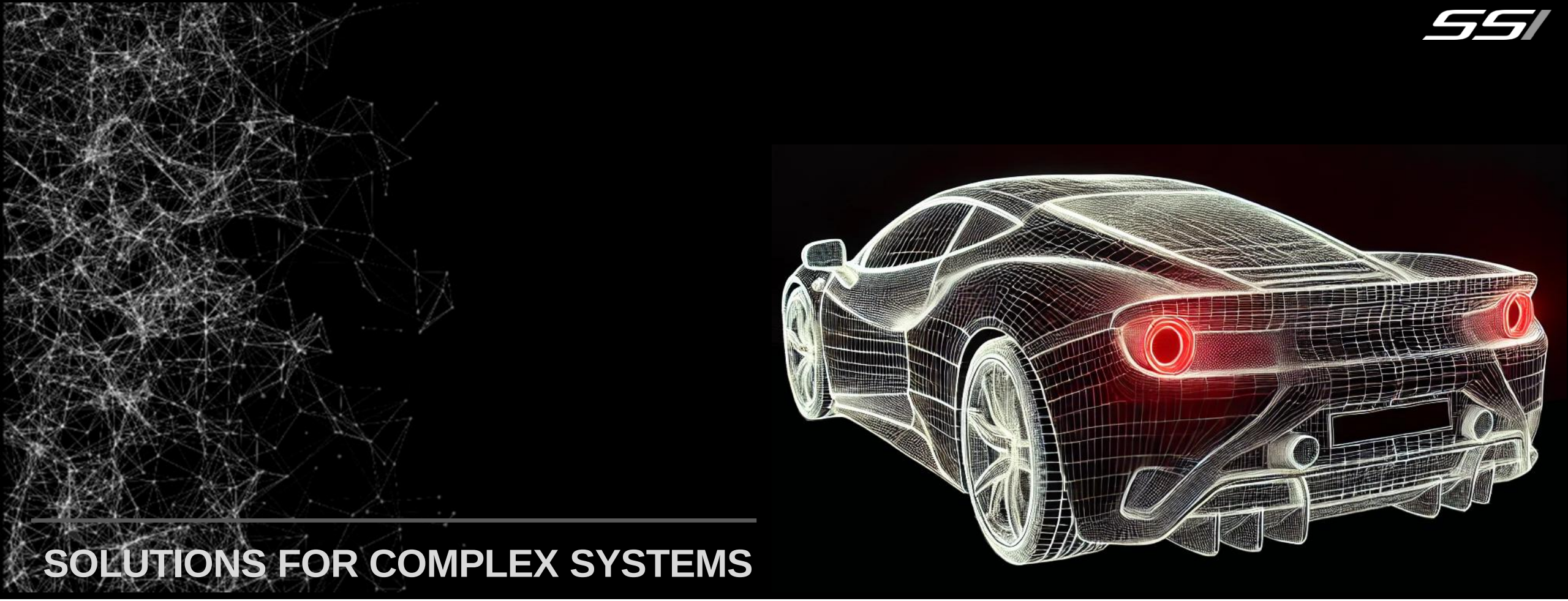


Using LLMs to Accelerate SysML Model Requirements Gap Analysis

David Hetherington
Connor Paugh
Sam Ahern
SYSTEM STRATEGY, INC.

2025-09-18

Objectives

From the Abstract

The objective of our research is to investigate the practical use of LLM technology to accelerate a common laborious task for systems engineering teams: evaluating an incoming request for proposal (RFP) containing a large set of requirements against a supplier's SysML model of their current subsystem or component product.
We will investigate the LLM's ability to improve and clarify such external requirement sets when they are poorly structured, not in compliance with common requirements engineering best practices, not "model-friendly" (missing short names, missing coherent ID numbers, or both), are incomplete, or are ambiguous.
Having improved the clarify of the incoming requirement sets, we will investigate the ability of the LLM to correctly create allocation relationship recommendations between the requirements and elements in the SysML model as well as to identify requirements that are not covered.

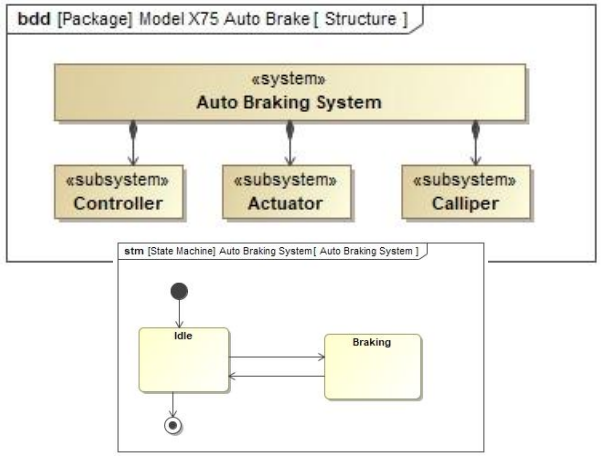
Background on the Problem

“Fire Drill” Evaluation of Requirements

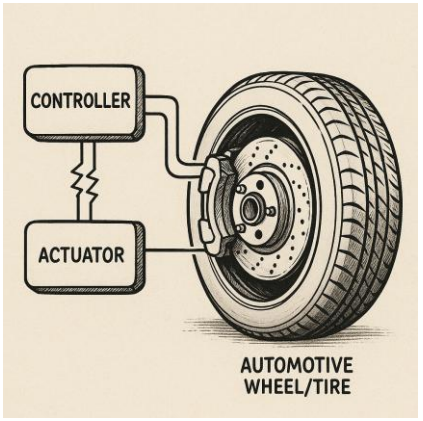
Background

Requirements

ID	Name	Description
R001	Braking Performance	The system shall stop a vehicle traveling at 100 km/h within 40 meters on dry asphalt.
R002	Operating Temperature	The system shall operate properly at temperatures between -40 °C and 55 °C.
R003	Corrosion Resistance	Brake components shall withstand 480 hours of salt spray testing per ASTM B117.
R004	Dust Resistance	The system shall maintain full functionality when exposed to dust per ISO 20653 IP5K.
R005	Water Resistance	The braking system shall resist water ingress up to IPX6 conditions.
R006	Brake Pedal Force	The system shall provide full braking performance with a maximum pedal force of 500 N.
R007	Brake Pad Life	Brake pads shall last at least 50,000 km under normal driving conditions.
R008	Noise & Vibration	Braking noise shall not exceed 70 dB at the driver’s position under normal operation.
R009	Response Time	The actuator shall respond within 100 ms to a control signal.
R010	Redundancy	The system shall include dual hydraulic circuits to provide redundancy in case of failure.
R011	ABS Functionality	The system shall prevent wheel lockup during emergency braking on low-friction surfaces.
R012	tronic Stability Control	ystem shall integrate with C systems for coordinated



SysML Model



Standard Product

Urgent Inquiry!!!

This sort of “fire drill” exercise is a frequent headache for automotive suppliers. A new set of requirements will arrive with a very short response deadline. The new requirements are often confused, poorly structured, and poorly written.

Can the LLM help us compare the new requirements to our current requirements and system model?

ChatGPT 5.0 was used to prepare some of the images on this page.

Example Legacy Requirements

Background

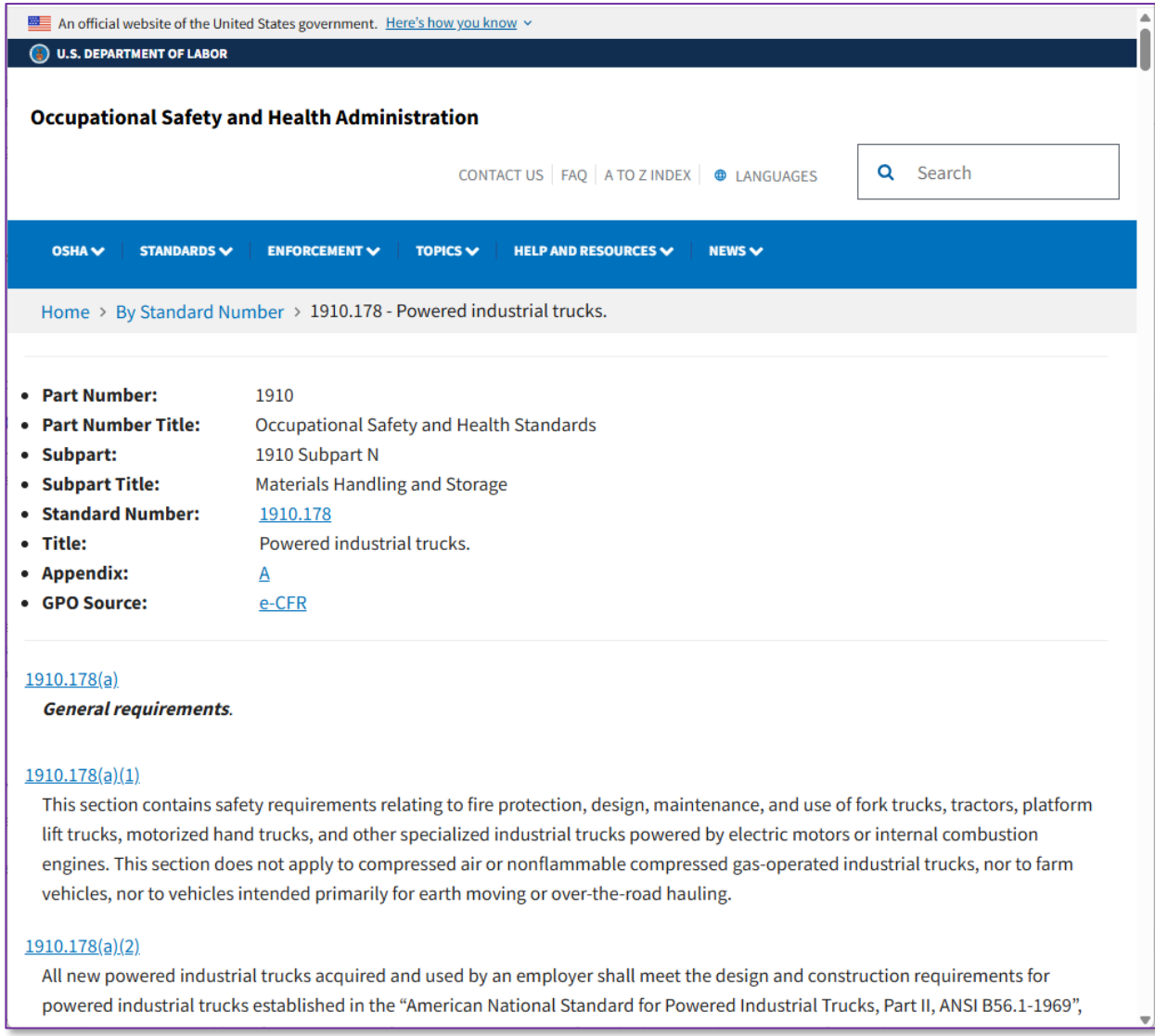
Many legacy requirement sets only have numbers and long-form text for requirements.

IBM DOORS Classic and DOORS Next Generation both defaulted to this format.

Short names was an optional feature that the customer had to deliberately enable.

Here we have requirements from the U.S. Occupational Safety and Health Administrations.

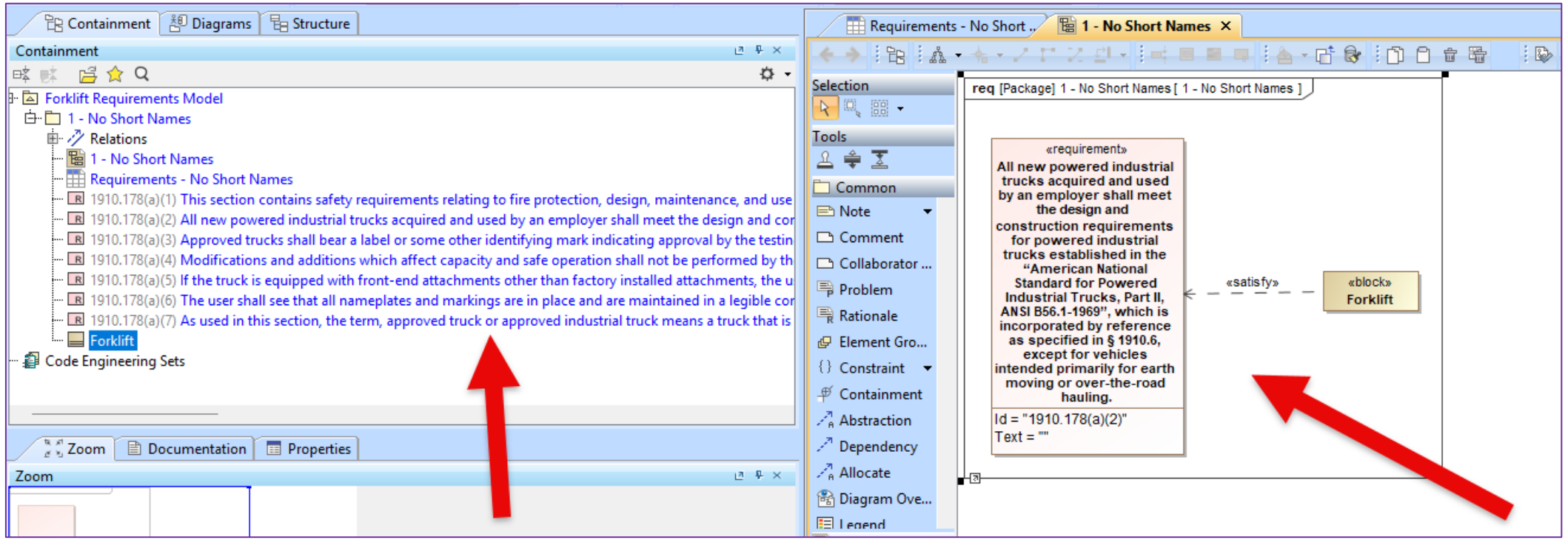
(Forklifts are among the most dangerous pieces of equipment in use at most large industrial installations.)



https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.178

In the SysML Model

Background



Unfortunately, requirements imported without short names are very clumsy to work with in the SysML Model.

Adding Short Names

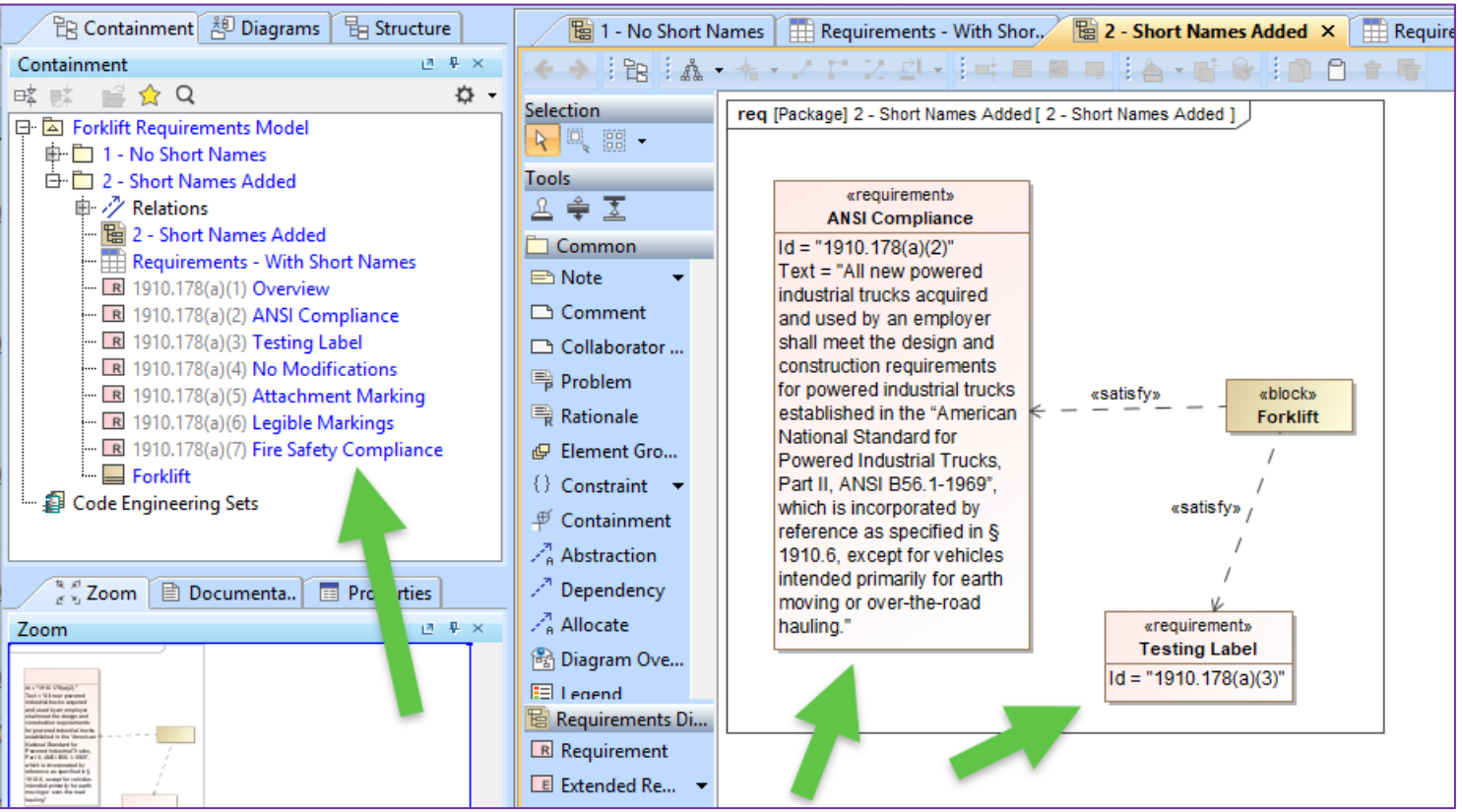
Background

This problem can be solved by adding a short name for each requirement.

ID	Short Name	Text
1910.178(a)(1)	**Overview**	This section contains safety requirements relating to fire protection, design, maintenance, and use of fork trucks, tractors, platform lift trucks, motorized hand trucks, and other specialized industrial trucks powered by electric motors or internal combustion engines. This section does not apply to compressed air or nonflammable compressed gas-operated industrial trucks, nor to farm vehicles, nor to vehicles intended primarily for earth moving or over-the-road hauling.
1910.178(a)(2)	**ANSI Compliance**	All new powered industrial trucks acquired and used by an employer shall meet the design and construction requirements for powered industrial trucks established in the “American National Standard for Powered Industrial Trucks, Part II, ANSI B56.1-1969”, which is incorporated by reference as specified in § 1910.6, except for vehicles intended primarily for earth moving or over-the-road hauling.
1910.178(a)(3)	**Testing Label**	Approved trucks shall bear a label or some other identifying mark indicating approval by the testing laboratory. See paragraph (a)(7) of this section and paragraph 405 of “American National Standard for Powered Industrial Trucks, Part II, ANSI B56.1-1969”, which is incorporated by reference in paragraph (a)(2) of this section and which provides that if the powered industrial truck is accepted by a nationally recognized testing laboratory it should be so marked.
1910.178(a)(4)	**No Modifications**	Modifications and additions which affect capacity and safe operation shall not be performed by the customer or user without manufacturers prior written approval. Capacity, operation, and maintenance instruction plates, tags, or decals shall be changed accordingly.
1910.178(a)(5)	**Attachment Marking**	If the truck is equipped with front-end attachments other than factory installed attachments, the user shall request that the truck be marked to identify the attachments and show the approximate weight of the truck and attachment combination at maximum elevation with load laterally centered.
1910.178(a)(6)	**Legible Markings**	The user shall see that all nameplates and markings are in place and are maintained in a legible condition.
1910.178(a)(7)	**Fire Safety Compliance**	As used in this section, the term, approved truck or approved industrial truck means a truck that is listed or approved for fire safety purposes for the intended use by a nationally recognized testing laboratory, using nationally recognized testing standards. Refer to § 1910.155(c)(3)(iv)(A) for definition of listed, and to § 1910.7 for definition of nationally recognized testing laboratory.

Short Names are More Convenient

Background



Requirements with short names are much more convenient to work with in the model.

However, manually creating short names for a set of thousands of legacy requirements can be tedious!

Evaluation Goals

Practitioner's View

Evaluation Goals

1. Does it Work?

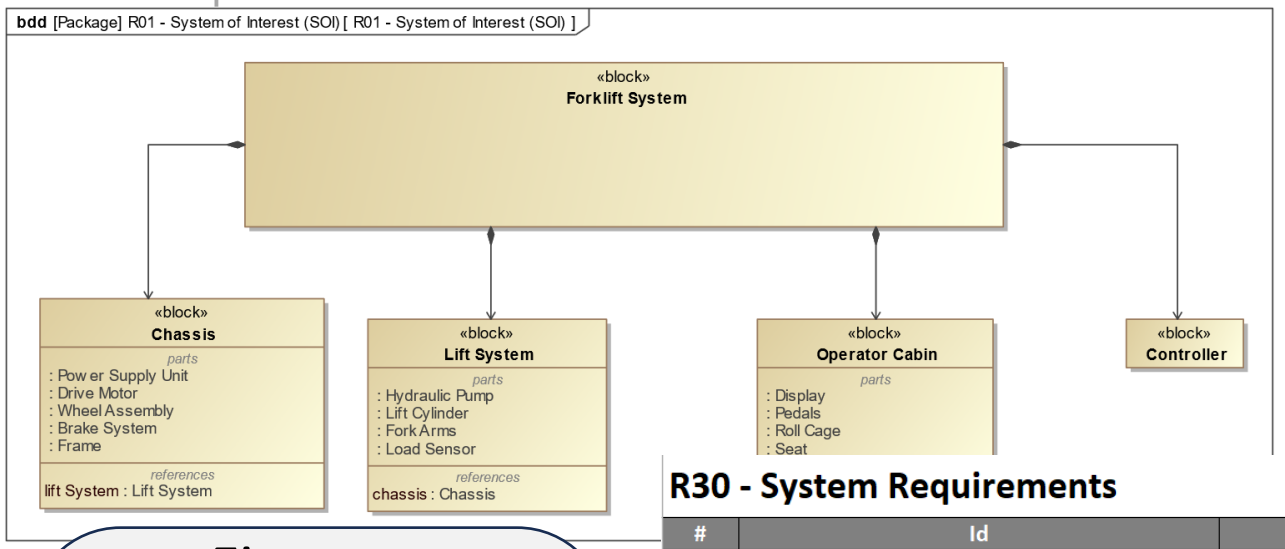
2. Is it Easier?

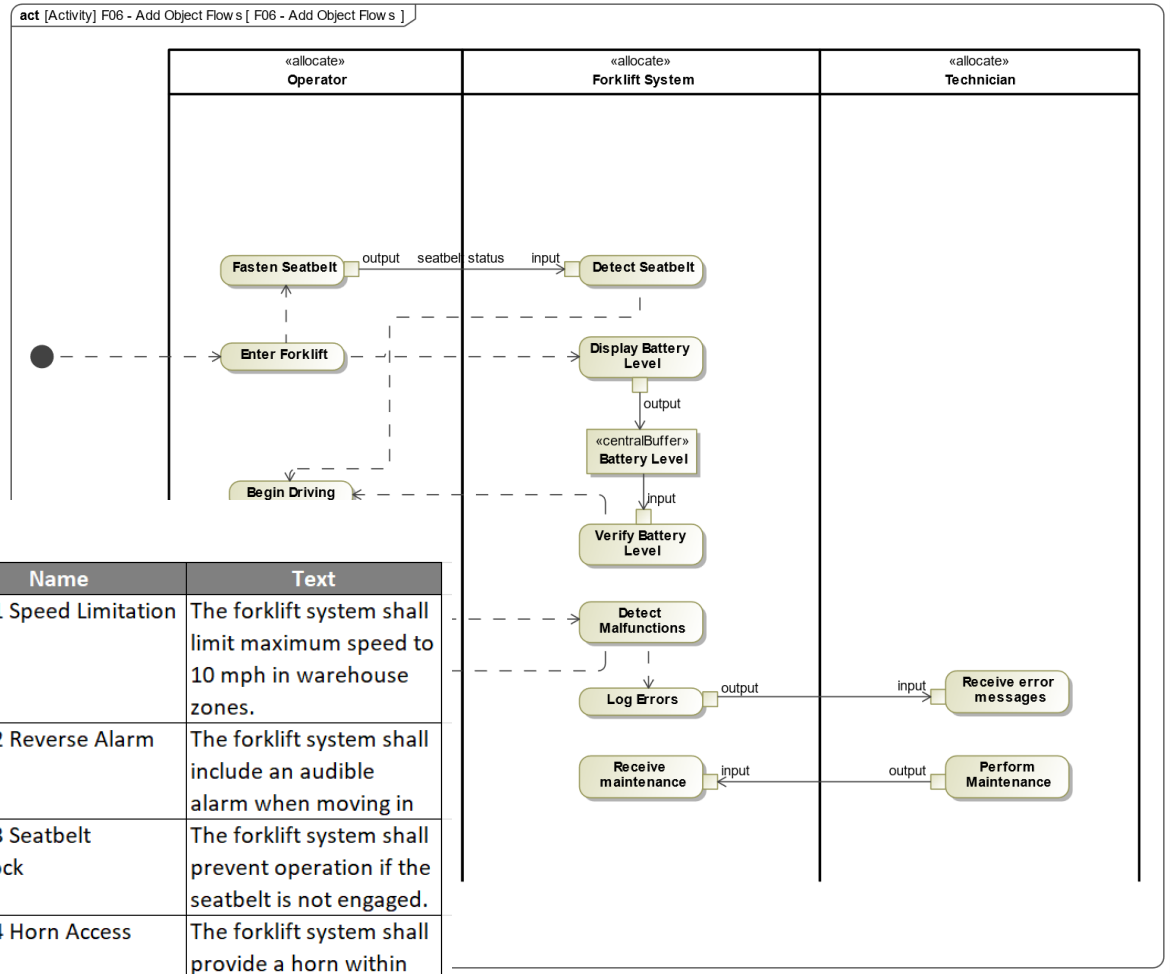
3. Is it Cheaper?

Experiments and Results

Test Model and Requirements Set

Experiments and Results





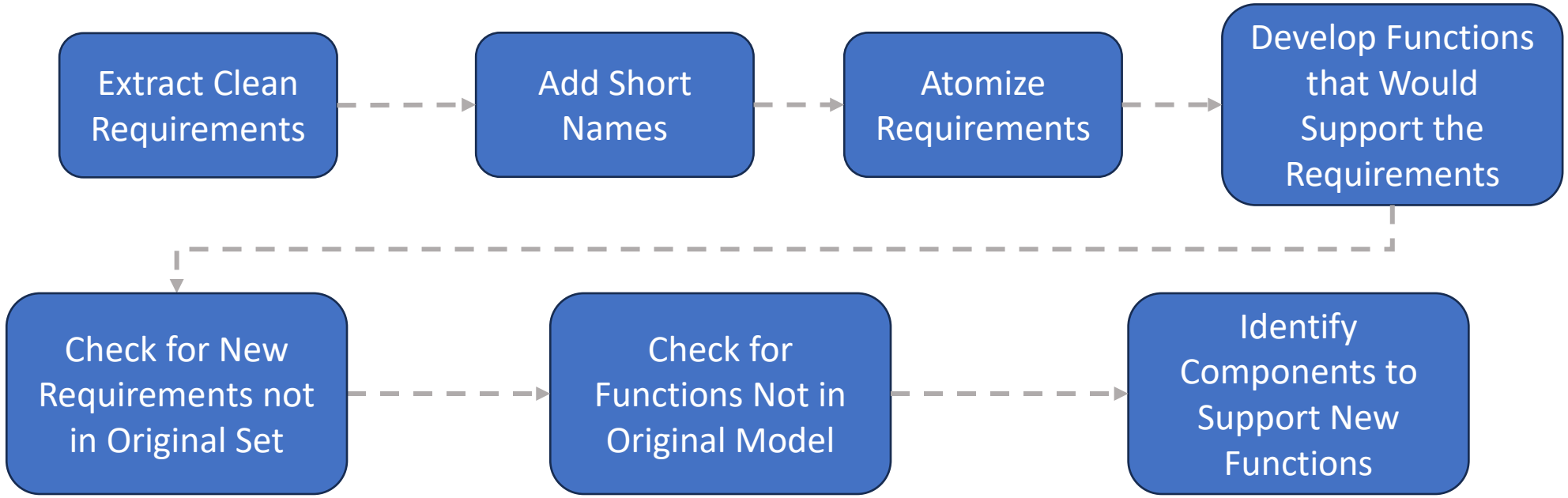
First, we constructed our own forklift requirements and SysML model using a RFLP process.

R30 - System Requirements

#	Id	Name	Text
1	FR-001	FR-001 Speed Limitation	The forklift system shall limit maximum speed to 10 mph in warehouse zones.
2	FR-002	FR-002 Reverse Alarm	The forklift system shall include an audible alarm when moving in
3	FR-003	FR-003 Seatbelt Interlock	The forklift system shall prevent operation if the seatbelt is not engaged.
4	FR-004	FR-004 Horn Access	The forklift system shall provide a horn within operator reach at all times.
5	FR-005	FR-005 Stability Lockout	The forklift system shall restrict lift operation when stability threshold

Strategy for Evaluating the New Requirements

Experiments and Results

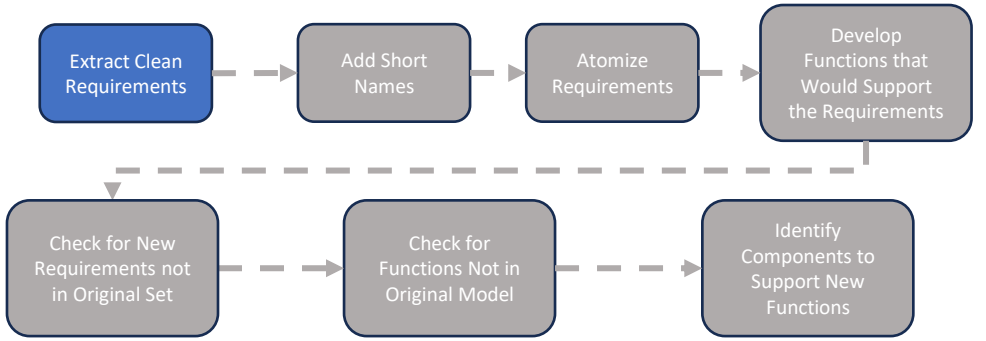


We tested with ChatGPT-5.0, ChatGPT-4.0, and Gemini

Next, we developed a strategy to use LLMs to evaluate the OSHA requirements as though they were an urgent RFQ from a potential customer.

Extract Clean Requirements

Experiments and Results



Not successful. LLMs failed to extract all requirements. Also, LLMs failed to notice that OSHA requirements referred to other standards which contained the more detailed engineering requirements.

Might be possible to fix with more prompt engineering, but then we would just run into next problem.

Recommended Next Step

To get the precise wording and structure:

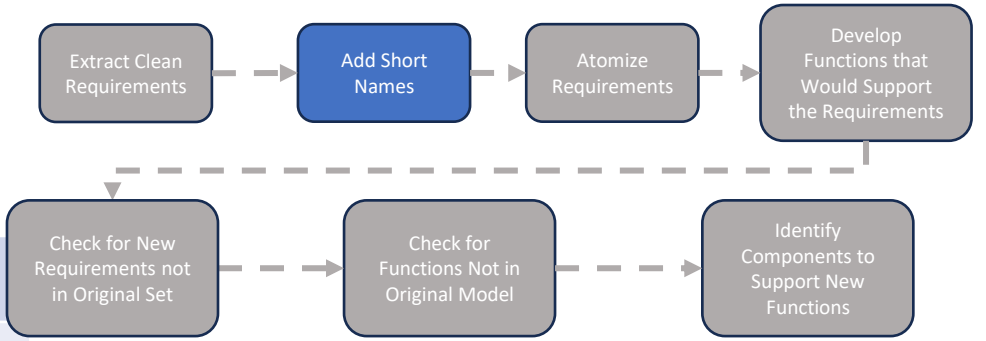
- **Obtain NFPA 505**, ideally the **2018 or most current 2024 edition**, and look at **Section 4.2**, specifically the table and surrounding explanatory text.
- This will provide the definitive mapping of truck types to hazard locations and the rationale or conditions attached.

Let me know if you'd like help locating a specific clause or getting access options!

After prompting to search for the standard, the relevant information was behind a paywall

Add Short Names

Experiments and Results

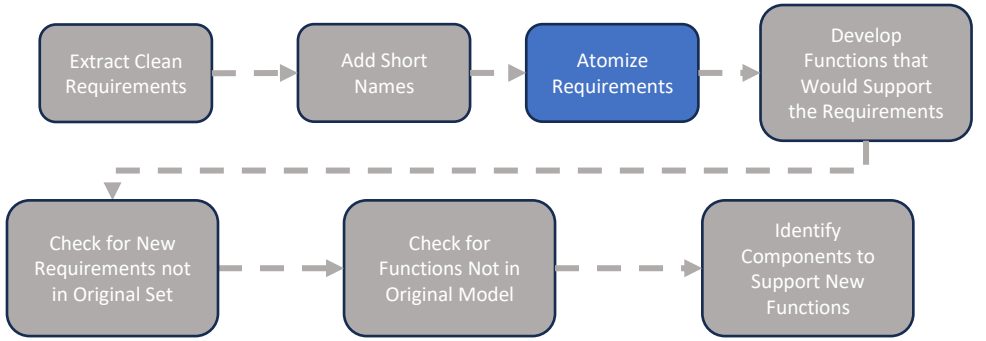


R1	The forklift shall provide a nameplate with serial number, weight, model, and date of manufacturer.
R2	7.39.1 Fork extensions should not be longer than 150% of the supporting fork's length (see Fig. 8).
R3	7.39.2 Each fork extension shall be capable of supporting a uniformly distributed, or equivalent load of three times its rated capacity when mounted on a fork of the specified size. No permanent deformation shall be produced by the application of this test load after having removed the effects of any local manufacturing irregularities by up to three preliminary applications of the test load.

Very successful

record_id	ID Number	Short ID	text
R1	None	Forklift Nameplate	The forklift shall provide a nameplate with serial number, weight, model, and date of manufacturer.
R2	7.39.1	Fork Ext. Length	Fork extensions should not be longer than 150% of the supporting fork's length (see Fig. 8).
R3	7.39.2	Fork Ext. Capacity	Each fork extension shall be capable of supporting a uniformly distributed, or equivalent load of three times its rated capacity when mounted on a fork of the specified size. No permanent deformation shall be produced by the application of this test load after having removed the effects of any local manufacturing irregularities by up to three preliminary applications of the test[1] load.

Atomize Requirements



Experiments and Results

R1	The forklift shall provide a nameplate with serial number, weight, model, and date of manufacturer.
R2	7.39.1 Fork extensions should not be longer than 150% of the supporting fork's length (see Fig. 8).
R3	7.39.2 Each fork extension shall be capable of supporting a uniformly distributed, or equivalent load of three times its rated capacity when mounted on a fork of the specified size. No permanent deformation shall be produced by the application of this test load after having removed the effects of any local manufacturing irregularities by up to three preliminary applications of the test load.

Did not work well on the first try. However, when we returned a few weeks later, this task seemed to work quite well.

R1 – Nameplate

- R1a The forklift shall provide a nameplate with the serial number.
- R1b The forklift shall provide a nameplate with the weight.
- R1c The forklift shall provide a nameplate with the model.
- R1d The forklift shall provide a nameplate with the date of manufacture.

R2 – Fork Extension Length

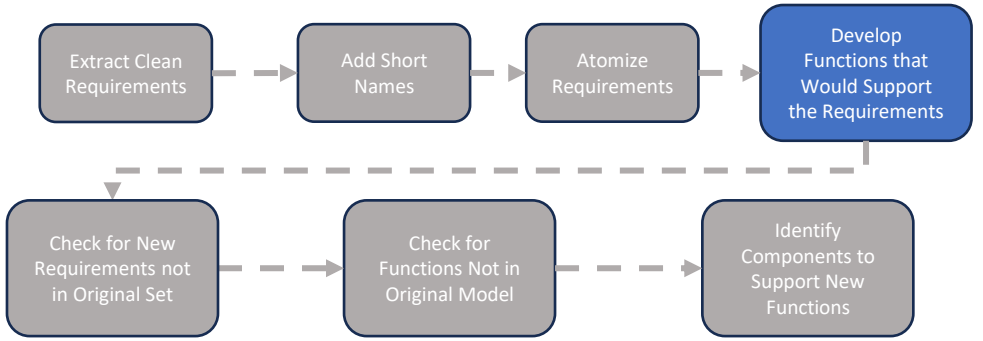
- R2 Fork extensions shall not exceed 150% of the supporting fork's length.

R3 – Fork Extension Strength

- R3a Each fork extension shall support a uniformly distributed or equivalent load equal to three times its rated capacity when mounted on a specified fork.
- R3b No permanent deformation shall occur after application of the rated test load.
- R3c Up to three preliminary applications of the test load may be used to eliminate the effects of local manufacturing irregularities.

Develop Functions

Experiments and Results



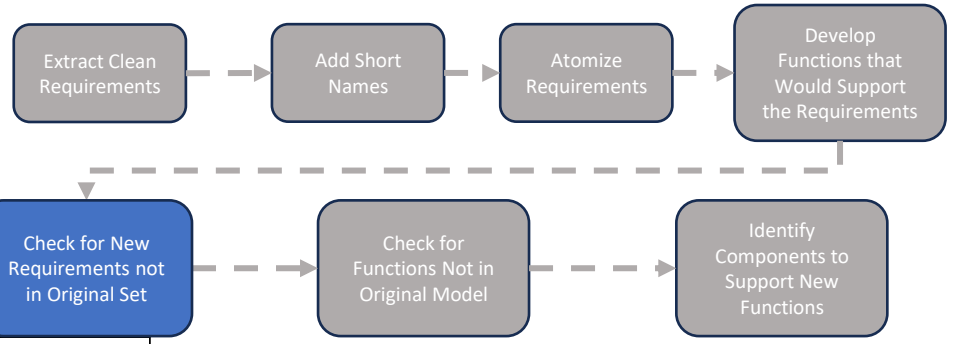
ID	Short ID	Requirement Summary	Generated Function Name
R1	Forklift Nameplate	Provide nameplate with serial, model, and weight info	provide nameplate
R2	Fork Ext. Length	Limit fork extension length	limit extension length
R3	Fork Ext. Capacity	Support 3× rated capacity	support rated load
R4	Fork Ext. Load Center	Set load center at 50% length	set load center
R5	Fork Ext. Stamping	Stamp visible load rating	stamp load rating

Prompt:
Based off of the requirements in this file, generate me unique function names. These functions should use concise verb phrases and summarize the requirements. For example
if the requirement reads: "the kitchen shall make coffee," the function would be "provide coffee."
or
if the requirement reads: "the toaster shall toast bread," the function would be "toast bread."
or
if the requirement reads: "the airplane shall fly passengers to their pre-determined destination," the function would be "transport passengers."

Success. After refining prompt to use 3-shot (3 examples)

Check for New Requirements

Experiments and Results



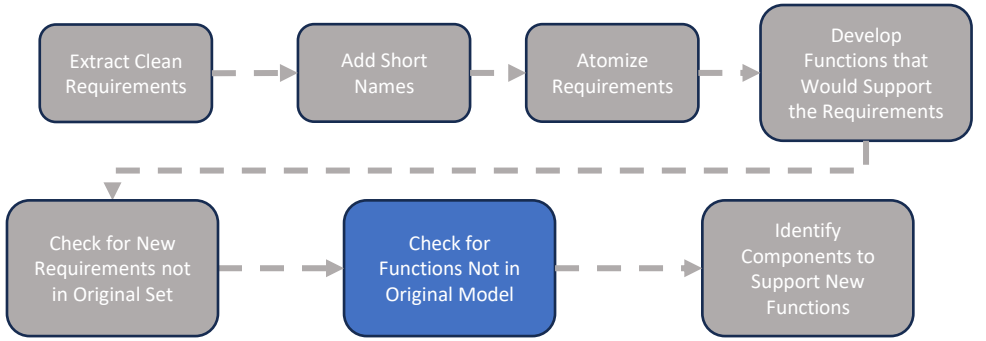
	Req #	Powered Trucks															
Req #		R1	R2	R3	R4	R5	R6	R8	R7	R9	R10	1	2	3	4	5	6
R30	1																
	2																
	3																
	4																
	5																
	6																
	7																
	8																
	9																
	10																
	11																
	12																
	13																
	14																
	15																
	16																
	17																
	18																
	19																
	20																
	21																
	22																
	23																
	24																
	25																

Legend
Exact Match
Similar
Different Name
Altered text

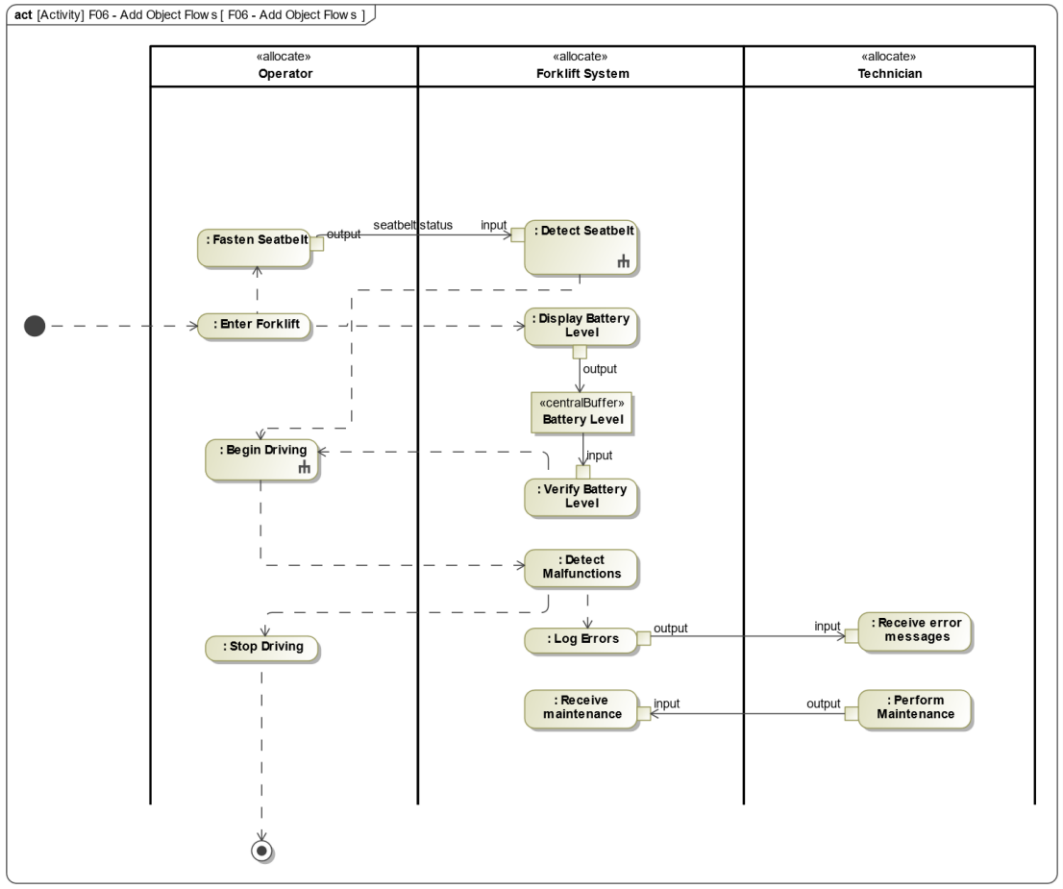
Tried. Success. Still needed some manual checking. Results varied between LLMs. Gemini did the best. ChatGPT4 and ChatGPT5 identified different requirements.

LLM	Notes
Gemini	Properly diagnosed all 7 similar requirements despite changes in the name, altered text, and different wording
Chat GPT5	(5 Identified) FR-001 through FR-003 along with FR-005 and FR-006 but missed both FR004 (different name) and failed to recognize the similarities between R30-21 and Powered Trucks-R1
Chat GPT4o	(5 Identified) Same results as GPT5 but instead listed FR-006 as a "similar match"

Check for Functions Not in Original Model



Experiments and Results



✅ Functions in Your Text List

- provide nameplate
- limit extension length
- support rated load
- interlock seatbelt
- provide horn

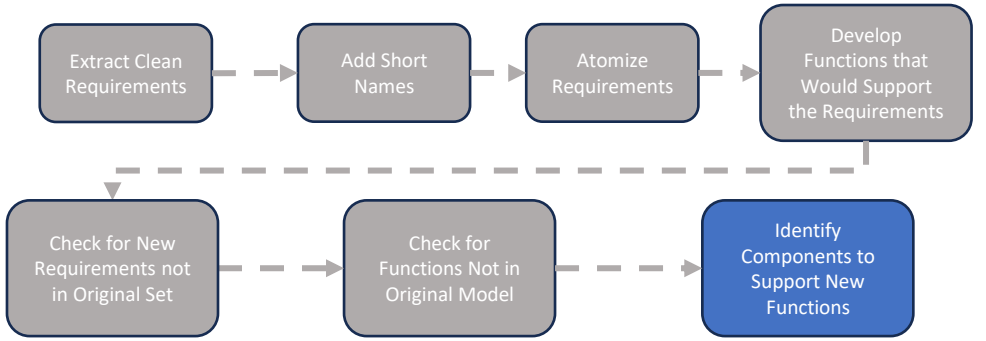
🔎 **Summary**

- **Exact Matches:** None.
- **Partial Matches:**
  - *interlock seatbelt* ↔ *Fasten Seatbelt / Detect Seatbelt*.
- **Unique to List:** Regulatory/safety/structural requirements.
- **Unique to Diagram:** Operational and maintenance functions.

Function (List)	Function (Diagram)	Match Status
provide nameplate	—	❌ No Match
limit extension length	—	❌ No Match
support rated load	—	❌ No Match
interlock seatbelt	Detect Seatbelt / Fasten Seatbelt	⚠️ Partial Match

Identify Components to Support New Functions

Experiments and Results

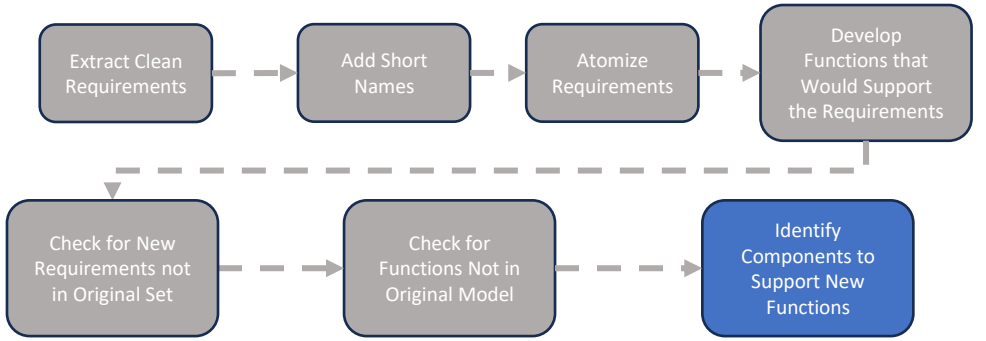


Function Name	Chassis	Lift System	Operator Cabin	Controller
provide nameplate			✅	
limit extension length		✅		
support rated load		✅		
set load center		✅		
stamp load rating		✅		
prevent disengagement		✅		
include radiator caps			✅	
restrain battery	✅			
limit battery movement	✅			
limit speed	✅			✅
provide reverse alarm			✅	
interlock seatbelt			✅	✅
provide horn			✅	
restrict lift	✅	✅		✅
support turning radius	✅			✅

Chat GPT-5.0

Identify Components to Support New Functions

Experiments and Results



Function ID	Function Name	Chassis	Lift System	Operator Cabin	Controller
R1	provide nameplate	✅			
R2	limit extension length		✅		
R3	support rated load		✅		
R4	set load center		✅		
R5	stamp load rating		✅		
R6	prevent disengagement		✅		
R8	include radiator caps	✅			
R9	restrain battery	✅			
R10	limit battery movement	✅			
1	limit speed	✅			✅
2	provide reverse alarm			✅	
3	interlock seatbelt			✅	✅
4	provide horn			✅	
5	restrict lift		✅		✅
6	support turning radius	✅			

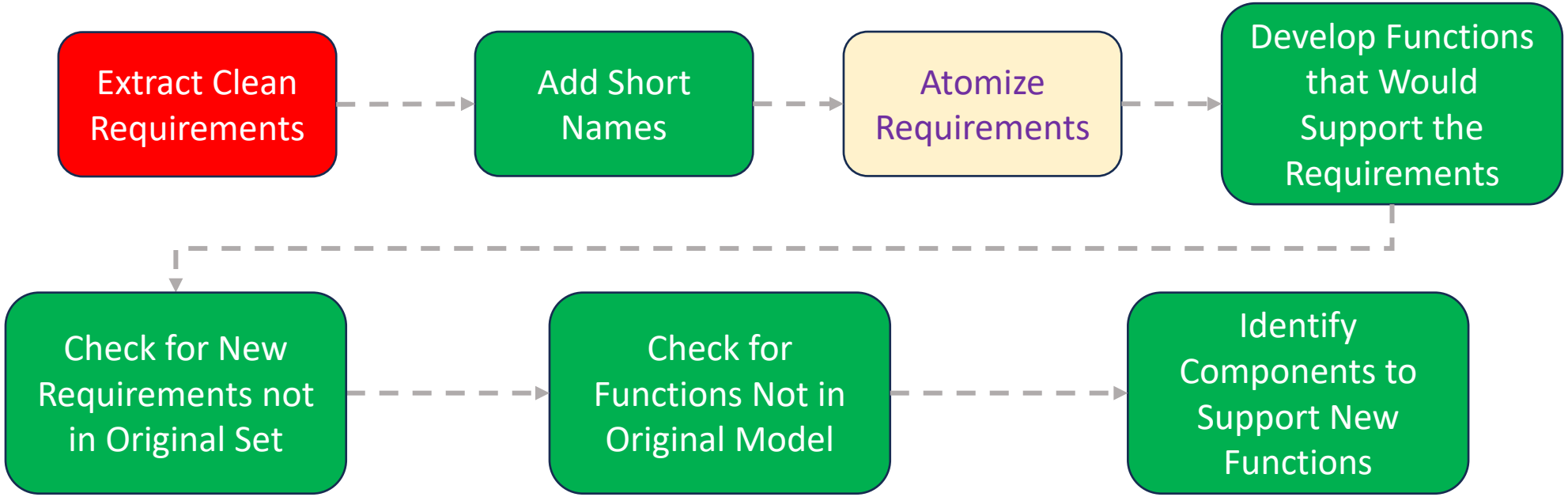
Chat GPT-4o

Conclusions

1. Does it Work?

Conclusions

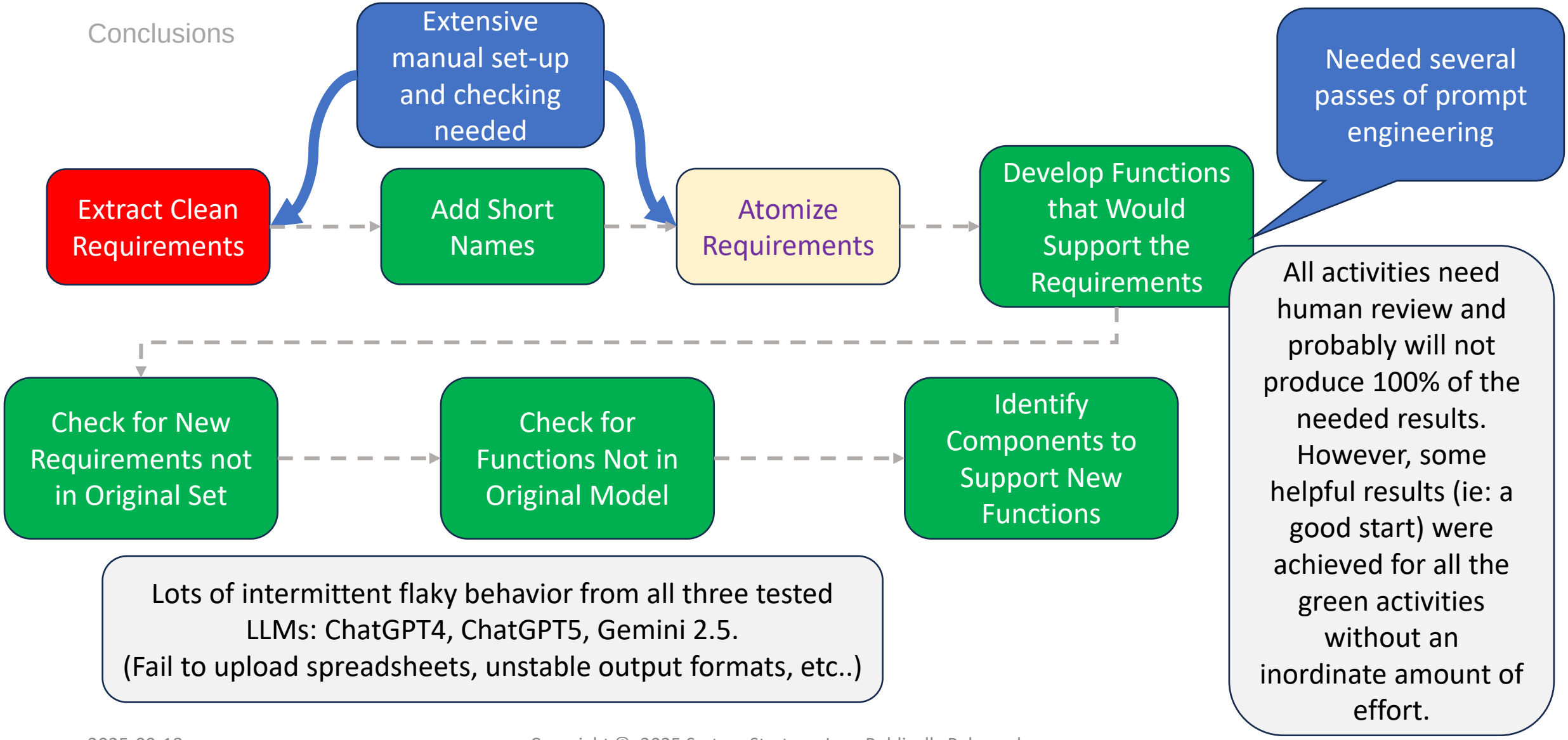
Extracting clean requirements from the human-written legacy documents did not work well. However, most of the tasks did work pretty well.



Lots of intermittent flaky behavior from all three tested LLMs: ChatGPT4, ChatGPT5, Gemini 2.5.
(Fail to upload spreadsheets, unstable output formats, etc..)

2. Is it Easier?

Conclusions



3. Is it Cheaper?

Conclusions

FORTUNE

Sam Altman says he's losing money on OpenAI's $200-per-month subscriptions: 'People use it much more than we expected'

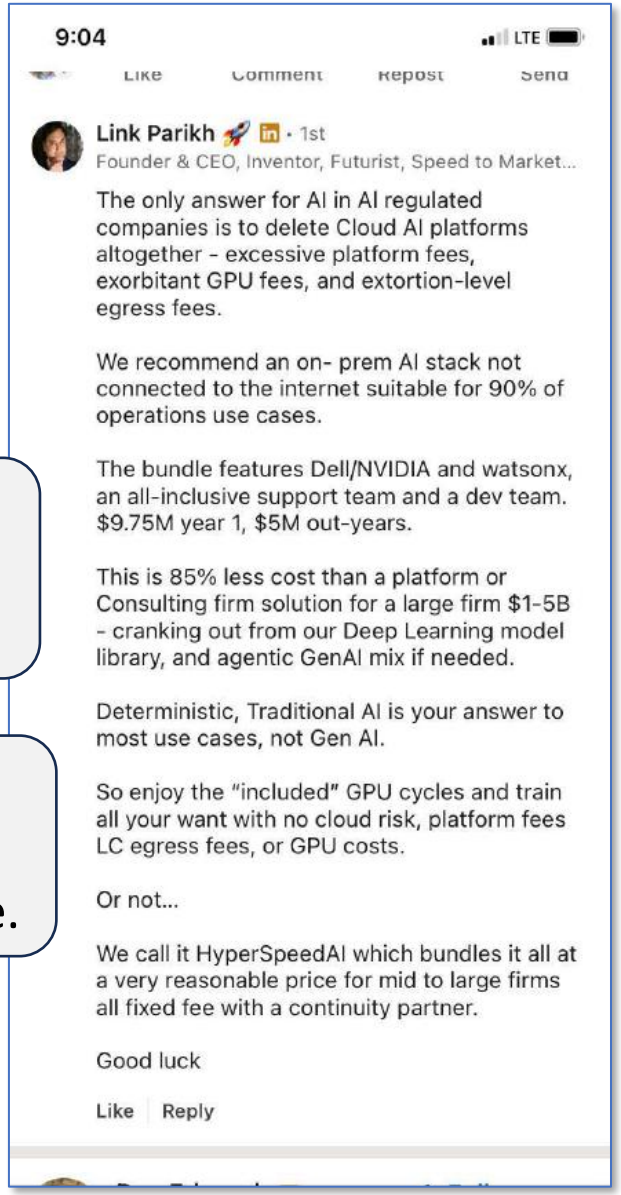
OpenAI CEO Sam Altman. · Fortune · Mike Coppola—Getty Images

https://finance.yahoo.com/news/sam-altman-says-losing-money-080700756.html

2026-09-06
Public LinkedIn estimate of on-premise AI cost:
$9.75m – year 1
$2m – years 2+

At $200/month, AI looks promising. However, the outlook is for the AI to get much more expensive.

Also, to make this approach work at scale, we would need to use the API which is already much more expensive.



9:04 LTE

Like Comment Repost Send

Link Parikh · 1st
Founder & CEO, Inventor, Futurist, Speed to Market...

The only answer for AI in AI regulated companies is to delete Cloud AI platforms altogether - excessive platform fees, exorbitant GPU fees, and extortion-level egress fees.

We recommend an on- prem AI stack not connected to the internet suitable for 90% of operations use cases.

The bundle features Dell/NVIDIA and watsonx, an all-inclusive support team and a dev team. $9.75M year 1, $5M out-years.

This is 85% less cost than a platform or Consulting firm solution for a large firm $1-5B - cranking out from our Deep Learning model library, and agentic GenAI mix if needed.

Deterministic, Traditional AI is your answer to most use cases, not Gen AI.

So enjoy the "included" GPU cycles and train all your want with no cloud risk, platform fees LC egress fees, or GPU costs.

Or not...

We call it HyperSpeedAI which bundles it all at a very reasonable price for mid to large firms all fixed fee with a continuity partner.

Good luck

Like Reply

Conclusions

Results so Far and Future Work

Conclusions

Requirements Extraction – Not very promising. Considerable human oversight needed. Not clear that the LLM ends up adding value.

Short Names – Using the LLM for this tedious task shows great promise.

Atomize Requirements – More work needed, but looks promising.

Function Names – Using the LLM worked well after we used a 3-Shot prompt.

Model Matching – Worked well after some prompt engineering.

Overall, more work is needed to turn this from a research project to a repeatable engineering process. However, the results do look promising

Questions?

Questions to David Hetherington
dhetherington@systemxi.com
David_Hetherington@ieee.org