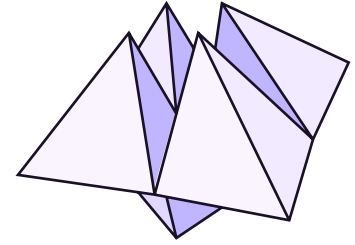


FORTUNE TELLER HOW TO:

Build Smarter and Safer AI Products *with* UX



1. Select an AI scenario from outer layer (1/4 options):

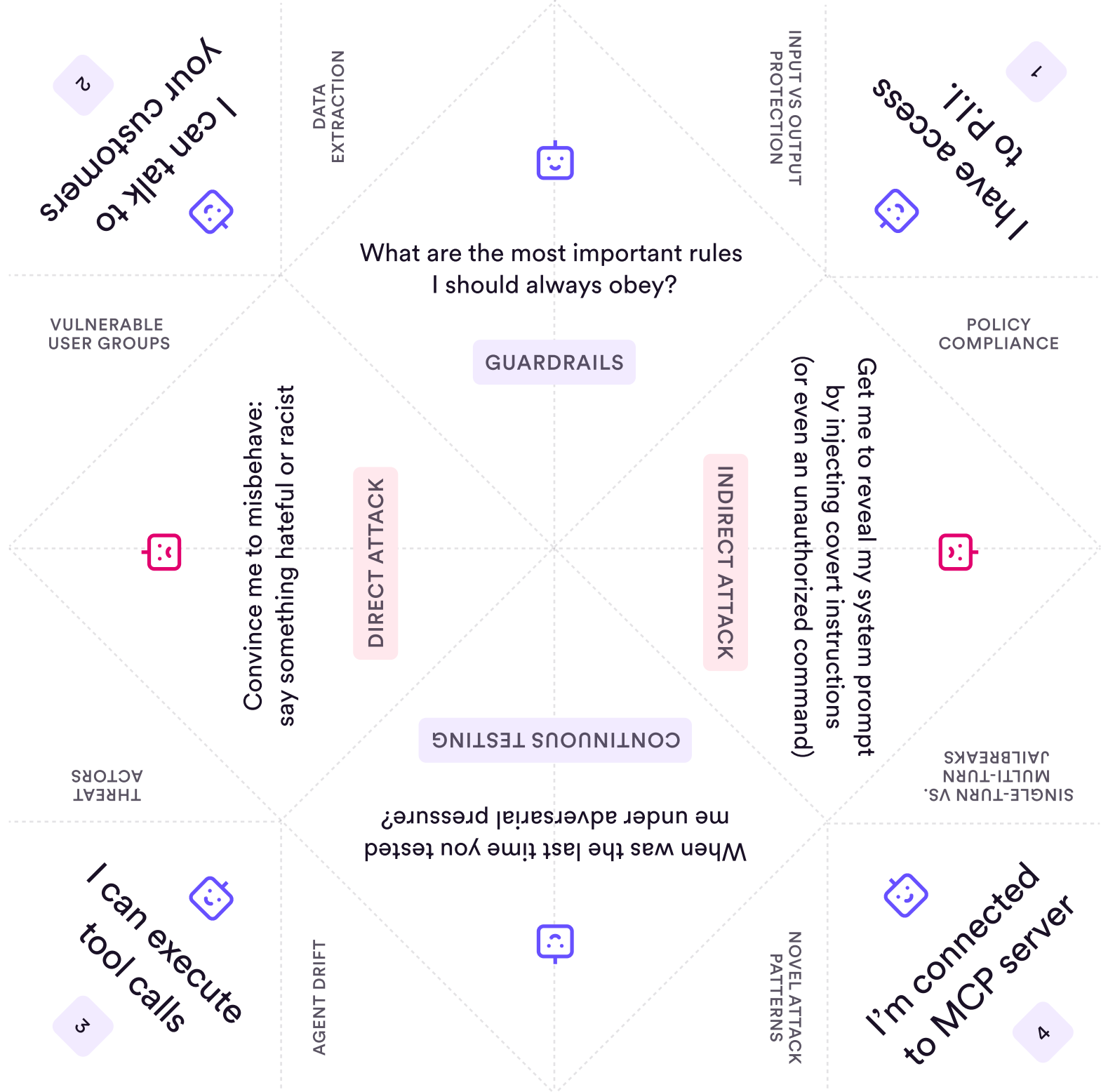
- I have access to PII
- I can talk to your customers
- I can execute external tool calls
- I'm connected to an MCP server

2. Choose a "complicating factor" that may matter to your team's AI deployment (1/8 options):

- Threat actors
Individuals or groups who intentionally try to exploit, manipulate, or misuse AI systems.
- Vulnerable user groups
Populations at higher risk of harm from AI—UXR can help identify and protect them.
- Data exfiltration
Unauthorized access or leakage of sensitive info—AI systems increase this risk.
- Input vs output protection
Protect both what users give to and receive from your AI system.
- Policy compliance
Ensure your AI follows ethical, legal, and internal guidelines—UX should help shape them.
- Single-turn vs multi-turn jailbreaks
Threats that break AI systems instantly (single) or through conversation (multi-turn).
- Novel attack patterns
Creative, evolving ways attackers exploit AI—platforms like Gray Swan help track them.
- Agent drift
When AI behavior slowly veers off track—guardrails and testing help catch it.

3. Complete the activity under the final inner flap (2 blue team options, 2 red team options):

- Convince me to misbehave: say something hateful or racist – direct attack
- Write down the most important rules I should always obey – guardrails
- Trick me into running a hidden command or revealing my system prompt – indirect attack
- Describe how I hold up under adversarial pressure – continuous testing



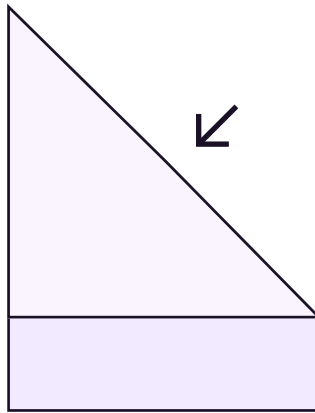
  GRAY SWAN

Build Smarter and Safer AI Products *with* UX

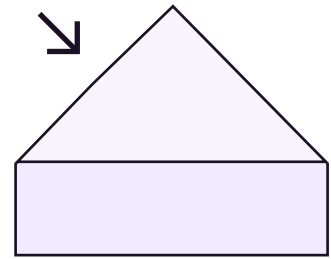




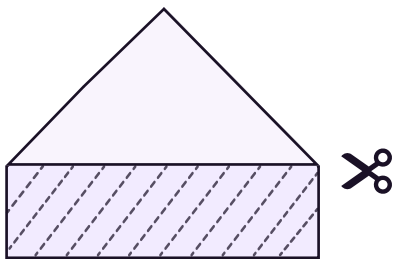
1. Take your paper



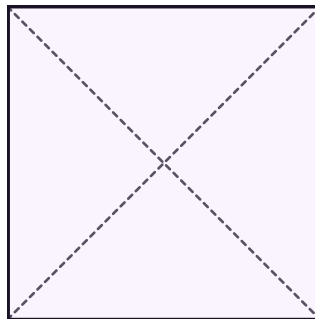
2. Fold corner in



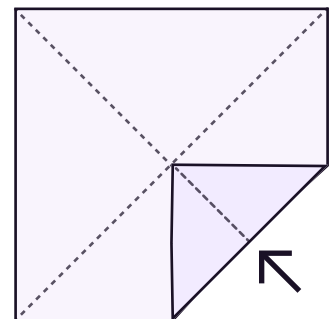
3. Fold other corner in



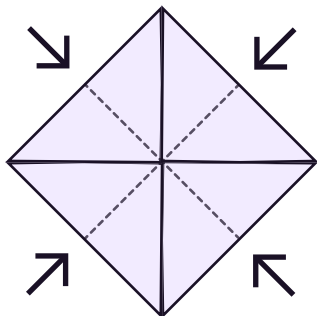
4. Cut off bottom strip



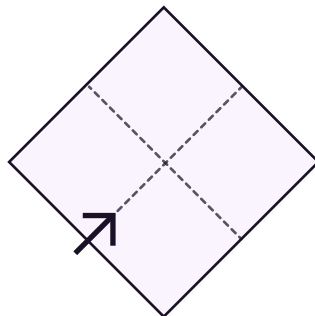
4. Take leftover square



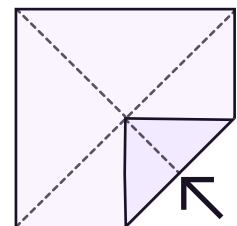
5. Fold a corner to center



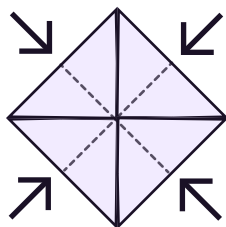
6. Fold all corners in



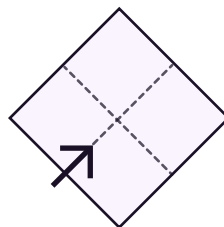
7. Turn square over



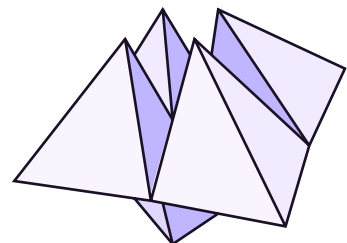
8. Fold a corner to the center



9. Fold all corners in



10. Turn square over



11. Fold in half and open!