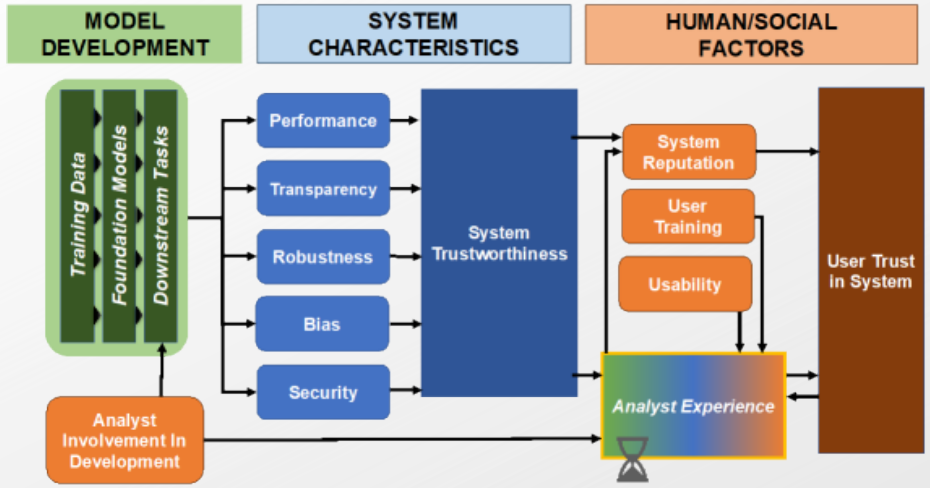
SAND2024-03918C

Trust Maturity Model for AI Systems

Prior work: Trust Framework



Goal: Operationalize applying this framework as a maturity model

Capability Maturity Models (CMM)
- Describe maturity of the software development process

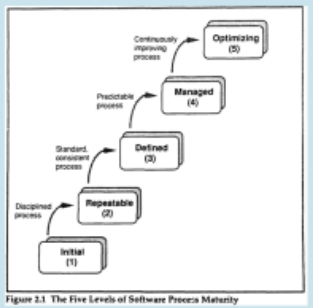


Figure 2.1 The Five Levels of Software Process Maturity

Technology Readiness Levels (TRL)
- Describe maturity of technology development

Predictive Capability Maturity Model (PCMM)
- Describes maturity of predictive simulation

Table 1: Table Format for PCMM Assessment

ELEMENT \ MATURITY	Maturity Level 0	Maturity Level 1	Maturity Level 2	Maturity Level 3
Representation and Geometric Fidelity				
Physics and Material Model Fidelity				
Code Verification				
Solution Verification				
Model Validation				
Uncertainty Quantification and Sensitivity Analysis				

Trust Maturity Model (summary)

	Level 1	Level 2	Level 3	Level 4
Performance Quantification	Unknown	Low confidence	High confidence accuracy	High confidence accuracy and uncertainty
Robustness	Not considered	Measured on single dimension	Measured on multiple dimensions	Continuous testing
Bias	No consideration	Some consideration	Well-defined	Proactive with minimized error
Transparency	Black box	Coarse mental model	Useful mental model	Accurate mental model
Security	Unknown	Awareness of vulnerability	Quantified vulnerability	Confidential with high confidence in integrity
Safety	Unknown	Basic guardrails	Broad misuse countermeasures	Potential misuse accounted for & tracked
Usability	None	Basic	Intuitive and well-targeted	Intuitive and adaptive to user/task

	Level 1	Level 2	Level 3	Level 4
Performance Quantification		✓		
Robustness		✓		
Bias		✓		
Transparency		✓		
Security	✓			
Safety		✓		
Usability			✓	

Example:
ChatGPT 3.5 (no RAG) rated on Science QA for high consequence settings

Poster #
Scott Steinmetz
Sandia National