno.com](https://os.agno.com).





**RamKrishna Bhatt** 1:08:30

OK.

I'm sorry.



**Tarun Jain** 1:08:39

Os.iagno.com.

Voice.



**RamKrishna Bhatt** 1:08:44

Yeah.



**Tarun Jain** 1:08:47

You run two times, yeah, Agno. So this one we can copy.



**RamKrishna Bhatt** 1:08:51

No, yeah, I just by mistake I just hit control C so that got stopped.



**Tarun Jain** 1:08:56

Oh, OK.

So you have to sign in first.



**RamKrishna Bhatt** 1:09:01

OK.

Is it fine with the Google right?



**Tarun Jain** 1:09:19

Or change current load.

You didn't get output yet.

For uh assembling for us then it would just work out-of-the-box. So now what you can do is click on the tag agent OS.



**RamKrishna Bhatt** 1:09:39

Mhm.



**Tarun Jain** 1:09:40

And paste your this thing the API endpoint.



**RamKrishna Bhatt** 1:09:47

Or this way.



**Tarun Jain** 1:09:48

Local host. So I hope you understood the difference between this. One is your UI and one more is your fast API endpoint. So fast API endpoint. If you want to verify always just use back slash docs. If you get swagger UI that means that is fast API.



**RamKrishna Bhatt** 1:09:50

OK.

OK.



**Tarun Jain** 1:10:06

And name you can give anything.



**RamKrishna Bhatt** 1:10:06

Oh, which means if I hit, if I hit over here and then.



**Tarun Jain** 1:10:10

Enter.

That is fine, you'll get this. So if you do back slash and then write docs.



**RamKrishna Bhatt** 1:10:15

OK.

Mhm.



**Tarun Jain** 1:10:19

Docs, huh. So you'll get this UI. Now this is API configuration, so come back to. So name can be anything. You can keep customer service or anything.



**RamKrishna Bhatt** 1:10:22

Got it.  
OK, so.

 **Tarun Jain** 1:10:35

Tags it is optional. Now connect. This can be anything. If not, you can also click on connect.

 **RamKrishna Bhatt** 1:10:39

OK.  
So what? What will be the use case of this text?  
OK.  
OK.

 **Tarun Jain** 1:10:51

So now if you see you have both the agents that you can chat if you want to do it individually, but if you want to chat with directly main orchestrator, you can use that.

 **RamKrishna Bhatt** 1:10:56

OK.  
OK. And what do we have to like, what is this task which we have to do now?

 **Tarun Jain** 1:11:03

So if you have anything related to customer service, you can answer question. It will redirect to that. If it is anything related to strategy and marketing, it will route to marketing agent. That's it.

 **RamKrishna Bhatt** 1:11:14

OK.  
OK.

 **Tarun Jain** 1:11:21

Cool. So now we can start with that blog part.

 **RamKrishna Bhatt** 1:11:22

Got it. Thanks.

OK.

**TJ** **Tarun Jain** 1:11:34

Yeah, does anyone want to try Langraph instead of for this thing? No.

Because I land graph is also same. If you see you just have to define create react agent, define your model, define your tools, define the prompt and name. That's it.

And instead of team you have OK, my screen is not shared.

So this is the Langra code. I'm sharing it in the chat.

Um.

Research a brighter and verified.

The thing is that I thought was.

Mhm.

So instead of team, you have supervisor agent in Langra. Probably we'll discuss this once it's done.

Thank you.

I think that what's the it's it's still core server only that is on error that I don't know. I know about the core server but The thing is like it should have been solved with this port.

It isn't. It is something fasting there. It is. It is what is, uh, I mean, what is this?

Uh, you said they said that you do this only what I did. Mm-hmm.

It's a listed that only with four. No, there's a different one that it's the report three.

Yeah, yeah, it is this one. Oh, if you run directly at this.

And later when it's a late finish.

So if I do it with that with the course and everything, it will open topic space.

OK, import is also not working. Can you? No, this works. It is just because it is pause.

Calling it from here won't work. Otherwise it is working. So to say which domain should be allowed to, you know, do this.

It could be something between you and UV con and for it.

Could be. I can use a lot, use a lot separately. What? Plot, plot, not yet. Paste your poetry dot this thing and also I'll see it. I'll see and then.

See what is there, OK?

OK.

Yes.

Domain or is performance, main or is from. So yesterday we discussed we have to

build a Blogger agent. So you'll have to create three agent. One is researcher, second writer.



**RamKrishna Bhatt** 1:14:56

OK, so Taran, what do we have to do now? So like, what is the task?



**Tarun Jain** 1:15:09

Writer doesn't have to have any tools, then verifier and then you need to have a team. In team you have to define instruction on how your entire article needs to be structured as.



**RamKrishna Bhatt** 1:15:13

OK.

OK, got it. I think yesterday I left early, so I was I didn't get time to revise the session, so.



**Tarun Jain** 1:15:25

Yeah.

OK, so three agent. For every three agent you'll have prompts and as well as role and instruction very detailed.



**RamKrishna Bhatt** 1:15:37

OK.



**Tarun Jain** 1:15:39

And when you define your main instruction right for everyone, make sure it's not telling to only pick researcher nor writer. It should be sequential. So that instruction needs to be given like researcher, writer, verifier. That is our main instruction for.

If you want to use full lab, we can use full lab. If you directly want to test it on UI, you understand.

We can do it on you.

What you is much simple.

And I will also show a showcase the code on how to use Landlab.

Same as, but The thing is names are different.

It's on middleware. The UI will go for private network. It will work fine and then try.

What did it do? If that works today, people will buy cloud subscription. Yeah, I can.

Yeah. Oh, OK then fine.

It's going not to have like.

All right.

No.

I just need it to run.

So once this is done we will have only human in a loop pending for multi agent systems. So human in a loop memory and MCP.

OK, to generate uh, do we also need to write agent from scratch?

What conference? Agent from scratch. Agent from. So it's not using. OK, 0 libraries.

Pure Python. No. Nothing. No. Pure agent. You want to write agent, pure pure Python. It's like tool calling, tool calling and.

And react. So we're calling in the end, right? So that I'll send the code. It's very simple.

OK.

OK to write so you want you have this thing right Matt? So tomorrow try to download cloud desktop. I will do MCP.

Yes, tomorrow, what time are we meeting? I'll check on flight now.

There is no clown test.

At least in the group it is fine like I'll give you another name or the same name. It's fine already completed you can do different.

You can keep Blogger Blogger, this Blogger.

How we can keep same time 11:00 to 12:30?

What time is? What time do you have to eat?

4:00 is light so.

What is your input? OK to write from with.

But verify it because most of the time prompt is not well drafted.

Uh huh. OK, go to work. Cloud work with poetry, with poetry, with that. So it created a middleware for adding one more header called excess control allow private network to through.

So I think what should have done that as well then?

See thinking, but OK, you know, that's Claude TJ.

And it becomes so easy, you know, when you are defending cloud in event. We are connect. If Jaminay is not talking, I'll go ask cloud connect in US, right? But you have to stop that and rerun it again. First you have to run this.

So if I go and chat with teams, I have a now go and create new refresh, refresh, refresh.

I'm new, yeah, yeah.

Delegate task to miss work.

Customer service assistant member ID the the user on access should give task as well.

Uh, it gives task as well. Uh, that is tool call.

And it gives expected output. Huh. What is expected output?

So that is part of agent call. Agent also will generate a response because here multi agent in the sense one agent has to give response to the HR. So treat main orchestrator as HR. So now HR is going and talking to a customer chatbot. So customer chatbot will give you output and HR will give the output.

To the customer. OK, so you have two outputs. If I give DB to this, it will just save all the convolution. We'll or else let's add DB, but let's complete BLOB.

Block is pending. I just he completed Vitesh. It's already too close. Oh my chat. I I'm not but but Nitesh should do one more project. He was he was asking everything to me and.

What do you call it? Ardip. OK, what is? And what is turning? Write a article on MCP.

We need what? Block generation? In that we need how many characters or teams?

It is three OK researcher, writer and verifier.

Yeah, if MCP full form is returned as something else upon from model context protocol or article is wrong.

Yeah.

Next 3 minutes. Take this. I mean very helpful. I even he come, he got the.

That works very good, but with very likely that won't show. OK, your prompt is not correct then for evaluator. What is the?

Coordinator forward some no no. Where is evaluator prompt? You can organize a out of the remarks come back to this thing.

Come back. Do you see any thoughts and feedback?

No test complete.

So click on this.

This is writer agent.

And this one is this.

OK, so now what do I have to do? I have to write a blog, right? You could make agent write a blog, anything which is trending, which is trending acknowledging a

company.

On we should not write any grammatical code for only focus on philosophy and targeting the Internet. Give me the events. I got done. Give the events.

No, this is not the output. It's like a researcher. No, this is wrong. So it's like it should go to what you call it.

First you should go to researcher, get the information, go to writer, get the information, then give to her. So he has written good points. So Madhav, the main instruction, you have to be very specific that researcher.

Hmm. So that information is to be in the main instruction. Like what we we uh, right there, right there. Now what should uh your inspector do? Gather all the main keywords. What are the submissions? What was what's not product?

So research researcher task is how should your outline of blog look like? You can ask anything. So now that's the follow up question saying a benefit of using MCP.

It said that.

No, no. So just tell for the other vertical you you can you can you can't OK there is no DB so you have to ask the question again huh.

For the first response key request in research papers.

So we give a blog for your agent.

It only was writer service. It didn't do a search. OK, it reviewed. OK, it reviewed it got. I got the blog, but did you do the Apple blog as well?

I didn't give. I gave the prompt that Ravi gave me that I want to create a block for the technology company, but it should not be little programming. It should only be philosophical and it should be targeting information, prospect and international and context being Indian.

I.

Yeah, always you can say somebody's think, give somebody right. So researcher should summarize all it's learning, then give it to a writer as instead of giving garbage. I would answer all the agent, yeah.

So, OK, we started using writer. Yes, we started. We will get a writer.

It's not so out of these three.

Up now, speaking. First use a researcher agent, give all this information, then use, then use.

When you verify agent, you have to define grouplets featured. He has defined certain grouplets. Now is not good. You will take agent to improve the structure.

But The thing is, memory is missing. We did our memory too. I got with as a codes



out of the. So what did you change? Just prompt me like the last. But that should be hard coded or no instructions.

GPA user data. So our coding should be done on port level, not here the main installation, right? Yeah, the main installation. So let's also add DB so that you can tell what happened earlier and you can follow up on it so you can use SQL Lite.

The only thing pending is now the feature leakage. What? For balance? For blocked? Yeah, you can use. I'll get this thing on. You have this thing, right? GPT file now.

OK, so let's add this multimodal everyone of a we have used right? So in multimodals you have images, so you have image model output so you can define that in one of the agent.

So it's like Gemini, Gemini 2.0 flash image generation. So I've add one more agent, one more agent for image generation and then tell make an image. So see here make an image then give one agent output.

And then you can copy this code OK.

So this is for if you want to add image as an output you can use directly multimodal.

So every time if you want to use anything apart from text that is multimodal. So text as an image as an input, image as an output. So if image is an input then what are we supposed to use from agno dot media input image?

Then define URL or file path. If you want to use audio you can use audio. Again same audio is input, audio is output. So if it is as an input you have to do from agno dot media input audio and then you have video. We can also use cursor and Open AI tools.

Huh. You can. So for open here, it will be Dali. Yeah, yeah.

Event generation but now we are doing the blocking the event generator. OK here he is also using GPT 5 meaning in this example and if you see just use this this line from agno dot DB dot SQL lite import SQL lite.

Then just define some path where you have to save your data and add this new parameter called add history to context to be true.

So we are adding one line, then two line, then three. This are the three new lines context. So one line is this SQL IDB and then we're doing DB OK you are.

Agent only. Are we going to only or this we can define this for team. No what you say define these two teams. Are you followed Ramakrishna? What you can do is everyone here are trying the blog agent.

Can anyone push the block agent to you can add your code was good. So Ayush will add the code. What you can do is in the teams we will add memory. So from agno

dot DB dot SQLite we have to use SQLite DB.

DB then DB dot SQLite. We are adding some path then add history to context. So what kind of memory is this? This is a long term memory.

Short term memory. This is long term memory. You're adding it to the SQL, you're only saving the session. So next time if you ask anything related to what you have already done after five days, that's gone. Preferences is not showed.

So now let's suppose you're creating new session and then you're asking any question. Your DB is gone.

So where do you add this under which agent? Researcher by teams?

OK, so if I click on teams and I'll have team session, you need SQL alchemy, SQL alchemy. So if you see that this is the syntax you have team, it's here the syntax for team. You are defining your SQLite any path then here.

You see you're defining DB equals to DB inside team. Thank you.

And make sure you also do that. For Azure also we need to huh. For Azure you also we need you need 2 parameters, Azure endpoint and Azure API key. And if your deployment is different, you have to define your deployment instead of modern name.

So 33 lines we need to add. One is SQLite, then this thing, then DB equals to DB and add add history. What was that?

It requires opening APIs. Where did we add that place? So add district, add district to context.

So let me add this so that it will not confuse. So I'll copy this line. I'll come back to DB. So where is my teams? So here I'll add this parameter.

And then uh DB equals SQL right?

So now these are the two new lines I've added and then copy this SQLite. That's it.

Is it clear? One is import statement and then these are the two new lines. OK, I've directly assigned it. OK, so these are the two new lines and this is import statement.

So import statement. I'll copy it, I'll add it here, but this should be the first time.

This should be your the starting few lines and what is the PIP install?

What for SQL it? Let's see if it use. OK, this is wrong. SQL I'll come in. Yes, this is on.

We need to discuss job.

So don't add this comment. This comment is just to make sure what import statement to add.

For that, but it is we're doing a custom pooling for it. So if you want to add DB for a specific agent, you can also add that and then test it separately just to see if it is

working or not.

And your open AI does not have the which model did you pick? I forum mini doesn't have no I I took for that. So here you need three things. One is Azure endpoint, then Azure API key and then deployment name.

Right. I have all those. No, I have all those set up. It's so we have only the endpoint in the I can, right? And now we are close to blocks.

It is using open EI tools, but it's not open EI here, no. So here instead of open EI, now what do we do here? That's it. It does not. Wait, wait. Once again, it's moving through a lot.

Oops, get get tool and model.

OK, yeah, probably Azure does not have it now.

Yeah.

So I'll send you the the model name is we can use.

So the model name for Nano is Gemini 2.5.

Gemini 2.5 Flash image.

So try to do one more thing, tell you to define the dimensions as well, because in blog you will need a four is to 312, not sorry 19.

It should be correct because by default it will do one is to one which is not a good image for block.

And second improvement. What we can try to do is for your writer agent define, add additional input, add additional context given block that you have written to copy this writing style.

That can move to add additional context.

And then you should tell don't verbose at the end.

So don't verbose and I've given you the my writing style use that. It can go in instruction or it can go in add additional context so that whatever blog you generate it won't be like Ki generated.

You can copy anyone who is a prior writer. So now let's suppose if you go to AI guys, AI guys, medium, these are PR authentic writers, whoever you see in AI guys.

Sure. And here I'll open Freedium, Freedium CFV. Why is it blocking?

Yeah, England.

Oh.

Freedom is blocked here.

OK.

If it is LS.

Like might be blocked. Something might be blocked. OK, yeah, I guess I actually these articles are paid, so usually I bypass this by freedom, but then it it's blacklisted. OK, so then no, I still have you're connected to a right now.

So you don't have our DNS result, we can go back out yet. OK, so if you're using our you can copy my articles. I don't use the content.

I'll put copyright claims if I see on the.

Don't know. Where is my? We'll use the.

But he'll be very good. I copy Varuma. What do you do? He's every on here so you are my copies. Uh so if you see my videos, I copy start starting. It's like full then B roll. This is good. He is also very.

He's crazy. He's my he's waste.

So you everyone tried the DV part as well? Just trying that.

If it's done, then we have you money in a loaf MCP and long term memory for tomorrow.

Yeah.

Mhm.

So for RAG project either you can use past API plus react for your demo, but if in case you want to use any SQL, I mean what do you call it stream, you can stream it for that report generation is a right use case.

You can use fast. This won't be useful. This you want to use agent plus your solution.

Yes then you can use this UI. But I was thinking to have two different UIs. So for agent use case you can use this thing.

But for the older one, report generation, you can have fast API. Just create fast API endpoints. Then just let cursor to add react code. If not cloud cloud code based, right? Right. That is all that we can do or fast API also it's not required. Just tell cloud it will also write fast API endpoints. It might not require that.

I don't require that much of work because you know it's already a process. It's only the content is coming and pulling and so it it might be a wrong job with this. I'll figure that out how I can do this properly, but React will be very better.

Computer stream it.

All my sessions, if it is recorded private sessions, I'll be banned. I told don't do stream data.

Last time, I guess when I give the first talk here only, I gave radio. Radio was competitor upstream. Yeah, yeah, after that we decided that we will. But it was greedy. I remember radio was.

Radio is competitive of, but both are open source.

Streamlit Snowflake is Streamlit, then Gradyo as a gameface. Snowflake later, but Streamlit is part of. I mean Snowflake purchased Streamlit and a gameface purchased Gradyo.

Can we see the history in this? Like how many tools were called? Yeah, so you have to add DB and then you have session at the side. Yeah, on right hand side you have the yeah, but it will show like which tools were called then.

I'm not sure. We have to click on it. Can you show that? I got it. Click on this. What is this? OK.

OK, this is the new one, yeah.

It's what a pyrogen's and a pics.

You probably it might be. It didn't get most of the yeah, correct us. Yeah, I got a question. Where are you? I use and agent and.

Yeah, yeah.

What do you sketch? So what is actually write a blog post to correlate and the agent and write a blog post about what do you sketch of?

So that can take all the problems, everything is there. We have a long question. We have something about hotly. We search on getting.

What is Gemina LLM calls we are using?

Oh, can you come back to your listing VS for something stop?

Uh, something was wrong here.

Click on memory. You use memory, right? The model is overloading that is we are using Gemini Pro for every this thing.

So switch back to Flash Flash and the last thing is there, right? Then you use Flashlight. Oh, section of memory. Memory. You know you have, right? Yeah, OK.

Why is it fail me something? I order my tool or write a image generation.

So for output we need to face report. So I did the the image generation. I used the which has a generation. It doesn't have image generation. I'm using this tool for image generation rule and.

So did you define that in main instruction like? Yes, yes. So this is the instruction.

Once the blog is verified, then send it to the image generator to create a featured image. Ensure the workflow is followed. Make sure you have at least three images. Me.

Uh, creators with your image and uh created for two feature images to create two feature images. OK, then when will it attach this like?

It it didn't called so this is the flow so the data research and writer and verifier.

Would then even delegate the task?

Probably once we have to test that agent separately if it is generating or not. OK, good. What is this not for? This is what you make. Oh, you make on standard, which is fine. So if we come back to the what do you call it documentation.

No. OK, yeah, this one will be us chat. So create an image for MCP.

Model working of.

OK, so you have got it. There was no model image from this link. It will be image, yes. I guess this is very.

We should then don't ask question.

So what tool was it? Yeah, it's storage from tools. I got that saved in DB the models.

I know how to stop it. I mean the future is already here. We want to stop. It isn't complete. We complete log creation team complete.

Click on this memory name name memory.

Refresh complete refresh memory. Can you come back to LSLS? Do you have it or not?

What happened?

So yeah, teams, probably something broke in between.

But then next time if you refresh, stop and refresh it again, it should be there teams.

So tools.

I will get.

Oh, this is the one right that you are using. Mm-hmm. Azure Open AI tools, right?

That is what I.

This one you have used. I like the deployer over it. I have to get this no no actually the same deployment name you can use. But what they've given is let's suppose if you define Azure LLM and.

Azure tools at same time this one Azure what you call Azure LLM deployment name might conflict with this tool so they have kept it as different.

How do we know if it is work? Huh. The complete. It's gonna be I guess we'll because the like complete it will tell the your agent is completed last line.

I'm getting like you will get a run completed run completed. I'm getting please proceed with evaluation.

You said it or did agent? No, I said nothing. Then it is executing the valid.

Um.

Yeah.

So before you run that, what what you call with teams, try it separately.

And also add 1 instruction saying that don't ask the following question directly generate.

For for Azure Open AI, it has clarifying question like what kind of image you need and you also have one more parameter called image quality. It is standard, you can keep it HD.

OK.

I can do that.

I guess it supports only Dali. It has a very good uh error. What's this? They spend two months for this.

I am stuck at this. Please proceed with your.

Is it showing 3 dot dot dot?

Read dot dot dot.

It is not show memory.

Yeah, I'm not getting, but here in the TV next.

Memory knowledge. Oh, you got the output? Yeah, Ramakrishna.

Hello.



**RamKrishna Bhatt** 1:53:11

No, I I am not following that. I have to review a stress session first.



**TJ Tarun Jain** 1:53:18

Can you give the code? Yeah, I already yesterday. OK, not the but the OK can you did you try this code? Yeah multiagent dot blog dot.



**RamKrishna Bhatt** 1:53:21

Sorry.

Yeah yeah I I I try I tried this code and like in local it is working fine but while while try to chat with it I am getting something went wrong error so.



**TJ Tarun Jain** 1:53:37

Yeah.

OK, you want to share your screen? Yeah, yeah.



**RamKrishna Bhatt** 1:53:41

Sure, sure, yes.

I haven't tried a DB part, but I just yeah, I have just take this code and run it in my local.

 **Tarun Jain** 1:53:50

No, DB fine. It's fine. You can ignore.  
Your screen is blurred.

 **RamKrishna Bhatt** 1:54:01

OK, let me share it again.

 **Tarun Jain** 1:54:09

How far are you?

So change writer agent and add 1 parameter called add additional context or copy some articles until copy this writing style.

It's still blood.

 **RamKrishna Bhatt** 1:54:25

OK, let me share my another screen then.

 **Tarun Jain** 1:54:30

Just share the tab, if not only tab.

 **RamKrishna Bhatt** 1:54:30

\*\*\*\*.

It is visible now.

 **Tarun Jain** 1:54:38

Now it's clear. Now it's clear. Can you Scroll down the call?

 **RamKrishna Bhatt** 1:54:40

OK, OK.

Yeah, I have changed the name.

 **Tarun Jain** 1:54:46



Huh. So you just ask one question saying what was your listing for a prompt. Now I'm asking ayush prompt. Write a blog post on one second. One second. Just ask one question. Write A blog post on serverless functions.



**RamKrishna Bhatt** 1:54:52

Sorry.  
OK, OK.



**Tarun Jain** 1:55:02

Serverless function.



**RamKrishna Bhatt** 1:55:07

Server less.



**Tarun Jain** 1:55:09

Functions. Functions. Yeah, that's it. Enter.  
So now this should first redirect to researcher. Hi, it's working.



**RamKrishna Bhatt** 1:55:23

Yeah, but I asked to write a blog on a like a quantum computing. Then it is giving me an error.



**Tarun Jain** 1:55:34

Oh, what did you ask?



**RamKrishna Bhatt** 1:55:36

Write a blog on quantum computing.



**Tarun Jain** 1:55:40

Quantum computing? Oh, that should also work, so you can just refresh and ask once.



**RamKrishna Bhatt** 1:55:46

Sure this was loading and after after some time sorry.

**TJ Tarun Jain** 1:55:48

Can you click on this number response?

Member Can you click on this member responses? So if you see our first it will go to researcher agent. It will gather all the information. Now this you have article. Now this will go to what you call writer agent. Then you will have your final output.

 **RamKrishna Bhatt** 1:55:54

Sure.

Mhm.

**TJ Tarun Jain** 1:56:10

Hello.

 **RamKrishna Bhatt** 1:56:10

OK.

**TJ Tarun Jain** 1:56:13

So you can try any this thing any article, but make sure if you give any short form right like write article on MCP it might take some wrong results. So give full form like model context protocol.

 **RamKrishna Bhatt** 1:56:20

M.

OK, got it.

Yeah, I got the output.

**TJ Tarun Jain** 1:56:33

Yeah.

 **RamKrishna Bhatt** 1:56:35

OK. Thanks.

**TJ Tarun Jain** 1:56:45

It should be. I didn't turn it memory. Should I end the meet? Yeah.

**TJ** **Tarun Jain** 1:58:45

Teen agent so it takes time. That's fine. One second I'll share my screen. All right, so we start with the from langra dot prebuilt import create react agent and then you have tably. So this one you can ignore the agent state. So these are three import statements.

So instead of agent, you know you have create react agent. Then instead of agno dot tools you have tab research. So this is a tool, agent and LLM. These are the three base components that you use when you define an agent.

And once this is done, you have simple how to define it. So create React agent, give a model which is LLM, then tools within search tool, then prompt is prompt.

Right, so now this is how you define a single agent. But if you want to define multiple agent, I told you there are different ways you can communicate, right? Supervisor agent is nothing but supervisor. It's similar to teams, but I'll show the workflow.

Where is the workflow?

So you have this one single agent. Single agent is nothing but LLM plus tools. The network is nothing but you're coordinating among yourself. Blog is kind of network. Then you have supervisor. So if you look at supervisor.

This is routing, so HR is coordinating with customer, HR is coordinating with marketer. Then you have hierarchal. Hierarchal is very messy so you can connect one agent to any other agent and then custom if in case you want to make any.

Huh. I RKL is kind of team to team. So if you see this is 1 entire team, this is 1 entire team. How would one just import statements. So here you have a hierarchical.

So if you see team one, team one supervisor, team one agent, team two agent, then add note.

Graph APIs, huh?

So now I'm going with prebuilt. So in prebuilt if you see you have a supervisor agent who is coordinating with order agent, file agent or any other agent. But in our case it is customer service, marketing service or tech service and open AI have something called a swarm.

So Swarm is nothing, but it's like a circular thing. So you're is this not very slow like it really happening there? Which one? This one? Swarm. Yeah, yeah, agents are always slow faster than humans, but still it's faster than humans. If you see blog article, you can't write so fast.

But now this second question is like if you just ask this question to LLM, LLM is faster

than agent, which is true because LLM is just doing function calling agent is adding the decision but the decision is it working or not working?

That's like, uh, you can't even guess. It's luck with agents. It's pure luck.

But can we add the control layer? As of now we don't have the valid response, we just add human in a loop. So if something is not happening properly, just have a human saying that if this is right or wrong.

So what we're trying to do now is we're using the supervisor agent.

I'll come back here. So define land graph supervisor agent. Same thing. You need 2 prompts, customer prompt, marketing prompt. So if you see this, this is similar to what we have already used before. Now define create react agent. You have model tools prompt name. That's it.

Same goes for marketing agent. Then just define supervisor agent and when it comes to land graph, once you have a graph, what are you supposed to do next? Compile.

So you just run compile. That's it.

And when you invoke, make sure you define messages.

Messages is your input variable for React agent.

And then if you see response of messages of second, so how does this work? You have a query you're asking an agent, it's doing a tool call. So when you do tool call, what is the model response?

So if I open Excel it wrong.

Where did we do that? OK.

Gemini LLM tool call diagram.

7.

So here if you notice you have a prompt, you're going to function definition and when when you're using the LLM, LLM is defining your function name and arguments, right? So this is happening on the function tools params. During this time, what is the output of the final LLM? It is empty.

That is empty. So now you have this function M and arguments. This is going to tool.

Tool will generate the context. So if you see here when I've defined response of messages as two, that means zeroth index.

Nothing is happening. First index you're doing the tool call. In the second index you have the action name that successfully you got the information from marketer service and if you see a name is transferred to market service, so it's transferring to a specific agent.

And last, which is your final index, you have your content.

So this is for identifying what tool was used and this is to identify the final response. Same goes for a different question. Here it is transferring to customer service and then this is the final response.

But this is hard coded in Agno. Agno is directly giving you a nice UI as well. So even when you do print response, it will give you that UI. So that is called PP print.

So this is a Python package again which will prettify the way you print your outputs. So this is by default in Agno.

And that library is PPP print, that's it.

And this is how you define in Langgraph.

Same goes for what you call LAMA index. Define agent, then instantiate it. I know it's fast, but I'm not seeing I know any case study like anyone actually using it like you're not going to use the UI. You already have the.

But I know it adds too much of wrappers.

I think that's, but that is simply to be. But if there is no latency issue, I know is best.

Because it's very easy to use, syntax is saved.

I'll schedule things and I don't need it to be easily at the right then huh then fine.

Then if in case I'm not doing demo to any anyone else, Agno works best because Agno has too many tools which is not even there in Landgraf.

Multimodal is 1 feature. Thank you. No, we'll build agent from scratch.

Really.

What was I saying? Lama index, Lama index agent.

Why do you know? OK, I'll show this last one. I'll I'll also show you my next one.

So for index we use something called as function agent. You are defining overflow then OK this will take tomorrow query engine in the sense this is agent. But The thing is this is how you define an agent. That's it 11 to 12 tomorrow.

Huh. OK, so you have agent function, agent tools, LLM and system prompt. That's it. These are common in every frameworks. If you use Google ADK, Google ADK is also same. What you define agent, you define name LLM and then you have prompt. So all frameworks are doing same thing.

The only one which is different is. So in this whole now of all of us, everybody is sending everything's done. We have the agenda agenda. Han, same. I've summarized here. Memory is pending. OK.

MCP is pending and is pending only this three. OK and this fine tuning. OK, this was like an optional part, but yeah.

OK, I think we're almost there now. These are the two pending topics. Agent also we

have already done. So whatever tools you have right tool, custom tool search, we define def search that itself is a tool. So we are not writing new code.

So we are just taking that search, then giving that as a tool. So rag is a tool now.

Yeah, so whatever that depth search is there, right? That search is now your tool.

That's it. You just have to add docstring apart from that entire code.

That is agent required where your knowledge base is a tool to an agent. Knowledge base is your vector DB and whenever you have to search from it. So right now I'm doing a lot of laborious work when I'm doing with this quality search and everything with this.

I just have to tell proper instruction and we go and do it. It will go to tool but if in case you need search it will pick search right? So understood that search is 1 tool knowledge base issue whenever whichever is required what I'm doing with right now. Because I don't need the latency partner, so I'll do it with it is so you can also search for agentic Agno.

That's it. Even we are wondering. So again, just search for agentic chunking in Agno.

That's it.

That is also just one line code. OK and then we have to look at this. I don't think what agent sub king equals root room. I guess it's boolean hybrid search. No hybrid search is different. Sub king is different.

Thank you.

● **Ajay Patel** stopped transcription