P24-112: Introducing SysEngBench: A Novel Benchmark for Assessing Large Language Models in Systems Engineering

Ryan Bell
Ryan Longshore
Raymond Madachy, PhD
Naval Postgraduate School

Acquisition Research Symposium
9 May 2024

Agenda

- Research Question
- Research Issue
- Research Methodology
- Results
- Recommendations and Future Work

Research Question

How do we assess, compare, and leverage the performance of Large Language Models (LLMs) in the field of Systems Engineering?

Research Issue

- Large Language Models (LLMs) such as GPT-4 have revolutionized the field of natural language processing (NLP) by demonstrating an impressive ability to understand and generate text and
  - Applications: Writing assistance, chat bots, code generation, summarization
  - Types: Open source and proprietary
  - Training Sources: public (GitHub, Wikis), private (textbooks, journals)
  - Varying levels of fidelity: Quantization and model parameter sizes

- How do we currently evaluate LLM proficiency?
  - Benchmarking
  - Early benchmarks focusing on foundational tasks such as work relationships and their semantic similarities to more recent, increasing complexity benchmarks such as College Medicine, Physics, Biology, Comp Sci, Math, Electrical Engineering, among others
  - We can see this progression with increased complexity and domain specific nature of the benchmarks over time

- Benchmarks for domain specific topics are sparse
  - Current benchmarks do not include system engineering specific
  - A domain specific benchmark is needed

- SysEngBench, a Systems Engineering LLM benchmark
  - Encompasses a comprehensive set of tasks derived from core systems engineering processes, including requirements analysis, system architecture design, risk management, and stakeholder communication
  - When complete, will leverage a diverse array of real-world and synthetically generated scenarios in addition to conceptual questions

Benchmarks Over Time

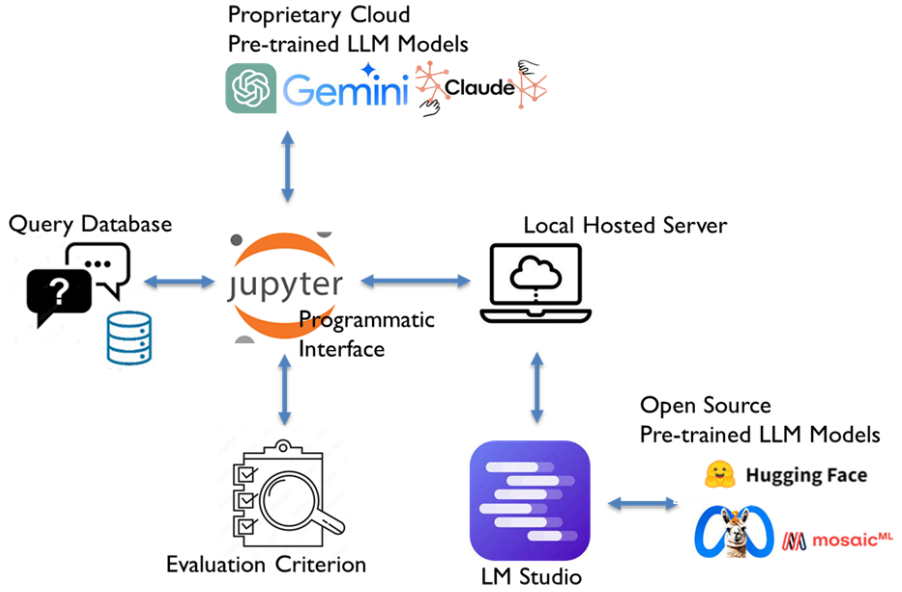
Benchmark Name	Topic	Released	Type of Benchmark
WordNet	Word relationships and meanings, foundational dataset for semantic similarity and language understanding	1985	Natural Language Processing
MNIST	Handwritten digit recognition, foundational for image processing and computer vision	1998	Image Processing
BLEU	Language translation quality metric, foundational for evaluating machine translation systems	2002	Natural Language Processing
Enron Emails	Recognizing names, entities, and information extraction from natural email datasets	2004	Natural Language Processing
ImageNet	Large-scale image recognition and classification, pivotal in advancing deep learning in computer vision	2009	Image Processing
LAMBADA	Understanding context and reasoning in text, focusing on predicting sentence endings (Paperno et al., 2016)	2016	Natural Language Processing
SWAG	Common sense reasoning and predicting plausible sentence endings in a given context (Zellers et al., 2018)	2018	Natural Language Processing
GLUE	A collection of diverse NLU tasks like question answering and sentiment analysis to advance language understanding across various contexts.	2018	Natural Language Processing
SuperGLUE	A successor to GLUE with more challenging tasks, pushing the limits of NLU models with advanced reasoning and co-reference resolution.	2019	Natural Language Processing
HellaSWAG	An extension of SWAG for more challenging common sense reasoning scenarios (Zellers et al., 2019)	2019	Natural Language Processing
ARC	"ARC evaluates an AI's ability to tackle each task from scratch, using only the kind of prior knowledge about the world that humans naturally possess, known as core knowledge." (Clark et al., 2018; Lab42, 2024)	2019	Natural Language Processing
DROP	Reasoning over paragraphs, requires numerical reasoning and understanding of natural language (Dua et al., 2019)	2019	Natural Language Processing
Winogrande	A large-scale dataset of winograd schemas designed to improve commonsense reasoning in AI systems.	2019	Natural Language Processing
XTREME	Cross-lingual understanding and translation across multiple languages, tests multilingual capabilities	2020	Natural Language Processing
MMLU	Measures professional and academic knowledge across various fields including College Medicine, Physics, Biology, Comp Sci, Math, Electrical Engineering, Professional Accounting, Psychology and worldly knowledge about Foreign Policy and Religions, among others (Hendrycks et al., 2021)	2021	Natural Language Processing
TruthfulQA	A question-answering dataset designed to evaluate a model's ability to produce truthful and factual answers.	2021	Natural Language Processing
GSM8K	Grade School Math 8K (GSM8K), a collection of math word problems aimed at evaluating numerical reasoning	2021	Natural Language Processing
BIG-Bench	Broad spectrum of tasks testing reasoning, common sense, professional knowledge, and language capabilities (*Google/BIG-Bench*, 2021/2024)	2022	Natural Language Processing

Research Methodology

Language Models Used

Source	Model	Size	Quantization
TheBloke	Orca-2-7B-GGUF	7.16GB	8 bit
TheBloke	OpenHermes-2.5-Mistral-7B-GGUF	7.70GB	8 bit
TheBloke	Llama-2-7B-Chat-GGUF	7.16GB	8 bit

- Open source models selected
- Standardized process for querying
  - In order of increasing structure for Q&A:
    - Method 1: OpenAI simple Q&A chat completions
    - Method 2: Langchain
    - Method 3:
      - Langchain + AI harness
      - Langchain + HELM
- Industry standard assessment with percent correct



LLM Evaluation Framework

T	Model	Average ⬆	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSM8K
🔶	abacusai/Smaug-72B-v0.1	80.48	76.02	89.27	77.15	76.67	85.08	78.7
🔶	ibivibiv/alpaca-dragon-72b-v1	79.3	73.89	88.16	77.4	72.69	86.03	77.63
🔶	cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_DPO_f16	77.91	74.06	86.74	76.65	72.24	83.35	74.45
🔶	saltlux/luxia-21.4b-alignment-v1.0	77.74	77.47	91.88	68.1	79.17	87.45	62.4

HuggingFace Leaderboards Screenshot

Results

- Performance Levels:
  - Mistral at 89%, Orca 2 at 79%, Llama 2 at 78%.
- Topic with largest differential between models:
  - Requirements questions where Mistral was a clear leader with 22 correct out of 22, followed by Llama 2 with 17 and Orca with 15
- Worst topics for each model:
  - Llama 2 by percentage was architecture
  - Mistral by percentage was functional analysis
  - Orca 2 by percentage was functional analysis
- Challenges and Limitations
  - Few LLM answers would have a letter selection followed by the choice verbiage and/or justification
  - Iterative refinement of the system message was required until the output was constant
- Going forward, tighter integration with LangChain and lm-evaluation-harness should solve these issues

			0.784482759	0.896551724	0.793103448
Row Labels	Question Count	Question %	20240320 LLaMA 2	20240320 Mistral	20240320 Orca 2
Fundamentals of SE	**116**	**100.00%**	**91**	**104**	**92**
SE Definitions	9	7.76%	8	9	9
Problem Definition and Stakeholders	11	9.48%	7	8	7
MBSE Overview	4	3.45%	3	3	3
Requirements	22	18.97%	17	22	15
Functional Analysis	11	9.48%	6	6	5
Value System Design	13	11.21%	12	12	13
Architecture	6	5.17%	3	5	4
Decision Making	10	8.62%	7	9	7
Risk	3	2.59%	2	3	3
System Integration, Qualification, Costs, Life Cycle Issues	27	23.28%	26	27	26
Grand Total	**116**	**100.00%**	**91**	**104**	**92**

Recommendations and Future Work

- Recommendations:
  - Insight into varying level of LLM performance in Systems Engineering.
  - A knowledge gap has been confirmed and needs to be fully quantified and baselined with SysEngBench.
  - Eventual implications include enhanced efficiency and reduction of cognitive load required for tasks like documentation review, compliance checks, and stakeholder communications enabling engineers to focus more on higher level aspects and navigating the available trade space of the complex system.

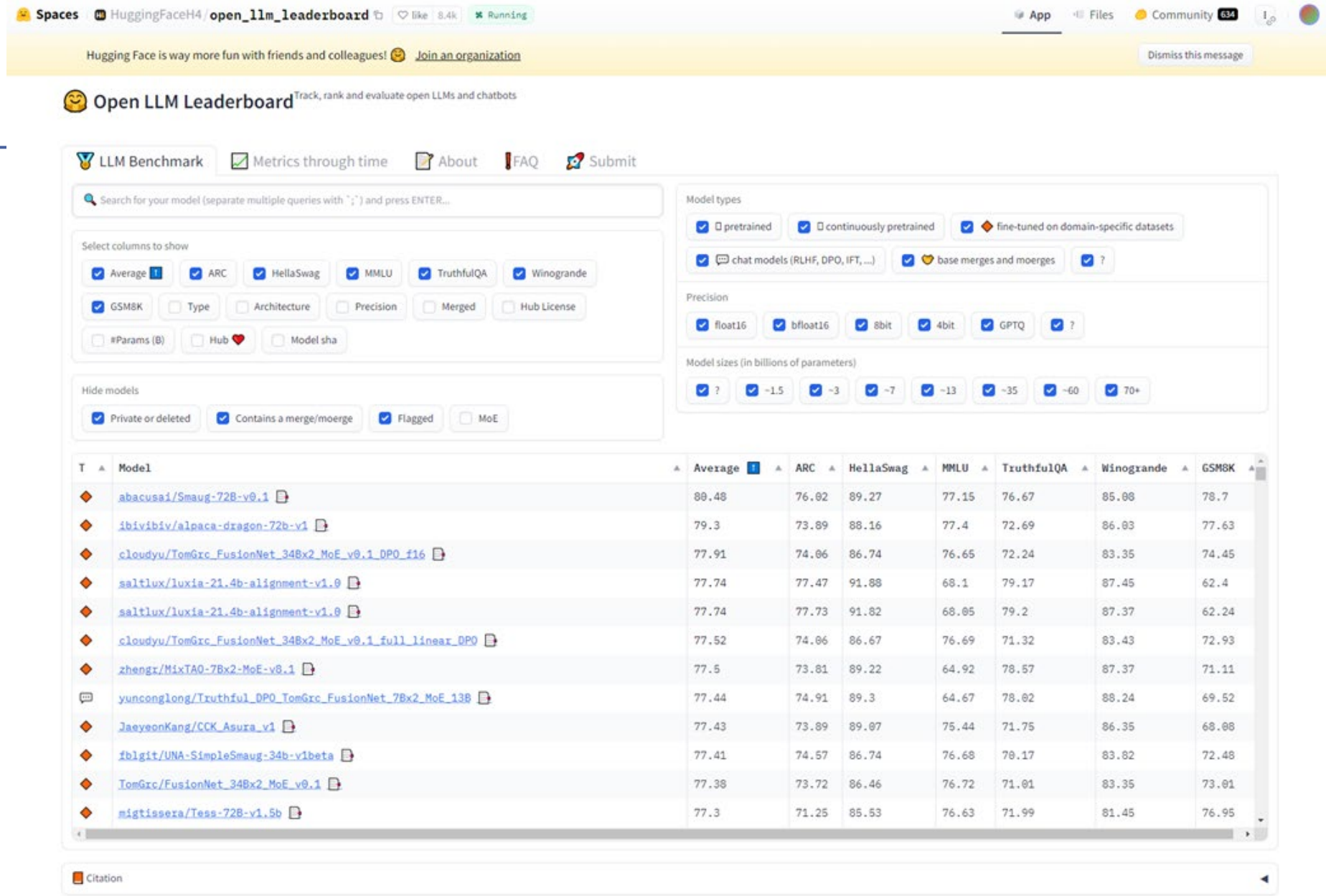
- Future Work:
  - Complex Question Expansion
  - Subfield Diversification
  - Evaluation by Practicing Systems Engineers:
  - Evaluation of Multiple Choice Question Bias within SysEngBench
  - Multimodal Input and Output Evaluation (e.g., diagrams, charts, and technical drawings)
  - Systems Engineering Domain Specific LLMs
  - Enabling Round Table AI Discussions with an AI Agent Systems Engineering Team

- Collaboration Efforts:
  - Ryan Longshore
    - Small Language Models for Domain Specific Knowledge
    - Evaluation of LLMs with SysMLv2 Queries
  - Dr. Raymond Madachy
    - Evaluation of LLMs for Modern Systems Engineering Cost Modeling with COSYSMO


QUESTIONS?

References

- Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., & Tafjord, O. (2018). *Think you have Solved Question Answering? Try ARC, the AI2 Reasoning Challenge* (arXiv:1803.05457; Version 1). arXiv. http://arxiv.org/abs/1803.05457
- Dua, D., Wang, Y., Dasigi, P., Stanovsky, G., Singh, S., & Gardner, M. (2019). *DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs* (arXiv:1903.00161; Version 2). arXiv. http://arxiv.org/abs/1903.00161
- *Google/BIG-bench*. (2024). [Python]. Google. https://github.com/google/BIG-bench (Original work published 2021)
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). *Measuring Massive Multitask Language Understanding* (arXiv:2009.03300; Version 3). arXiv. http://arxiv.org/abs/2009.03300
- Lab42. (2024). *About ARC*. https://lab42.global/arc/
- Longshore, R., Madachy, R., & Bell, R. (in press). *Leveraging Generative AI to Create, Modify, and Query MBSE Models*. 21st Annual Acquisition Research Symposium.
- Madachy, R., Bell, R., & Longshore, R. (in press). *Systems Acquisition Cost Modeling Initiative for AI Assistance*. 21st Annual Acquisition Research Symposium.
- Paperno, D., Kruszewski, G., Lazaridou, A., Pham, Q. N., Bernardi, R., Pezzelle, S., Baroni, M., Boleda, G., & Fernández, R. (2016). *The LAMBADA dataset: Word prediction requiring a broad discourse context* (arXiv:1606.06031). arXiv. https://doi.org/10.48550/arXiv.1606.06031
- Zellers, R., Bisk, Y., Schwartz, R., & Choi, Y. (2018). *SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference* (arXiv:1808.05326; Version 1). arXiv. http://arxiv.org/abs/1808.05326
- Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., & Choi, Y. (2019). *HellaSwag: Can a Machine Really Finish Your Sentence?* (arXiv:1905.07830; Version 1). arXiv. http://arxiv.org/abs/1905.07830


Spaces HuggingFaceH4 / open_llm_leaderboard Like 8.4k Running
App Files Community 634
Hugging Face is way more fun with friends and colleagues! Join an organization
Dismiss this message
Open LLM Leaderboard Track, rank and evaluate open LLMs and chatbots
LLM Benchmark Metrics through time About FAQ Submit
Search for your model (separate multiple queries with ";") and press ENTER...
Select columns to show
Average ARC HellaSwag MMLU TruthfulQA Winogrande
GSM8K Type Architecture Precision Merged Hub License
#Params (B) Hub Model sha
Hide models
Private or deleted Contains a merge/moerge Flagged MoE
Model types
pretrained continuously pretrained fine-tuned on domain-specific datasets
chat models (RLHF, DPO, IFT, ...) base merges and moerges ?
Precision
float16 bfloat16 8bit 4bit GPTQ ?
Model sizes (in billions of parameters)
? ~1.5 ~3 ~7 ~13 ~35 ~60 70+
T Model Average ARC HellaSwag MMLU TruthfulQA Winogrande GSM8K
abacusai/Smaug-72B-v0.1 80.48 76.02 89.27 77.15 76.67 85.08 78.7
ibivibiv/alpaca-dragon-72b-v1 79.3 73.89 88.16 77.4 72.69 86.03 77.63
cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_DPO_f16 77.91 74.06 86.74 76.65 72.24 83.35 74.45
saltlux/luxia-21.4b-alignment-v1.0 77.74 77.47 91.88 68.1 79.17 87.45 62.4
saltlux/luxia-21.4b-alignment-v1.0 77.74 77.73 91.82 68.05 79.2 87.37 62.24
cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_full_linear_DPO 77.52 74.06 86.67 76.69 71.32 83.43 72.93
zhengr/MixTAO-7Bx2-MoE-v8.1 77.5 73.81 89.22 64.92 78.57 87.37 71.11
yunconglong/Truthful_DPO_TomGrc_FusionNet_7Bx2_MoE_13B 77.44 74.91 89.3 64.67 78.02 88.24 69.52
JaeyeonKang/CCK_Asura_v1 77.43 73.89 89.07 75.44 71.75 86.35 68.08
fblgit/UNA-SimpleSmaug-34b-v1beta 77.41 74.57 86.74 76.68 70.17 83.82 72.48
TomGrc/FusionNet_34Bx2_MoE_v0.1 77.38 73.72 86.46 76.72 71.01 83.35 73.01
migtissera/Tess-72B-v1.5b 77.3 71.25 85.53 76.63 71.99 81.45 76.95
Citation

	Category	Sub-Category	Tags	Source	Generation	Question	Choice A	Choice B	Choice C	Choice D	Answer	Justification	Notes	QA Review #1
100	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	What best describes a system?	A singular eler	A group of ele	A random coll	An isolated tec	B	A system is defined as a group of elements or components that work together towards a specified pu		
101	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	Which of the following best encapsulates the basic meth	Focus solely o	Understand th	Start solving t	Ignore alternat	B	Systems Engineering (SE) is characterized by a disciplined approach to problem-solving, emphasizing		
102	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	What are the core aspects of Systems Engineering (SE)?	Focusing solel	Design, produ	Ignoring const	Concentrating	B	Systems Engineering is an interdisciplinary field that focuses on the design, production, and mainten		
103	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	According to INCOSE, what is the focus of Systems Engin	Concentrating	Focusing only	Integrating dis	Ignoring custor	C	INCOSE defines Systems Engineering as an interdisciplinary approach aimed at realizing successful sy		
104	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	What is the primary role of Systems Engineering?	To focus exclu	To ensure cust	To limit the de	To avoid testin	B	The primary role of Systems Engineering is to study, define, and specify the operational, functional, a		
105	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	What does Systems Engineering offer in the technical de	An undisciplin	Solely a meth	A disciplined a	Ignoring the id	C	Systems Engineering provides a structured and logical methodology for the technical development o		
106	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	How is "architecture" defined in the context of systems	The random pl	The arrangem	Solely the phy	Ignoring the str	B	In systems engineering, "architecture" refers to the deliberate and strategic arrangement of element		
107	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	What best describes the role of systems architecting?	Ignoring client	Only focusing	Planning and I	Acting as a gen	C	Systems architecting is a discipline that merges the theory and practice of architecting with systems t		
108	Fundamentals	SE Definitions		SE 3100 1SE+Definitic	Semi	What encompasses the process of management accordir	Solely focusin	Engaging in pl	Ignoring the o	Exclusively lea	B	According to Dym & Little, management is defined as the process of achieving organizational goals th		
109	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which of the following best describes the initial trigger f	Completion of	Discovery of a	An emerging r	Random inspira	C	Systems Engineering processes often begin in response to an emerging need that can be categorized		
110	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which question category best aids in identifying the targ	Why	Who	What	Where	B	In the context of Systems Engineering (SE) analysis, the "Who" category is essential for identifying th		
111	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which of the following techniques is primarily used in Sy	Five Why's	Ishikawa Fisht	Causal Loop D	SWOT Analysis	A	The Five Why's technique is a systematic problem-solving method that involves asking the question "		
112	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	In Systems Engineering, which technique is specifically c	SWOT Analysis	Five Why's	Ishikawa Fisht	Causal Loop Dia	D	The Causal Loop Diagram is a graphical tool used in Systems Engineering for visualizing the interactio		
113	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which problem-space exploration tool in Systems Engin	Functional Ana	Stakeholder A	Scenario-base	Operational Co	B	Stakeholder Analysis, which includes research and interviews, is a critical tool in Systems Engineering		
114	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	In the context of Systems Engineering, which of the follc	Beneficial Stak	Engineers, Tec	Suppliers, Cor	Internal Stakeh	A	Stakeholders in Systems Engineering can be broadly categorized based on their interest and involven		
115	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	What is the first step in conducting a stakeholder analysi	Develop a list	Conduct resea	Identify relev	Consolidate inf	C	The first step in conducting a stakeholder analysis in Systems Engineering is to identify relevant stake		
116	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which question is critical for understanding the primary	What are the r	What are the c	Who are the u	What functions	D	Understanding the primary functions and intended use of a system from a customer's perspective is c		
117	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	What is the primary purpose of the elicitation process in	To finalize the	To identify pa	To determine	To assign tasks	B	The primary purpose of the elicitation process in Systems Engineering is to deeply understand the sta		
118	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which elicitation technique in Systems Engineering invo	On-site key sta	Active stakeh	Written surve	Focus groups	D	Focus groups involve gathering a group of stakeholders to discuss and provide feedback on system re		
119	Fundamentals	Problem Definition and Stakel		SE 3100 2problem+ar	Semi	Which of the following is a direct outcome of the require	A comprehens	A detailed pro	A bounded sta	Finalized syste	C	A bounded statement of scope for the system or product is a direct outcome of the requirements elic		
120	Fundamentals	MBSE Overview		SE 3100 SE3100-MBSE	Semi	What is true about a model in the context of systems en	It is a detailed	It includes eve	It is built with	It emphasizes c	D	In systems engineering, a model is an abstract representation of a real-world system. The purpose of		
121	Fundamentals	MBSE Overview		SE 3100 SE3100-MBSE	Semi	Which of the following is NOT a direct benefit of modeli	Visualization c	Enhancing phy	Facilitating co	Generating anc	B	In systems engineering, modeling provides several benefits including visualization, communication,		
122	Fundamentals	MBSE Overview		SE 3100 SE3100-MBSE	Semi	What is essential for a model to be well-formed in the c	A modeling to	Syntax or gran	A single softw	Elimination of	B	In Model-Based Systems Engineering (MBSE), a well-formed model is governed by a specific syntax o		
123	Fundamentals	MBSE Overview		SE 3100 SE3100-MBSE	Semi	What best describes the distinction between model veri	Verification er	Verification pr	Verification ei	Validation conf	C	In systems engineering, verification and validation (V&V) are critical activities that establish the mod		
124	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What distinguishes effective needs from primitive need	Effective need	Primitive nee	Effective nee	Effective need	C	In systems engineering, effective needs are differentiated from primitive needs based on the suppo		
125	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	Which tool is NOT typically used in need analysis for syst	Stakeholder A	Functional Hie	Pareto Analys	Randomized Cc	D	Need analysis in systems engineering employs a variety of tools to understand and prioritize the nee		
126	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	Which of the following is an example of an effective nee	"Increase the s	"Provide the c	"Improve over	"Design a stora	B	Effective need statements in systems engineering are characterized by their specificity, support by ev		
127	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What is a key element of a Concept of Operations (CONC	Detailed techr	Use of scenari	Focus solely o	Financial analy	B	The Concept of Operations (CONOP) document in systems engineering is crucial for outlining the inte		
128	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	In the development of an Operational Concept using sce	To detail the s	To document t	To assign own	To create a con	C	Partitioning the system threads based on who owns the behavior is a critical step in developing an Op		
129	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What is the main purpose of focusing on the futurity of e	To specify the	To facilitate cc	To conduct a d	To provide hist	B	The primary purpose of focusing on the futurity of events in scenario planning within systems engine		
130	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	In systems engineering, what is the primary purpose of c	To document t	To focus on th	To define exte	To provide a co	C	The primary purpose of defining the system boundary in systems engineering is to define external in		
131	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What is a crucial aspect of functional requirements in sy	Defining the a	Establishing in	Planning the s	Selecting the o	B	A crucial aspect of functional requirements in systems engineering is establishing input, output, and		
132	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	In the context of systems engineering, how are "What" r	"What" requir	"What" requir	"How" require	"How" require	B	"What" requirements in systems engineering describe the system's purpose from an external viewpc		
133	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	How do functional requirements (FR) differ from non-fu	FRs describe tl	NFRs specify t	FRs describe v	NFRs outline th	C	In systems engineering, functional requirements (FRs) describe what the system needs to do. They fo		
134	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	According to the Systems Engineering Guide Book, what	Designing the	Determining s	Planning the s	Selecting the c	B	The primary goal of requirements analysis, as outlined in the Systems Engineering Guide Book, is det		
135	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What is the primary purpose of validating requirements	To ensure the	To confirm the	To make sure	To determine t	C	The primary purpose of validating requirements in systems engineering is to make sure that the requ		
136	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What is a key aspect of requirements management in sy	Selecting the k	Carefully cont	Organizing an	Designing the c	B	A key aspect of requirements management in systems engineering is carefully controlling changes to		
137	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What does requirements traceability mean in the contex	Tracking the fi	Monitoring th	Describing an	Recording the p	C	Requirements traceability in the context of systems engineering refers to the ability to describe and		
138	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What makes the requirements definition process challer	The process re	It involves onl	The process re	Solutions and p	C	The requirements definition process in systems engineering is challenging because it requires interd		
139	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What does the 'M' in SMART requirements stand for, and	Marketable: Tl	Measurable: T	Movable: The	Motivational: T	B	The 'M' in SMART requirements stands for Measurable, implying that a good requirement can be quar		
140	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What characteristic of a SMART requirement ensures it is	Flexible	Specific	Generic	Broad	B	A Specific requirement in the SMART criteria ensures that the requirement is clear, concise, and strai		
141	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What does the 'M' in SMART requirements signify regard	Marketable	Movable	Measurable	Motivational	C	The 'M' in SMART stands for Measurable, meaning that the requirement must be quantifiable or asse		
142	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	In the context of SMART requirements, what does 'Attair	It is ambitious	It is attainable	It is abstract.	It is arbitrary.	B	'Attainable' implies that a requirement is realistically achievable within the existing constraints and r		
143	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What does the 'R' in SMART criteria emphasize about the	Revolutionary	Redundant	Realistic	Random	C	The 'R' for Realistic emphasizes that a requirement must be practical and sensible, taking into accoun		
144	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	What aspect does 'Time-bound' in SMART requirements	Timeless	Temporary	Time-bound	Time-consumir	C	'Time-bound' specifies that each requirement should have a defined timeline or deadline, clarifying		
145	Fundamentals	Requirements		SE 3100 3+needs-ana	Semi	Which of the following examples illustrates a guideline	Use condition	Ensure require	Encourage the	Requirements	B	Poorly written requirements often suffer from the use of words that change conditions, such as "if," "		
146	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	What is a key component of Functional Architecture in sy	Detailed cost a	Hierarchical m	Legal compliai	Marketing strat	B	Functional Architecture within systems engineering focuses on defining the system's functional and p		
147	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	In the context of flight, what is essential according to bo	Legs and Take	Eyes and Sens	Wings and Pro	Brains and Nav	C	To achieve flight, both physical and functional decompositions emphasize different yet complement		
148	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	Which of the following is an example of functional deco	Wings	Produce Verti	Legs	Eyes	B	Functional decomposition involves breaking down a system's operations into its basic functions, expl		
149	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	Which of the following is an example of physical decom	Navigate	Produce Horiz	Wings	Sense Position	C	Physical decomposition focuses on identifying the physical components that make up a system, detai		
150	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	Which statement best reflects the critical realization in t	Identifying tha	Understanding	Recognizing th	Discovering the	A			
151	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	In IDEF0 functional modeling, what role do arrows (arcs)	Represent the	Indicate the hi	Carry the infor	Symbolize the	C	IDEF0 is a method designed to model the decisions, actions, and activities of an organization or syste		
152	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	In the context of process modeling, what element speci	Entry Conditio	Exit Condition	Controls	Mechanism	D	Process modeling involves detailing the various components that contribute to the execution of a pro		
153	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	Which IDEF0 rule ensures that every function within the	Conservation c	Every function	Every function	Outputs from c	C	The IDEF0 framework provides a structured methodology for modeling a system's functions, emphasi		
154	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	According to IDEF0 guidelines, where should feedback b	From the top l	From the bott	Along the vert	In a circular pat	B	IDEF0 is a structured method used for modeling an organization's functions, processes, and systems, p		
155	Fundamentals	Functional Analysis		SE 3100 4+Functional	Semi	How does the Functional Flow Block Diagram (FFBD) mo	FFBD focuses c	FFBD is primar	IDEF0 exclusiv	FFBD and IDEF0	B	The Functional Flow Block Diagram (FFBD) and the IDEF0 model are both tools used in systems engine		

Requirements: Presentation Requirements

- Details
  - Panel #20: Enhancing Acquisition with Artificial Intelligence
  - Date: Thursday, 9 May 2024 - 2:15pm - 3:30pm PT / 5:15pm – 6:30pm ET
  - P24-112: Harnessing AI Tools for Enhanced Risk Identification, Analysis, and Management
  - NOW P24-112: Introducing SysEngBench: A Novel Benchmark for Assessing Large Language Models in Systems Engineering

- Presentation Requirements / Preparing Your Presentation
  - The time allocated for a presentation is no more than 15 minutes.
  - Think in terms of the following slides:
    - A title slide (name, title and affiliation)
    - One slide with the research question
    - 2 or 3 slides covering research issue and methodology
    - 2 or 3 slides covering results and recommendations
    - Target no more than 7 slides (roughly 2 mins. per slide)