AI NODs	NOD Description
NOD-AI-1	Vendor holds rights to County FTT/FST trained model
NOD-AI-2	AI FST employed to train a LLM model
NOD-AI-3	FST/FTTtraining data not archived and no model checkpoints saved
NOD-AI-4	No re-training plan when vendor releases a new LLM base model (after prior LLM model FTT trained by County)
NOD-AI-5	No protections in County FFT/FST trained vendor model for potentially compromising sensitive data contained in the model
NOD-AI-6	County data will be ingested into Vendor LLM model
NOD-AI-7	DLP not to be used on County LLM prompt input
NOD-AI-8	No data access enforcement for RAG
NOD-AI-9	DLP not used on LLM output
NOD-AI-10	Vendor using County LLM prompt input/output for other purposes

NOD-AI-11	Incomplete mitigations protecting against unilateral generative AI decision making

Comments/Guidance

Explain how the County can be protected from the vendor using the County trained model

Explain why a base vendor trained LLM could not be used with Peraton/County providing FTT. Costs to support FST are much larger than FTT; those costs, and risks, must be justified

Explain why data and model checkpoints are not be archived. Employing such techniques can make debugging LLM models much easier.

The County will not be able to take advantage of new vendor training updates to a vendor baseline LLM model they are using. This may later potentially negative impact the solution leveraging the the vendor modek

Explain why protections will not be established to ensure users of the model only have access to data they are entitled to access

Explain why this is necessarry. As a matter of policy the County discourages choosing vendors who will use County data to train their LLM models.

Explain why DLP not being used when Peraton/CoSD FFT/FST trained models are being used or when vendor's LLM will be trained on CoSD Data

Explain why controls will not be put in place to ensure entity receiving RAG data has access rights to it

Explain why DLP is not being used on LLM output when LLM output will leveraged for downstream processing without human review of the LLM output

Explain why the County should allow the vendor to use County LLM prompt input; or LLM output for other purposes.

Explain why the County should allow the LLM to make unilateral decisions when only incomplete mitigations can be put in place in the solution to protect against bad decisions

Category	Item #	Checklist Description	Compliance Status	Compliance Guidance	Comments
AI Governance: AI-Gov		**Generative AI Model Data Issues: Input/Output/Training**			
	1	Does the proposed solution truly require a generative AI (LLM) application to achieve desired County functionality? In other words, could any other type of AI technology (e.g. expert system, predective, etc.) be used without deleterious impact to solution end functionality?		A NOD is not required to use generative AI/LLM technologies---but such technologies do introduce additional governance processes--for instance this spreadsheet must be completely filled out to classify a vendor generative AI solution that is to be used in a new County sokution.	
AI-Gov-Model_Data	2	Will County need to train the Vendor LLM Model?		If compliant, skip to item 9	
	3	Will vendor hold any use rights to the FTT/FST County trained model?		If non-compliant, a NOD (NOD-AI-1) will be required	
AI-Gov-Model_Data	4	If training an existing vendor model input FTT (Fine Tuning Training) in Compliance Staus Field; otherwise input FST (From Scratch Training)		FST is not encouraged and will require a NOD (NOD-AI-2)	
AI-Gov-Model_Data	5	Will input training data be archived, along with selected LLM checkpoints?		If non-compliant, a NOD (NOD-AI-3) will be required.	
AI-Gov-Model_Data	6	If FTT, do you have a plan for re-training the vendor LLM, when they release a new version of their BASE model? Skip to Item 10		If non-compliant, a NOD (NOD-AI-4) will be required	
AI-Gov-Model_Data	7	As County data is being ingested into the LLM via FTT or FST; is any data being ingested whose sensitivity must be protected on output?		If no/compliant, proceed to to item 10	
AI-Gov-Model_Data	8	How would it be insured that sensitive County data is not output to an entity/process which does not have authorization to access the sensitive data?		If no protections are in place to mitigate this risk, a NOD (NOD-AI-5) will be required and this requirement shall be marked as non-compliant.	
AI-Gov-Model_Data	9	Will the vendor model integrate County input prompt data into its LLM (i.e., use County Data be used to train the vendor LLM)?		If compliant/no, skip to item 10. If non-compliant/yes a NOD (NOD-AI-6) will be required.	
AI-Gov-Model_Data	10	Will DLP be used on prompt input?		DLP on input is generally required for Peraton/CoSD FFT/FST trained models. DLP required when vendors LLM will be trained on CoSD Data. NOD (NOD-AI-7) required in such circumstances.	
AI-Gov-Model_Data	11	Does the AI solution being utilized use Retreival Augmented Generation (RAG)?		If no skip to question 13	
AI-Gov-Model_Data	12	Is there a plan in place to ensure the entity interacting with the LLM will not gain access to data retrieved by the LLM via RAG that they othwerwise would not be given access to? Please explain how access to such data (via RAG) will be achieved and enforced		If non-compliant, a NOD (NOD-AI-8) will be required	
AI-Gov-Model_Data	13	Will DLP or any other form of validation/inspection be used on LLM output? Please explain.		Explain why DLP will not be used on output. If LLM output is going to be fed to downstream application without human inspection, a NOD (NOD-AI-9) will be required	
	14	Regardless of whether CoSD data is being ingested into a vendor LLM or not; excluding the LLM; will the vendor be doing anything else with the County data either input or output from their LLM model? For instance will they profile it? e.g., scan input and output; attempt to categorize it and sell this information to third parties, etc.		If the vendor is further touching or accessing the County data IN ANY MANNER a NOD (NOD-AI-10) will be required	
AI-Threats: AI-Th		**OWASP Threats**			
AI-Th-Prompt	15	Prompt Injection: This manipulates a large language model (LLM) through crafty inputs (e.g., "ignore previous instructions," etc.), causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources		Explain how the vendor guards against these attacks. Or whether customer must leverage their own tools (e.g. DLP) to guard against such attacks. As appropriate, please refer to your answers in 2-14 above; if vendor has no native mitigations	
AI-Th-Output	16	Insecure Output Handling: This vulnerability occurs when an LLM output is accepted without scrutiny and forwarded to downstream applications without inspection. Consider for instance code or scripting that might be generated and then passed on to an application that would execute that code. This can expose backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution		As appropriate, please refer to your answers in 11-14 above	
AI-Th-ACL	17	Assurance that AI enabled application and/or the entity making the prompt request will not have access to sensitive information it is not authorized to access as defined by ACL's and/or information security data labelling.		This may be moot if leveraging a vendor LLM not trained with CoSD data. However it would be germane for applications which leverage a Peraton or CoSD source for RAG and or a Peraton or CoSD FFT FST model is used. As appropriate, please refer to your answers in 2-14 above.	
AI-Th-Training	18	Training Data Poisoning: This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior. Sources include Common Crawl, WebText, OpenWebText, & books.		What methodlogies has the vendor employed to mitigate this potential threat?	

AI-Th-DOS	19	Model Denial of Service: Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs	Related ot injection attacks--as this attack vector focuses on prompts that unduly burden the LLM--making it non or less responsive to others using the LLM
AI-Th-Supply	20	Supply Chain Vulnerabilities: LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre- trained models, and plugins can add vulnerabilities	Has the vendor thoroughly vetted their supply chain supporting the entirety of the LLM training and deployment activities?
AI-Th-Leak	21	Sensitive Information Disclosure: LLMs may inadvertently reveal confidential data in their responses, leading to unauthorized data access, privacy violations, and security breaches. Its crucial to implement data sanitization and strict user policies to mitigate this.	The DLP strategies discussed above in 10-14 can mitigate this risk. But has the vendor themselves tried to address this issue with unique security features in their AI solution?
AI-Th-Plugin	22	Insecure Plugin Design: LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution	Related to the supply chain issue discussed above to some extent---if one knows where their plug-ins are coming from and trust the vendor--that is a start. Of course, the plug-in still must be configured and managed in a manner that is consistent with cybersecurity best practices.
AI-Th-LP	23	Excessive Agency: LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.	This is very important--does the AI solution provide more than an answer? And then attempt to take unilateral actions based on it? What mitigations will be in place to protect against the BAD actions that might get taken on comprmised (e.g. hallucination, etc.) output? A NOD (NOD-AI-11) may be required if mitigations are considered incomplete.
AI-Th-HIL	24	Overreliance: Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLM	The trust issues discussed below(items 26-32) can help strike a balance with this. However; generally, the AI solution should be "providing an answer" and additional elements of the solution should be figuring out how to best use the information in a safe manner"
AI-Th-MT	25	Model Theft: This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information	Like a data security issue--one does not want what belongs to them to be stolen/exfiltrated, etc. If the vendor is hosting the LLM; they need to convince Peration/COSD that if we train a model on County data (which generally the County is not inclined to do) the vendor will be able to ensure us the custom model will not get stolen--of course Peraton/CoSD has a part in this as well.
AI-Trust: AI-Tr		**Model Trustworthiness Security Issues**	
AI-Tr-Safety	26	AI systems should "not under defined conditions, lead to a state in which human life, health, property, or the environment is endangered" (Source: ISO/IEC TS 5723:2022).	See WARNING in Item 33 below
AI-Tr-SecResilience	27	Security and Resilience: 1) AI systems may be said to be resilient if they can withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary (Adapted from: ISO/IEC TS 5723:2022). 2) Security includes resilience but also encompasses protocols to avoid, protect against, respond to, or recover from attacks.	See Items 15-25 above.
AI-Tr-Explain-Interp	28	Explainable and Interpreble 1) Explainability refers to a representation of the mechanisms underlying AI systems' operation. 2) Interpretability refers to the meaning of AI systems' output in the context of their designed functional purposes. Said another way user's of an AI system optimally need to understand why an AI solution has responded to a prompt in a manner in which had done so (e.g., like asking an "old" AI based expert system to explain why it made a decision) and actually be able to understand what the AI solution is telling its user.	Important for Item 24 above
AI-Tr-Privacy	29	Privacy-Enhanced. Privacy refers generally to the norms and practices that help to safeguard human autonomy, identity, and dignity. These norms and practices typically address freedom from intrusion, limiting observation, or individuals' agency to consent to disclosure or control of facets of their identities (e.g., body, data, reputation). The implication here is that trainers of an AI solution need to control PII/PHI, or other sensitive information which might be contained in training data sets. While users of an AI solution need to control the PII/PHI, or other sensitive information which might come out as a result of prompt input	See Items 2-14 above
AI-Tr-Fair	30	Fair With Harmful Bias Managed. Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination. Standards of fairness can be complex and difficult to define because perceptions of fairness differ among cultures and may shift depending on application. Organizations' risk management efforts will be enhanced by recognizing and considering these differences. AI systems becomes biases because their training data was such.	Primarily a vendor LLM issue.

AI-Tr-Accountable	31	Accountable and Transparent.The Meriam Webster Dictionay defines accountable as: 1) capable of being explained : EXPLAINABLE; 2) subject to giving an account : ANSWERABLE. Transparency reflects the extent to which information about an AI system and its outputs is available to individuals interacting with such a system	Important for 24 above.
AI-Tr-Valid-Reliable	32	Valid and Reliable. Validation is the "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" (Source: ISO 9000:2015). Reliability is defined as the "ability of an item to perform as required, without failure, for a given time interval, under given conditions" (Source: ISO/IEC TS 5723:2022)	Important for 24 above.
AIRisk Management Framework: AI-RMF		**Will Ongoing RMF Actions be Required?**	
AI-RMF	33	Based on answers to questions in 1-31 above, are active measurement and management actions (from NIST AI Govern/Map/Measure/Manage RMF) required at some interval to verify security posture of the AI product in the solution? If yes, please describe product/solution elements to be measured and managed and at what intervals. **WARNING: ANY SOLUTION LEVERAGING GENERATIVE AI TECHNOLGOGIES MUST EMPLOY AN ONGOING RMF GOVERNANCE PROCESS IF THE SOLUTION USING THE GENERATIVE AI TECHNOLOGY IMPACTS HUMAN SAFETY.**	