# GAI Profile Adoption

## for End Users and ‘Other AI Actors’

September 24, 2024

Presented by Jurnell Cockhren

# This Talk’s Primary Goal

Proposes a contribution format where **Other AI Actors** can facilitate the adoption of relevant GAI profiles for **End Users**.

CIVIC HACKER

# Possible Contribution Format

**GOVERN 6.2:** Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.

| Action ID | Suggested Action | GAI Risks |
|---|---|---|
| GV-6.2-001 | Document GAI risks associated with system value chain to identify over-reliance on third-party data and to identify fallbacks. | Value Chain and Component Integration |
| GV-6.2-002 | Document incidents involving third-party GAI data and systems, including open-data and open-source software. | Intellectual Property; Value Chain and Component Integration |

**MANAGE 4.2:** Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI Actors.

| Action ID | Suggested Action | GAI Risks |
|---|---|---|
| MG-4.2-001 | Conduct regular monitoring of GAI systems and publish reports detailing the performance, feedback received, and improvements made. | Harmful Bias and Homogenization |
| MG-4.2-002 | Practice and follow incident response plans for addressing the generation of inappropriate or harmful content and adapt processes based on findings to prevent future occurrences. Conduct post-mortem analyses of incidents with relevant AI Actors, to understand the root causes and implement preventive measures. | Human-AI Configuration; Dangerous, Violent, or Hateful Content |
| MG-4.2-003 | Use visualizations or other methods to represent GAI model behavior to ease non-technical stakeholders understanding of GAI system functionality. | Human-AI Configuration |

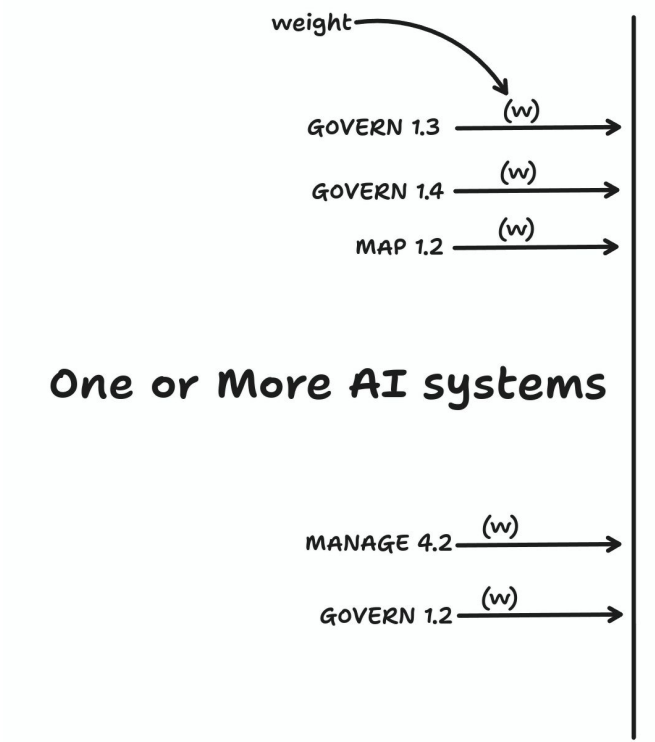
**AI Actor Tasks: AI Deployment, AI Design, AI Development, Affected Individuals and Communities, End-Users, Operation and Monitoring, TEVV**

Selection Methodology

(3 paragraphs max.)

CIVIC HACKER

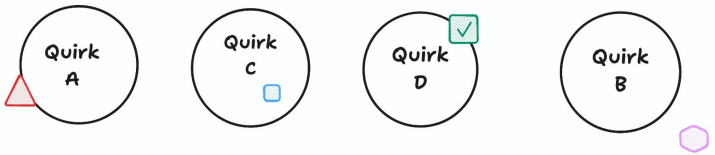
# Entities with the bandwidth




- Can develop comprehensive GAI profiles using VRTs
- Will use spreadsheets, may be motivated by scoring
- Can repurpose existing setups used in similar compliance efforts (CSF, SOC)
- Can purpose software to help gather evidence

CIVIC HACKER

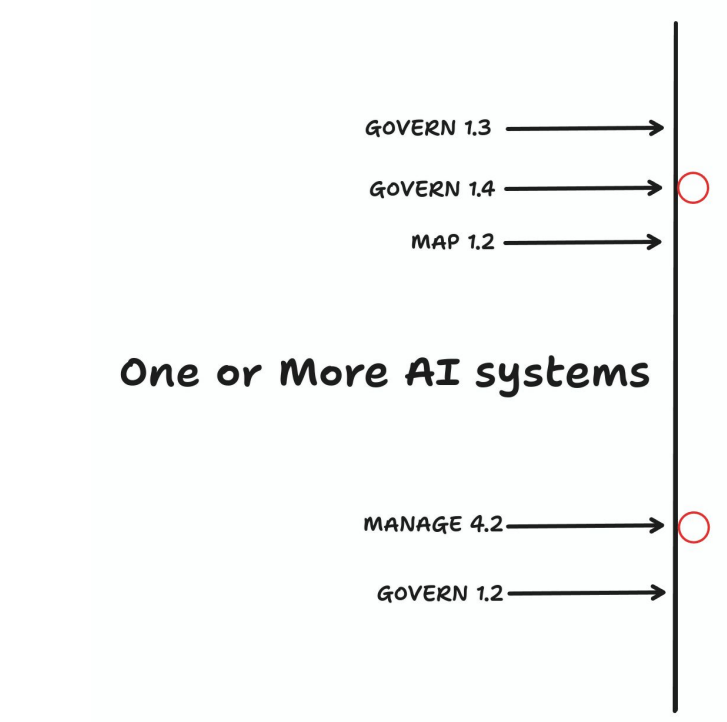
# Entities that may need help




- Don’t have resources to crawl through 600-1 document
- Motivated by their differing uses/intents/risks/needs within their domain
- small changes can yield massive impact
- Can readily determine if first & third-party actors GAI profile fulfills their concern(s)

CIVIC HACKER

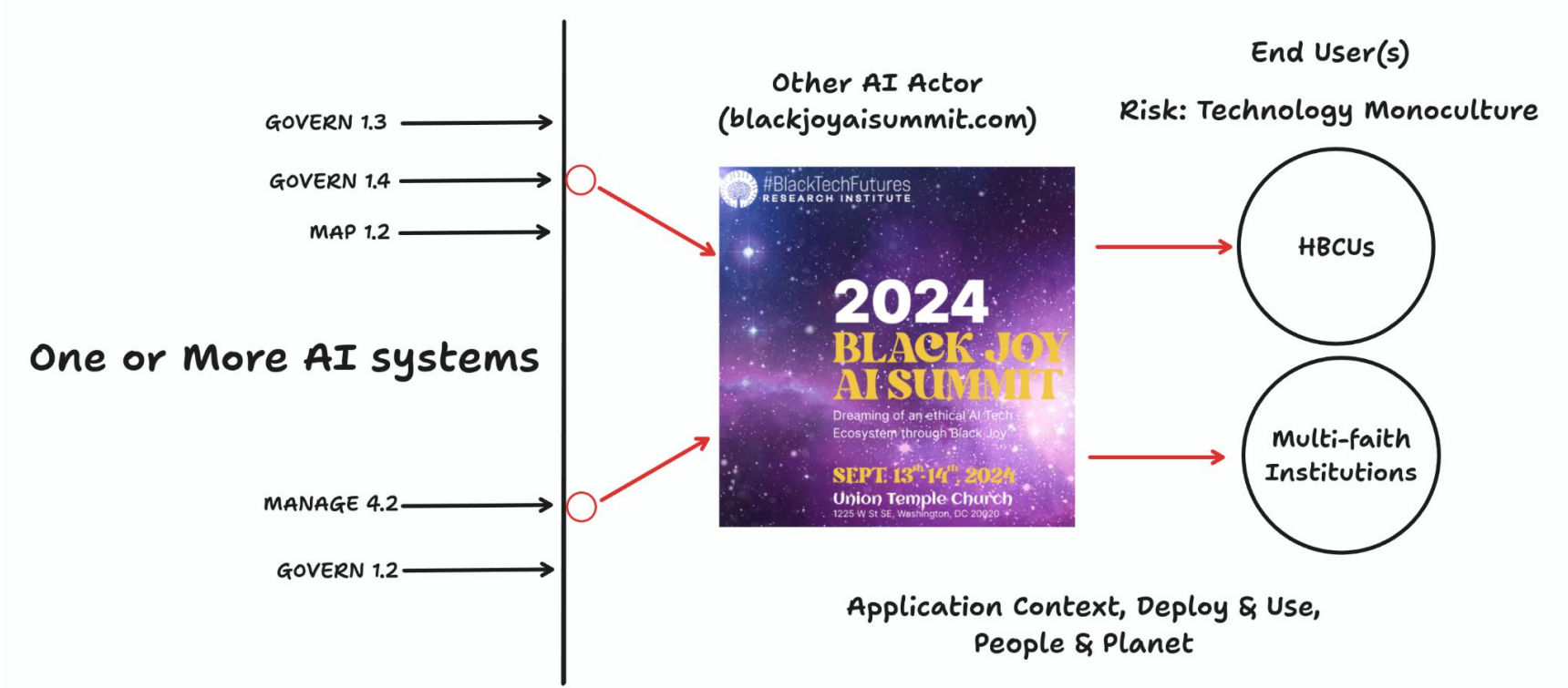
# ‘Other AI Actors’ can help carry the load




- Can readily determine **feasible suggested actions** for End Users
- subcategories can serve as **docking mechanism** for End Users
- Can readily ‘**make it make sense**’ to their community

CIVIC HACKER

# Ex: Communities of Practice




CIVIC HACKER

# Possible Contribution Format

**GOVERN 6.2:** Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.

| Action ID | Suggested Action | GAI Risks |
|---|---|---|
| GV-6.2-001 | Document GAI risks associated with system value chain to identify over-reliance on third-party data and to identify fallbacks. | Value Chain and Component Integration |
| GV-6.2-002 | Document incidents involving third-party GAI data and systems, including open-data and open-source software. | Intellectual Property; Value Chain and Component Integration |

**MANAGE 4.2:** Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI Actors.

| Action ID | Suggested Action | GAI Risks |
|---|---|---|
| MG-4.2-001 | Conduct regular monitoring of GAI systems and publish reports detailing the performance, feedback received, and improvements made. | Harmful Bias and Homogenization |
| MG-4.2-002 | Practice and follow incident response plans for addressing the generation of inappropriate or harmful content and adapt processes based on findings to prevent future occurrences. Conduct post-mortem analyses of incidents with relevant AI Actors, to understand the root causes and implement preventive measures. | Human-AI Configuration; Dangerous, Violent, or Hateful Content |
| MG-4.2-003 | Use visualizations or other methods to represent GAI model behavior to ease non-technical stakeholders understanding of GAI system functionality. | Human-AI Configuration |

**AI Actor Tasks: AI Deployment, AI Design, AI Development, Affected Individuals and Communities, End-Users, Operation and Monitoring, TEVV**

Selection Methodology

(3 paragraphs max.)

CIVIC HACKER

# Key Takeaways

- ‘Other AI actors’ are inherently incentivized to participate.
- Contributions as 2 to 3 suggested actions (subcategories)
- Include the Selection methodology (w/ intended End User)

CIVIC HACKER