

AI VALIDATION READINESS CHECKLIST

10 Questions to Help Security Leaders Define AI Risk and Readiness

AI readiness often spans security, engineering, operations, governance, risk, and external technology providers. These questions can help create a common view of risk when evaluating internally developed AI as well as models and agents provided by external vendors.



1. What data can the AI access?



Does the AI have access only to the information required for its function, or can it reach additional systems, repositories, applications, and data sources?



2. What tools can the AI interact with?



Can it view information only, or can it modify configurations, execute workflows, trigger actions, or interact with security controls?



3. What permissions has the AI been granted?



Have least-privilege principles been applied, or does the AI have broader access than necessary?



4. How does the AI behave when it receives incomplete, incorrect, or conflicting information?



Has the organization evaluated how the system responds when the data it receives is imperfect, ambiguous, or misleading?



5. What are the AI's known failure modes?



Can the organization identify situations where the AI struggles, performs inconsistently, or produces undesirable outcomes?



6. How will AI activity be monitored?



Can security teams clearly see what actions the AI is taking, what decisions it is making, and what resources it is accessing?



7. What requires human approval?



Has the organization clearly defined where human-in-the-loop oversight remains mandatory?



8. How will performance be measured?



Can the organization determine whether the AI is achieving desired outcomes, operating consistently, and performing as expected over time?



9. Where are the boundaries of acceptable risk?



Has the organization defined clear limits around autonomy, authority, access, and operational decision-making?



10. How will validation continue over time?



What process exists to evaluate performance as threats, environments, workflows, and AI systems evolve?



If these questions cannot be answered, the organization may be trusting AI before it has been properly validated.