Work Status

===========================================================================

Week-1

**8/18** – Discussed the project main goals, started the codebase of crawler.

**8/19** – MVP 3 example urls were crawled and reported first.

**8/20** – Fixed the errs and kept the correct code structure.

**8/21** – Completed the MVP 3 example urls crawling and started the crawling of 160 AMBA    list

**8/22** – Completed 160 AMBA list crawling

**8/23** – Had a rest.

**8/24** – Discussed the project issue and received the Task 2 with call.

===========================================================================

Week-2

**8/25** – Updated full raw text and CSV with my own web crawler.

**8/26** – Shifted approach from code crawler to GPT. Analyzed Task 2 (“News Search”).

**8/27** – Updated Raw1–30. Implemented GPT chat workflow for raw text updating and news    search.

**8/28** – Had a rest.

**8/29** – Updated Raw31–90. Reported first result of Task 2, and Raw text(1~30).

**8/30** – Analyzed Task 3 using video files. Fill the Empty Raw Text.

**8/31** – Completed updating raw text(31~154). Reported second result of Task 2.

===========================================================================

Week-3

—----------------------------------------------------------------------------------------------------------------------------

**9/1**

**Task1**

Reported the Task1 (MoldMakersCSV3). Once completed it perfectly, I will update MoldMakersCSV1 as MoldMakersCSV3.  
Task1 is not completed yet, (I spent 7 hours to generate theMoldMakersCSV3).  
Now I’m fixing the uncorrect data, structure in Task1.(Yellow part in  MoldMakersCSV3).  
It will be completed until Tomorrow.

[https://docs.google.com/spreadsheets/d/1dgZNpyC\_WQ\_iFawcx9DXG4qMaUesserpk9mLuppqf10/edit?gid=0#gid=0](https://docs.google.com/spreadsheets/d/1dgZNpyC_WQ_iFawcx9DXG4qMaUesserpk9mLuppqf10/edit?gid=0" \l "gid=0)

—----------------------------------------------------------------------------------------------------------------------------

**9/2**

**Task1**

 Completed Task1.   
[https://docs.google.com/spreadsheets/d/1dgZNpyC\_WQ\_iFawcx9DXG4qMaUesserpk9mLuppqf10/edit?gid=0#gid=0](https://docs.google.com/spreadsheets/d/1dgZNpyC_WQ_iFawcx9DXG4qMaUesserpk9mLuppqf10/edit?gid=0" \l "gid=0)

As you can see the csv.  
Row29 company is not in US

Yes just commenting on Row 29, now in Red.

DEMGY is a perfect example of companies NOT in our scope. The GPT prompt can be modified as follows: (feel free to edit and integrate into your current version of task 1)

‘We seek US and Canada-based companies that are typically privately held and family or partner run. They are small or medium businesses, typically with a single location (sometimes two, rarely 3+,  less likely in our scope) .

An example is DEMGY, multiple locations in multiple countries, headquarters outside US.  If you encounter a company in a list we present, preserve the name and URL, and mark in “Products and services offered”. “Outside of project scope”.  No further analysis of that row is necessary.’  
  
I got it. then I can delete it?  Leave it in with the “outside of scope” comment.  Documents that we looked at it.

See my comment below.

Outside of scope conditions can be:

Non-US owned company.

Subsidiary of larger business, possible  roll-up.  (This could be US owned, but would still be flagged as outside of scope. Useful information for the analysis team.)

We may discover other outside of scope conditions as our analysis expands to lower probability lists than AMBA.

Row151 I can’t visit the company site in more detail. about-us or contactus.

I’ll look at 151 today.

**Task2**

https://chatgpt.com/share/68b71ea5-21d8-8008-9039-1a7f002a3aa6

As you can see, my chat with AI isn’t perfect, but I’m really trying, and I’ve captured the main goal. I wrote each prompt with GPT several times, carefully considering your requirements.

I definitely want to see your GPT prompt development, post to coach you and for my own learning!

Here is my detail prompt for making more correct prompt.

- The candidate AMBA company list includes each company’s name and website URL.

- For each company, search Google News or Google Advanced Search using the company name in quotes plus local context, include articles from any year only from acceptable sources (local/regional outlets, small-town newspapers, trade publications, or official economic development sites), and exclude junk links, directories, social media, About pages, job postings, or promotional content.

- From each valid article, extract the headline, publisher, publication date (if available), full article text, and any photo captions.

- You(GPT) should produce concise bullet-point summaries that include the headline, publisher, and date, key details from the article, all verbatim quotes with speaker’s name, title, and source attribution (or explicitly state “No direct quotes appear in this article” if none), and, if available, a bullet beginning with “Photo:” followed by the caption text.

- Create a new CSV file  with the columns Company Name, Company Website URL, News Story URL, and GPT Summary (a short, factual, CSV-safe bullet list that fits within one cell), allowing each company to have zero, one, or multiple rows.

This is generating the prompt

I have a list of AMBAt(change to small, privately held high tech metal manufacturing companies) and need to create a prototype for news research that focuses on these firms. The idea is to use ChatGPT to extract and summarize authentic news stories mentioning each company. For this, each company name will be submitted in quotes into a Google News or Google Advanced Search query, which should return genuine news articles rather than junk links, directories, or irrelevant content. The focus should be on local and regional news, small-town newspapers, and trade publications that report on the company or companies like the one I submitted. Also use your own research to find news stories for the company.  When a real news article is found, the full text will be extracted and provided to ChatGPT, which should then generate a concise summary in bullet points. The summary must include verbatim quotes from company officials, management, or other individuals, with the speaker and the source or publisher clearly attributed. If the article contains no direct quotes, the summary should explicitly state that no direct quotes appear. The summary should also include contextual details such as the article’s headline, publisher, and publication date when available. Instead, the results will be placed into a dedicated news table, (in CSV format?) where each row will contain the company name, the company website URL, the news story URL, and the GPT-produced summary (also date and source into columns). The required output format is concise bullet points,t one verbatim quote with speaker and source attribution whenever possible. I will provide the AMBA company urls.  
  
This is prompt for generating the news table(not completed 😓)

I follow your requirement and reflect.

OK, I think your prompt is excellent. Please note my edits. Feel free to question anything.

The main edit is that the project is not seeking AMBA companies  
!  AMBA a list is where we started

Sorry is not necessary. We are learning as we go as fast as we can Sorry is not necessary. We are learning as we go as fast as we can.   
  
The only other point is to remind you that I am using a technique, but I really believe in here.

Good.

Before running a prompt, I give it to ChatGPT.  I asked ChatGPT to reform the prompt for conciseness and completeness.

And before running to ask any questions and resolve any ambiguities.

Finally, I encourage it to reformat the prompt so the execution will be most efficient

That I read the prompt that gives back to me and see what improvements it made.

OK so now we are in agreement on the prompt.  We captured all the ideas. The editing of the prompt is a little ragged, but we’re going to have GPT fix that. We’re not asking GPT to run this prompt yet. We’re asking for what is in the blue highlight.   
  
I understand, so you want me to update the prompt and report it here.  
Then you will update it and add or fix.  
  
so when we agree the perfect prompt. JUST RUN?

Generally, that will be great practice for both of us.  With this prompt, since we both looked at it and discussed the topics to get it aligned, I’m pretty confident the GPT response will be ready to run. If I’m not available, you can go ahead.

Not sure how much longer your workday will be today.

Communicating in here with you might seem crazy, but it’s much easier to be in focus with you when we’re both online.

Good Alan this is really special method for discussing.  Especially since we can jump  between topics and go backwards if necessary. That’s a good idea  
  
But in my opinion, it’s good. the point what you want  is very clarify better than slack.

Slack is getting hard for me because of the linearity.  I find myself repeating things I’ve said previously because going back and finding them as a burden. And then, if I say something slightly different, it becomes a further point of confusion.

If we work like this for a couple of hours, you can then go back and clean up what is redundant, answered, or not needed.   
right, I agree. This is also good method.

Great

With regard to the general process of prompt creation, please copy in your comment that I made in Slack about a four step process.

I want to make sure we are following that

Role definition ( for this project: expert marketing and competitive analysis, and industry news researcher for privately held US based companies)

Definition of the task using paragraphs. Request for GPT to edit the prompt for clarity, efficiency of execution, and identifying ambiguities and posing questions.

Reviewing GPT’s version of the prompt and running it

Asking GPT, to analyze it’s result in terms of clarity, completeness, and processing time. How could it have been better?

I would like to capture the input and output prompts in this text file as you go. That will allow me to quickly review work that you are doing when I am not available or sleeping.  I expect your skill to expand exponentially, and for you to quickly spot weak points, and places for improvement.

Yes, Alan.  
I understand your mind enough now….   
And you are not only my client, but also teacher.  
  
Now I have a good experience in AI and GPT.

Good.  I hope you understand **these are all skills I learned in August**.  Things are changing really fast.  A lot of people are climbing the learning curve at the same time. And also, there is a tremendous amount of noise on YouTube, as well as the good advice.

What we are doing used to be called “Agile development”.  AI has accelerated it so much that in the case of a project like this, the client will be getting major benefit and making high value project with our work product **before** MVP.  he can take the first 20 companies and put his team to work developing the telephone and contact strategy while we do the rest of the project.

This method of prompt development has a lot of benefits

First, it allows me to relax while I am gathering my ideas, that it doesn’t have to be perfect.

Another important benefit for you to understand about working this way is that ChatGPT changes how it operates from version to version and it will change again

At some point in the near future, I imagine it will start “charging” for execution time.

The graphics AI are charging by “tokens.”  so I don’t want to send it on compute-intensive tasks unless I know I have a reason. <—- this is the similar discipline to writing type code that supports efficient execution.  But now we’re just doing it in thoughts and commands in paragraphs.

I’m now going to comment on your question about task 4, and then I’m going to be off-line for a while.plz

Excuse me Alan, could you retell the Task3 main goal? I’m confused a little becasue AMBA…   
  
And you want to discuss our project here? no slack?

Re task 3:  This is not an AMBA project.  It is a project about small privately held companies mostly US based. AMBA is a membership organization that **some** of them belong to.  (“ American Mold Builders Association”)

Not all!

**we are now looking for techniques to find the others that are not AMBA members.**

Task 3 is one strategy ( there may be others, the client may think of some) to find such companies.  In this case, the source field becomes. “Job listing search”   All the rest of the crawl processing is identical.  Both the row for the CSV table and the news search.

So far, every record/row you’ve seen has source of “AMBA” because all the companies you looked at are from the same AMBA membership list.

I will send you other lists of other organizations.  The general process will be:

checking against our existing list for duplications.

Running the new ones through the crawl process.  Changing the “source “field as we add them to the table CSV and table.

Does this make more sense? yes source column not “AMBA”. Yes, then the client will know **how** we got that company for them to look at.

So AMBA is also MVP for our scalable project? Yes  exactly. I got it

They are able to look at the AMBA list and provide valuable feedback.  The AMBA list is the.” high value ore” … the most concentrated group of targets of interest.

Other groups will have a lower concentration of targets.  But they will be of higher value because they are harder to find.

Questions?  Does this make sense to you?

Task1 is only based in AMBA,   Yes  MVP

Task2 (news query and sum) is for ALL present and future targets

and Task3 (using job listings in LinkedIn and Indeed to identify candidate companies) is not limited to AMBA. In fact, we can make GPT more efficient by telling it to ignore AMBA members because we already have them .  Task 3 casts a wider net for candidates. It will have a lower percentage concentration of targets among those candidates.

But they will be really important targets, because they were harder to find.

I will also have other lists that can be run through the task 1 crawler identical except for a different “source” column for each list.

I understand. like AMBA? another source group you have?  Yes I Will have other source groups.

Additionally, I implemented the GPT development env for our project.

  
  
So I believe our project quality will be better.

Great!

**Task3**

I started extracting LinkedIn companies related to AMBA using GPT-5

Let me clarify.

We don’t need companies who are members of AMBA because we have the AMBA list.

We seek companies like I described above that are **not** AMBA members.

I have 3 questions for Task3

Do we need to use **Amplify** and **Make.com** for our project?

Should the exported table follow the same format as the **Task1 table**?

Should the extracted company information be added to **CSV1**?

 I don’t know what **amplify** and **make.com** are**,** so I’m not qualified to answer this question yet.   Can you get me links or info ? ??  
the platform?

2) Yes

3) Yes

The source field will be “Jobs listing search” instead of “AMBA”.  Otherwise identical instructions for Task 1 crawl.  We are looking for companies that are not members of any organization, and might or might not come up in a Google search.

But they are small privately held tool makers, and mold makers like I have described above. They typically always need skilled machinists, especially once who can work with the new computer, controlled machine, machines, and robotic machines.

I have one more suggestion for your prompt above. I’m going up there now to edit it in with a green highlight.

so you want to my prompt for Task 3?yes going up there to add one suggestion   
I C… but it’s not completed and I made the prompt for the future prompt for Task3. Yes I made suggestions and I have one more.

Going there now.

From the website text heaps we already collect, we will extract the jobs and careers sections and combine all job postings into a single Google Doc. GPT will then process this document to de-duplicate and normalize the job titles, producing a ranked list of the most common positions companies are hiring for. Using these top job titles, we will search LinkedIn Jobs and Indeed, focusing specifically on small high-tech mold makers, tool and die shops, and precision machining companies. From the relevant postings, we will capture the company names and their official website URLs, clean and de-duplicate the results, and merge them into our candidate company list. Any new companies not already in CSV1 will be added, and the crawler from Task 1 will be run on them to gather full site text for analysis. The ultimate goal is to expand CSV1 beyond the AMBA list by using job postings as discovery signals to identify additional high-value target companies.

I think our understanding on this is good now. If you’ve got all the ideas captured, you can delete the above because it’s not necessary to read once the prompt incorporates all these ideas.  
  
these sentences are organized with your requriement and I’m updating it for the prompt and good result. but as you can see,,,,, It also based in AMBA  
  
And I’m gonna add extracting part of Linked in and indeed

Wait here now  
  
**Task4**

Could you please provide me tih the interface for Task4, or clarify it further?  
Honestly, i’m a little confused about the requirements for Task4

**OK, here are some comments about task 4.**

It’s really about the context of the whole client need and requirement.

The simple requirement is so client can see all the relevant data for a company at one glance and put edits in a notes field, which will be stored in the Gsheet in its own column

Put CSV1 on your screen.  Imagine you are a support person for a private equity research researcher.  You are reviewing candidates that the crawl has identified as high value.

“CY” status.  You want to get a snapshot of the company in your mind, so you look at the website because the URL is right there.

You want to look at all the information we have extracted.  But you are wrestling with your screen because you have to scroll scroll scroll scroll scroll to see it.

And if you place your mouse on the wrong row you get lost.

You want to look at all the news for this company.  You wish there was a fast way to do it. You need a menu option or a button or hot key that pops up the list of stories showing the date and the headline.

You could click URL to go to the whole story or just look at our summary — bullet points and quotes —which pops up in a window.

Suddenly, within three minutes, you know quite a lot about this company, which has been sifted and sorted for you.  Your job is to make a strategy of how to approach them.  For example, perhaps you see that they have a “Makino” high-tech metal making machine on their floor.  You happen to know the sales rep for that company because you can see what state or region they’re in.

So you call him, and because he’s a sales rep, he picks up his phone.  And you learn a long time CEO (maybe owner, check the news quotes)  is retiring because his wife is sick.

So now he wants to make a strategy to reach out to ownership. We have the owner’s name in a field, which he can see without scrolling.

So he makes his notes in the note field about the first call. If someone else makes the call, they can enter their notes about what happened on the call.

The whole purpose is to set up a meeting to find out if they are candidates for being bought out in a special way that benefits the owner, the employees, and the community.

My job is to make that process as easy as possible for them to do without having to do a lot of training in a CRM product, **or writing a lot of code.**

Because our universe of candidates is small (600-1000?), and the universe of targets will be even smaller, maybe 150 or 200 max, GPT can do all the processing.

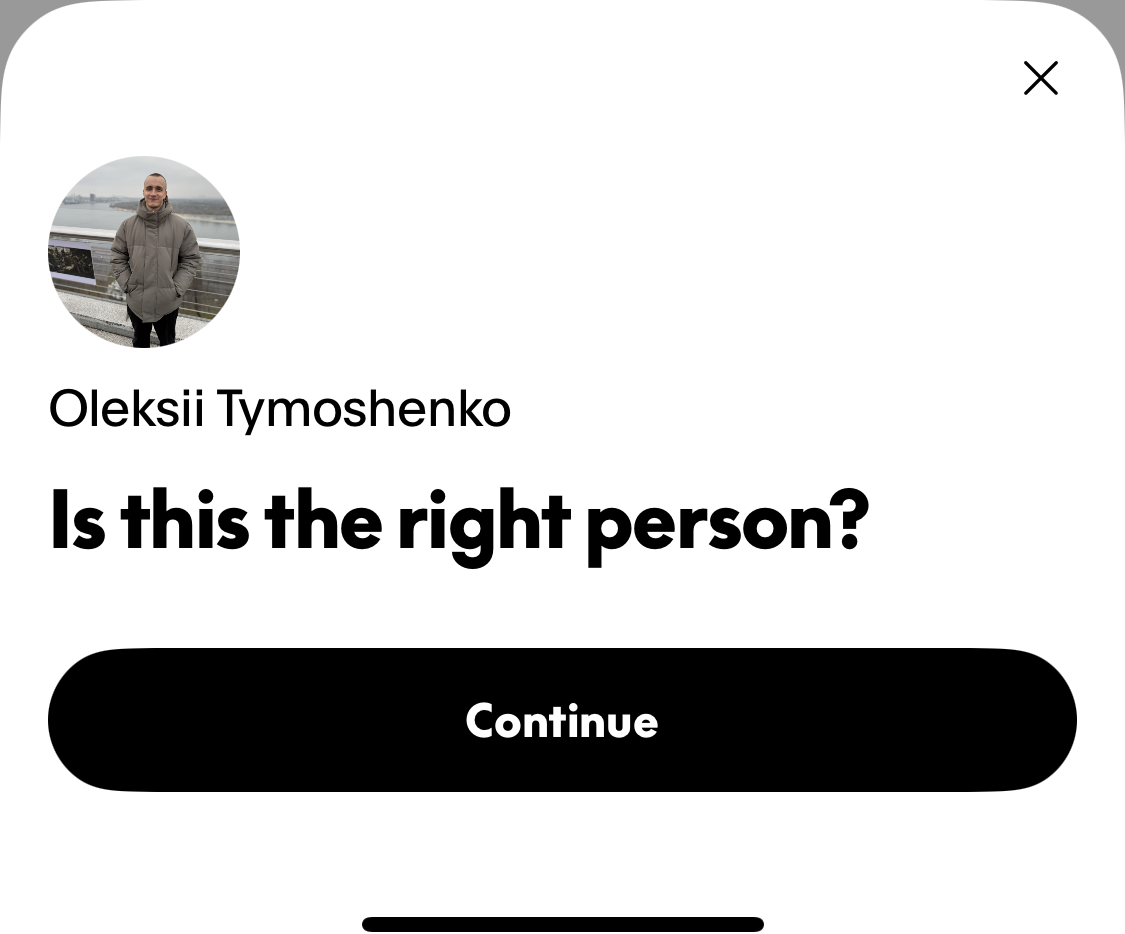
My goal is to get this usable ASAP and then make a few adjustments based on their feedback, so they can get on with the task as quickly as possible.

CSV1- my own crawled result table.  
CSV2- Task2 news table mvp

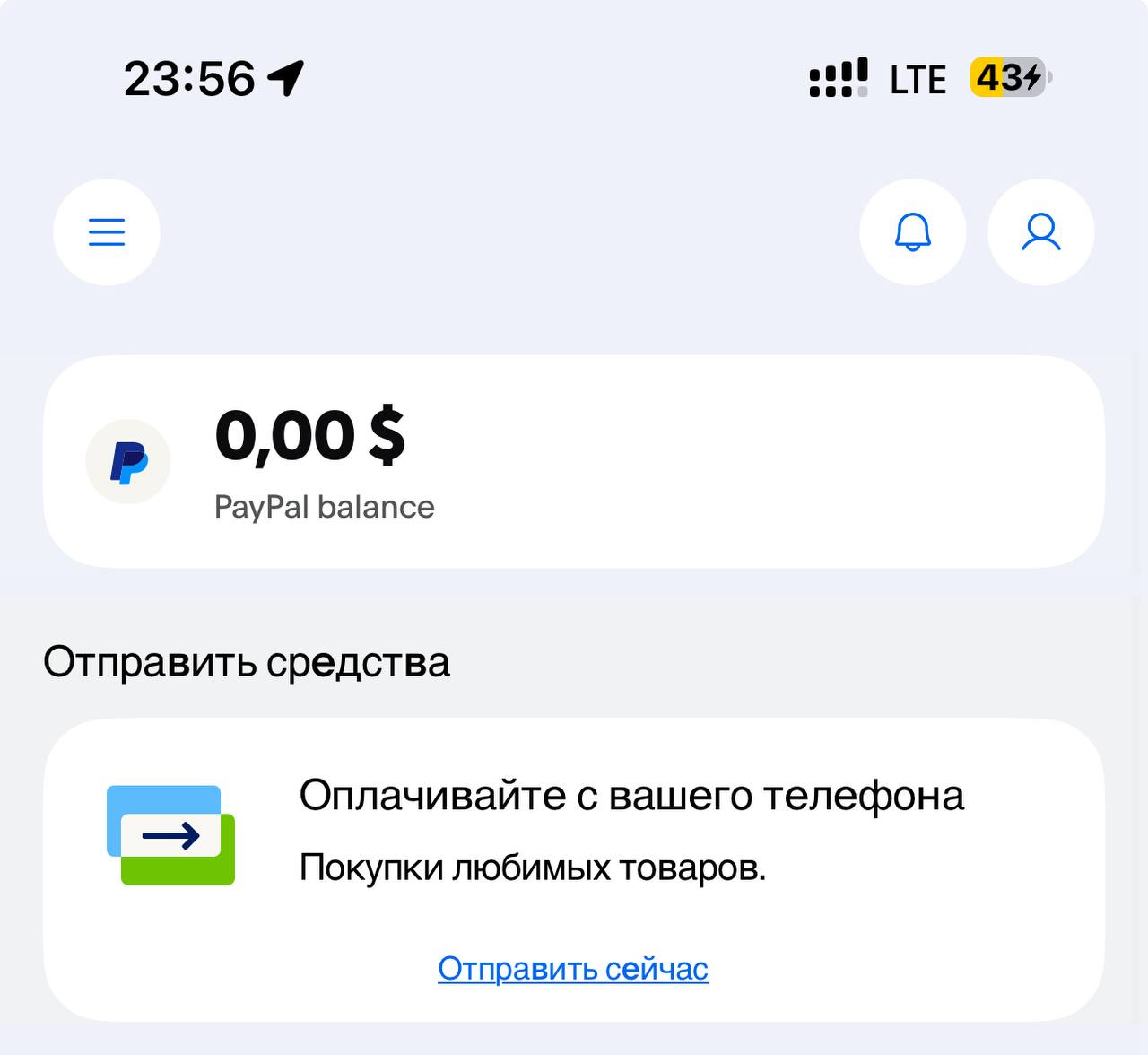
CSV3- Raw Text data table

**olekseytumoshenko@gmail.com  and this is PayPal address. Hope to pay to this.**

**PayPal asks:**

****

**I paid this email $200. at 1 PM PDT.  It matched the email you gave me.**

**please resend it. now is bedtime here, and the receiver is waiting.  
  
  
  
  
I received that’s past Let’s down to the 9/3 and Let’s start our Agile Development**

Are my comments about task 4 good enough for now?  Or tomorrow, I’ll show you what GPT gave me to try to implement task 4, including some cut and paste code.

Meanwhile, you can finish the task 2  prompt generation and start on task 3.

Maybe GPT can integrate the queries to LinkedIn and Indeed.  Maybe it can’t maybe they’re behind a paywall.  I don’t know.

I’m sure you’ll figure that out!

Your GPT prompt can start with “We want to find more companies identified as “\_\_\_\_\_\_” what we told it in prompt for task 2.  To accomplish that, we want the top 10 job description phrases that we extracted in column.\_\_\_\_\_ of prompt \_\_\_\_.

Prepare queries to LinkedIn and indeed defining the universe of companies and the job titles we are seeking.

Provide to me a list of company names, and URLs for companies seeking job candidates for these job titles.

(something like that, using the prompt creation disciplines we discussed above , do it in your own words. I’m just free associating)

Bye for now!

 Alan, plz listen one thing…ok but   
fast   
okay, no worry .  Task4 could I develop my own mode with Python or HTML for seeing well??

We’ll see what ChatGPT gives us, and decide after we get feedback from the client.  If we want to program control then yes, if they’re happy no.

okay, then I will focus on Task2 and Task3..

yes, because we will review those companies and then feed them back in through the prompt for task

—----------------------------------------------------------------------------------------------------------------------------

**9/3**

**Now I go to Dinner ( Here is Dinner time ), I will be back in 15min**

**Alan, I’m back.  
  
Let’s start our Agile development!!!**

**Alan, this is my suggestion for discussing here:**

**Here now y  
My report text will remain in the normal style.   
During project discussions, your text should be highlighted in yellow, while my statements will be shown in green.   
  
YOUR TEXT**

**MY TEXT  
  
If you want to emphasize something, please use red comments, and for my emphasis, I will use blue comments.**

**YOUR EMPHASIZED TEXT**

**MY EMPHASIZED TEXT**

**How about it?**

Good to start thx

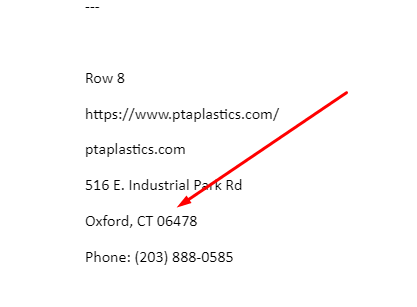
**Task1.**

**I updated CSV1 as CSV3 (Fixed the Row151).**

Should I look in here now or after we finish this review?[MoldMakersCSV1](https://docs.google.com/spreadsheets/d/1R7Oa9_tIKlt1qQQ1eBqCTNN79dCMDCdpTvpLxnd2rHk/edit?gid=0" \l "gid=0) yes

[https://docs.google.com/spreadsheets/d/1R7Oa9\_tIKlt1qQQ1eBqCTNN79dCMDCdpTvpLxnd2rHk/edit?gid=0#gid=0](https://docs.google.com/spreadsheets/d/1R7Oa9_tIKlt1qQQ1eBqCTNN79dCMDCdpTvpLxnd2rHk/edit?gid=0" \l "gid=0)

Zipcodes need a little work.  Quite a few have only 4 digits.   
US Zipcodes are in two forms:  99999  or 99999-9999.  The first 5 is fine for our list.    
Canadian Zipcodes are 1A1-B2B

Yes leading if leading zero is the problem, easy to fix.  Make sure that that is the problem, and figure out the source.  If GPT is extracting zipcodes, it should never make this mistake.  
For simplicity, we can drop the last 4 of the zipcode after the dash.  Not useful for our purposes.    
  
  
**Do you know**

as u can see Raw text data was crawled correctly, the 4 zip code is just style of Google sheet.  
Zipcolumn confused the style as Number, so 06478 -> 6478   
Ah style, correct.  GSheet is recognizing this as a data type number.   Data type “character” works best for zip codes.

Row 30 has a strange problem.  **Elba Tool and Die**   
The column Industries served says:   
Automotive; aerospace; agriculture; medical; consumer & electrical; defense; heavy equipment; energy Precision machining; die-cast dies; plastic injection molds; tool repair & maintenance; 5-axis CNC; EDM; value engineering; QC & inspection.

This data should be split up.  :  This part is Products and services offered:

Precision machining; die-cast dies; plastic injection molds; tool repair & maintenance; 5-axis CNC; EDM; value engineering; QC & inspection.    
And “5-axis” is right there in 5-axis CNC, but the 5-axis column doesn’t say “Yes”

<http://www.injectionmoldmaking.com> ? See above Elba Tool and Die  [Injectionmoldmaking.com](http://injectionmoldmaking.com) is row 31 because of header row.  I was reading 30 in the GSheet, it probably is “Row 30” in the data crawl.   
  
Question for you:  Do you know how to programmatically assert formatting in the GSheet?  
Or does it have a macro-recording capability to create a script?  Some basic formatting would be good to script.   Ofc, I have experienced in Word and Excel control with Macro

**General formatting for CSV 1 listings**

 Row 1 bold, centered.   
All data cells vertically top and aligned left.    
Products / services offered and industries served word-wrapped, a bit wider.    
Yes / No columns narrow , etc.

This data looks good!  Plenty good enough to show to client.

The 2nd tab Record View, is my attempt to introduce a no-code view provided by GPT.  I ran out of gas after several problems trying to work with GPT to do this the night before last.  I am now working on a 2nd method with GPT.  When I get it done, I will present this to the client.  I expect them to make some changes and suggestions..

I C

sorry. Row29  
  
what does mean?

Row 29

http://www.elbatool.com

elbatool.com

220 Covington Dr.

Bloomingdale, IL 60108

Phone:  (847) 895-4100

Company: Elba Tool & Die

About Us:

- Supplier of precision machining, die-cast dies, plastic injection molds, tool repair & maintenance.

- Combines modern CAD/CAM tech with old-world craftsmanship.

Capabilities:  Question:  Is GPT doing “Capabilities” instead of or in addition to “Products and Services” Is this alternative language, or can it figure out the difference, which is pretty subtle?

I’m fine if we label our column “Capabilities”.  Or do we add a column because it can distinguish “Products and Services” separate from “Capabilities”?   
These are both valid concepts, but they are quite similar.

Shall we just call this column “Products and Services / Capabilities” ?  
I think the client would be fine with this.  The distinction is very small and subtle, and doesn’t really matter much. If GPT is getting us good descriptive words, its not confusing at all.

Make sure our prompt is adjusted to say for this column (fuzzy search) extract words to describe “Products, services and/or capabilities”.  GPT can figure that out.  Very important the prompt is clear and explicit, because the prompt becomes the documentation.  Its like the principle of programming comments.     
←- This is an important example of a general principle.

Yeah, but my prompt for Raw Text data, just crawl the whole pages.

As you know, each company site has About US, Contact US, Services, Equipment.  
And so on     Alan if you are uncomfortable yellow comment, no worry let’s change.  
  
you should write normal style and I will only use comment without report okay?

Not every company has the same format -- Any data can be located on any page. Yes, just make sure the prompt states clearly how we want this summarized for both meanings in the same column . 

That’ right. Yeah, Good…   so the product servicec can be crawled, and capabilities can be crawled , anyway both in some case.

Do you understand me?

Good, I agree

GPt can recognize the context and adjust accordingly.  
  
Sometimes, they overlap, but in other s ‘Capabilities’ is broder than just ‘Product and services”

GPT is providing the Raw Data automatically, and I implement the GPT chatroom for Task1.  
  
Capabilities can somtimes be used as alternative language for “Products and Services”, but..  
it’s not always identical. ‘Products and ZServices” usuall

- 5-axis CNC machining

- EDM

- Value engineering

- Quality control & inspection

Industries:

- Automotive

- Aerospace

- Agriculture

- Medical

- Consumer & Electrical

- Defense

- Heavy Equipment, Energy

This is an excellent example of Industries served.

OFC

Are we crawling for “Jobs offered” ?? We need a column for t his, we’re going to use it to develop Task 3.

Alan if you are uncomfortable yellow comment, no worry let’s change.  ← I’m OK with highlight for now, when I go back and read it will be clear who said what, after I’ve forgotten. .   
Thanks for your effort, I thought you are uncomfortable about it…

Its worth it, at least for now  I’m hoping this method will be easy to go back to.

OK, you will adjust prompt 1, and we’ll get a column for “Jobs offered “ ?

Or a separate crawl for Jobs Offered? Yes some sites will not have.    
That’s OK, all we want is a standard list of the most common job titles these types of companies are hiring for.  That info is going into the prompt for Task 3… So maybe a separate crawl?   That info should be in our data heap if I understand correctly?

I can impement it, but I’m not sure that all company rows has it.

I understand, doesn’t matter.  I just want the most common job titles.  I don’t think the client even wants it as a column in the table.

okay no worry, I can do it soon.  OK we move on👍

But Alan I think this is very rare, It can provide  or not depends on company urls.

you should write normal style and I will only use comment without report okay? GOOd use

normal style  
  
  
Contact:

Phone: (847) 895-4100

**Yes go on  
<https://docs.google.com/document/d/1k9vEwS_2SmnbKjSy87_mCzswzP6ODQhfeWrInSjVFf8/edit?tab=t.0>check it this is source  
I see, 6455 = 06455,  I will fix it right now**

**Task2.**

**Here is completed draft prompt for Task2.**I’m here now .  I’m going to put comments for you in the prompt so you see context.  Then after you read, move them to below or delete them.   
  
hi, yes. I agree

Help me create an optimal ChatGPT prompt for the following need.  
  
First:  What role are you asking GPT to take ?   
  
I’m assisting with a research prototype on small, privately held high-tech metal manufacturing companies,   
  
I thought this part is Role. You ask me 4 things. So first I emphasize my role

Your emphasis is your role, but tell it something specific so it routes us to heavier thinking mode.

For this task you are an MBA-trained expert with 5 years of experience in research in US industrial companies that specialize in precision metal making ….

This step is SO IMPORTANT to what GPT does, I don’t mind being un-subtle and redundant.

 This narrows:  “precision mold building and tool and die manufacture”

Okay,    
We can review its results and if it gives us too many, in areas we don’t want, we can narrow more.  But the principles here are important for all your future tasks of this type. Tell GPT you want it to be a specific expert, and define the scope as focused as you can.

Yes

Yeah honestly, yesterday, you asked me to scale the scope from AMBA ….. do you remember?  
So I thought my companies must be huge scope so….. after considering your point, I wrote like that😥  
Its OK, this is our first example of applying the principle.  We may have dropped a bit of understanding along the way.  No worry.

Is this clear now?  Esp the parts in red?

Yes

This tells GPT how much smarts it needs to bring to the task. I’m using all this language so it  can narrow the scope as it goes.

I understand my mistake.  
  
Yeah….. I see. So the portion is just based in Our Project? No  whole tech metal manufacuturing companies???  
  
privately held high-tech metal manufacturing companie => precision mold building and tool and die manufacture

I’m assisting with a research prototype on small, privately held high-tech metal manufacturing companies who specialize in precision mold building and tool and die manufacture. For each company, use the company name in quotes with local context such as city, state, or region to simulate a Google News or Google Advanced Search. Prioritize local and regional newspapers, small-town news outlets, trade publications such as Plastics News, Plastics Technology, and MoldMaking Technology, and official economic development sites. Reject junk links, directories, social media, About pages, and promotional content. There is no time limit, so include valid stories from any year.  As well as Google, u**se your own knowledge and reasoning to identify relevant authentic news coverage.**

For each valid article, extract the headline, publisher/source, publication date if available, full article text, and any photo captions. Summarize the article in concise bullet points, always including headline, publisher, and date. Extract all verbatim quotes and attribute them with the speaker’s name, title or relationship, and the source of the quote. If no quotes exist, add a bullet that reads “No direct quotes appear in this article.” If a photo caption exists, add a bullet beginning with “Photo:” followed by the caption text. Keep all bullets short, factual, and CSV-safe. Each article must be returned as one CSV row with the following columns: Company Name, Company Website URL, News Story URL, Publication Date, Publisher/Source, and GPT Summary, where GPT Summary is the concise bullet list. Each company may have zero, one, or multiple news rows. Your task is: given a company name, website URL, and one or more authentic news articles, return the results in the exact CSV row format described above and do not include anything else outside of the CSV row(s).

No GPT?? I just said that you & your -GPT  in prompt . Please explain more, so I can ignore it with “GPT”  word?

Doesn’t this make sense?--->    
We may give this prompt to perplexity or Grok.  So when you use “you” in a prompt it knows what you mean.  (Just makes it easier to cut and paste. )

Are we OK so far?  Any questions?  This prompt is generally excellent.

 (Just makes it easier to cut and paste. ) so we shouldn’t use “You”?   Yeah…… I see.  
Okay I understand, GPT is very very smart 😅  LOL  I know its weird to call software “You” but it isn’t worried about that. okay, let’s go on

Its fine to say “You”  …. If you get in the habit of saying GPT, then Grok or perplexity won’t know what to do.  Part of our learning here will be to try some of these prompts in other engines, and compare the execution speed and differences.  That’s part of why I want the prompts to be self-contained and complete.  Not assuming the AI engine knows what we mean from prior prompts.

OK I’m ready to look at the result in your link below. 👍

**I got the first result(CSV2) with this prompt, and as I see… it looks good.**

**<https://chatgpt.com/share/68b858f4-c264-8008-b0db-0d7935592491>https://docs.google.com/spreadsheets/d/1Jrrkc3\_m95TdFwNZdzhu-H7cJhIviOm\_yNRi9N0\_hGU/edit?gid=0#gid=0**

Do you have the download file?   GPT going very slow on retrieving it for me here.   
  
The CSV2 is for you.  
**https://docs.google.com/spreadsheets/d/1Jrrkc3\_m95TdFwNZdzhu-H7cJhIviOm\_yNRi9N0\_hGU/edit?gid=0#gid=0   
  
Can you see it. ??? This file is shared by you.**

**No, its probably thinks I’m you. Getting “File Not found”**

For now just share your copy with me.

Or copy / paste in here.

**wait,  can you see CSV3?**

**My CSV3 is blank**

**see again**

OK this is great!    Here is my feedback:

GPT asks if we want a blank row for companies with no news. I’m not sure.   
Need your advice.  The way we will use this is have a macro, menu option or hotkey when client is looking at a row or record view.  What’s easier -- to show a row that says “No news”, or logic to detect no news rows found for that company?

OK then answer GPT -- “Valid question, for now just do a row for every company whether news or not, we will adjust later if needed.”  ←- Do you agree ?   
  
If you are good, That’s okay. but in my opinion… we have to fill the all rows.  That’s fine, at least we know its trying each company.

**I think that there is no companies with news.  
so I will fix it soon with updated prompt**

**at least one news will be included in company.**

Also we want a column for “headline” ,

one for “Source” for now AI engine or “Google search”,  ←- Let’s learn who is finding these stories.

**I discovered that err, I gave 5 examples but the result only 3 rows, this is BUG I think.**

**In the field with all the text, can we get some formatting?  Soft CR/LF between quotes, Basic outlining to appear in the cell?**

**yes in google sheet?** Yes  
  -- And the headline in its own column, not in the text summary.  Adjust the prompt.

I’m very excited about this.  Its obviously going to work great!

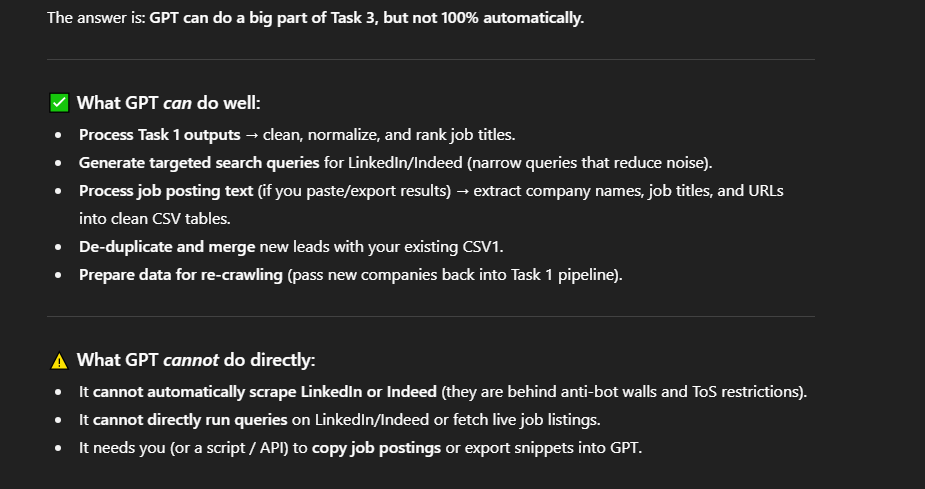
**I will do it tomorrow, no worry.. I think our main target is just GPT prompt quality  
Others are no problem for us.**

**Alan I want to discuss  about it in Task4😉**

I want to comment on task 3 next.

**yes**

**Task3.**

**LinkedIn and Indeed don’t have open feeds that GPT can just pull from directly. Most of their job data is behind restricted APIs or paywalls.   
GPT can’t integrate with them natively, but it can help by optimizing search queries, setting up alerts, or working with other open job boards that provide APIs.  
Looks here (Now I’m moving to be a GPT-Expert from an origin code Expert🙂)  
  
**

**As you can see, GPT can’t handle all of Task 3 with a prompt, so I’ll manually take care of the warning part. Do you agree?**

I understand this limitation. The purpose of this task is to see if we can identify new candidates that don’t belong to membership organizations or aren’t in other queries.

Using GPT to get us the most common job titles from the companies we know is good, as you have listed.  Possibly we do these queries to LinkedIn and Indeed one time, by hand.

We should probably process a 2nd list besides AMBA first.  I haven’t yet gotten that list ready, but I can by tomorrow.

So you would crawl a second list, with a different source, then try an Indeed or LinkedIn query by hand, and see if any companies were returned looking for any of those jobs that is a legitimate candidate.

**I can make the queries for LinkedIn and Indeed, then I have to scrapy the searching result.  
This is difficult for us to catch with GPT.**

**If you have the company and URL, as returned by Indeed or LinkedIn, we need to de-duplicate it before submitting it for scraping, correct?  Or tell GPT to skip it if its in a certain list already.**

**Okay**

**yes,**

**Alan, give me the important point for query.**

**Query is   
We are seeking job listings for   
Job 1  
Job 2  
Job 3  <-- all these are the most common job titles we get from scraping text in Task 1   
Job 4  
Job 5**

**We only want to apply to**

**“**small, privately held high-tech metal manufacturing companies who specialize in precision mold building and tool and die manufacturing” located in the United States

Indeed and LinkedIn may have their own ways of letting you scope their searches.  They may charge for this.  Find out how much.  (You may need to establish an account to run such a search)

**Task 4.**

**Please share the GPT link or your codebase for Task4.   
After analyzing it and I will hand in my thinking or feedback about Task4.**

 okay , Alan.   No worry. IF then….. Let me know the Task4 interface requirement correctly.  
Then I will make Frontend project for this.

My idea is to show a very rough version to the client and see how far they want to go in polishing the interface.  I’ve designed and implemented 100’s of these from decades ago.  It depends on their focus.  I think they’ll be thrilled by the data and want to get on with the task of analyzing and contacting these targets.   
**I will do this today, after I play with the current approach a bit more.**

**As you will see when I share my GPT thread, it hit a bug which it tried 5 times to fix, including giving me debugging instructions.  It was very time consuming and was a blind alley.  You might spot the problem right away, but for now let me try to finish its 2nd iteration, then I will turn it all over to you.  It wrote hundreds of lines of scripting and at least 100 lines of HTML.**

**I don’t want you to waste the time I did.  We almost had it … and then we almost had it … and then …. etc., etc.  You know the story.**

So you want me to make the prototype for this??

We’ll decide tomorrow.  Please adjust the news per my notes and when we get the summary a bit more formatted, run that news query for the whole task 1) list we have.  I’ll try to get you another fairly small list to add / merge / purge.   
  
Good

**Alan, so you want to me to control the GSheet for convenient interface?**

**I think so, we’ll get some adjustments from the client.  Think about two scenarios:**

**We come up with a new source and add rows to the client’s working table  
We come up with a suggestion for a new column, which may throw off the import function if its not coordinated properly.   
In both these scenarios, we want to have as little fiddling and communication from client to me to you , and you to me to client, to make this go smoothly.**

**The work amount of this is very very huge😅**

**I know LOL**

**I think that’s it for tonight.  Unless you have any other questions?**

**You can proceed on 2 and 3.  Task 2 just clean up text formatting and a column for “Headline”.**

**Task 3 Queries to Indeed and LinkedIn by hand, with my instructions above, based on job description extracted from the crawled heaps.**

Yeah no worry, I will hand in the first result of Task3 to you.  
  
And I will fix and update the prompt or Row of Task1 and Task2.  
  
And Task4…. I will ignore it until you got the correct requirement and ciient’s discussing.

**Great!  Have a great night I hope! See you tomorrow.**

**Bye Alan, Have a nice day.  🙂👍**

**If you want to emphasize something after I’m sleeping or offline. plz write here  
I will consider it.**

—----------------------------------------------------------------------------------------------------------------------------

**9/4**