

AI Working Group

Status Update

November 2024

Today's Agenda



- Goals and Objectives
- Revere Meeting Summary
- Recommended Next Steps
- Action List
- Implementation Schedule
- Questions / Discussion

Overall Objectives for AI Implementation



- Understand "guiderails" for use
- Understand potential pitfalls
- Identify initial focus areas
- Develop an implementation and testing plan
- Develop an implementation and testing schedule
- Continuous improvement

Who is Revere?



- Similar size company to AASI
- Similar use cases, workflow and business environment

Revere Meeting Summary

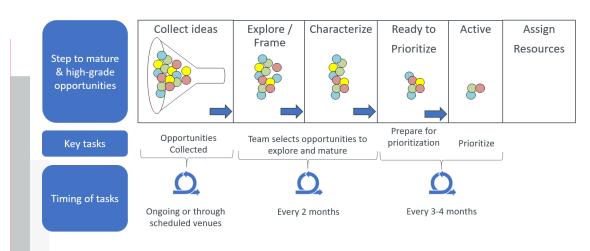


- Learned the workings of LLM-based AI models
- AI Adaptability: The tool itself is less important than how it's used. Focus on enhancing roles and training people
- Token Management: AI performance depends on token limits; tools like Gemini handle large documents by splitting them into manageable sections.
- Accuracy Testing to gain confidence in outputs
- Future Potential: Explore areas like data analysis, accuracy testing, and enhancing QA through AI.



- Most AI initiatives fail because of failure in selecting the right projects
- AI project is 15% tech, 85% people
- "Buy-in" across the company
- Confirmation bias can be dangerous
- Seek low-hanging fruit impact, likelihood of success

Continuous process to mature opportunities







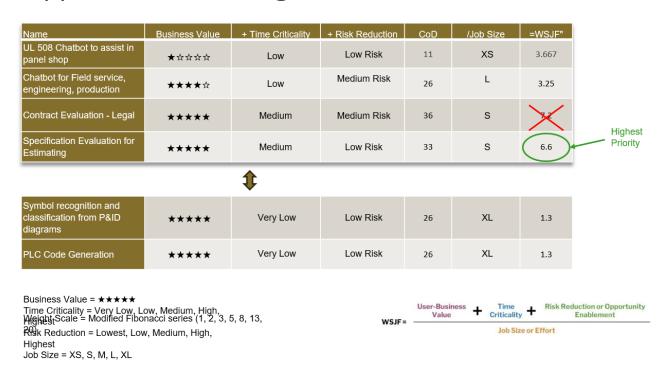
• Portfolio of potential focus areas

Name	Phase/ Dept	Champion	Definition	Type of application	End User	Bussiness Value	"+ Time Criticality	Reduction/Oppor tunity Enablement	CoD	/Job Size or Duration	"= WSJF'
Contract Evaluation	Estimating	Edward		NLP		****	Medium	Medium Risk	36	M	4.5
Specification Evaluation - Hardware recommendations 2nd gen	Estimating	Edward	2nd generation of the row 4			****	Lowest	Low Risk	26	M	3.25
Sage produces the pdf cutouts, of specs. of customer specs. with comments,	Estimating	Noah				****	Medium	Low Risk	33	M	4.125
Estimating Schedule Completion	PM					****	High	Medium Risk	41	XL	2.05
Historical Change order indexing	PM					★☆☆☆☆			#N/A		#N/A
Forecasting job performance using based of prior perfomance in Spectrum/GP Swing	Estimating, PM		Avoid loosing GP			****		ľ	#N/A	XL	#N/A
Automated Invoicing		Katie	Automation not Al					7	#N/A		#N/A
Symbol recognition and classification from P&ID diagrams	Engineering		May wait on developing tech			****	Lowest	Low Risk	26	XL	1.3
Symbol recognition and classification from One-line diagrams			May wait on developing tech			****	Lowest	Low Risk	26	XL	1.3
O&M manuals. Automatically generating a manual out of BOMs		Noah	We believe that this is a good product			****	Low	No Risk	14	S	2.8
Indexing drawing archive. The Go-by.	CAD, Engineering, Estimating	Noah	ptor of the application and BOM to our archive. Doe	snt need to be drawing	3	****	Medium	Low Risk	33	XL	1.65
Pruchasing and Inventory control	Panel Manufacturing	Katie	follow up action item						#N/A		#N/A
UL 508 bot to assist in panel shop	Panel Manufacturing	Bert				★★☆☆☆	Low	Low Risk	15	XS	15
NEC bot (i.e Wire, breaker sizing)	Panel Manufacturing	Noah				*****	Low	Low Risk	15	XS	15
PLC Code Generation	PLC Scada		May wait on developing tech			****	Lowest	Low Risk	26	XL	1.3
Migrating program. PLC and SCADA infrastructure, live values	PLC Scada	Bert	May wait on developing tech			****	Lowest	Low Risk	26	XL	1.3
Adaptive Loop tunning	Site Services		MPC does this.			****	Lowest	No Risk	17	L	1.30769
AR tool taking a visual record and archiving service calls	Site Services	Bert							#N/A		#N/A
Chat bot for Field service, enginieering, production	Site Services	Edward	t. Keep records of field service, eng, production defic	ciencies and solution. C	Quality issue	****	Low	Medium Risk	28	S	5.6
HR recruiting tool. Pair the job description with the resume of successful current employees	HR	Katie	Concensus is that there are tools developed					ſ	#N/A		#N/A
Find patterns in Spectrum data to gain insight into loosing and winning jobs		Katie							#N/A		#N/A
HVAC management		Noah				* 拉拉拉拉	Lowest	No Risk	3	S	0.6
Cyber security using Al		Bert	erabilities related to products that we already work	on. Penetration test.		★★★☆☆	Lowest	Medium Risk	17	M	2.125
Disaster recovery		Bert				****	Lowest	Low Risk	14	M	1.75
Open chatgpt to shop and personnel to just use it.		Edward	Looking at pricing for GPT enterprise						#N/A		#N/A
Soho data - Find patterns to determine good estimation vs bad estimation			Not having data increaes complexity			* 拉拉拉拉	Medium	Low Risk	14	XL	0.7



• "WSJF" method to prioritize

List of opportunities – Weighted Shortest Job First





• Employee survey

MVP Survey Results

What did we learn from the MVP Survey?

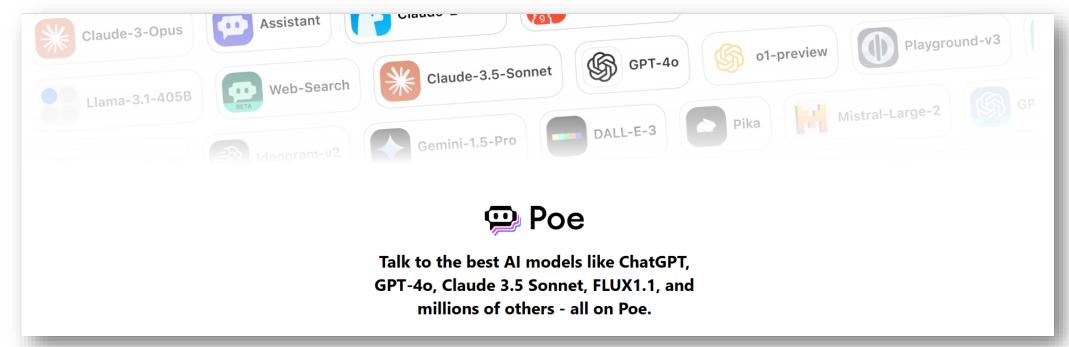
- Out of 27 participants the acceptance was 77% Positive, 15% Middle, and 8% Negative.
- Efficiency and possible quality improvements were the positive themes.
- Missing relevant information was the predominant fear.
- Features needed
 - High level list of deliverables
 - Summary

A chatbot can satisfy the features needed

Poe App



- Allows for use of many different AI for same monthly fee
- Each AI has its strengths and weaknesses
- AI models are evolving rapidly don't just choose 1
- Different AI's handle document types and prompts differently (i.e. engineering drawings)



Prompting and Questionnaires



Prompts

- Training people to use effective prompts is critical
- "Start fresh" old query chains can skew output by perpetuating bias
- Training is in the user, not the model
- Trial and error

Questionnaires

- Frame the required output from AI (i.e. summarizing elements of a document)
- Developing robust context-based questionnaires to query AI is critical
 - Subject matter experts for each group
 - Living documents that are maintained and refined over time to become very reliable

Effective Prompts



- Be specific
- "Act as if"
- Specify output presentation
- Use "do" and "don't"
- Use examples
- Specify tone and audience
- Build on previous prompts
- Save for future use

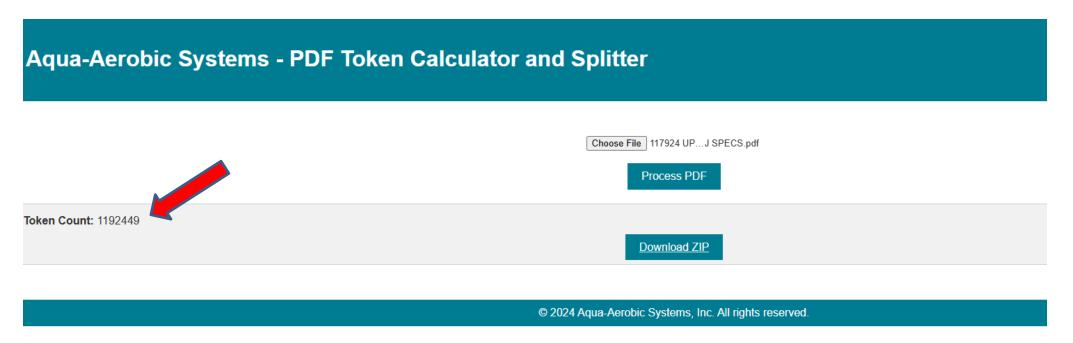
Good resource for prompting best practices: https://huit.harvard.edu/news/ai-prompts

Use AI to build an AI prompt!

Token Calculator and PDF Splitter



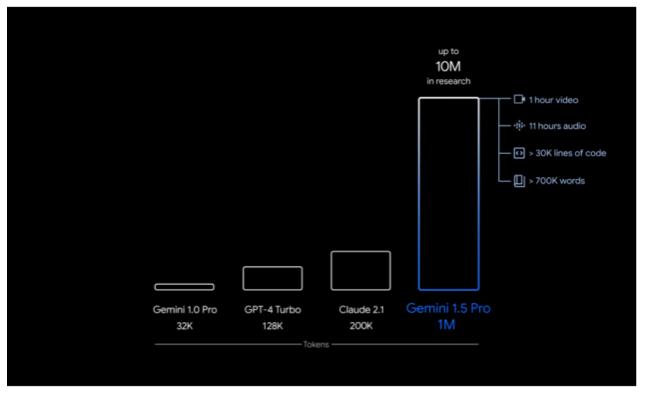
- Calculating token "weight" and splitting PDF is key
- Pairing with appropriate AI "context window" to prevent coherency loss
- Files with images have large token counts



Context Window



- # tokens a model can process in a single interaction
- Gemini 1.5 Flash 1M



Context lengths of leading foundation models

Revere - Spec Review Process



1. Poe Al Toolbox:

- Offers various AI models for different use cases.
- Uses Gemini 1.5 Flash 1M Model (1M refers to the number of tokens handled in one session).
- Poe manages API tokens in the cloud.

2. Run PDF through Token Counter:

Example: PDF contains 700k tokens, Gemini model's limit is 1M tokens per question.

3. Run PDF through Splitter:

- Splits PDF into sections to preserve tokens.
- Smaller sections are easier for processing.

4. Load Question Set:

- Uses predetermined prompts from a word document.
- Starts fresh each session to avoid contamination.

5. Load Spec Review:

Copy and paste a premade prompt, such as "You are an engineer," to focus LLM's perspective.

6. Generate Output:

Save the output into a text file and organize responses into a .csv file for further analysis.

Confidentiality and Copyright Protection

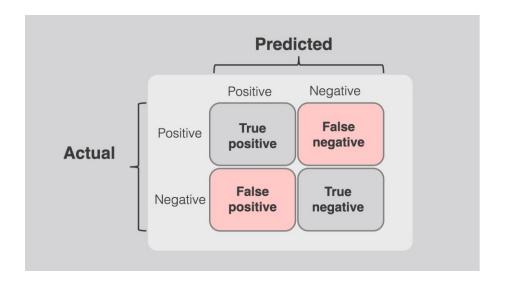


- Best practice is to scrub docs of sensitive info. Can write a program to do this
- Generally speaking, AI output is <u>not</u> copyright protected as it is computed by a generative model, therefore should be "original". Many universities and company now allow it with proper citing
- Regulations and legal analysis is evolving rapidly; may change in future
- Enterprise AI versions have ability to create "closed" networks, but lose access to the global database (and costs more \$)

Accuracy Testing



- How to trust and test the model?
- Key part of implementation
- Benchmark AI result vs human effectiveness (humans are not 100%)
- Accuracy vs efficiency how accurate does it need to be?
 - Depends on focus area and associated risk vs efficiencies gained
- Confusion matrix approach
- Subject Matter Experts to develop answer keys for scoring



Accuracy Testing



- Define accuracy requirements.
- 2. Assess the relevance of items missed by AI and humans.
- Evaluate time lost due to inaccuracies.

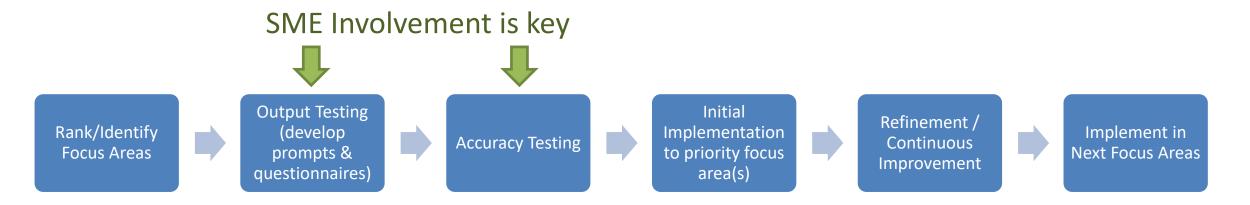
Overall Recommendations



- Working Group should remain intact through testing, implementation, continuous improvement, etc.
 and meet regularly
- Initial steps (testing, implementation, refinement) will take employee's time; the Company needs to allow for this, plan appropriately and allocate resources
- Select initial focus area(s)
- Poe app should be rolled out in lieu of individual databases
- Focus areas need a group representative to drive actions and identify sub-tasks

General Plan





Working Group to identify sub-tasks

2-4 weeks

2-4 weeks

4-6 weeks

1-2 weeks

ongoing

After 3 months of use

AquaAI Web App

https://aquaai.app/





Open Questions



- Determine focus area selection methodology
- Determine accuracy testing methodology