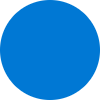
**Transcript**

August 5, 2025, 10:07PM

0:14  
OK.

 **JIA Shuyue** 0:14  
It.

 **William Hsu** 0:16  
Hi, Bruce.  
How are you?

 **JIA Shuyue** 0:26  
I really, I'm, I'm good. Just finished the robot of Manu's paper and we decided to redraw the paper because of the low score and we will submit the paper to the Journal of American Informatics Association, which was, you know, a journal.

 **William Hsu** 0:30  
Good to see you.  
Huh.

 **JIA Shuyue** 0:45  
Yeah, we are also working on multi agent system recently. You know I'm I I started to build the multi agent system based on the open source large modules.

 **William Hsu** 0:46  
Good luck with that.  
I have some ideas about that, building on the ones I shared with you last time, but I wanted to go over a couple of things. The 1st is that we now have actual services.  
Uh, stood up for our Uh.  
Not just the large language model services, but the but also the data collection. So I am asking Wesley and Logan to provide a extraction.  
Um.  
Or rather a data collection service for the.

 **Jahid Chowdhury Choton** 1:42  
Sign it.

 **William Hsu** 1:44  
Uh, actually, surely.  
Uh, thanks for joining us, Logan. Um.  
OK, um, so Logan, I'm gonna clone the um uh sponsored project page. Um.  
So that the data collection API is um.  
Basically scrubbed of all.  
Sensitive data items and I'm going to.

 **Logan Struck** 2:14  
So it it really shouldn't have any sensitive data items, it should just you it. It'll pretty much just contain the link to go to the API web client and you'll need to have credentials to get it anyway. But if if you wanna still remove it anyway, that's totally fine.

 **William Hsu** 2:28  
Quite OK.  
I'm just going to rename so that's good to know. So the we will have a different actual service or a different URL. If you want it to be AC name or anything like that we can but.

 **Logan Struck** 2:35  
But.

 **William Hsu** 2:48  
The idea here is that we'll we'll actually have it for these other other projects. So let me go ahead and share this.  
Disregard the.  
Sponsored projects URL. We will give one that's approved to for everyone to use and I'll I'll discuss security follow up with you Leslie, but I'll just give a a actually.  
Can you can you talk about what's what this is a a placeholder for and?

 **Jahid Chowdhury Choton** 3:26  
Not sure why I.

 **William Hsu** 3:34  
At what the purpose of this API is open.

 **Logan Struck** 3:37  
So a bunch of these placeholders are just showing you the eventual data that's gonna be there. So if you see like the configuration at the bottom, that's gonna be how it how it's configured. The post and get and delete and all that sort of stuff is how you interact with it. So that's where.  
You should be able to see it in the future. It's just, uh, kind of a rough guide for, you know, like roughly to expect what's gonna be there. Uh, actually at the top of the page, what is actual concrete information is the IP address and the URL to get to it. Those have been set in stone now.  
And the.

 **Jahid Chowdhury Choton** 4:15  
ENB work such.

 **Logan Struck** 4:17  
Actually the the URL doesn't work right now. We do have a we do have the IP address set up to the proxmox at the moment. I'm actually I'm actually working on getting it it properly running right now.

 **William Hsu** 4:20  
OK.

 **Logan Struck** 4:33  
But an overall view of the services is that the that data collection API will scrape as much information from all the social media sites as possible, whether it's YouTube, Reddit or.  
Whatnot. Just so that everyone is aware, Cobalt has been having issues gathering videos from YouTube. So if anyone has any issues scraping that, that's gonna that's just some bot detection going on. So expect some issues with YouTube, but otherwise the rest are fine.  
Uh, every other service is serving pretty much that data collection service. So uh the the Postgres database is for that specifically uh our uh LM client which.  
It's basically just for every model that we can throw in there. It's really easy to mess around with. And then with Whisper and OCR, that's for data collection and that's just for transcripts and image recognition.  
Uh, let me know if I'm missing anything off the top of my head. Um.  
Is that roughly what you wanted?

 **William Hsu** 5:47  
Yes, we're going to also request a video llama.  
Service for.

 **Logan Struck** 5:57  
And that'll be a part of the the VLLM, right?

 **William Hsu** 6:01  
Yep. Um.

 **Logan Struck** 6:02  
OK.

 **William Hsu** 6:04  
So this will be for description text. This is video to text.  
So these are video to text or V2T tasks include everything other than other than.

 **Jahid Chowdhury Choton** 6:22  
Yeah.  
It by R key changes. I took no chance to take this.

 **William Hsu** 6:31  
Yep. Um.  
So basically action, action and event recognition.

 **Jahid Chowdhury Choton** 6:35  
Take that out.  
The train one there.

 **William Hsu** 6:43  
Object captioning and captioning rather.  
Entity actually.  
And another pattern.  
And finally, um.  
Sound audio.  
Analysis. So this is song titled matching and or genre.  
File recognition and then.  
Uh, so.  
Jot these down. These are the profitable tasks and.  
Any questions about these tasks?  
I'll have a little bit more explanation of this on the actual page for the service in terms of each of these services if it's tied to an implementation that is a.  
Published paper. We'll go ahead and ask you to reference that, and this is also why today's.  
Meeting outs there saying.  
There's an agenda item for Um.  
References. OK, so progress updates. I've created the page I've.  
Yeah.  
In the process of uh tasking. Um. So if you look at the um uh, the actual tasking um.  
Repository under repositories. I'm going to put the this one.  
So.  
There is an actual repository for this and I'm gonna go ahead and I'm working with Kyle that actually.  
Import this, but this is the GitHub project and under the GitHub projects you'll be able to see under team capacity any assignments.  
And just to give you a concrete example.  
Uh, what that looks like. These also don't have any sensitive terms, but um.  
These are the ones that I've set up for myself. Um, for the sponsored project. Um.  
So under this project, uh, depository, you'll see uh.  
Uh.  
Um.

 **Jahid Chowdhury Choton** 10:13  
Uh.

 **William Hsu** 10:14  
Sorry, um, go ahead. I think you're unmuted.  
OK, um, so here, here's mine. Um.  
OK, and so it consists of a top level task, which is to add session links for the five primary large language models. Llama deepseek.  
ChatGPT, Gemma, sorry, Lama, Deepseek, Gemma, ChatGPT and Claude. OK, first three are open, open source. We are going with the open source ones first, giving priorities to those.  
And then for each of these, we have a.  
A set of a large task which is for a particular task. So the tasks that I am focused on are the.  
Data collection task, the learning to rank task and the smart crawler.  
Right now the focus is just on on query generation, so I have a couple of references for you all on query generation just so you can understand first of all how to work with me on that and use the tools that I'm.  
Creating to get posts. We're starting with seed posts. So this month my goal is to just go through the three main test beds and give you all the tools of seed posts and the platforms that we're going to create posts from that are opened and.  
Not resource constrained. Don't require paid accounts or they have generous enough open minutes, which may just be blue sky and Reddit for the most part, but the terminology extraction, table extraction and relationship extraction projects are the ones that Timothy.  
Um.  
Jahid and Ahsan are working on, but each of you is assigned to one of these and then the ones that we were talking about, optical automatic speech recognition, optical character recognition and then descriptive video captioning. So this is the breakdown of tasks.  
All of these are discussed in these channels that I've linked to. These are the Natural Language Division channels. I'll link up the project road map, but that this will all be under this conference page.

 **Jahid Chowdhury Choton** 12:48  
Oh.

 **William Hsu** 13:00  
And so first of all, are there any questions about the overall project organization and what?  
But this um.  
What the team capacity of?  
Once populated like and what it means. So I'm going to go ahead and this is the this is what this is the team capacity tab of.  
Of the ID project. Open ID project. Um.  
We're using GitHub projects for project tracking GitHub projects. I always abbreviate GHP.  
Ask the tracker. Um.  
There's also a Gantt chart like visualizer for progress in terms of um, the project road map and the deadlines for things. Um.

 **Jahid Chowdhury Choton** 14:11  
Period.

 **William Hsu** 14:12  
And um.  
So.

 **Jahid Chowdhury Choton** 14:19  
So.

 **William Hsu** 14:38  
No.  
Any questions about um?  
Yeah.  
What we're what we're doing. Um, let let me go back to the um.  
Sorry, let me make sure the repository's make up first. OK, so the GitHub repository. This is a shared repository called. This is the one for profile tools for ECSCR for automatic speech recognition. These are all shared.  
And then there's this, there's the one for Olama and.  
Check.  
One that this is bound to for.  
So all of these repositories. These are not GitHub projects, but actual repositories. Um.  
All of these repositories are open except for the sponsored project one.  
Um.  
Go back.  
Continues.  
Speech.  
See.  
OK, uh.  
Oh.  
Here.  
Ask factor. OK, yeah, that's the.  
Speech recognition one.  
And there's another one for. There's a few other services.  
Logan, do you have a directory of the?  
GitHub repositories.  
for these services.

 **Logan Struck** 17:22  
No, they're just within our GitHub organization.

 **William Hsu** 17:25  
OK, so you can browse that from from either this menu that's in the bookmarks file or the one for or by just going to the KDD.  
Um.  
So there's the OCR one and there'll be a this one or.  
OK, uh, I mean for the uh descriptive video. Um.  
Any questions about the services?

 **Patrick Stingley** 18:22  
I have a question, not quite about that. Does everybody have this directory structure of your bookmarks?

 **William Hsu** 18:23  
Yes.

 **Patrick Stingley** 18:34  
'Cause um.

 **William Hsu** 18:34  
Yes, yes. Um.

 **Patrick Stingley** 18:37  
Did you put it somewhere? My computer that you gave it to me on the motherboard went up in smoke so I could use another copy. Could we put a copy of that in confluence so that people who are similarly challenged can just go there?

 **William Hsu** 18:44  
Oh yeah, sure it's.  
Yes. Thank you for. Thank you for reminding me. Yeah, I'll do that right now. Let me.

 **Patrick Stingley** 18:55  
OK, that way we can all go like follow the same, have that tree and we can all follow along in the same way.

 **William Hsu** 18:59  
Yeah.  
That's right.

 **Patrick Stingley** 19:04  
At least for a while, at least while there, at least initially.

 **William Hsu** 19:08  
Yes, so that what Patrick is referring to is this file here and I'll I'll link to it for now because right now Confluence just points to where things are.

 **Patrick Stingley** 19:19  
OK.

 **William Hsu** 19:24  
So we're talking about.  
Thing that's in the KDD group under KDD members and.

 **Patrick Stingley** 19:34  
OK.

 **William Hsu** 19:37  
Here under bookmarks. So this one the directory is here I think.

 **Patrick Stingley** 19:51  
Yeah, if you just put that in the chat, it'd make it easier to find my way back because.

 **William Hsu** 19:53  
Put that in. Yep. So I'll I'll put it in the confluence 1st and update. Make sure you can see it. This is the.

 **Patrick Stingley** 20:02  
Yeah.

 **William Hsu** 20:22  
I check.  
SharePoint directory and then under the SharePoint directory is on the actual so the most recent one that I have.  
Uploaded is the July 28th one.  
Here um.  
OK, now that that will always be there and um, browse through the directory and download it. Um.

 **Patrick Stingley** 21:22  
Right.

 **William Hsu** 21:23  
So I.  
Make sure you can grab that.

 **Patrick Stingley** 21:26  
You mind just pasting it in the chat so that I can find my way to it?

 **William Hsu** 21:30  
Sure.

 **Patrick Stingley** 21:32  
Alright, that'll be easier for now.  
Thanks.  
OK, it's all on the same page.

 **William Hsu** 22:03  
All right. Um.

 **Patrick Stingley** 22:05  
Hope so. Not yet, but I'm I'm sure I will.

 **William Hsu** 22:08  
Can you see it?  
OK, uh, so that can import into Edge or in or Chrome and uh.

 **Patrick Stingley** 22:20  
So you know.

 **William Hsu** 22:21  
Equally important under confluence, you can see all the hosted services. There'll be a new one for the description for video for LMS.  
When you log in, when you log into the VPN.  
You'll see this.  
If you're logged into the VPN.  
You should get.  
This prompt and be able to share.  
Sessions from Mistral and all of the others. Thanks to whoever set up Mistral, I set up Lava and deepseek and Gemma were already there.  
Yihong and I are looking at Quen and at Lava and at Video Lava. So we are looking at we're discussing with Logan Wesley how to actually run the video analyzer as a service.  
The purpose of this project is to extract content of.  
Social posts about scientific topics. I'm going to click on the.  
So this is our Yep. OK.  
This is the the poll. If you haven't completed the poll, we're gonna go ahead and submit this. I mean, I'm gonna go ahead and commit this, but this is the poll one more time.  
In case you haven't voted in the poll topic poll.  
Right here.  
Right, so the topic poll is just this, which of the following open IU test beds are you interested in? OK and.  
We're interested in your input.  
And um.  
2.  
OK, that's on this form. OK, so that's the current results of the poll. There's only been 6 responses, so I know that not everybody's actually given input, but some subset of the.  
Six of you hopefully have actually replied already.  
But here's the.  
Purpose.

 **Patrick Stingley** 25:58  
Bill, I hate to bother you again, but the the last thing I see in my chat is GitHub is the GitHub projects task tracker. It says I can't send messages because I'm not a member of the chat.  
I don't know. It was. I was getting stuff, but I'm not anymore.  
Can't send any messages here.

 **William Hsu** 26:24  
Let me see. Can you share screen?

 **Patrick Stingley** 26:30  
Yeah, sure.

 **William Hsu** 26:31  
OK, go ahead. Let me see what's going on.

 **Patrick Stingley** 26:37  
Well, it just disappeared. Uh.  
It just disappeared. The whole Teams thing I was looking at went away. Where did it go? Here. There it is. There we go. This is what I see.  
And so it was.

 **William Hsu** 26:54  
Oh, hit X on meeting chat. That is a quirk of of Teams. If you hit X on meeting chat, it'll take you back to the hit chat again.  
That's a weird little reply environment that it goes into. Yeah, no, it's that that happens to me all the time. It's it's a weird quirk. I can't tell you exactly why it happens. I just know how to get out of it.

 **Patrick Stingley** 27:12  
Oh, there it is. OK, sorry to bother everybody.

 **William Hsu** 27:24  
OK, uh.  
Alright, um.

 **Patrick Stingley** 27:28  
Access denied. You can see that I I still can't see it, but the action you're trying to perform is blocked by the organization. That's.

 **William Hsu** 27:29  
OK, so.  
Wait, you still can't see what?  
Oh, sorry, could you share again? I accidentally.

 **Patrick Stingley** 27:42  
It's I think I still am.  
I'll try again.

 **William Hsu** 27:48  
I think I I think I um, yeah.

 **Patrick Stingley** 27:49  
Oh, you took control? Yeah, sure.  
So you know, I clicked on that thing you gave us and there's that also, but they get, you know.

 **William Hsu** 28:04  
OK, which thing? Which thing are we talking about?

 **Patrick Stingley** 28:08  
Hang on, where? Where? I'm gonna go back to full screen here. Bookmarks, right? I'm just gonna try to get to the bookmarks.  
Access denied.  
All right.

 **William Hsu** 28:22  
That's probably.

 **Patrick Stingley** 28:27  
Oh, let me go to the point.

 **William Hsu** 28:28  
How you're signed into the SharePoint? Um.

 **Patrick Stingley** 28:30  
Well, at least I can do the poll.

 **William Hsu** 28:33  
So if you can oh um.

 **Patrick Stingley** 28:38  
At least I can get to the pole.

 **William Hsu** 28:40  
Yeah, I can attach the the uh.  
This this has something to do with how you're coming into Teams 'cause that is an actual share from files. Let me just check something.  
That's something right here, cause Bruce probably can't see it either. In that case, let me just choose.

 **Patrick Stingley** 29:26  
Didn't work.

 **William Hsu** 29:31  
OK, I just sent you. Um.

 **Patrick Stingley** 29:36  
OK.

 **William Hsu** 29:36  
A invite to the bookmarks directory.

 **Patrick Stingley** 29:40  
How would that be?

 **William Hsu** 29:42  
P. Stingley at ksu.edu See if you click on that and yeah, so if you open your KSU e-mail and just make sure that you can actually get to that.

 **Patrick Stingley** 29:48  
Oh, it's in e-mail.  
OK, yeah, I can get to that. Hang on. Outlook, Outlook. I'm not there. e-mail. I don't know.

 **William Hsu** 30:03  
OK, Logan, please remind me tomorrow to fill in the request paperwork with central IT for Bruce to get an EID.

 **Patrick Stingley** 30:14  
Tell me too.  
So did you get the message about that I couldn't get into VPN? I need something from you on that.

 **Logan Struck** 30:18  
Oh yeah, I.

 **William Hsu** 30:25  
Oh, you too. OK, uh.

 **Patrick Stingley** 30:26  
Yeah, yeah, I I have an EID and all that, but I called up, you know, I tried to do it myself and then I I didn't get it anywhere. So I talked to tech support and they said that you have to send some certification over to them.  
To get me. OK, I didn't know what to do. OK, that's fine. Sorry to bother everybody.

 **William Hsu** 30:42  
OK, I know what to do cause I just did this for for one of my alumni. So I'll I'll do that for you and so Logan, Patrick and Bruce.

 **Patrick Stingley** 30:56  
Yeah, there's a dear Patrick Cuba selected to fill out short survey.

 **Logan Struck** 30:59  
OK, I get that.

 **William Hsu** 31:03  
OK, um.

 **Patrick Stingley** 31:04  
OK.  
K-State support just sent that to me Sunday.  
Hmm.  
I didn't get you your e-mail was my point.

 **William Hsu** 31:15  
Um.  
And get my e-mail. There it is.

 **Patrick Stingley** 31:20  
Well, wait, yeah, you just sent me one. You just sent me an e-mail with, um, the link to the.  
Um.  
The the bookmarks.

 **William Hsu** 31:33  
Uh, is it's not the um.

 **Patrick Stingley** 31:38  
OK.

 **William Hsu** 31:38  
The one where I mentioned you, OK.

 **Patrick Stingley** 31:42  
Stingley at kosu.edu.

 **William Hsu** 31:44  
They make sure it's not then get sent to John.

 **Patrick Stingley** 31:48  
To whom? Oh, junk.

 **William Hsu** 31:49  
Uh.  
Um.

 **Patrick Stingley** 31:57  
No.  
Inbox. I don't know. You and I are gonna talk Wednesday tomorrow, right? We can do this without bothering everybody else.

 **William Hsu** 32:07  
OK, um, let me let me do one last thing and uh, then you can uh.  
I don't like.

 **Patrick Stingley** 32:22  
Not sure.  
Stop. I will get rid of that.  
Stop sharing.

 **William Hsu** 32:51  
OK, uh.  
OK.

 **Patrick Stingley** 32:54  
Mm-hmm.

 **William Hsu** 32:55  
Try this one.

 **Patrick Stingley** 32:57  
Oh, in e-mail.

 **William Hsu** 32:58  
Thank you.  
Yeah, it should also be an e-mail. OK, let me know if that works.

 **Patrick Stingley** 33:01  
OK.  
Yeah, I got that one.  
I got something.  
Let's see open.  
This invite will only work for you when people are existing.  
No, it's not not doing anything. I really, I don't want to take everybody else's time. You and I can talk about it.

 **William Hsu** 33:29  
OK, all right. We'll we'll follow it up on that offline. OK, everybody should have access to the Confluence itself. I'll try attaching the bookmarks and doing anything directly in Confluence, but if you are actually having trouble accessing our Teams SharePoint, that's probably an EID issue and.

 **Patrick Stingley** 33:33  
Tomorrow.

 **William Hsu** 33:50  
You know what I can get you is VPN access and and then if anything else is not working, I I will have to diagnose something. See offline. OK, back to the LLM.  
Uh, to install things, uh, you go to uh, you go to the uh.  
Settings and.  
Under general you can.

 **James Chapman** 34:42  
You're not sharing Doctor Xu.

 **William Hsu** 34:43  
Oh, thank you. Um.  
Let me try logging back in. Um.  
What do I do to log into the?  
Into this.

 **Logan Struck** 35:21  
Like, uh, into the service itself. I got the admin account.

 **William Hsu** 35:29  
Yes.

 **Logan Struck** 35:34  
So that's that's currently the only thing there. The access management is going to be an update later.

 **William Hsu** 35:44  
OK, because Wesley showed me how and I thought I knew how to install new new models and that was.

 **Logan Struck** 35:54  
New models, yeah.  
Well, let me look again, 'cause, yes, I do remember talking about it.

 **William Hsu** 36:04  
That user.  
Um.  
OK, so it keeps logging me back in. Um, automatically. Um.

 **Logan Struck** 36:49  
Uh so you you hit uh user uh in the bottom left you go to settings and then in settings uh you wanna go to the bottom one which is admin settings.

 **William Hsu** 36:54  
Mhm.  
Mhm.

 **Logan Struck** 37:03  
And then there you could go to models, yes.

 **William Hsu** 37:03  
Oh, there it is. OK, OK. And if you all have not actually downloaded models before, you have to actually name the model explicitly. It's not a dropdown menu. So if you go to olama.com, it'll give you the browsable list of all the models right here.  
OK, so go to olama.com to see models, read their specs, choose the size to get the actual the exact name of the model and that will allow you to do things like.  
Specify a model by name and it'll it has to be a valid model name. There's a similar one for for VLLM and.  
Bruce, I'll go over with you how to do that. Uh, how to use it. Uh.  
Later, when we have.

 **Patrick Stingley** 38:00  
So, uh, Logan, you set this up, right?  
Is that right?

 **Logan Struck** 38:07  
No, uh, this was mostly Wesley's doing. Uh, Yep.

 **Patrick Stingley** 38:11  
OK, So what you know, what memory footprint and stuff should we does this provide? Like I I you could. What I do when I choose a model is I look at the computer that I'm using and I say well.  
Do I have a GPU? Do I have a CPU? How much shared memory does it have? And then I go to ChatGPT or one of these things and I say, well, what's the best model given the size of of my hardware? So what? What? What is the?

 **Logan Struck** 38:32  
Thank you.

 **Patrick Stingley** 38:42  
Does anybody have the constraints of what the hardware is that we have?

 **Logan Struck** 38:44  
Are are you you're just asking me what?

 **William Hsu** 38:47  
It depends on where you're running the actual service. Yeah, go ahead.

 **Logan Struck** 38:53  
So you're just asking me what is the hardware configuration? Is that it?

 **Patrick Stingley** 38:59  
Yeah, well, what's the hardware configuration availability?

 **Logan Struck** 38:59  
Is it OK? It's it has four, it has four A 40s and I believe like 132 gigabytes of RAM or something. So whatever models that you have at it, you should be able to throw at it. It has a lot of capability.

 **William Hsu** 39:17  
So the spec sheet, I just put the spec sheet of the NVIDIA 840 our GPU servers.  
Have eight each. There are two GPU servers. We've reserved 4840s for large language model.

 **Patrick Stingley** 39:37  
OK.

 **William Hsu** 39:37  
11 qualifier to what Logan is saying, we do not have NV link, so we can't actually pool a 40 memory.  
To the best of my knowledge, we can only use um 48 gig at a time.

 **Patrick Stingley** 39:52  
48 gig at a time.

 **William Hsu** 39:53  
Yeah, Logan, Logan, do you know any different? Because I'm not aware that we have.

 **Patrick Stingley** 39:54  
OK, well, it's a constraint.

 **Logan Struck** 39:57  
Oh.  
So that that virtual machine is using four of them, so it should be. It should be pooled with the with the path through. So so that VM should really be viewing one graphics card, but in reality it it's four, so it should have all that.

 **William Hsu** 40:10  
OK.  
OK, so but it's not. It's not using the NB link backplane, it's using the fiber channel backplane.

 **Logan Struck** 40:21  
I mean, I think it's using something else for the pass through. I honestly can't remember.

 **William Hsu** 40:26  
OK, OK. Patrick, you know what we're talking about. I don't know if anybody else is familiar enough with.

 **Patrick Stingley** 40:27  
OK.

 **Logan Struck** 40:28  
OK.

 **Patrick Stingley** 40:34  
Well, basically it's a virtual machine, so you you have it set up for four A 40s and 180 gigs and it'll give it give us whatever it's got, but it's a multi-user machine.

 **William Hsu** 40:34  
How are you doing?

 **Patrick Stingley** 40:47  
Yeah, the footprint will look the same, but it'll slow down if other people are using it at the same time.

 **William Hsu** 40:52  
Right, right. I I was asking about the backplane because we have fibre channel cross connects, but there's still not NV link speed, but we don't actually have NV link. Just just FYI here.

 **Patrick Stingley** 40:59  
Mhm.  
Run.  
Right.  
So you're probably right. We could probably throw a lot of stuff at it without it really worrying about it.  
Yeah.

 **William Hsu** 41:15  
Yes. And Bruce, you you asked if if I could get you access to this. I'm working on it. This is one of the things that I'm asking voted to help me get filled out for both Patrick and you.

 **Logan Struck** 41:18  
And.

 **William Hsu** 41:32  
The the grads already have access to these, but the other thing that we need to do is is see about.  
Oh.  
Pittsburgh Supercomputer Center. Jahid, do you still have the PSC application information?

 **Jahid Chowdhury Choton** 41:53  
Uh, yes, I have.

 **William Hsu** 41:56  
OK, can you put that in the chat please?

 **Jahid Chowdhury Choton** 42:00  
OK, yeah, I'm putting it.

 **William Hsu** 42:27  
Here we go.  
OK, this is what I got. I'm gonna go ahead and Oh yeah, OK. Yeah, you just pasted it already. So yeah.  
Uh, for those of you who are um.  
Have key accounts. Take a look at.  
This.  
Please say plain text. It's this is the same message, but it's on our mail server. So if you have KDD, if you're on the KDDL mailing list, you should be able to get directly to the message, but.  
In here is the how to actually sign up for access. OK, let me know if you have any questions about that. Meanwhile, we're working on getting everyone access to our.  
Uh.  
GPU cluster, but you have very limited. You'll have 16 GP US. Most of them are in use for a couple of papers right now, so PSC resources are your best bet.  
And if if I can get you access to.  
VPN I can also get you access to Bay OK. So that's that's how we're going to primarily do this question for you Logan, you asked me about.  
CPU too. You said you were gonna run Proxmox on that. How will users?

 **Logan Struck** 44:49  
Not necessarily. We are going to run proxbox on it. I was just if we weren't using it for anything else, I just would like to use it for something, whether it be for your classes or or whatnot, OK.

 **William Hsu** 45:02  
Oh yes, please. Yes, please. Yes, for for my classes is great and we don't need a remote RDP servers and.  
I definitely established that remote dot CS meets our needs so.

 **Logan Struck** 45:15  
Yeah.  
OK.

 **William Hsu** 45:22  
First priority goes to running the front end for this and.  
By in 731 by 31 class. OK, so told you all how to install models. Later you'll need to be signed in to install models. This is just a essentially.

 **Logan Struck** 45:36  
Yes.

 **William Hsu** 45:45  
1.  
Sandbox.  
Playground for hosted moms are now the inferencer we're going to.  
You can currently only run it from the from the front end prompt, but.  
We're going to look at.  
Actually uploading files and what's gonna take the place of that? So that's that's a key question.  
I have not done extensive testing for vision language models with attachments, but that's that's one thing I'm gonna be doing both with the free version and other things. OK, so.  
I just wanted to show you in terms of multimodal input, how we get from social posts to crawled posts, what we call raw posts to index posts, which are the five text blobs. So let me let me sort of review the definition here if you go to.  
Go to the IE page. Um.  
You'll see a definition of.  
Ideas for improvement and these multimodal vision to text tasks and what else we have is.  
Uh, social post.  
Uh.  
Are initially downloaded and use Cobalt and we call the downloaded version raw, so.  
Those are the ones that are saved from the platforms as files currently on Google Cloud. These are free cloud or free tier cloud by the data collection team and they're used to make index codes. These this is the original source data for information extraction and it has.  
That they're multimodal. They're intrinsically multiple. Index posts are the ones that have gone undergone preliminary video to text tasks, including speech recognition, optical character recognition to produce a Postgres database entry that we've been referring to as JSON tuples or fortified text blogs.  
Those four or five text blogs be title body.  
Speech transcript, OCR transcript, and descriptive text. So those are the input to information extraction. And then what the sponsored project calls annotated posts are those that are output by the information extraction project. So these are currently committed to a different.  
Table of the Postgres database and there's an inference output for extraction. So a question for you Logan, is there going to be a way? Is there going to be a writable API for us to actually add?

 **Logan Struck** 49:23  
What API?

 **William Hsu** 49:25  
A writable API.

 **Logan Struck** 49:28  
Uh, there should be an incoming feature. Uh, are you talking about for Postgres?

 **William Hsu** 49:37  
Yes, for posting.

 **Logan Struck** 49:39  
So I think what's gonna happen is for the data collection API you will go into our Postgres database within CS, but for any additional Postgres things that you guys want.  
I'm gonna I'm gonna try and make a a separate one on our own servers just because that one on the CS network is is is shared and we only have one account for it that has access to everything on the Postgres database. So it's a major security issue if everyone.  
Can go in there. So we're working on ways to limit it and where you can actually have a couple separate postgres things and not have it all get tangled up with with all the the Max access and going around. So we and one of the things is we we just have to be mindful of everyone else's storage on there again because.

 **William Hsu** 50:29  
OK.

 **Logan Struck** 50:34  
It it's shared so.

 **William Hsu** 50:38  
OK, OK. Um.  
That that makes sense. So however you want to regulate access, just tell us if you wanted to populate something, something local, something on a different server.  
You want if it's not a firewall issue, but a authorized user issue, I just want to.  
Verify with you how you want.  
Essentially heat warming to work right so that if this can be managed using group lead grouper, I would like to just say here are the IE project members in grouper. Let's take their.  
SSOI mean take their credentials from Grouper and then.

 **Logan Struck** 51:39  
So we can probably just do a single sign on for it and use a a Web API for you guys to be able to, uh, sorry, a web client, yeah.

 **William Hsu** 51:46  
Oh, that'd be perfect. That's perfect.  
This is our shared target data log. Annotated post is our shared target data log.

 **Logan Struck** 51:57  
Actually, uh.  
I think with the S3 database client that we're using, we can combine that and everyone can also see the Postgres through there too. So we just make it one link to set too so.

 **William Hsu** 52:16  
OK.  
Um. And uh, I just wanted to, uh, let you know that I am um.  
On.  
I'm refining this prompt that I previously shared with you, but this is the prompt.  
It's the one that gave me the list of suggested domains, and I'm going to actually specify the first name, so the upshot of this.  
This whole is that we are now committed to the climate task plus the neuroscience and the task. So in terms of this is our primary domain.  
And then the secondary and tertiary domains are these two based on this cold.  
Yeah, so our sort of current leaders are in climate change, mental health, neuroscience, space and astronomy and consumer tech and A I tools are kind of preserved. So right now the the leaders are like.  
I think climate has a lock, but the others, uh, there's a there's a sort of relay time, so.  
Two and then three-way tie on the rest.  
So uh.  
It's fine. It's fine.  
1.  
This is going to depend on what platforms there exist. Extend seats already, so I'm going to go into an interactive prompt mode.

 **Jahid Chowdhury Choton** 54:27  
Everything just now.

 **William Hsu** 54:37  
This is also why I'm asking you all to complete this poll so it's help break this time, but there it is so.  
I know some of you have expressed this specific personal interest in one or more of these. So these are these are the ones that we're.  
Starting with and sorry to be determined, we're going to proceed in parallel a single team and look at crawling these platforms. So in terms of data access.  
Climate is the leader because it has there are existing data sets for all platforms. So for the climate domain, this is what we're actually doing to see.  
And just to remind you, the the actual tasks are dialogic tasks because it's if you look at the actual, there's already an existing data set of social posts, there's a video sharing platform.  
There's an API or free tier from which we can harvest data using existing research licenses or agreements fits with STEM themes, adjacent domains such as health and environment, and if you see here the climate discourse.  
The actual discourse analysis is the the shared theme and so misinformation detection is our sort of showcase task, but in terms of multimodal reasoning.  
We all have interests in these. This happens to be a shared in happens to be our shared interest. So Bruce and I have been talking about agentic reasoning and agentic.  
Analytics. So I know a little bit about agentic analytics because I've worked with different kinds of APIs. I have a question for you, Patrick. You have mentioned long chain. Have you actually tried any other?  
Tools including and beyond one.  
Or managing.

 **Patrick Stingley** 57:12  
Yeah, so um, what is it called? Uh.  
Flow wise. I tried Flow wise that worked pretty well. N8N seems to be the market leader right now in terms of connecting up the dots and.  
I don't know if you know that out what is it? Open AI added their agentic capability yesterday night, like last night and I I don't know if you know how to get to it yet. Can you bring up Open AI?

 **William Hsu** 57:43  
Yes.  
Call Sam's.  
Sure.

 **Patrick Stingley** 57:51  
You bring up Chao GPT.

 **William Hsu** 58:04  
Yep.

 **Patrick Stingley** 58:05  
OK, down there under tools.

 **William Hsu** 58:08  
Oh yeah, tools.

 **Patrick Stingley** 58:10  
All right. Well, look at that.  
Mine does. Mine has agentic.  
Try the plus sign, but I think it's under tools. No on the left with the little plus thing. Plus is plus like files and yeah, that's not it. Tools isn't showing on your.

 **William Hsu** 58:25  
Oh, sorry.  
OK, uh, were you on a wait list? Uh, sign up list?

 **Patrick Stingley** 58:33  
No, but I I do pay the 20 bucks a month, so that must give me that.

 **William Hsu** 58:35  
Oh, OK. OK. So it may be only available in the page, yeah.

 **Logan Struck** 58:37  
E.

 **Patrick Stingley** 58:40  
Yeah, yeah. So I'm just saying that the agentic dropped last night and I don't know how, you know, if everybody will get it, but I've got it. I wasn't.

 **William Hsu** 58:45  
Yeah.

 **Patrick Stingley** 58:53  
I I really couldn't see much difference between that and just regular ChatGPT. I I wasn't. I tried to get it to do some stuff, complicated stuff with flow charts today. I mean with Excel spreadsheets in it. It it choked and died just the same as it used to die. So you know they were going on about.  
About how brilliant it is with spreadsheets, but I didn't. My experience with it today did not prove that out. N 8 N is what I would recommend, and here's a couple of reasons. One of them is that you can host it locally. Secondly, it is like the market leader.  
For doing flow uh workflows.  
I did use the the other one like I said flow wise and it was good to learn how to do it, but it's cloud hosted. It was kind of hard to get it to work. I think we'd be better off with an 8 in and and like I said we can host it locally so.  
We can contain any changes. That's the thing I think I like about it the best is that you can host it locally. You can use it on the cloud too, but I I like the idea of hosting it locally because then we we're not going to get anything broken because they decide to upgrade somehow.

 **William Hsu** 1:00:11  
OK.

 **Patrick Stingley** 1:00:11  
That's been the biggest problem I've had with Lang Chang was Lang Chang has gone in four different directions and every time they change it, the code just doesn't work. And I I don't think we want to go with Lang Chang because it's just that unstable.  
Every time I do it, I end up spending a lot of time trying to get a this version to work and then I have to put it in containers because it's independent libraries don't work with other stuff and and overall it's just too much of a pain in the neck.  
We want to process, we want to, I would think we want to make progress and and we'd be better off to do it, use a tool like N8N and just do the job rather than be computer scientists about it.

 **William Hsu** 1:00:59  
Uh, are you recommending N8N overflow wise or um?

 **Patrick Stingley** 1:01:02  
Yeah, I am. I'm recommending it as the I if I had to choose, that's what I would choose today. That's what I would choose. It's you can pick it up. It's easy to learn. You got stuff. You can pick it up off YouTube how to use it and it you know you'll you probably use it in a later life if you leave.  
You know, get a job in private industry. Um.  
Yeah, I would recommend that. I don't recommend hacking Langchain or any of its sister products. It's just too too difficult to keep up with it, and they're usually not compatible with each other, and it's just terrible.

 **William Hsu** 1:01:43  
OK. I will take that under advisement and check it out and confirm that this is the one that we're going to.

 **Patrick Stingley** 1:01:47  
Yeah, you could ask around.  
But I I I would, I would go with that.

 **William Hsu** 1:01:53  
OK.  
OK. Thank you. I appreciate it. And Bruce says that he's using Quinn 30P.

 **Patrick Stingley** 1:02:04  
Oh yeah, Quinn's another good one too. Yeah, it's. I haven't used it, but I'm. I'm familiar with it. That's not a bad choice.

 **William Hsu** 1:02:10  
Quinn is Ali Baba's, right?

 **Patrick Stingley** 1:02:15  
I don't know.

 **JIA Shuyue** 1:02:16  
Pardon me?

 **William Hsu** 1:02:17  
I think when was developed by Alibaba.

 **JIA Shuyue** 1:02:20  
Yeah, Alibaba, yes, yeah, you list the Alibaba group and it's open source and we can deploy this module on our local workshop or work spaces.

 **William Hsu** 1:02:23  
Yeah, yeah.

 **Patrick Stingley** 1:02:30  
Choice too. I just haven't used it. That's a good no, no problem with that.

 **William Hsu** 1:02:37  
OK, um, I will do a sort of feature comparison. Um, I think um.

 **Logan Struck** 1:02:37  
OK.

 **William Hsu** 1:02:44  
We just need to see what meets our.

 **Patrick Stingley** 1:02:51  
You know, in terms of a feature comparison, I would spend, you know, 1/2 an hour on YouTube and see which one has a good, you know, how to use this tool. Because to be honest with you, OK, yeah, that's what I would do because you know, feature by feature doesn't really matter if it's inconvenient to use.

 **Logan Struck** 1:02:54  
Uh, excuse me.

 **William Hsu** 1:03:01  
That's exactly what I meant by feature comparison. Yeah, exactly.  
No, no, I don't mean, um, like Consumer Reports. Um, I mean, um.

 **Logan Struck** 1:03:11  
OK.

 **Patrick Stingley** 1:03:12  
Yeah.  
Yeah, yeah, yeah. Just watch the YouTube and see what it says. That's a great idea.

 **William Hsu** 1:03:21  
OK, thank you. Let's so, so Quinn has an agentic builder, I mean other than the fact that we're already trying up the the mobs.

 **Logan Struck** 1:03:21  
Next.  
OK.  
Sir, I I have another obligation to go to. Is there anything else that you guys need from me?

 **William Hsu** 1:03:36  
Does it have an agent? Um.  
OK. Thank you. I've actually Bruce's comment about this basically in order number one BLLM will go over that the actual BLLM sort of timeline and.  
And an agentic interface. But before that we definitely need video llama hosting beyond O llama. We need the actual service.

 **Logan Struck** 1:03:56  
OK.

 **William Hsu** 1:04:10  
Stood up the same way that OCR and PSR are. So if you could just let Leslie know and and we'll go over that Friday.

 **Logan Struck** 1:04:15  
OK.  
All right, can do.

 **William Hsu** 1:04:21  
Thank you, Logan. Thanks for coming. Thank you. OK, any questions about tools? Because I'm going to cut straight to the chase on task definition if there aren't any other.

 **Logan Struck** 1:04:23  
Yeah, no problem. Thanks. Thanks everyone.

 **William Hsu** 1:04:42  
So OK, this is the point major thing, OK.  
So is is this? Did you say you're using this client agent builder, Bruce?

 **JIA Shuyue** 1:04:55  
Oh William, actually I'm actually using VLLM to serve this. I didn't use this package, but this package is also one option and I'm just using the VLLM and with the toy.

 **William Hsu** 1:05:01  
OK, OK.

 **JIA Shuyue** 1:05:10  
Uh, with a different, uh, a predefined toss and uh, you know the agent can uh, ask you to execute uh one to or a batch of toss and uh then execute some um some stuff. You know the AI agent is just a.  
You have the, you have the, um, you have the memory, you have the planning and you have the tall execution and uh, finally you and you have.

 **William Hsu** 1:05:27  
Right.  
Yeah.

 **JIA Shuyue** 1:05:40  
Yeah, I forgot that to work.

 **William Hsu** 1:05:40  
But I have done work on uh.  
Plan recognition.  
And and also and also HTN planning.  
Before, before my current work in reinforcement, the reason that's important is we've been talking about what is better feedback than RLHF and GRPO. And one of the things in learning to rank is you actually want to look at query responsiveness.

 **JIA Shuyue** 1:05:59  
OK.

 **William Hsu** 1:06:14  
So in computer vision, query responsiveness is is key to the visual identification task. If you are trying to do optic classification, you look at mean average precision loss on a validation data set. If you're trying to do segmentation, you look at intersection over union loss on a mask or a box.  
Right. So for and those can be integrated over. That's a saliency, right? That can be integrated with a contrastive lost weight or interceptual similarity weight.  
To a to a reference and in other words, we can do super resolution, crop the image to the actual mask to saliency mask. We can focus attention using.  
Fusion decoding. Like in Grad Cam, we can use a BLM to actually back.  
Inference based, visual identification based.  
Loss. Um.  
But for agentic, for an agentic system, we might have multi-step interventions, right? So if we're doing data augmentation and we're trying to do use that gradient to do directed augmentation like first crop to the mask, then do.  
Brightening or you know if there's multi step. Traditionally that was look ahead policy search and there's an MDP or looking at policies. Today we can actually treat that as a planning task and this is where the agentic agentic task comes in.  
So this is data augmentation, multi multi stage directed, gradient directed.  
Directed data augmentation will decline.  
Or a CV task, even a BLM not so not as well defined for reasoning tasks and learning to write. So this is This is why.  
Yeah, he knows that I because he's been working with me on the first one.  
You know why I I am trying to do this with the social platform.  
The short reason is because we already have image and video tasks that are defined as classification and they're defined as hierarchical classification. So the look ahead policy search is already well defined there.  
Um, so if they're well defined here, this is, um, look ahead. Um, policy search. Um and I have a reference for you to take a look at. Um.  
Bruce that it's this one.  
This is the original.  
It's actually from our workshop back in 2002, but.  
So the paper that he and and that Lebner and Ross Griner.  
Published and then eventually they came up with this agentic system. So as you can see, this is from 20 years ago, 22 years ago and it was early work on.  
Basically, MDP solvers and RL for image interpretation. Very, very simple kind of filters today completely superseded. The actual ground filters are completely superseded by convolutional neural networks.  
And yet the interventions at a higher level of abstraction in terms of super resolution imaging, propping, autofocus, they are actually still salient and and attention is.  
And salients are more relevant than ever in in that agentic setting. That's why I think it's time to kind of provide that direction. That's my overarching motive for doing that, my overarching purpose for doing that.  
But the purpose with respect to information extraction is what if you actually need reasoning and you need a mixture of experts kind of model, right? How do you resolve?  
Dialogue and discourse, so resolving.  
Multiple opinions.  
So dialogue and discourse model.  
And argumentation.  
So I have a had a PhD student who did her dissertation on question generation in legal in computational.  
The the original vision was to look at legal documents more generally. The scope of her dissertation was completely focused on privacy policy.  
Primarily the kind of privacy document that you get from a Internet service provider and specifically a mobile Internet service provider, so policy.  
Documents from wireless providers, which is very, very, very narrow, but it was a kind of a showcase for this, so.  
This is the.  
This book that I was, I've been reading that I picked up at.  
I think it was.  
The AI I think it was either HCHI 2020 or.  
It's by 2020 or.  
2019 but um, but AAAI 2020 is kind of.  
OK, so I'm gonna wrap it up here. This is the argumentation book that is by IE focused because just doing IE.  
Even Hung Ji, when I spoke with her about do any of your students work on IE these days, she said. Since Jing Ring.  
It's hard to get any of my students to work on IE anymore, so I don't know if you know him, Bruce, but he he's at College of William and Mary. He's one of.  
Recent alumni, along with Emily, we worked with him very briefly before he graduated and his main paper is this one, Scientific Multimodal Science Foundation Models with Critical Thinking. And he's also done more with research agents and planning and so.  
This is this is where his work impinges on both hot topics of 2022 and hot topics of 2026. So this is the paper that I'm referring to.  
Is.  
This one. Um.  
Very prolific as we can see.  
This.  
OK, so this paper is really great. We worked with him on trying to basically leverage this or.  
For our sponsored project from Lawrence Livermore National Labs on recipe extraction, procedural information extraction, I'm still very interested in procedural information extraction, but the the reasoning.  
Specification there is much more nebulous than than a dialogic model and much less agentic, right? So that that's why that's why I want to kind of go hard to the boot on.  
Um, basically a debate agent. Um.  
With a specified QA. So if you have a specified QA task, what evidence is there for or against anthropogenic?  
Sources or causes of.  
Global climate change as manifested by ocean temperature, mean ocean temperature increase. That is, that's a crisply defined statement. And then you can from that specification you can look at OK, here's all the.  
Here's all the relevant sort of debate team, you know, form up the debate team and and look at the look at the sources. And you can also critique the source quality, which is which is kind of the point if you're doing misinformation, right?  
Um, so um.  
I'm interested in the agentic reasoning aspect as a way to.  
Do a source.  
Validation. So in a nutshell, that's where this idea comes from. I've been talking with a lot of my students about how this is the this is the paper.  
He jumps.  
Oh yeah, I think this is the one I've seen. Yeah. OK, so if you look at this paper, they mine wiki hub or.  
Recipes and they actually extract recipes by using large language models. They have retrieval log meta generation which gives a certain amount of improvement. They have a multimedia selected encoder which basically says how much of the recipe is actually in the wiki how diagram.  
If we caption the diagram, how much can we enrich the actual recipe? So this is the retrieval augmented decoder and the retrieved step encoder and the base model is actually blit.  
Uh, blip 2 um click bars versus flip 2 and they do better with flip 2, but significantly better using the additional.  
Improvements. They also use a contrastive loss, which I was excited by, but does not really do very much. It's like an epsilon increment attribution.  
The margin of victory and then there's loss functions. I thought this would actually provide more of the improvement. It doesn't, but that's the that's the synopsis.  
Of the of the task. If there's interest, we can discuss this in reading group or you can just take a look at this particular finding and you'll see this is the state-of-the-art in.  
Generative summarization of procedural information and.  
It it basically made showed that it was feasible to to extract recipes from a general source as of 2-3 years ago.  
So it's clearly that's not a solved problem.  
But what's a much more timely and interesting problem, given LLM limitations, is source attribution and and source evaluation. So open problems related open problems.  
Source attribution as Patrick notes also.  
General common sense reasoning.  
General common sense reasoning.  
Agentec.  
Mixtures mixture of experts you've probably seen like the biggest buzzword these days is a mixture of agents like like that's new. It's it's not really but because agentic.  
Reasoning is a hot topic. People have kind of rediscovered mixture of experts or agents, but actually fusing the the votes.  
Supporting evidence, supporting arguments of agents. That's of course the problem. OK, so I've summarized what I'm trying to get at. I'm trying to get us to the door of agentic reasoning as a worthwhile.  
Way to do better at extracting information about controversies, debate topics, topics for which there is.  
Mixed. There's no consensus or there may be consensus, but there's a lot of noise, like very, very high, very low signal to noise ratio because of all the conspiracy theories and disinformation and people doing quote UN quote, their own research and and all that stuff, so.  
Humans have some critical thinking skills to wade through those. We're not.  
All very consistent about using those, but agents have more reproducible methods, but those methods are also fallible.  
And as as the security research is showing that they're also susceptible, they're vulnerable to attack because of they're vulnerable to adversarial policy induction and adversarial basically being.  
Street party process, right? So.  
Rather than try to.  
Attack that open problem that is that runs up against computability limitations. I think it's more fruitful to look at how humans actually use evidential reasoning.  
And attribute reliability to a source beyond trusted reputation system. So in terms of information trust, OK, this source is good because it cites.  
Sources, right? Like a page rank kind of. It cites reliable sources, so it that confers certain trustworthiness, right? So the the aspects of trustworthiness.  
Uh, that I'm interested in are uh, source Betty.  
What's the evaluation? Um.  
Information fusion and resolution of conflicting.  
Answers or viewpoints?  
Out on.  
Textual entailment, in other words, interpretation or.  
Evidential reasoning.  
In other words, site your sources.  
And uh.  
This is our this. This has been our approach to graph graph rag anyway, right? This is this underlies our approach to graph rag.  
So this is why I I would like to take this.  
Um and GMN um and I'll grab it like uh.  
For five reasoning also.  
So I'll I'll tell you more about this part, but the basically there's I I've seen a darts of modals, a shrink base.  
And.  
Basically type-based consistency chain, which is very surprising. I asked a question the other day. I said what was the best? What was the last time a DC movie topped the box office charts?  
Before Superbang.  
If you ask anything about Superman right now, it'll.  
That isn't time constraint. It will actually say I assume you mean Superman 2025, but and of course that is what I meant. But if you ask it things that it does not have very high fertility on, it will suddenly switch to saying I assume you mean Superman 1978.  
OK, so I asked that question. It it said the last movie before Superman, which is taken to me in Superman 1978 is Superman 1978 itself. OK, which means it doesn't have any form of representation of what before means.  
It doesn't have any thing to check that before it's not a reflexive, right? It's an irreflexive modality modal operator and it also doesn't have a good type based unification because there's an occur check problem in terms of.  
Anything that you say, what's the best other one besides this one or before this one should never be that thing itself, right? Besides the besides the here reflexivity, there's there should be a tight.  
Constraint too. So it should fail on all of those and obviously it doesn't use any of those. So can we actually use plausible reasoning to give us clue as to those? And can we do sort of optimalistic? I don't know anything I I I know.  
A bit about probabilistic reasoning, or a bit about modal logic and constraint based logic. A bit more about probabilistic reasoning, which I've been working on for 30 years. But.  
Nothing about modal, probabilistic modal logics or or even that much about possibilistic logics. But the point is that LLM at least give us the the logits.  
The basic units of computation. If we want to use an LLM to populate marginals, we can do that. So why not actually try using agentic reasoning with these improvements?  
So that's that's the the quickest way that I know to get at that is to actually take a debate task basically and sort of like a debate scoring task, right?  
And maybe maybe a good side effect would be that it actually does flag misinformation or specious arguments. Yes, Patrick.

 **Patrick Stingley** 1:27:30  
Yeah, I um.  
I hear you. I don't really. So I I think I I'm kind of in agreement, but I'm going to talk about it a little differently when when the when the human brain.  
It takes, it takes in memory all day long into the short term storage and then to process it into long term storage when you go to go to sleep, it connects the your memories into existing contact.  
Concepts typically, unless you have something that's really novel, in which case it makes up something new. But typically for low, you know, just to make it to keep the low energy state, it takes your memories and says, Oh well, that's connected to that.

 **William Hsu** 1:28:06  
Mm.

 **Patrick Stingley** 1:28:20  
And so I would say that rather than an adversarial approach, more of a conversational approach rather, you know, rather than like again, which is an adversarial approach, will this it where it contrasts and it battles to find.  
You know, battles to find a a winner. I I don't think that that's necessarily the approach that I would recommend. I would think that more of a conversational approach where you're translating memories into concepts.  
That would allow you later to if if a concept's not in keeping with the other concepts, this discordance would would make it hard to remember, but it also would give you the tool you need to be able to identify fake news.  
OK, because of it, because the this concept, the idea that it's taking data and and converting it into concepts for long term storage is it's doing a a thing where it's seeing well what's.  
Copacetic with with existing things, and if with fake news or something like that, if it's not copacetic, it'll spot that and eventually if there turns out to be a conflict.  
In the in the grand scheme of things, like if it takes in a bunch of concepts that it thinks is one one way, but eventually you find something that contradicts it, it it'll be able to, it should be able to identify the whole works as well that was.  
All fake news. So anyway, I'm just saying rather than a rather than a adversarial approach, a conversational approach or it might be more helpful because adversarial is just kind of.

 **William Hsu** 1:30:06  
Mhm.  
Yeah, yes, I agree. I agree.

 **Patrick Stingley** 1:30:20  
Yeah, I can see where you would like adversarial because of your Bayesian background, but I I don't think that that I think it's not on and off.

 **William Hsu** 1:30:28  
Mhm.  
I I don't. I don't mean it in this of of assuming that assume the worst behavior in terms of policy induction. I mean in the sense of there's opposing objectives in terms of proving right there's there's one position that's held by one party and.  
And the model of argumentation is based on, as you say, dialogue. But it's there's it's not it's it's not necessarily to try to negotiate a shared resolution.

 **Patrick Stingley** 1:31:07  
Quick.

 **William Hsu** 1:31:08  
And achieve a meeting of the minds. It may be just to convince the other party or defeat the other party's argument.

 **Patrick Stingley** 1:31:16  
What was the and what I'm saying is it's a different approach instead of adversarial. It's it's the yes and approach. It's the I've forgotten what the word is that they used.

 **William Hsu** 1:31:26  
Hmm.

 **Patrick Stingley** 1:31:33  
For for that in like Saturday Night Live skits and so forth, there's a there's a word that they use for the yes and as opposed to.

 **William Hsu** 1:31:38  
Mm.  
Oh, and then.

 **Patrick Stingley** 1:31:46  
Adversarial as opposed to adversarial, which is more yes or no. This is a yes and and because I think it takes less energy to store something when you can add on to something new that exists.  
Rather than something that rather than just simply true or false like a like a Bayesian approach. It's like yes and well can can this work and yes and and yes and now you can get to a point where yes and doesn't work and then you have to go back down the tree to figure.  
Where it fell apart and and people do that. But I'm saying that that give that can give you a continuity of context which makes finding false news easier as opposed to.

 **William Hsu** 1:32:23  
Yeah.

 **Patrick Stingley** 1:32:39  
Because it continues. It's a it's a continuous context. I forgot what that word is. But anyway, just I was listening to what you're saying. I'm thinking that.

 **William Hsu** 1:32:43  
Mhm.  
In in looking at additional ways of resolving disagreement, I I think that there's room for us to explore and compare both. So I'm not. I'm not. I don't have my heart set.

 **Patrick Stingley** 1:33:03  
Yeah.

 **William Hsu** 1:33:06  
On a formal model of argumentation, I rather want to make sure that context includes modality, constraints and types as we can infer them probabilistically.  
I there there's this is a, I would say controversial position because the folks who do formal verification is in this domain want what we in probability call plain beliefs.  
Right 01 probabilities or Kappa 1 minus epsilon probabilities and they don't want anything else. But I think in terms of plausible reasoning this is this is fully Bayesian. It is not.  
It's not a Dempster Schaefferist representation.

 **Patrick Stingley** 1:34:01  
OK.

 **William Hsu** 1:34:01  
If we can actually bound, well, we can also look at variance bounding and in terms of Jahid will tell you the reason why I'm taking this line of inquiry is because.  
I have a project that I'm trying to start on value of information in classification and it's simply a if you can, if you are, if we are already committed to learning to rank and we're already committed to looking at top K, can we look at value of information with respect to top?  
OK, right. So if we want to say, OK, here are the top two theories, top two competing theories entail this position and this position, and these are the conclusions, how can we look at which one has is better supported by evidence?

 **Patrick Stingley** 1:34:52  
Yeah, it gets back to that true or false Bayesian thing. What I'm the the pattern I was thinking of is what is it called improvisational rather than.  
Rather than confrontational, improvisational and over that would retain the context so that you could find if it's astray.

 **William Hsu** 1:35:10  
Mhm.

 **Patrick Stingley** 1:35:20  
That's all. It's just a, yeah, improvisational approach. The yes and so I I just.  
I think I like a lot about what you were saying about the the two different components evaluating against each other and and I'm saying rather than battling each other a yes and approach where maybe even with an eigenvector approach, you know where you go, OK, yes and where's this going?  
Where is this going? Is it going in the right direction? It could be that you'll find your false news because it's the vector is just going in a slightly different way. It's not false, it's just going in a way that's.

 **William Hsu** 1:35:50  
Mhm.  
Mhm.

 **Patrick Stingley** 1:36:05  
Askew.

 **William Hsu** 1:36:08  
That's an interesting um concept. Um, I've uh uh, we've we've been here about an hour and 45 minutes. So uh, I need to to draw uh our meeting to a conclusion. Um.

 **Patrick Stingley** 1:36:14  
Sure, sure.

 **William Hsu** 1:36:20  
I want to make sure that everyone has access to the tools that I've laid out. I'm going to bring you all into the GitHub project, make sure you have access to the repository, assign you all to the the tasks that will let you tap into the data and the actual.  
Seeds so that we can get started on crawling, crawling new data and running the data collection services that Logan and Wesley set up for and that.

 **Patrick Stingley** 1:36:51  
I think that's a great idea.

 **William Hsu** 1:36:54  
I think will be a good start to this task of.  
Trying to explore verifiable reasoning and explore agentic reasoning in.  
Information extraction models or let me rephrase that as used as it as they use information extraction for as a as a test bed for.  
The kinds of things you can do with agentic.  
IE so.  
That's what I think is the unifying theme behind these, behind these domains and I'm.  
Happy to.  
Um.  
Are happy to discuss any follow-up questions that you may have offline. So first of all, thanks to everyone for joining us today and I Patrick, I will follow up with you about the dialogue.

 **Patrick Stingley** 1:38:06  
OK.  
Sure.

 **William Hsu** 1:38:19  
Thank you.

 **Patrick Stingley** 1:38:22  
Talk to you tomorrow.

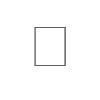
 **William Hsu** 1:38:22  
I'll see you tomorrow too.

 **Patrick Stingley** 1:38:26  
All right.

 **JIA Shuyue** 1:38:28  
Yeah, goodbye, William. I'm gonna talk to you tomorrow.

 **William Hsu** 1:38:32  
Sounds good. See you.

 **JIA Shuyue** 1:38:35  
The William.

 **William Hsu** stopped transcription