

filtering, hybrid search, long-term memory integration with Zep, multi-tenant support, and comparisons between LightRAG and Graphiti. The conversation also covers practical deployment concerns such as GDPR compliance, handling large document volumes, and troubleshooting specific nodes.

 TB**Theodoros Belas** 8h

...

Hi Alan,

thank you for the great work you shared with us. I have a question regarding long-term memory. In the workflow, changes in the analysed documents are taken into account. If a document changes, the stored information is deleted, and the new content is integrated into both the vector store and the graph store.

But what happens if the old information is also stored in the long-term memory through a previous user conversation? That information might ...

See more

Like Reply

1 like

 Alan Walsh Admin 7h

...

AI Automator

Theodoros Belas Hi Theo, you're welcome! Yes, this is a good observation. Yes, there is the potential for this to happen. You could just push the user's messages to Zep and not the assistant messages (see screenshot below). That way, it's only remembering information about the user and what they entered into the chat. You could even have another call to an LLM that checks to see if the information is relevant for long-term memory, and it only calls Zep if so....

See more

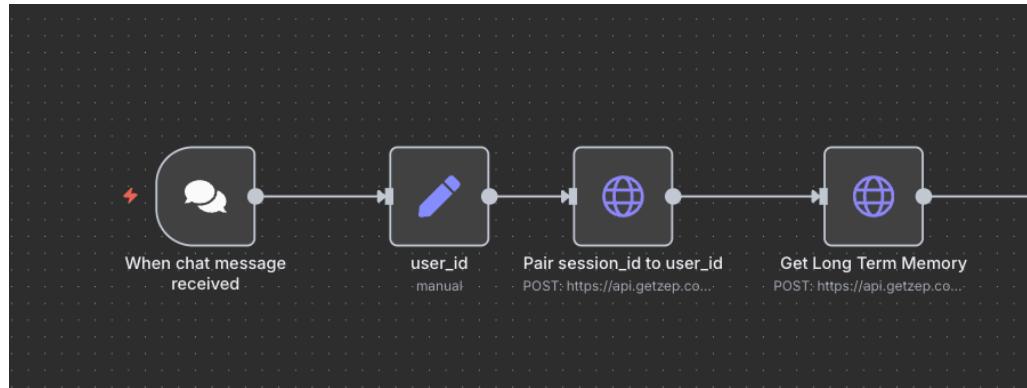
Like Reply

 TB**Theodoros Belas** 6h

...

Yes, i understand, thank you.

What i remarked is that i had to pair the session_id with the user_id to have all user specific conversation stored together, otherwise zep is creating always a new user when the session si changing.



And if the message that shall be stored in zep exceeds 2.500 characters the i shall use zep.graph. in the meantime i stored a ...
See more

Like Reply

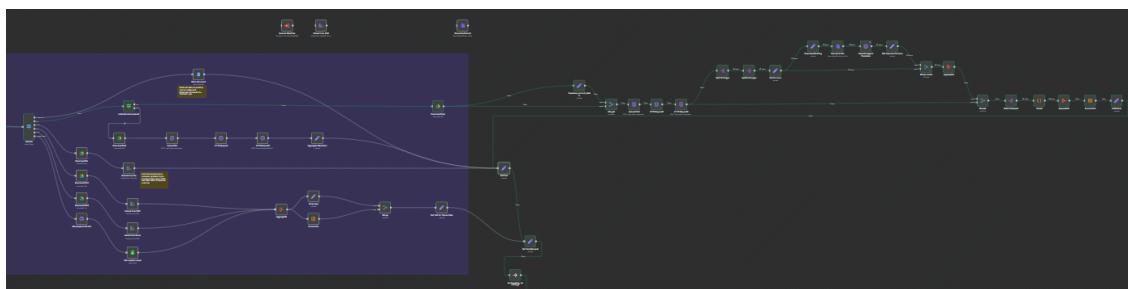


Delmar 2d

Pumped about AI

...

Instead of using Multimodal as a sub-workflow, I just wired it in directly. Does this look correct? The ingestion seemed to work fine... images are stored in the Supabase bucket; the chunks have image descriptions and URLs...



Like Reply

1 like



Alan Walsh

Admin

2d

AI Automator

...

Delmar Hey Delmar - it looks fine to me at high level, two things to keep in mind: 1. There is a switch node here, I think it might be based on the older RAG workflow where we're checking if the PDF is machine readable. You could just delete that switch and route all PDFs through the multimodal flow. 2. If you're getting out of memory errors with this, then using the sub-workflow approach from the current blueprint may resolve that, because memory is freed up once sub-workflows finish executing....

See more

Like Reply

1 like



Delmar 2d

Pumped about AI

...

Alan Walsh I forgot about the memory issue. I'm self hosted; no problems yet but I will keep it in mind.

Like Reply

1 like



Andrew 3d

...

Alan Walsh and **Daniel Walsh** ... Curious if you guys have considered swapping LightRAG with Graphiti (from Zep).

I'm really excited to get this up and running, but a bit confused trying to do that on my own. I've also been building something along the lines of your NotbookLM version, so this is a big step for me as well.

What I'd like is to set up Graphiti as an alternative to GraphRAG—mainly for the temporal capabilities and the more robust options.

...

See more

Like Reply

1 like



Daniel Walsh Admin 2d

...

AI Automator

Hey [Andrew](#), great questions.

Yes, it's definitely on my list to dive deeper into Graphiti - yes, we've included Zep for long term memory in this SOTA agent however I'd like to get under the hood into the framework itself to see could it be used instead of LightRAG.

My "evolving" understanding is that Graphiti and LightRAG are not easy like-for-like replacements.

Yes, they are both knowledge-graphs, however the way they inge...

See more

Like Reply

2 likes



[Andrew Johnson](#) 4d

...

Thanks [Alan Walsh](#). Would this work with all of Hormozi's content (pdf, videos etc) in a Google Drive? 🤔

Like Reply

1 like



[Ethan Davidson](#) 3d

...

[Andrew Johnson](#) Doing this as we speak lol 😅😂, also using the videos I've seen on youtube about ACQ AI to rebuild it. Will let you know when its done! 🤝

Like Reply

2 likes



[Andrew Johnson](#) 2d

...

[Ethan Davidson](#), nice one 🙌 I would love to see what you have created when you are accomplished.

Like Reply



[Alan Walsh](#) Admin 2d

...

AI Automator

[Andrew Johnson](#) 😊 yeah you can index vast amounts of information with these vector DBs, and a few thousand video transcripts would probably result in 10'000s of vectors which should be pretty reasonable. You could use apify to scrape the YT transcripts similar to what I did here in this older Make blueprint:

[The AI-Powered YouTube Researcher \(Summarize LOTS of Videos at Once\)](#). You could get the apify actor to create new Google docs for each video, though you might need to slow down processing ...

See more

Like Reply

1 like



Guillaume LCE 5d

...

"Dare to innovate, create the future." 🚀

hello [Alan Walsh](#), please, in your details, specify the name for hybrid function to use = hybrid_search_v2_with_details

...

See more

Like Reply

1 like

AS

Azmie Sally 5d

...

Great work as always, i tested and build this - everything works smooth, except for one node - Trigger Hybrid Search- the URL for superbase is https://YOUR_SUPABASE_URL/functions/v1/hybrid-search, however, i think the function should be hybrid_search_v2_with_details, this is as per the file - hybrid_search_database_function.txt. - because i do get the error- **The resource you are requesting could not be found** Requested function was not found. Any thoughts on this ? I do also, want to see, what happens if you load 1000 documents ? how do you guys think thi...

See more

Like Reply

2 likes



Alan Walsh Admin 5d

...

AI Automator

Azmie Sally Hi Azmie, thank you! You should replace that full URL with whatever you see here within your supabase *edge functions*, *this should include your full supabase project URL (see below screenshot)*. The “hybrid_search_v2_with_details” is actually the *database function* for this, whereas the edge function makes the API endpoint available that we can hit from n8n. This edge function then calls the database function (which includes all the hybrid search logic), and the edge function then relays its response bac...

See more

Like Reply

1 like

Azmie Sally 5d

...

I did managed to fix it by Directly calling the RPC - /rest/v1/rpc/hybrid_search_v2_with_details

Like Reply

AS

Azmie Sally 5d

...

Alan Walsh Great thanks, let me make the necessary changes, as well and see how it turns out. Furthermore, I feels, though we both lightrag and hybrid search most of the documents i uploaded use the vector rather than the knowledge — the data sources i tried were mostly Samsung appliances, just want to see how it handles. What is the logic behind choosing vector or graph or both ? Is it part of the prompt saying look for information from graph, or does the ai make intelligent choice. Or and one more think, using Azur...

See more

Like Reply

1 like

Show more replies

A

Alex 5d

...

OOOOO that's my weekend full of fun :-)

Like Reply

2 likes

 **Robert Tellier** 5d

...

Follow up request 😊 **Alan Walsh** and **Daniel Walsh** - Love the work you are putting in.

Multi tenant follow up question

I am investigating how to update the state of the art - modular RAG system to multi-tenant at organization level (client_id). Are you planning to do that soon? I love to include the system in our content creation platform but need it for a multi client approach ...

Like Reply

1 like



Alan Walsh Admin 5d

AI Automator

...

Robert Tellier It depends on how you're planning to set this up. We might add some instructions for multi-user at some point - whereas multi-tenant adds a layer of complexity, especially if you want users from those tenants interacting with the agents directly.

If you're looking to set up multiple tenants with multiple users per tenant then one approach is that you could have a lookup table that maps users → and what tenants they're part of. In that case, you could use Daniel's approach of getting the User ID in this video: ...

See more

Like Reply

1 like



Robert Tellier 5d

...

Thanks **Alan Walsh**, I will go for a unique system setup for every customer first (max 5) to see and experiment what works and what doesn't. Also I didnt know that the GrapghRAG is not multi tentant

Like Reply

1 like



Kelly Cromwell 6d

...

Your RAG masterclass is what brought me to the community and the fact that you keep leveling it up is much appreciated, definitely think you are well ahead of other influencers in the space, for sure.

For me, one major hiccup that is getting in the way of deploying a system into production for our agency is the known issue with the google doc node that seems to struggle reading/saving documents from shared drives. Apparently it's existed for months, does anyone know if there is ...

See more

Like Reply

2 likes



Daniel Walsh

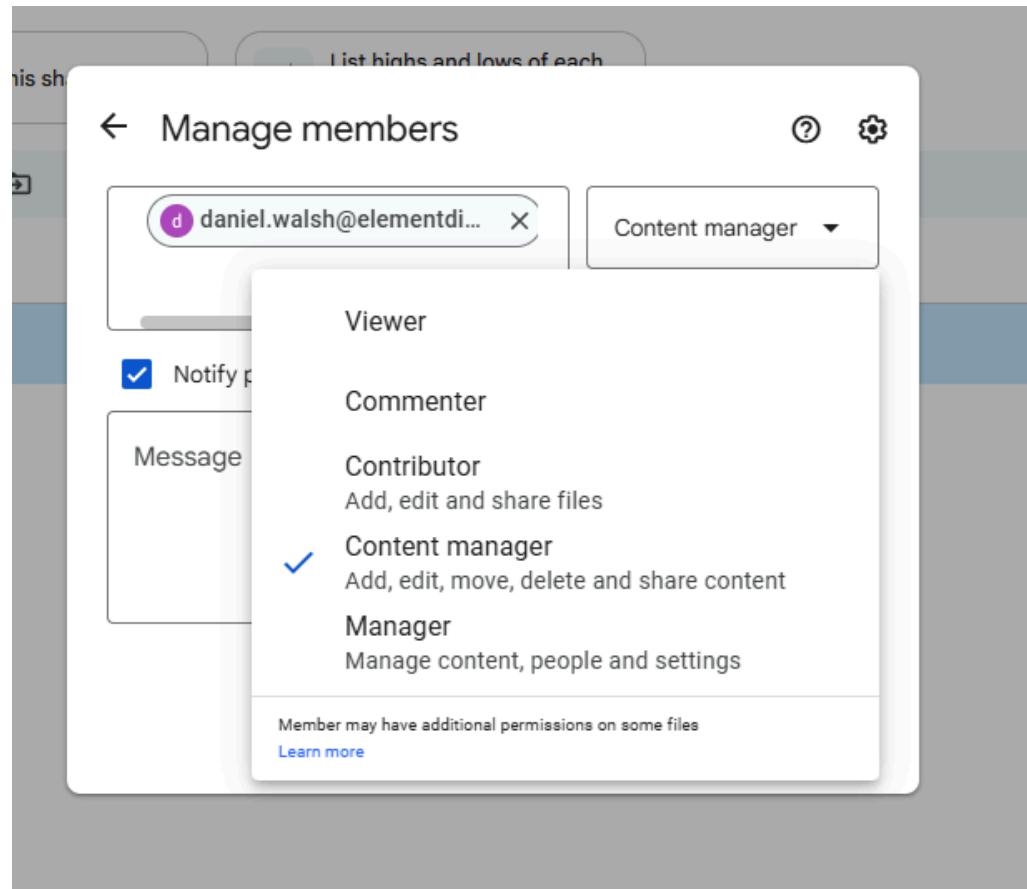
Admin

5d

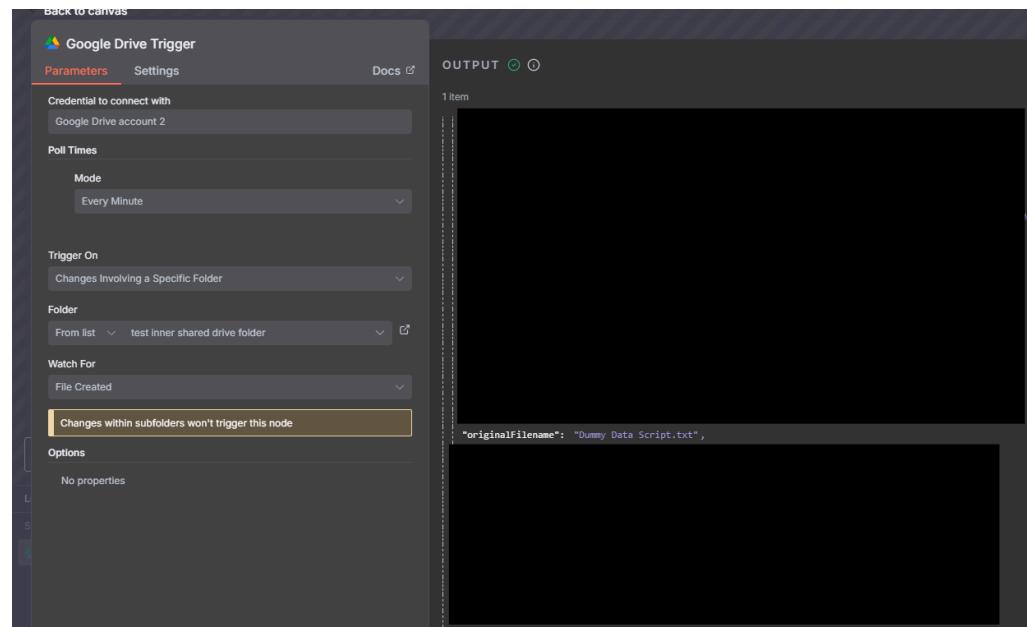
AI Automator

...

Hey [Kelly Cromwell](#) - I think this may just be a permissions issue of the member in the shared drive.



I just tested the Google Drive Trigger on a Shared Drive I created and it worked for me (with the action set to File Created)





Dominik

Hey,

- + New
- Home
- Activity
- Workspaces
- My Drive
- Shared drives

first of all, thanks for the upgrade. Great work 😍

I haven't updated the older version yet because I'm currently planning how to set everything up.

Our main constraint is GDPR compliance: we need to host all components either in our own datacentre or Azure-only.

Right now I'm experimenting with ZEP. I haven't found a straightforward ...

See more

In this case, the account that is configured as the connection in ... 1 like



Alan Walsh

Admin

6d

AI Automator

...

Hi [Dominik Heiß](#) - good question! Yes, the Zep community edition is no longer supported. Graphiti is the open-source framework that I believe Zep's memory service uses, so if you're looking to go the open-source self-hosted route, that could potentially be an option, but I haven't used it. [Daniel Walsh](#) tried out using Graphiti when he was doing a deep dive into knowledge graphs, so he might be able to advise further.

We're using LightRAG as the knowledge graph for the document ...

See more

Like Reply

1 like



Daniel Walsh

Admin

5d

AI Automator

...

Alan Walsh Dominik Heiß I didn't get too far in to trying out Graphiti actually. I was only using the Zep interface and checking out the Graphiti documentation and chatting to the codebase in Cursor!

I don't think LightRAG and Zep are interchangeable. Zep is a temporal knowledgegraph for tracking users preferences and memories over time.

LightRAG is built for generic document ingestion with entity and relationship extraction - while it does have some level of de-...

See more

Like Reply

Vincent Niou 6d

...

Hello [Alan Walsh](#) and [Daniel Walsh](#). First-off, thanks so much for this. Seriously great stuff.

I have a specific question regarding combining hybrid search and contextual embedding, specifically when it comes to keyword vectorization and the contextual summaries appended to each chunk. For example, here's a chunk that includes the contextual summary:

```
1 This chunk is part of the Business Strategy Implications section, focusing
  on enhancing partnerships and accelerating feature development to improve
  eSIM upsells and competitive positioning in the South-East Asia travel eSIM
  market.
2
3 ---
4
5 Songkran, Eid) to maximise share of voice when overseas bookings surge.
6
7 ### Business Strategy Implications
8
9 - Deepen alliances with airlines, OTAs and digital wallets to embed eSIM
  upsells into booking flows.- Accelerate development of competitive features
  (VPN, eSIM sharing, validity trackers) to match or exceed rival offerings.
```

See more

Like Reply

1 like



Alan Walsh Admin 6d

...

AI Automator

Vincent Niou Hi Vincent - glad you like the workflows! In Postgres Full Text Search (FTS), it breaks these up into stemmed words and they are stored with their positions as they appear in the document. For example, 'esim':21,33,59,70 means the word "**esim**" appears 4 times at those positions in the text (see below image). So it looks like this is capturing both the contextual summary and the chunk. Does this answer your question?

```
1 This chunk is part of the Business Strategy Implications section, focusing
2 on enhancing partnerships and accelerating feature development to improve
3 eSIM upsells and competitive positioning in the South-East Asia travel eSIM
4 market.
5 Songkran, Eid) to maximise share of voice when overseas bookings surge.
6
7 ### Business Strategy Implications
8
9 - Deepen alliances with airlines, OTAs and digital wallets to embed eSIM
upsells into booking flows.- Accelerate development of competitive features
(VPN, eSIM sharing, validity trackers) to match or exceed rival offerings.
```

It appears that words that are in the contextual summary (but not in the actual content chunk) are being vectorised in the 'fts' column as well. See below:

```
'acceler':16,64 'airlin':52 'allianc':50 'asia':31 'book':44,62 'busi':7,46 'chunk':2
'competit':24,67 'deepen':49 'develop':18,65 'digit':55 'east':30 'eid':36 'emb':58
'enhanc':13 'esim':21,33,59,70 'exceed':77 'featur':17,68 'flow':63 'focus':11
'futur':12,13 'implic':12,13 'intend':11,12,13 'maxim':11,12,13 'minim':11,12,13 'overseas':11,12,13
```

[See more](#)

[Like](#) [Reply](#)



Daniel Walsh Admin 6d

AI Automator

...

Hi **Vincent Niou**, great question. As Alan mentions above, the FTS column reflects both the chunk and the contextual snippet (not just the snippet)

On the question of does having the contextual snippet materially included in the FTS field reduce hybrid search performance; I honestly don't know. This would probably take some testing and evaluation to figure out.

I can imagine scenarios where it may improve accuracy. In your ...

[See more](#)

[Like](#) [Reply](#)

VN

Vincent Niou 6d

Appreciate the quick response both! Very helpful 👍

Like Reply

...

1 like



Guillaume LCE 7d

"Dare to innovate, create the future." 🚀

...

Hello, first thank you [Alan Walsh](#) and [Daniel Walsh](#) !

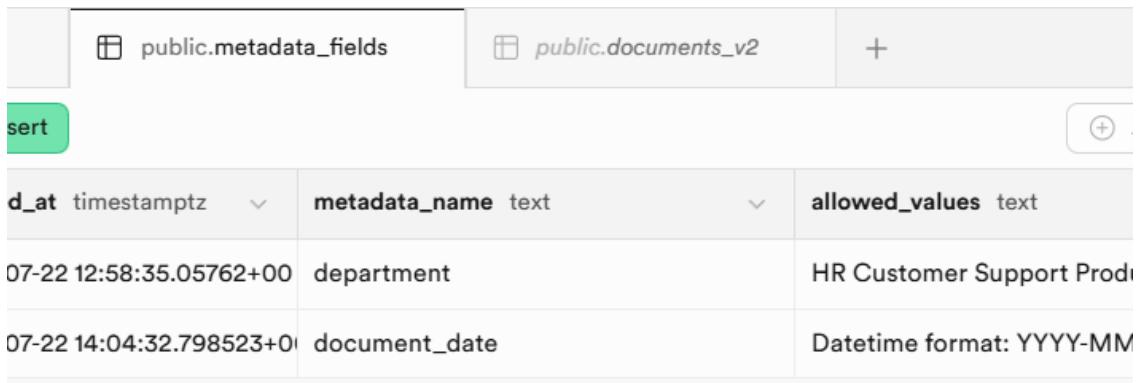
If I'm right, in this node, delete metadata "season" and "motors_category", and write "department" (name and value)

The screenshot shows the 'Default Data Loader' tool interface. It has tabs for 'Parameters' (selected) and 'Settings'. There are two property configurations:

- Name:** doc_type
Value: `fx {{ $('Set Data').first().json.doc_type }}`
[ERROR: No execution data available]
- Name:** department
Value: `fx {{ $($('Enrich').item.json.output.department }}`
[ERROR: No execution data available]

At the bottom, there is a button labeled 'Add property'.

because no (season and motors_category)



public.metadata_fields		public.documents_v2	+
sert			
d_at timestampz	▼	metadata_name text	▼
07-22 12:58:35.05762+00		department	HR Customer Support Prod
07-22 14:04:32.798523+00		document_date	Datetime format: YYYY-MM

Like Reply

3 likes



Alan Walsh Admin 6d

AI Automator

...

Guillaume LCE Hi Guillaume - yes you're right. Thanks for the feedback on that. I'll make a note to update this in the next minor version of the blueprint. Those metadata fields in the database are just there as an example, but it would be good if the template were consistent with it!

Like Reply

1 like



Robert Tellier 5d

...

Guillaume LCE **Alan Walsh** Thanks for the question - as I'm learning how the system (metadata filtering) works this did give me some confusion... 😊

Like Reply

1 like



David Gikandi Aug 20

...

oooooh this is sooooo exciting! been waiting for this for ages, so glad i found you. thank you!!! digging in now. this is a game changer for sure. thanks

Like Reply

2 likes



Sam Patel Aug 19

...

Awesome sauce, everything I need to make the abusive dietitian of my dreams.



Like Reply

2 likes



Mike Montefusco Aug 19

...

This is incredible honestly. This would have taken me thousands of hours to figure this all out.

Like Reply

2 likes



Philip COUBOURA Aug 19

...

Awesome [Alan Walsh](#) ! This alone is worth the annual membership. Btw you should think about delivering a certificate for those who manage to through the whole masterclass and implement it properly, a kind of Lab with a certificate of completion 😊 .

Like Reply

3 likes



Alexandr Aug 19

...

Very cool!

This is what I need!

How do I start studying these materials?

Should I follow the prompts or will there be a video that will allow me to put this flow together?

Like Reply

1 like



Alan Walsh Admin Aug 20

...

AI Automator

Alexandr Thank you! If you're looking to understand the concepts in this template as well as how everything fits together, I'd recommend you first go through our [#n8n RAG Masterclass](#). This template is a follow-on from those, so the masterclass should put you in a good position to be able to set this up. You could treat this template as Lesson 10 of the RAG Masterclass, but I've uploaded it as a standalone system template outside of the masterclass because this is a self-contained template....

See more

Like Reply

1 like

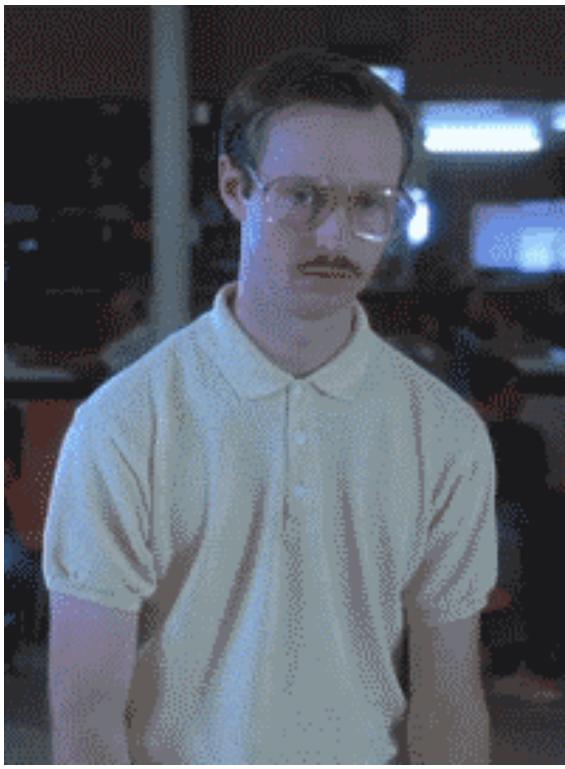


Delmar Aug 19

...

Pumped about AI

Wow



Like Reply

2 likes



Stuart Asta Aug 19

[Delmar](#)

...



Like Reply

2 likes



Robert Tellier Aug 19

...

Will start working on it first thing tomorrow. Cool

Like Reply

2 likes

KB

What are your thoughts?



Post