The “Who”, “What”, and “How” of Responsible AI Governance:

A Systematic Review and Meta-Analysis of (Actor, Stage)-Specific Tools

Blaine Kuehnert*
blainekuehnert@cmu.edu

Rachel Kim*
rachelmkim@cmu.edu

Jodi Forlizzi

Hoda Heidari

Carnegie Mellon University

Motivation

LinkedIn's search algorithm apparently favored men until this week

FaceApp forced to pull 'racist' filters that allow 'digital blackface'

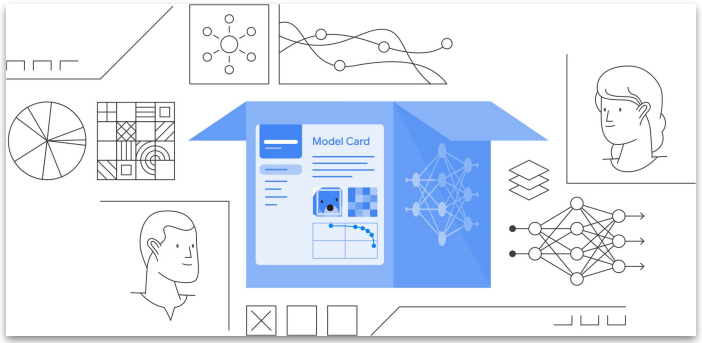
What Went Wrong With Tay, The Twitter Bot That Turned Racist?

OpenAI confirms threat actors use ChatGPT to write malware

Tools for Responsible Governance of AI

Fairlearn

Datasheets for Datasets



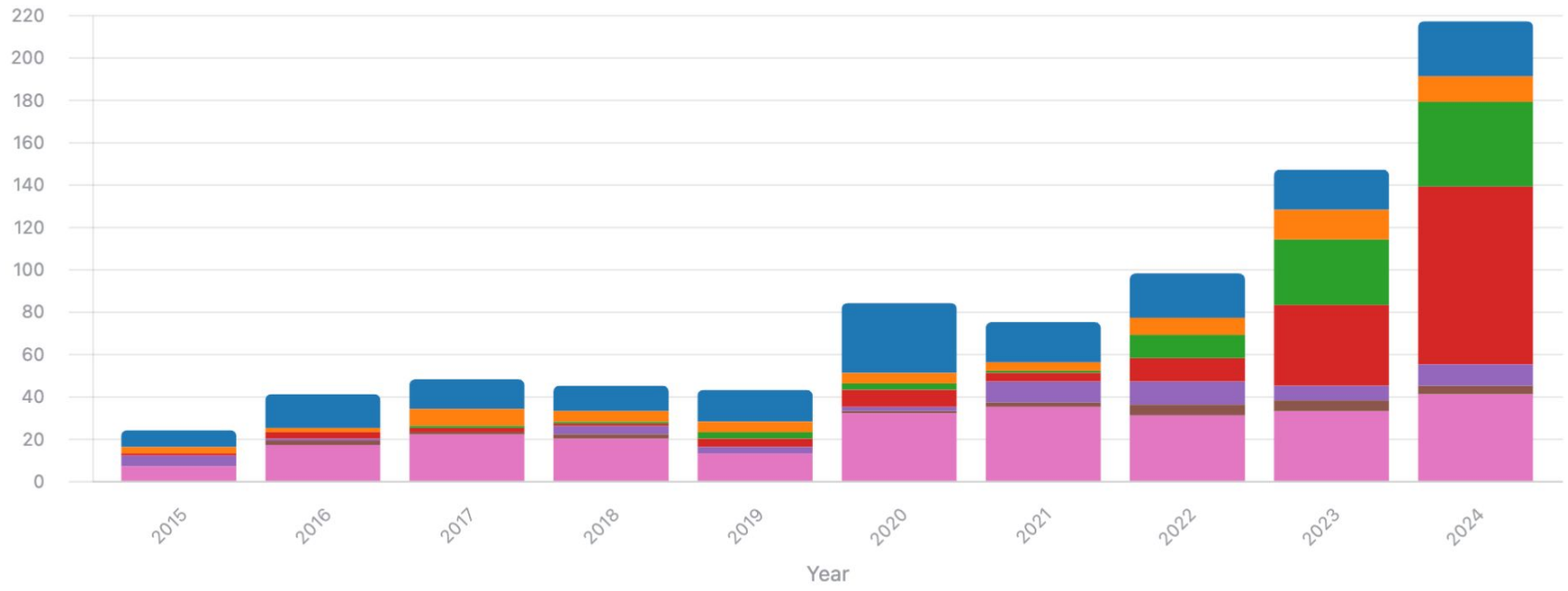
Model Cards for Model Reporting

AI Fairness 360

Incident Count
Total number of incidents analysed
869

Report Count
Total number of reports processed (most incidents have multiple reports)
4,006

Risk Domain (Incident count)

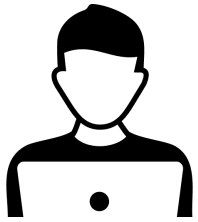


Why is responsible governance of AI hard to operationalize?

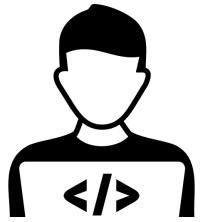
Why is responsible governance of AI hard to operationalize?

1) A **diverse** set of **AI stakeholders** often has a **diverse** set of **goals**

Leaders

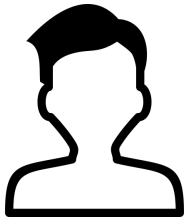


Designers



Developers

Deployers

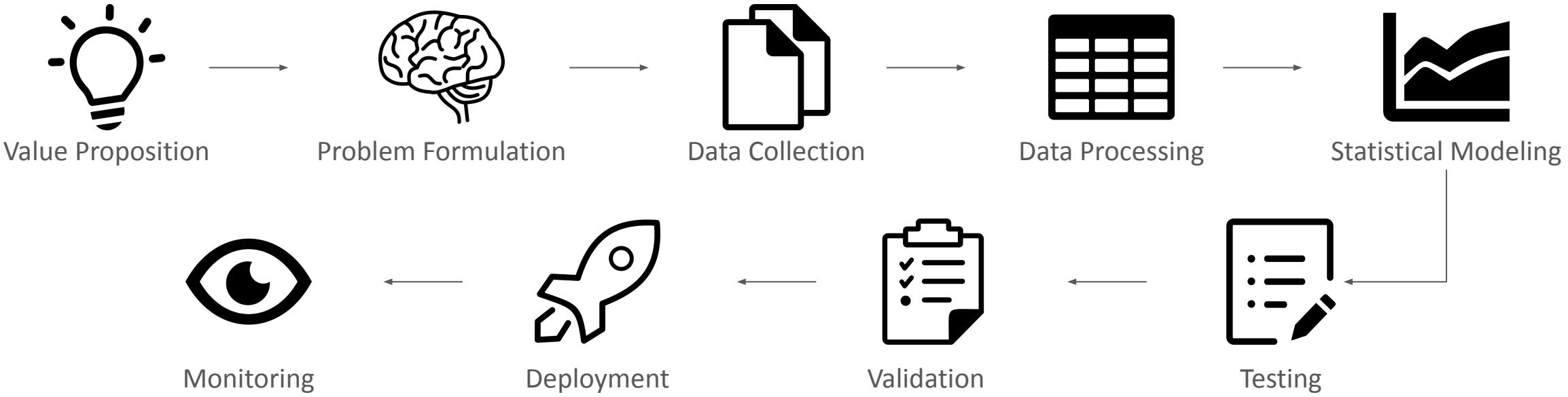


End-users

Impacted Communities

Why is responsible governance of AI hard to operationalize?

1) A diverse set of stakeholders often has a diverse set of goals
   - Leaders, Designers, Developers, Deployers, End-users, Impacted Communities
2) A **complex AI lifecycle** comes with **numerous decisions at each stage**



Why is responsible governance of AI hard to operationalize?

1) A diverse set of stakeholders often has a diverse set of goals
   - Leaders, Designers, Developers, Deployers, End-users, Impacted Communities
2) A complex AI lifecycle comes with challenges at each stage
   - Value Proposition, Problem Formulation, Data Collection, Data Processing, Statistical Modeling, Testing, Validation, Deployment, Monitoring
3) **Ambiguous** allocation of **responsibilities** and **best practices** for compliance.

Why is responsible governance of AI hard to operationalize?

1) **Who** are the **stakeholders** of the AI system?
   - Leaders, Designers, Developers, Deployers, End-users, Impacted Communities
2) **What** are the **responsibilities** of each role at various **stages of the AI lifecycle**?
   - Value Proposition, …, Statistical Modeling, Testing, Validation, Deployment, Monitoring
3) **How** should AI stakeholders discharge their responsibilities to comply with RAI?
   - Our Work

Toward Operationalizing Responsible Governance of AI

Roles	Leaders
	Designers
	Developers
	Deployers
	End-users
	Impacted Communities

Toward Operationalizing Responsible Governance of AI

	AI Lifecycle Stages								
	Value Proposition	Problem Formulation	Data Collection	Data Processing	Statistical Modeling	Testing	Validation	Deployment	Monitoring

Toward Operationalizing Responsible Governance of AI

		AI Lifecycle Stages								
		Value Proposition	Problem Formulation	Data Collection	Data Processing	Statistical Modeling	Testing	Validation	Deployment	Monitoring
Roles	Leaders									
	Designers									
	Developers									
	Deployers									
	End-users									
	Impacted Communities									

Our Research Questions

1) What **tools** are available for ***each stage of the lifecycle*** and for ***each role***?

> **Tool:** a **specific** and **practical** instrument, technique, or process offering **concrete** steps and methods to complete a specific task.

2) Which of these tools are ***validated*** in any way?

> **Validation:** **empirical** (or even suggestive) **evidence of usability** and **efficacy in practice** (e.g., via a case study, experiment, or a pilot program)

Methodology

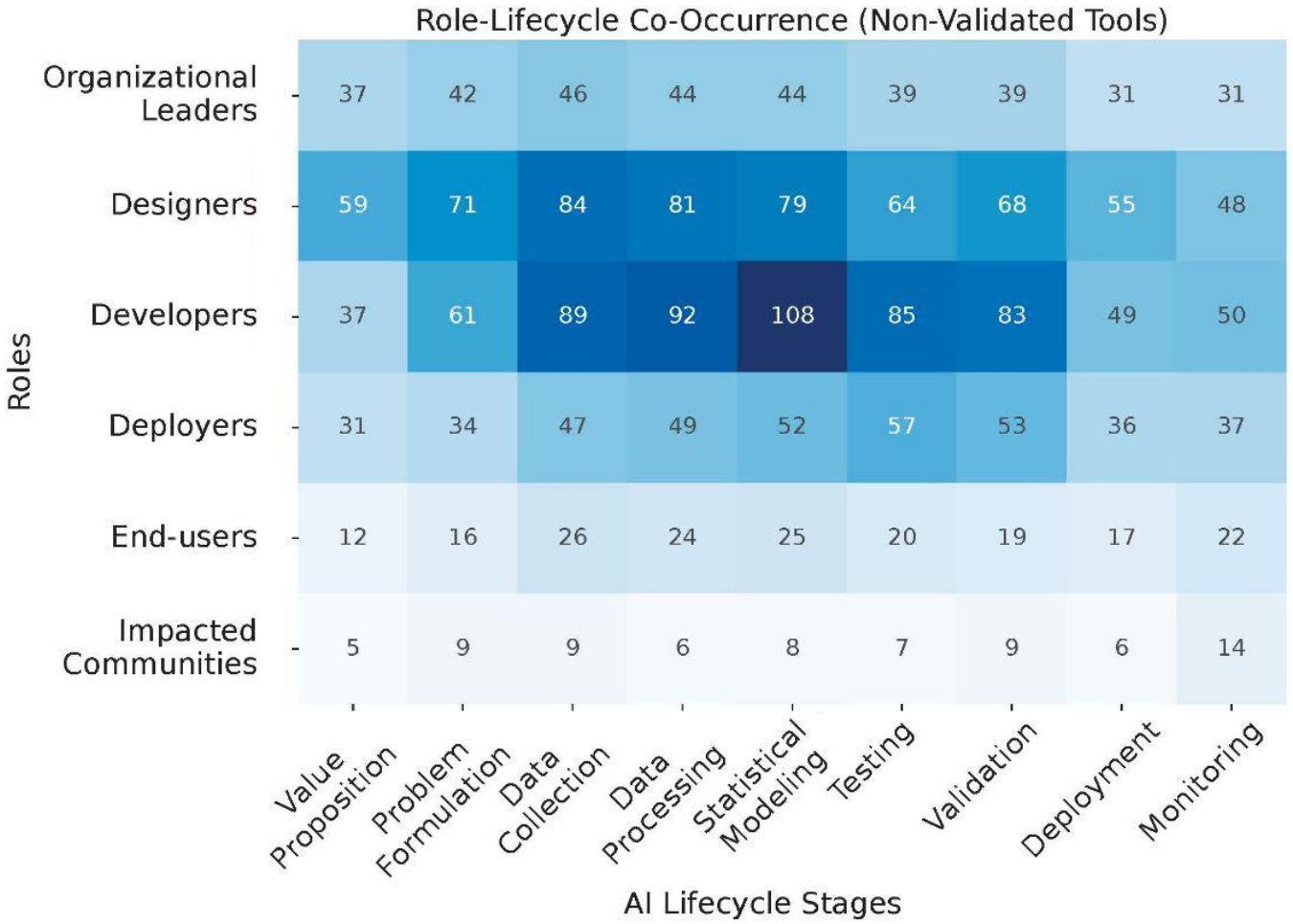
Methodology – Systematic Literature Review

1) Collected over 1300 papers
2) Final dataset of **over 220 papers and tools**
   a) Academia
   b) Industry

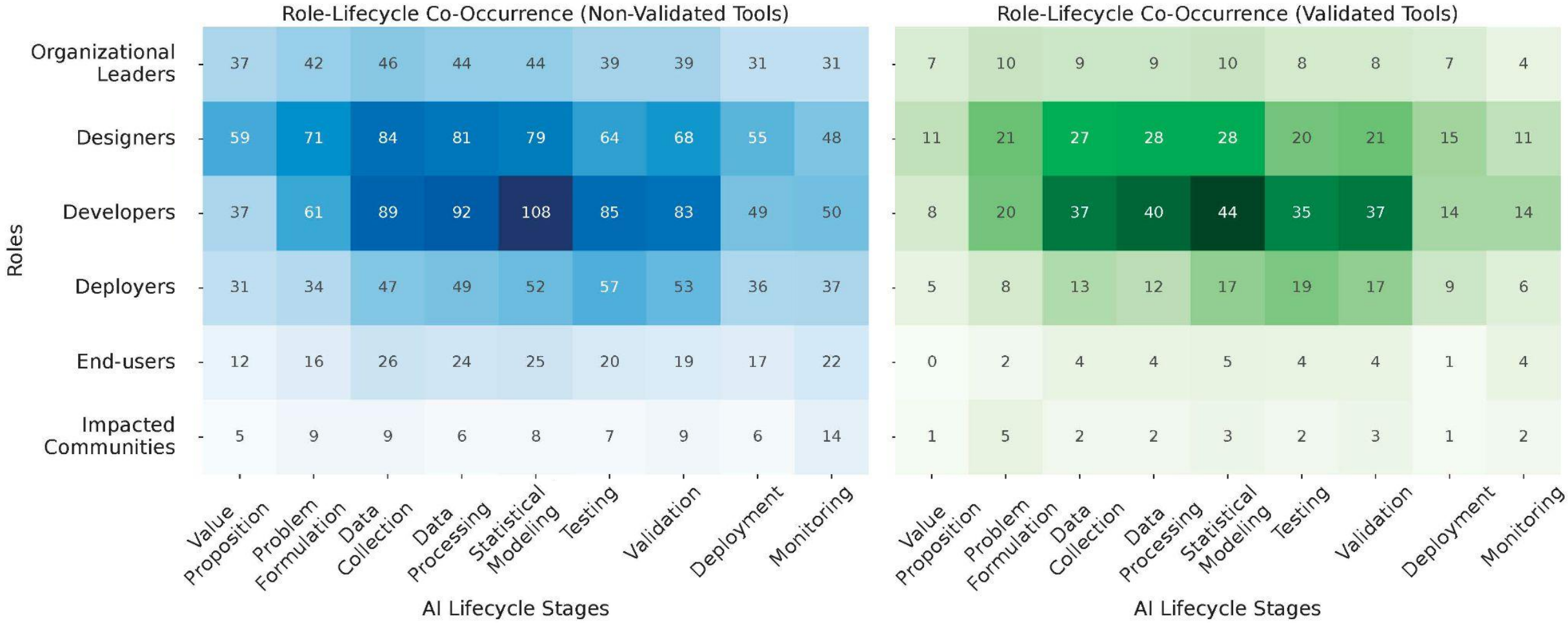
Assessment – Categorized based on roles and lifecycle covered, and whether validation was included

1) Roles
2) Stages
3) Validation

Findings: Outsized Focus on Technical Roles and Stages



Findings: Lack of Validation Efforts for Existing Tools



Carnegie Mellon University

ACM FACCT ATHNA 2025

Implications

1) Lack of tools designed for specific (role, task)-pairs → **infrequent/improper use**
2) Lack of validation for Responsible AI (RAI) tools → **ineffective/problematic use**
3) Lack of tools addressing all the stakeholders and stages of the AI lifecycle → **fragmented approach to AI governance**

Recommendations

1) Lack of tools designed for specific (role, task)-pairs → **infrequent/improper use**

Develop RAI tools in close partnership with **AI stakeholders** and to address their specific **tasks and responsibilities**.

- Focus on under-studied cells of our matrix
- Need-finding & co-design as the appropriate methodology

Recommendations

2) Lack of validation for Responsible AI (RAI) tools → **ineffective/problematic use**

Validate existing and new RAI tools.

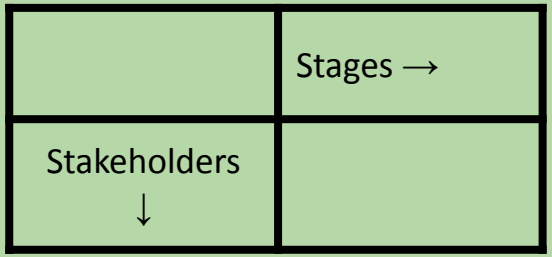
- Document the use and efficacy of RAI tools in practice (observational studies)
- Conduct controlled experiments to assess the efficacy of tools (experimental studies)
- For new tools, include validation as part of the tool design process itself

Recommendations

3) Lack of tools addressing all the stakeholders and stages of the AI lifecycle → **fragmented approach to AI governance**

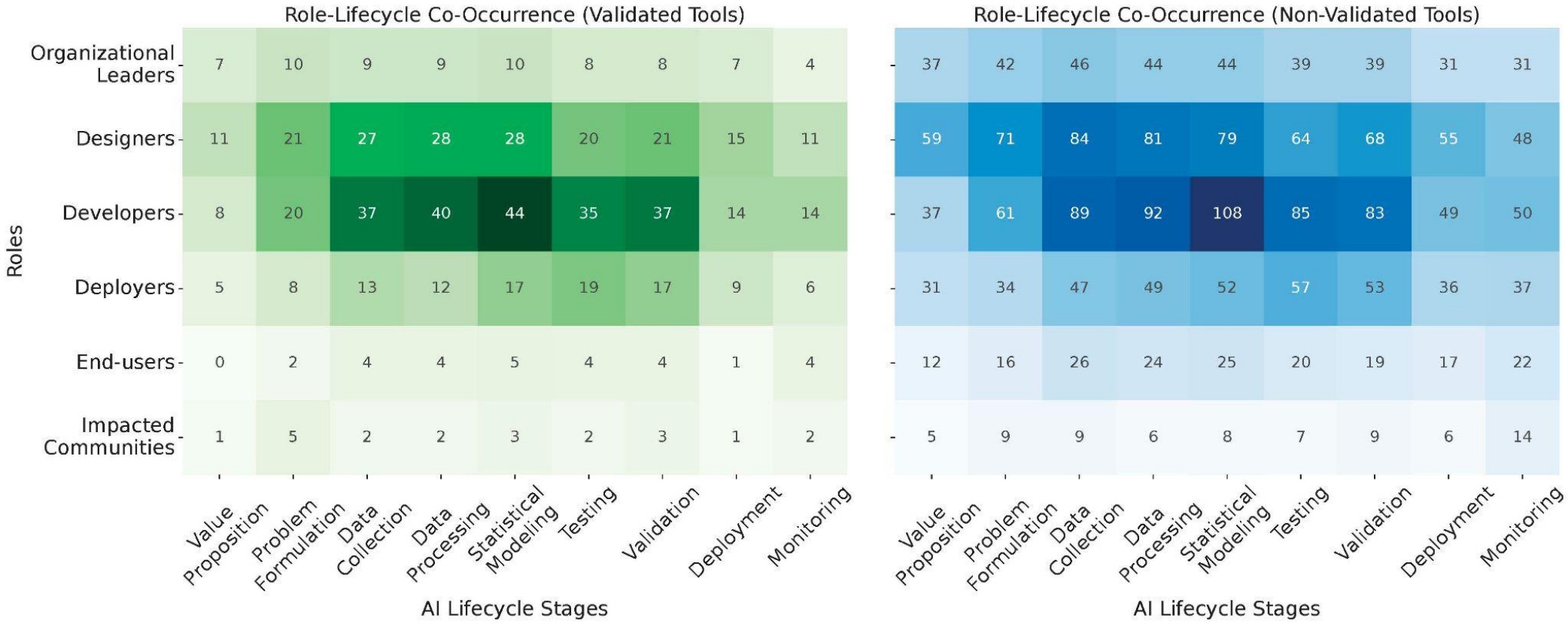
Use (stakeholder, stage)-matrix as a **blueprint for AI governance** in organizations.

- Clearly delineate who is responsible for which decisions
- Offer concrete tools to AI stakeholders
- Coordinate efforts across stages and stakeholders



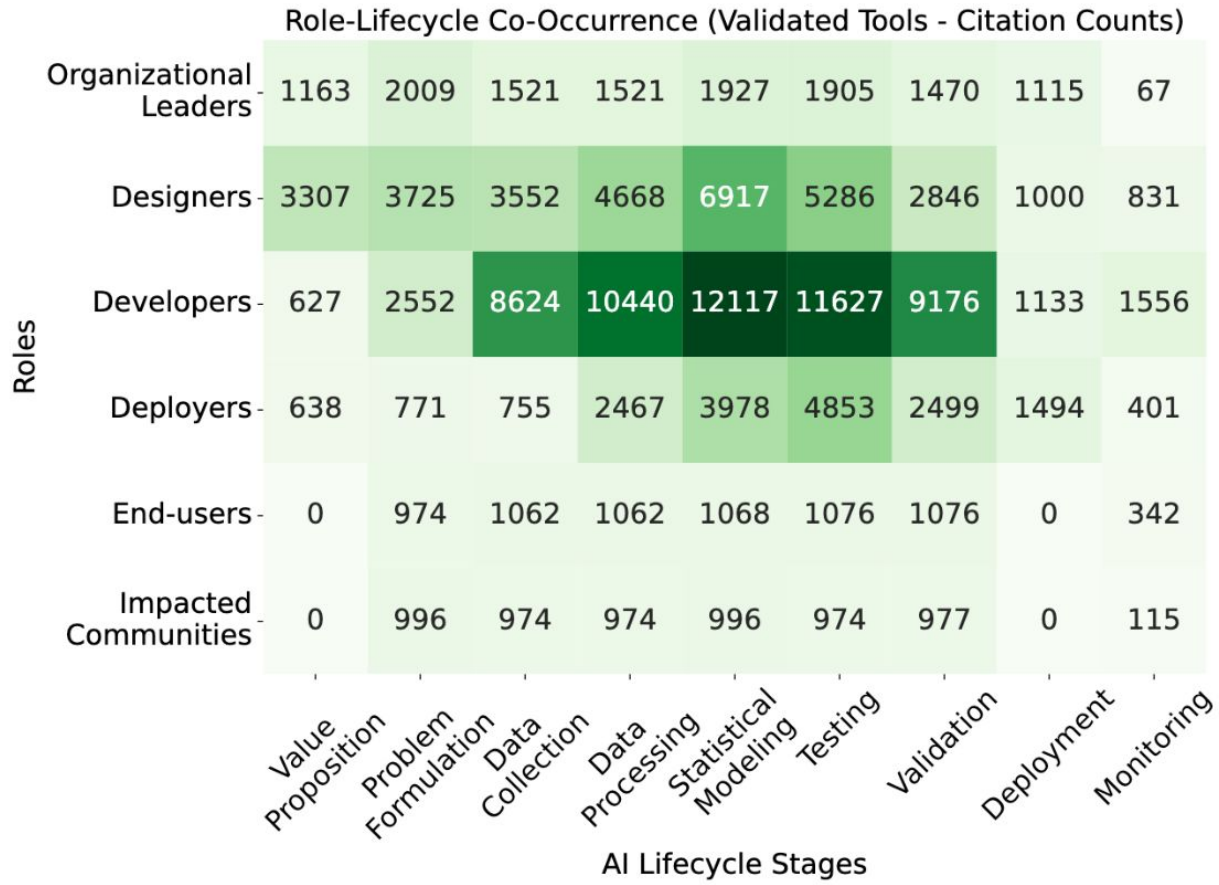
Thank you! Any questions?

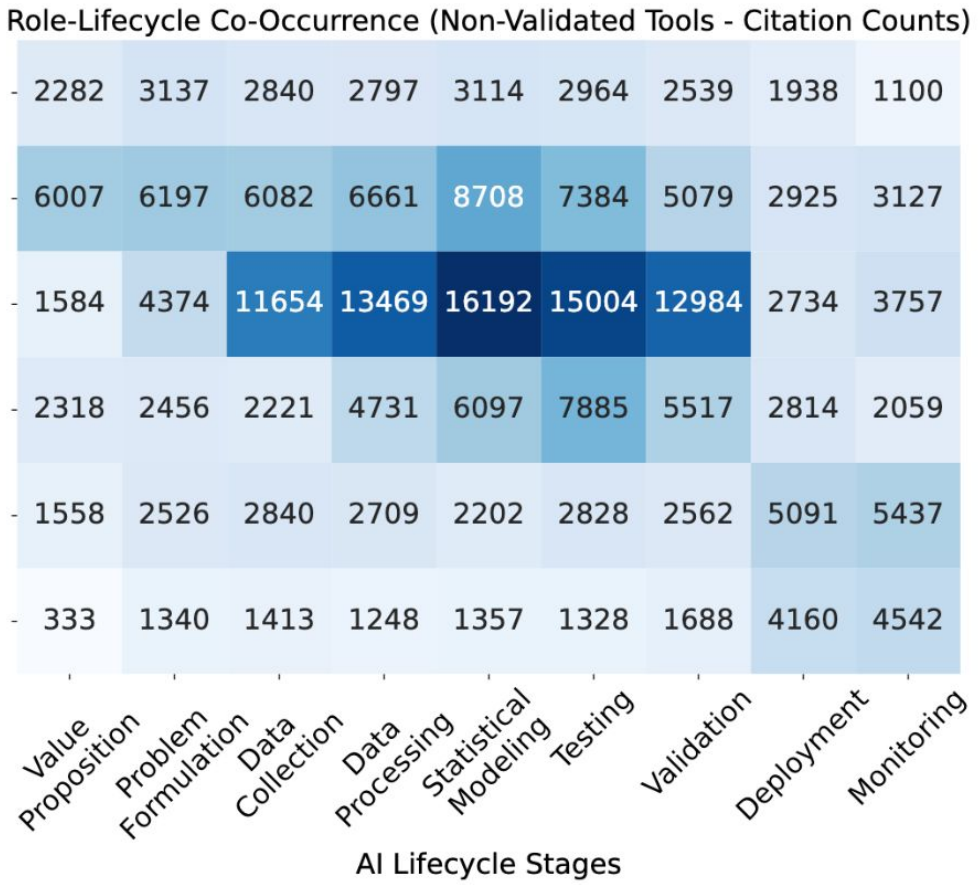
Email: blainekuehnert@cmu.edu, rachelmkim@cmu.edu



Carnegie Mellon University

(Stakeholder, Stage)-Matrix by Citation Count





	Value Proposition	Problem Formulation	Data Collection	Data Processing	Statistical Modeling	Testing	Validation	Deployment	Monitoring
Leaders	[10, 78, 87, 102, 106, 112, 119, 129]	[10, 39, 62, 78, 84, 85, 87, 102, 106, 119, 125]	[39, 56, 84, 85, 87, 102, 106, 119, 125]	[39, 56, 84, 85, 87, 102, 106, 119, 125]	[39, 56, 62, 84, 85, 87, 102, 106, 113, 119, 125]	[39, 56, 85, 87, 102, 106, 113, 119, 125]	[39, 56, 84, 85, 87, 102, 106, 119, 125]	[39, 56, 78, 87, 102, 113, 119, 125]	[102, 112, 114, 119, 125]
Designers	[6, 10, 38, 47, 80, 87, 91, 103, 106, 119, 129]	[2, 6, 8, 10, 35, 38, 39, 47, 48, 62, 80, 84, 87, 90, 91, 103, 106, 119, 121, 125]	[2, 4, 16, 24, 27, 38, 39, 58, 59, 70, 77, 80, 84, 87, 90, 91, 103, 104, 106, 109, 116, 119, 121, 124–126]	[2, 4, 13, 15, 24, 27, 35, 38, 39, 58, 59, 70, 77, 80, 84, 87, 90, 91, 104, 106, 109, 116, 119, 121, 124–126]	[2, 4, 13, 15, 16, 24, 27, 35, 38, 39, 58, 62, 77, 80, 84, 87, 90, 91, 103, 106, 109, 113, 116, 119, 124, 125, 131, 132]	[2, 4, 13, 15, 27, 35, 39, 80, 87, 90, 91, 95, 106, 109, 113, 119, 124, 125, 131, 132]	[2, 4, 13, 35, 38, 39, 70, 77, 80, 81, 84, 87, 90, 91, 95, 106, 109, 119, 124, 125, 131]	[25, 27, 38, 39, 77, 80, 87, 91, 96, 113, 119, 121, 125, 126, 131]	[27, 48, 70, 77, 80, 116, 119, 124–126]
Developers	[6, 10, 38, 80, 87, 102, 103, 106, 112]	[2, 6, 7, 10, 12, 32, 35, 38, 39, 62, 80, 84, 85, 87, 90, 102, 103, 106, 121, 125]	[2, 4, 7, 12, 14, 16, 27, 29, 32, 38, 39, 49, 56, 58, 59, 70, 72, 73, 77, 80, 84, 85, 87, 90, 102–104, 106, 109, 118, 121, 122, 124, 125, 128, 134, 137]	[2, 4, 7, 13–15, 27, 29, 32, 35, 38, 39, 49, 55, 56, 58, 59, 70, 73, 77, 80, 84, 85, 87, 90, 102, 104, 106, 109, 111, 118, 121, 122, 124, 125, 128, 133, 134, 136, 137]	[2, 4, 7, 9, 13, 15, 16, 27, 29, 32, 35, 38, 39, 50, 55, 56, 58, 62, 73, 77, 80, 84, 85, 87, 90, 100, 102, 103, 106, 109, 111, 113, 118, 120, 122–125, 128, 130–134, 136]	[2, 4, 7, 13, 15, 27, 29, 32, 35, 37, 39, 46, 49, 56, 73, 80, 85, 87, 90, 95, 100, 102, 106, 109, 113, 118, 120, 122–125, 128, 130–132, 136]	[2, 4, 7, 13, 29, 32, 35, 37–40, 46, 49, 56, 70, 72, 73, 77, 80, 84, 85, 87, 90, 95, 102, 106, 109, 118, 120, 122–125, 128, 130, 131, 134, 136]	[27, 32, 38, 39, 56, 60, 77, 80, 87, 96, 102, 113, 121, 125, 131]	[7, 27, 32, 40, 60, 70, 76, 77, 80, 102, 112, 124, 125, 134]
Deployers	[6, 78, 80, 87, 102, 119]	[6, 12, 78, 80, 87, 102, 119, 121]	[4, 12, 27, 72, 80, 87, 102, 119, 121, 122, 124, 128, 137]	[4, 15, 27, 80, 87, 102, 119, 121, 122, 124, 128, 137]	[3, 4, 9, 15, 27, 80, 87, 100, 102, 113, 119, 120, 122–124, 128, 130, 131]	[3, 4, 15, 27, 37, 46, 80, 87, 89, 100, 102, 113, 119, 120, 122–124, 128, 130, 131]	[3, 4, 37, 46, 72, 80, 81, 87, 89, 102, 119, 120, 122–124, 128, 130, 131]	[27, 78, 80, 87, 96, 102, 113, 119, 121, 131]	[3, 27, 80, 102, 119, 124]
End-users	[]	[39, 85]	[39, 85, 116, 118]	[39, 85, 116, 118]	[9, 39, 85, 116, 118]	[39, 85, 95, 118]	[39, 85, 95, 118]	[39]	[20, 76, 116, 117]
Impacted Communities	[83]	[39, 48, 62, 83, 85]	[39, 85]	[39, 85]	[39, 62, 85]	[39, 85]	[39, 81, 85]	[39]	[48, 117]