Visualizing Structure in an Unstructured World

written by DaltonRuer | February 1, 2025

Background

30 years ago I moved my family to the Atlanta, Georgia area to help start a new company. Yeah .com era. My wife was a dental hygienist and quickly found a job with a fantastic dentist who was beginning her practice. Over the course of time our families grew very close. Eventually my wife transitioned from doing dental hygiene for the wife and began consulting with the husband to help other dentists capture missing insurance payments. At one point I even helped him. He later started an appointment reminder company and my wife joined that organization. While she has retired recently, we are still in touch with this amazing entrepreneurial couple. So, when he reached out to me and said “Can you help me visualize some data?” I was like “Heck yes.”

We had a brief conversation and I found out he was now running yet another dental focused business called Practice Matchmaker. He had a message board where people posted messages wanting to sell a practice, or buy a practice or find an employee or search for a job. “That seems easy enough Allen, please just send me some sample data and I should be able to build a dashboard in an hour or two.”

How hard could it be?

Guessing that like me, you assumed the message board had well structured forms that accepted inputs and the sample data I was expecting to see would have a myriad of columns. One table of seller data values. A table of buyer data values. A table of job posting values. Etc. I guess I should have known if it was that easy he probably wouldn't have asked for my help.

Low and behold when I opened the Excel file all of the data was in a single worksheet.

No separate tables, and worse yet no myriad of columns. The worksheet had only 12 columns in it. If that doesn't seem like enough data to you, it's ok … I was scratching my head as well. Until my worst fears were realized: 1 of the 12 columns was simply labeled "Content." It was a free-form, unstructured column where buyers, sellers, job applications etc., simply hand typed their messages. "Oh yeah it is from a message board where people just write to each other with no regards for Allen or the Qlik Dork's desire to analyze and gain intelligence from their chicken scratch."

Unstructured

Notice: Don't panic I have massaged the data values and there no real names, numbers, emails or anything but I did want you to see the context of the situation.

Here are some of these sample messages:

- Alex Carter: Hello! I've been practicing for 7 years and am now ready to take the next step into practice ownership. I'm looking to purchase a practice in Los Angeles with a minimum of 3 operatories. Prefer FFS/PPO (No Medicaid/HMOs). I'm also open to a part-time associateship that could lead to acquisition. Please comment below or send me a private message with any

leads. Thank you! #looking #buying #losangeles #socal #california 2023-03-05 14:01:54

- Jordan Blake: I’m interested to see if there are any associate doctors looking to work anywhere from 1-5 days a week in Santa Barbara, California. We are a general dental office looking to expand and strengthen our team. #associate #hiring #santabarbara I’m also interested in purchasing another dental office anywhere from Goleta to Ventura if anyone is considering selling. #buying 2023-07-20 2:21:54
- Taylor Morgan: Looking to purchase a pediatric dental practice in Sacramento, Placer County, or surrounding areas. If you have any leads or are considering selling, please reach out. Thank you! #buying #california #sacramento #roseville 2023-07-20 17:55:07

Like me, you can read the messages and your brain can figure out what is going on. We could easily put this text into a comment field and let an end user read it. Right?

But how would you create structured data visualizations out of these unstructured comments? You can’t. Neither can I. So, my heart kind of sank. This close personal friend that I’ve known my entire life in Atlanta, who had basically provided job opportunities for my wife’s entire working career, needed help making sense of these chaotic and ill-formed unstructured text messages and I couldn’t do it. Well … maybe … what if I …

Reading and Inference

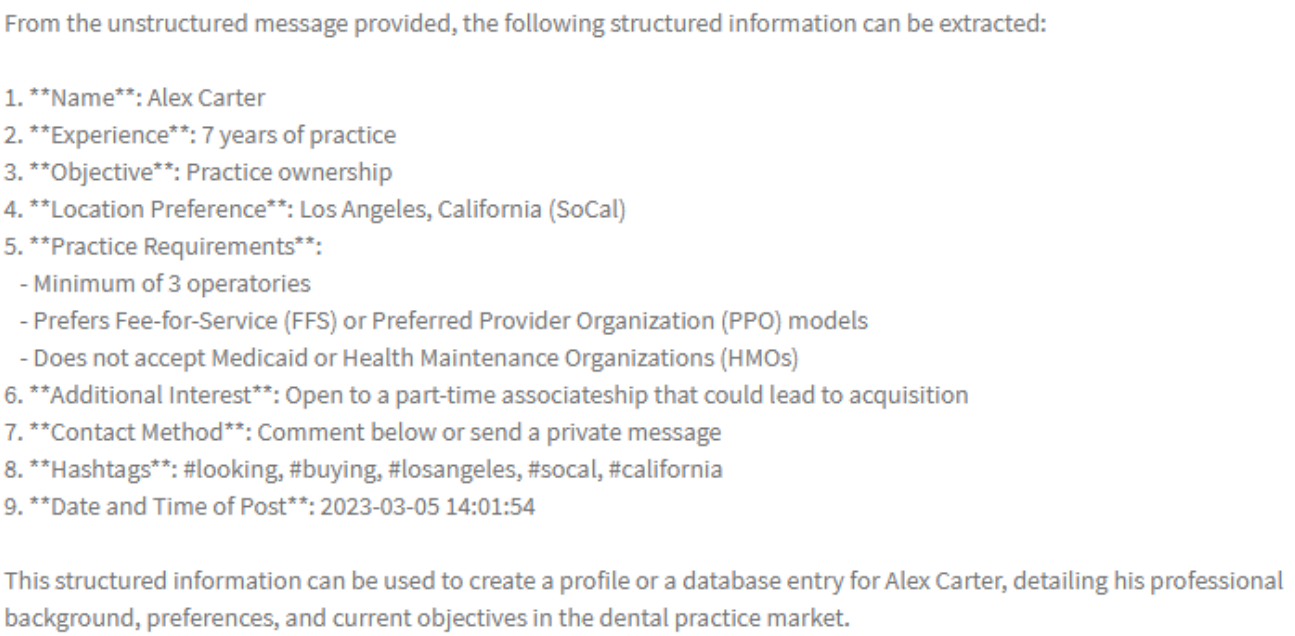
Quick Question: You know what is good at inferring information from text just like humans?

You got it my friend … Large LANGUAGE Models (LLM’s).

The Qlik Dork could just pass in Allen’s nasty data into an LLM with a prompt like:

What structured information can you pull from this unstructured message?

Bing-bang-boom I would get data in the way I could use it. The response I got was:



From the unstructured message provided, the following structured information can be extracted:

1. **Name**: Alex Carter
2. **Experience**: 7 years of practice
3. **Objective**: Practice ownership
4. **Location Preference**: Los Angeles, California (SoCal)
5. **Practice Requirements**:
   - Minimum of 3 operatories
   - Prefers Fee-for-Service (FFS) or Preferred Provider Organization (PPO) models
   - Does not accept Medicaid or Health Maintenance Organizations (HMOs)
6. **Additional Interest**: Open to a part-time associateship that could lead to acquisition
7. **Contact Method**: Comment below or send a private message
8. **Hashtags**: #looking, #buying, #losangeles, #socal, #california
9. **Date and Time of Post**: 2023-03-05 14:01:54

This structured information can be used to create a profile or a database entry for Alex Carter, detailing his professional background, preferences, and current objectives in the dental practice market.

Figure 1: First response for a message asking it to give me the structured information contained in a free form text field

Look at that baby! It is so much easier to read. Right? Then it dawned on me … while it is easier to read, that is still very much unstructured data. Worse yet, was the fact that when I asked the same question for other messages I got different returns each time. Oh man, I should have known I would be foiled. All this LLM stuff is just hype anyway:

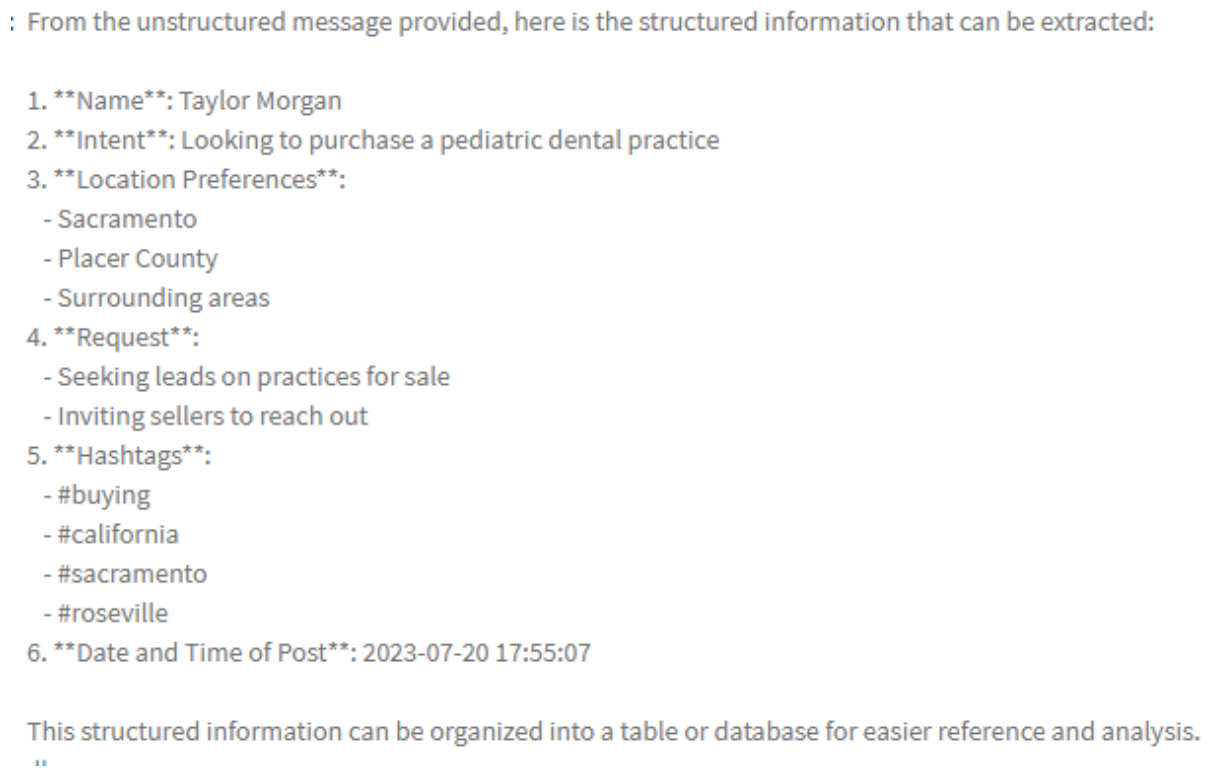


Figure 2: Second response for a message asking it to give me the structured information contained in a free form text field which is good but different than the first one. It's visualizing structure in an unstructured world but not in a form I can use

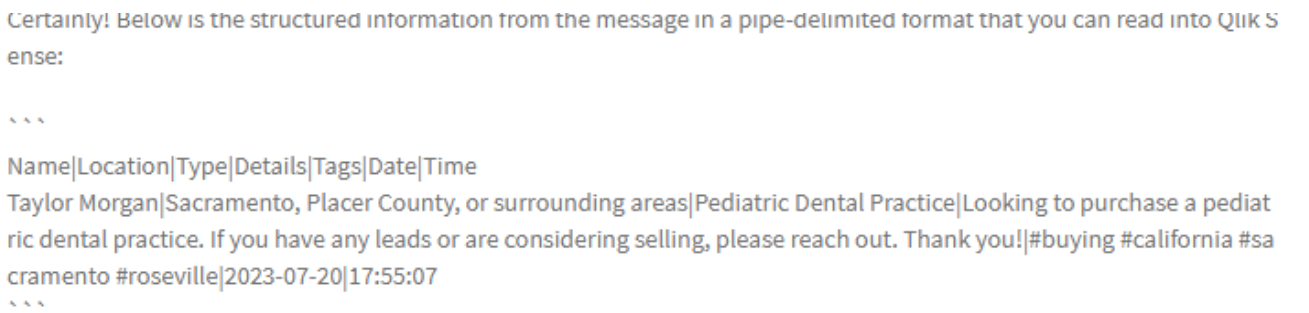
Ask a better question

As a [Chief Question Officer](), I know that solutions are often uncovered when you ask better questions. In this case I was simply asking for the structured information, but wasn't clear on the format. I had a very narrow expectation of what the technology could do. Like many, I got excited at the thought it could 1 simple task to make my life easier but wasn't really pushing it. I needed to be able to load the "structured" data into a Qlik Sense application if I was going to create some spinning 5 dimensional pie charts that gave Allen a back rub. But I hadn't asked for that. So I tried again:

Can you pull the structured information from this message

> ***and put it into a pipe delimited formatted that I could read into Qlik Sense?***

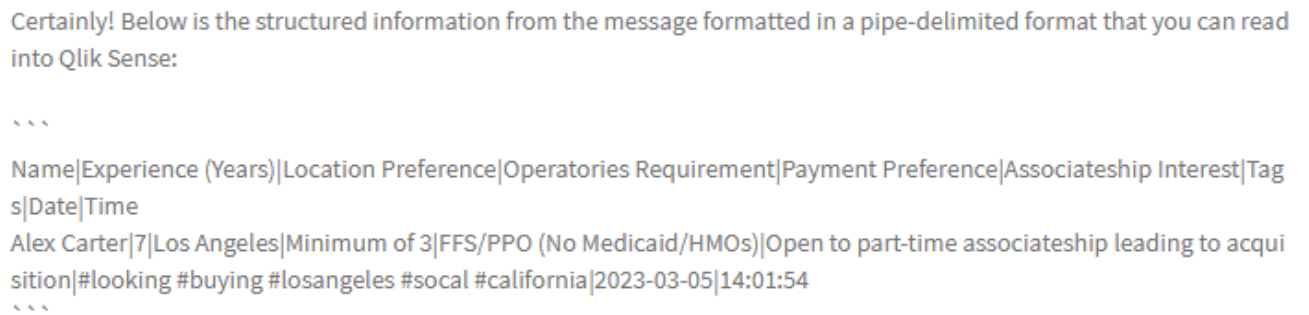
Woohoo … now we are cooking with gas.


Certainly! Below is the structured information from the message in a pipe-delimited format that you can read into Qlik S
ense:

```
Name|Location|Type|Details|Tags|Date|Time
Taylor Morgan|Sacramento, Placer County, or surrounding areas|Pediatric Dental Practice|Looking to purchase a pediat
ric dental practice. If you have any leads or are considering selling, please reach out. Thank you!|#buying #california #sa
cramento #roseville|2023-07-20|17:55:07
```

Figure 3: With a better question the response is returned in a pipe delimited format to ingest into Qlik Sense

But shoot, again I had an issue with a different record:


Certainly! Below is the structured information from the message formatted in a pipe-delimited format that you can read
into Qlik Sense:

```
Name|Experience (Years)|Location Preference|Operatories Requirement|Payment Preference|Associateship Interest|Tag
s|Date|Time
Alex Carter|7|Los Angeles|Minimum of 3|FFS/PPO (No Medicaid/HMOs)|Open to part-time associateship leading to acqui
sition|#looking #buying #losangeles #socal #california|2023-03-05|14:01:54
```

Figure 4: Great response again with a pipe delimited format to ingest into Qlik Sense, but still different than figure 3

Try again

It’s almost like this rusty bucket of bolts needed me to be even more specific on the DATA MODEL that I needed. Shouldn’t LLM’s be smart enough to read our minds and know we have thousands of messages and we need it to return the same thing for us every time? That’s not how it works my friends. You

need to share all of the information that is in your head. Give it your hopes and dreams but more importantly provided the needed CONTEXT and SPECIFICITY.

> ***Can you pull the structured information from this message and put it into a pipe delimited formatted that I could read into Qlik Sense? The fields I need are Buyer Name, Dentistry Type, Goal Type, Desired Location, Post Date, Transition Timeline, Transition Preference and Operatory Room Requirements***

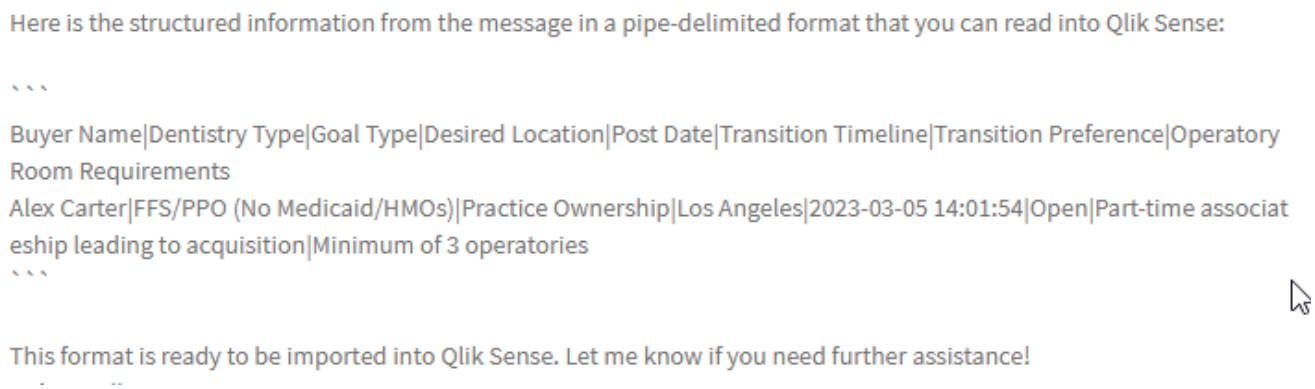


Figure 5: A more consistent response driven by a more clearly defined question. Visualizing structure in an unstructured world requires better prompts.

That worked out great and in fact was very consistent as I tried message after message after message. While I could easily parse that out in Qlik Sense, I didn't want the hassle of the 10 seconds needed to remove the header row. Qlik Dork is a busy man you know.

Then it hit me, I needed my response in more of a code friendly manner. I know what you are thinking … "Qlik Dork, LLM's are really good at generating code." You are so right … I should have thought about that. Perfect technology for inferring information from language and really good technology for most coding type tasks. Glad I have you here to push me to go further:

> ***You are a data modeler and I need the following structured information that must be in a well formed JSON format that can be loaded by an application so no extra text above or below the json, and the field names should be in upper case but the data values should remain as is. The fields I need are Buyer Name, Dentistry Type, Goal Type, Desired Location, Post Date, Transition Timeline, Transition Preference and Operatory Room Requirements. If there are no values for Transition please state No Preference. Also please add a recommendation string on what you think might be needed to enhance the post so it would stand out and call that field RECOMMENDATION***

For the sake of your time I will skip a few of the iterations I went through in terms of step by step but didn't want to share them with you. Before asking for the field names in upper case, I had results that used 3 different formats for casing on the field names. I needed consistency so I asked for it. When no transition information was found sometimes I got null, and sometimes I got something else. I also needed to include the in my visualizations so Allen and his team could select those with no preference. So I asked for it. One of the things that Allen and I discussed when I talked about the solution I was proposing was actually asking for a recommendation that could be shared with the message poster to aid getting more attention for their post. So I asked for it.


```
{
 "BUYER_NAME": "Alex Carter",
 "DENTISTRY_TYPE": "FFS/PPO (No Medicaid/HMOs)",
 "GOAL_TYPE": "Practice Ownership",
 "DESIRED_LOCATION": "Los Angeles",
 "POST_DATE": "2023-03-05 14:01:54",
 "TRANSITION_TIMELINE": "No Preference",
 "TRANSITION_PREFERENCE": "No Preference",
 "OPERATORY_ROOM_REQUIREMENTS": "Minimum of 3 operatories",
 "RECOMMENDATION": "Consider adding more details about the desired practice size, patient base, and financials to at
tract more targeted leads. Highlighting any specific equipment or technology preferences could also help."
}
```


Figure 6: Response formatted as a JSON block with consistent

field names in an easily ingestible format that Qlik Sense can utilize for visualizing data in an unstructured world

Finish it off

How slick is that? This magical LLM was able to set me up for success in visualizing structure in an unstructured world. Now all I had to do was read those field values.

You might be thinking “But Qlik Dork that JSON code isn’t any better than the plain text you were getting originally. It’s still unstructured and it’s not like Qlik Sense allows you to easily pull the values from the JSON blocks.” Aha! You are wrong my friend. Qlik Sense does indeed have functions to deal with JSON, cause you know it’s awesome like that. Easy peasy lemon squeezy, I was able to pull each value directly out of the JSON block returned by my LLM. By using the JSON structure, you can also use that response anywhere else you might want, which is why I handed it to you on a silver platter.

```
50  Left Join (Buying)
51  Load [Question#],
52      JsonGet([content], '/BUYER NAME') as [Buyer Name],
53      JsonGet([content], '/DENTISTRY TYPE') as [Dentistry Type],
54      JsonGet([content], '/GOAL TYPE') as [Goal Type],
55      JsonGet([content], '/DESIRED LOCATION') as [Desired Location],
56      JsonGet([content], '/POST DATE') as [Buying Post Date],
57      JsonGet([content], '/TRANSITION TIMELINE') as [Buying Transition Timeline],
58      JsonGet([content], '/TRANSITION PREFERENCE') as [Buying Transition Preference],
59      JsonGet([content], '/OPERATORY ROOM REQUIREMENTS') as [Buying Requirements],
60      JsonGet([content], '/RECOMMENDATION') as [Buying Recommendation]
```

Figure 7: Qlik Sense code using JSON functions to get the values out of the response so that the data can be visualized, filtered, aggregated on etc.

LLM Choice

I have created many different Qlik Community Posts and Qlik Dork videos on many different LLM's out there. If we were speaking privately I could share with you my preferences for different tasks. But for this particular task I selected an LLM that is currently receiving a ton of hype … Deepseek just for the fun of playing with it. "Look at the Qlik Dork on the bleeding edge."

Figure 8: Deepseek logo

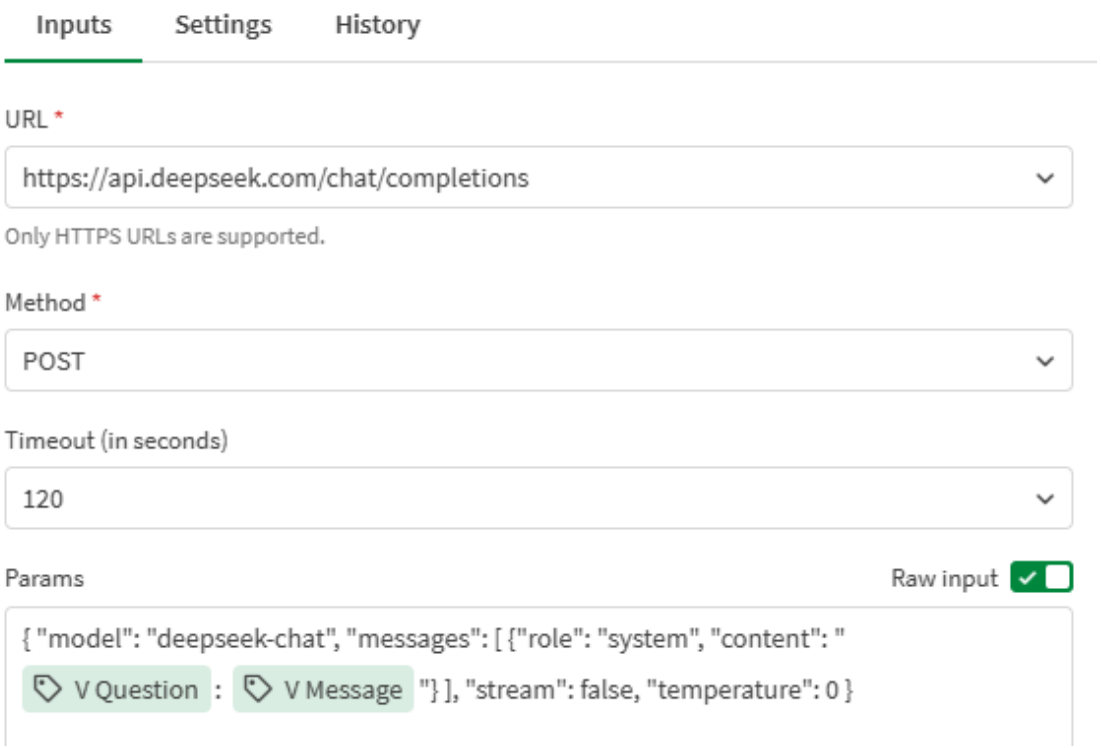


Figure 9: Qlik Application Automation Call URL block calling deepseek api to get responses

You would probably expect more of me than to do it just because I wanted to play with it, and to stay above the “hype”, and you would be right. So bear with me for a second and understand:

I am not recommending Deepseek. Their policy is to keep everything and that is scary with real data. I was passing faked names and other fake information just to test. The reason I wanted to make you “*gasp*” when I mentioned it was so I could bring up the one point that is driving everyone in the world crazy.

The low low low price of this thing. While you may not want to use it in the open internet, you very well may want to take the open source and run it in some secured network of your own making. The low processing requirements are blowing peoples minds. While I’m still not advocating for it, the fact that AWS and Snowflake now offer Deepseek in secured environments for you, after being tested and proven, might eventually become something you may consider.

So, I wanted to conduct the test for this fun one off process for my friend Allen using this because it was really cheap (and that dude is frugal). Plus I know how to manipulate and mask the data so sharing faked data didn’t bother me.

How low priced are we talking about? I got 1,000,000 tokens for $2. I meant I could have also bought a Double Big Gulp of Mountain Dew for that same price, but the price was worth it for a test. My prompts ranged anywhere from the low 300’s to the low 400’s for tokens used for each question.

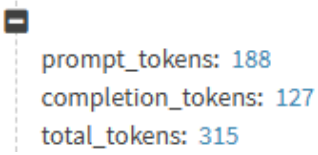


Figure 10:

Screenshot of the fact it took 315 for one of the prompts

Let’s be conservative for our cost calculations and error on the high side, and say 400 tokens per question. That comes out to 2,500 messages that could that I could visualize structure for for $2. That’s 11.5 messages processed for 1 penny. That’s affordable. That meant I didn’t need to smack myself in the head, or sacrifice a second or third Double Big Gulp of Mountain Dew trying figure out how to tune my prompt to give me exactly what I needed. It also afford Allen, and Practice Matchmaker, the chance to visualize structure in their unstructured world comprised of free form messages. By asking for recommendations it affords them to ponder how they could become PraticeMatchmaker.ai and utilize AI in a way to guide their clients to succeed with an affordable price point.

Switching Gears

I know my audience, and I know you aren’t buying or selling a dental practice. But my friend Allen’s “problem point” is the same as in your organization. More than likely you have “unstructured notes/comments/feedback/surveys” but the manual effort to pull useful structured information out of them is more than you wish to deal with. So, that unstructured data just accumulates and collects dust. At the same time you want to utilize AI, in a real and meaningful way. I want my post to be like a matchmaker to help you practice thinking about AI in those situations. (See what I did there?)

So, let’s say you take my recommendation to heart to think about using an LLM and you just happen to be in the healthcare space. You utilize a partner for surveys and they do a great

job for of you aggregating the NUMBERS on the surveys. But what about all of that unstructured data the comments they write by hand and their is no structured # to aggregate? Maybe something like this:

> **The nursing staff was incredibly rude and completely dismissed or ignored questions me and my family had about my care. In some cases I actually saw them roll their eyes, one to the other. The clinical care I received was excellent and I was healed and out of the hospital faster than my surgeon originally told me it would take.**

You've heard about Sentiment Analysis but let me ask you for real, does a value like .62 negative do much for you? Oh sure you can aggregate the numbers up so someone gets a bonus for "65% of our patients reported positive surveys." But does the .62 help you know what actions you need to take internally to ensure you drive that number higher? Not a chance.

Are you on the edge of your seat waiting to see what response I get? I sure am so let's get going. First thing I need to do is change my question.

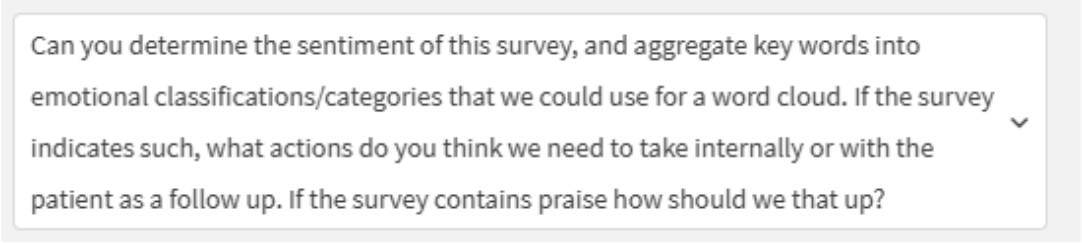


Figure 11: Healthcare question asked to deepseek to analyze a survey comment

Then I changed my API call and instructed it to use the "deepseek-reasoner" model instead of just the chat model. I also told it to pretend it was a quality improvement analyst. You know, not just some bean counter tracking .62's all day long.

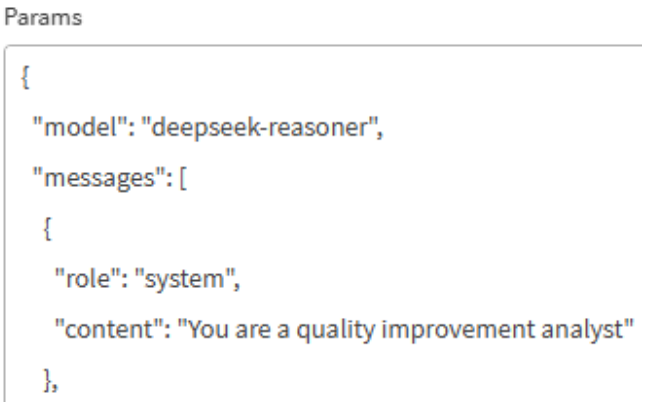


Figure 12: Screenshot of Qlik Application Automation Call URL block to call the deepseek reasoner model and act as though it is a quality improvement analyst

I might be bias, but I think these results are insane. I've repeated this particular exercise on nearly every LLM I have access to, and truth be told, this reasoning model generated the best response I've seen to date. I just asked the question directly, one time, without any tuning. I'm sure as I get more familiarity with deepseeks "reasoner" model, I will understand just how far I can push it.

Sentiment Analysis

The survey reflects a **mixed sentiment** with **strong negative emotions** toward the nursing staff's communication and bedside manner, contrasted with **positive sentiment** about the clinical outcomes and efficiency of care.

Key Words & Emotional Classifications

Negative Sentiment Categories

1. **Frustration/Disrespect**: "incredibly rude,"

"dismissed," "ignored," "roll their eyes."
2. **Lack of Communication**: "ignored questions," "dismissed concerns."

Positive Sentiment Categories

1. **Gratitude/Satisfaction**: "excellent clinical care," "healed faster," "surgeon's expectations exceeded."

Word Cloud Suggestions

- **Negative focus**: Rude, dismissed, ignored, roll eyes, frustration.
- **Positive focus**: Excellent, healed, faster, clinical care, gratitude.

Recommended Actions

Addressing Negative Feedback

1. **Staff Training**:
   - Conduct **empathy and communication workshops** for nursing staff to address dismissive behavior and improve patient-family interactions.
   - Implement **role-playing scenarios** to practice responding to questions respectfully.
2. **Patient-Family Communication Protocol**:
   - Establish a structured process for nurses to acknowledge and address patient/family concerns (e.g., scheduled Q&A times).
3. **Internal Follow-Up**:
   - Investigate the incident with the nursing team involved and provide constructive feedback.
   - Consider anonymous patient satisfaction

surveys post-discharge to identify recurring issues.

Patient Follow-Up

- **Direct Apology**: Have a patient experience manager contact the patient to apologize for the negative interactions and acknowledge their concerns.
- **Commitment to Improvement**: Share steps being taken to prevent recurrence (e.g., training initiatives).

Amplifying Positive Feedback

1. **Recognize Clinical Team**:
   - Publicly praise the clinical/surgical team in internal communications (e.g., newsletters, team meetings) for delivering efficient, effective care.
   - Consider awards or incentives for teams that exceed recovery expectations.
2. **Share Success Stories**:
   - Use anonymized positive feedback in staff training to highlight ideal patient outcomes.

Word Cloud Design Tips

- Use larger fonts for impactful negative terms (*rude, ignored*) to emphasize areas needing improvement.
- Balance with positive terms (*excellent, healed*) in smaller sizes to reflect gratitude while prioritizing actionable feedback.

This approach addresses systemic issues while reinforcing strengths, fostering a culture of accountability and

continuous improvement.

Are you kidding me?

Incredible response. Right?

So, you know, that response used 1,227 tokens (reasoning uses more tokens) for a rough cost of $0.002592, and it is a GAME CHANGER for us. Notice how I stayed with the theme of visualizing structure in an unstructured world. I actually had it roll things up emotionally and it gave me advice on how to build a well designed Word Cloud.

My next steps would be to pull all of this back … from each survey. Pull out the emotional classification words so I can build a real word cloud where I size the words based on how many surveys the were included in. That way I know if “rude” is a 1 off, or if we have a lot of patients that report it. I take the recommended actions and send them through a workflow engine and route them to the appropriate person on staff in whatever manner using Qlik Application Automation for example.

Sure glad my old friend Allen started me on this journey. Seems like his message board use case, for visualizing structure in an unstructured world, is actually really applicable to the rest of the world.

Summary

I am not advocating for Deepseek.

I am not recommending you throw it all your data.

I am not recommending you throw it any of your data.

All I wanted to do was use if for this test case to highlight the fact that at the price point, and fact that it can run with much less CPU power is something that just might change

the ROI needed. Which might justify use cases for Allen, and for you.

At $1 per prompt, Allen might not want to use it. He can pay a human less than that. But at a cost of .000864 per prompt, he can afford it. If it were 2 cents per prompt, would he still find it worth it, so that he can utilize a safer LLM? He's evaluating that.

For my healthcare example … I think we would all agree that bringing that level of reasoning and presenting an action plan, would absolutely be worth several dollars for every survey comment. And would be a steal at just 2 cents per survey comment with a more trusted LLM. Because seeing RUDE, DISMISSED on a chart would tug at people's hearts and make them want to get out of their seat to take action, or see it and realize a change is needed on their end.

If your organization is looking for an immediate win with AI – consider using it for visualizing structure in your unstructured world.