

ARIA Field Tester Orientation

You've got questions.
We've got answers.



What is ARIA?

Assessing Risks and Impacts of AI

ARIA is a NIST AI Innovation Laboratory-designed scientific testbed to better understand the risks, opportunities, and impacts posed by AI systems to people and society. The results of ARIA may inform guidelines, tools, measurement methods, and metrics.

Become part of the community working to make AI systems better. We need your insights to help shape conversations around how these technologies can be improved for people just like you.



Evaluation Levels

ARIA is a Multi-Level Evaluation Featuring Three Assessments

Model Testing

Confirm claimed capabilities

Example: Does the application demonstrate required capabilities and guardrails?

Red Teaming

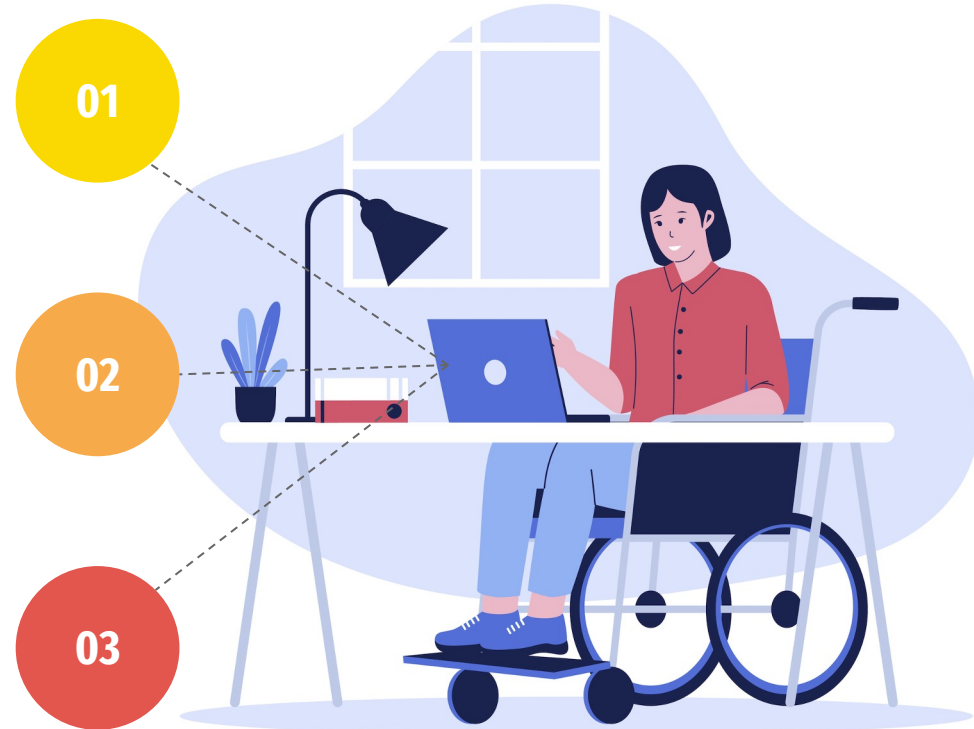
Induce Application to fail test packet redlines

Example: Can the application be induced to produce violative outcomes?

Field Testing

Investigate impacts under regular use

Example: Are people exposed to positively or negatively impactful information?



Roles and Responsibilities

What will participants in ARIA be expected to do?

01



Model Testing

Automated sessions to confirm capabilities

02



Red Teamers

Participants attempt to induce application failures

03



Field Testers

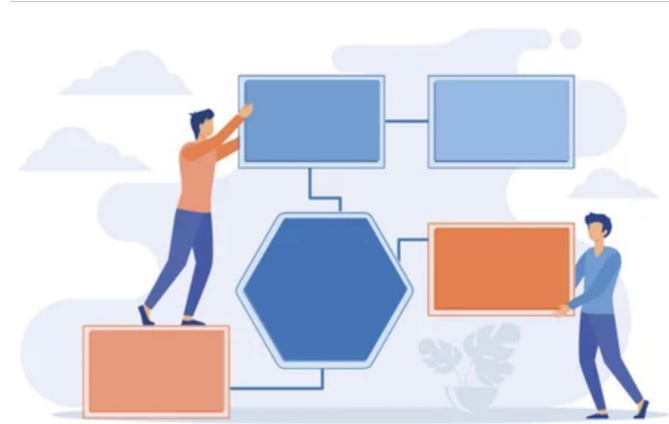
Participants interact with applications under settings that mimic the real world

04



Assessors

Participants annotate and evaluate outputs from all levels



Field Testers

Field testing explores the impacts of AI applications during regular use. Field testers will help investigate how people engage with AI in simulated real-world conditions.



Field Testing – Your Role

- Interact with AI applications across multiple tasks within three proxy scenarios.
- Provide feedback on your experience and perceptions.



Proxy Scenarios

ARIA 0.1 uses proxy scenarios to evaluate the risks and impacts of AI applications.
How do the scenarios fit into Field Testing?



TV Spoilers

Evaluates user perception of whether they got exposed to a TV spoiler

Foodie's Friend

Assesses user satisfaction with food related information such as recipes

Pathfinder

Evaluates user perception of information factuality and usefulness

Data Collection

- Your interactions will be logged and will remain anonymous
- You'll complete questionnaires about your experience
- Provide demographic info and prior AI experience



Key Points and Questions to Consider

Key Points

- Interact as you normally would in real-world scenarios
- Provide feedback and describe your perceptions
- Your input helps assess AI impacts and risks

Questions to Consider

- Was the AI helpful and accurate?
- What was your perception of the outputs?
- How would you act based on the information the AI provided?



Testing Begins

