SoK: Mapping Threats to Defenses in Online Survey Fraud

Shiza Ali, ***Wellington Barbosa***, Matthias Fassl, Aditi Ganapathi*, Jaron Mink*, Adam J. Aviv

The George Washington University
*Arizona State University

22nd Symposium on Usable Privacy and Security · August 26th, 2026

Online surveys are prevalent in data collection



Survey Protocol

Online surveys are prevalent in data collection



Survey Protocol

qualtrics XM

SurveyJS

Implementation

Online surveys are prevalent in data collection



Survey Protocol

qualtrics XM

SurveyJS

Implementation

amazon mechanical turk

Prolific

reddit

SOUPS 2026

Distribution

However, it requires careful data cleaning





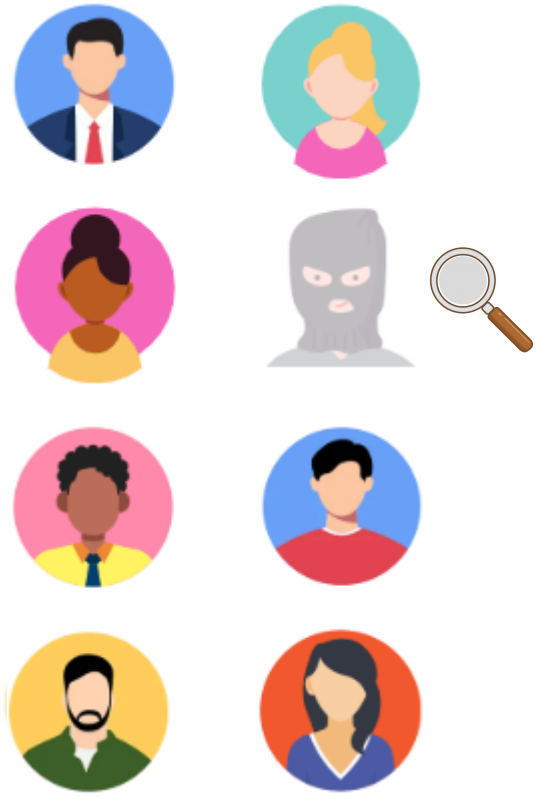
Survey data	
Submission date	**Email**
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 14, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com

However, it requires careful data cleaning



Survey data	
Submission date	**Email**
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 14, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com

However, it requires careful data cleaning



Survey data	
Submission date	**Email**
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com
May 7, 2035	hello@reallygreatsite.com

GO **Continue data analysis**

How is fraudulent participation *actually* detected?

Different - and often reactive - approaches to detect and mitigate fraud

- Attention checks?
- Response time?
- Platform controls?
- reCAPTCHA?
- Quality of open-text responses?

How is fraudulent participation *actually* detected?

Different - and often reactive - approaches to detect and mitigate fraud

- Attention checks?
- Response time?
- Platform controls?
- reCAPTCHA?
- Quality of open-text responses?

What about LLMs?

How is fraudulent participation *actually* detected?

Different - and often reactive - approaches to detect and mitigate fraud

- Attention checks?
- Response time?
- Platform controls?
- reCAPTCHA?
- Quality of open-text responses?

What about LLMs?

What is even *considered* fraud and what *methods* have researchers used to combat it?

Specifically, we investigated these questions:

RQ1 [Fraud Definition]

How is survey fraud conceptualized and defined across research domains?

Specifically, we investigated these questions:

RQ1 [Fraud Definition]

How is survey fraud conceptualized and defined across research domains?

RQ2 [Detection & Mitigation]

What detection and mitigation strategies are proposed, and how do they align with specific fraud vectors?

Specifically, we investigated these questions:

RQ1 [Fraud Definition]

How is survey fraud conceptualized and defined across research domains?

RQ2 [Detection & Mitigation]

What detection and mitigation strategies are proposed, and how do they align with specific fraud vectors?

RQ3 [Reporting Practices]

How are fraud mitigation practices reported, and what are their implications for interpretability and reproducibility?

Methods

1. Corpus built following the PRISMA 2020 protocol

Methods

1. Corpus built following the PRISMA 2020 protocol

2. Inclusion criteria
   - Papers published between **2010 - 2025**
   - *Plus* at least one of the following:
     - Conceptualization of Fraud
     - Detection or Mitigation
     - Guidance or Implications

Methods

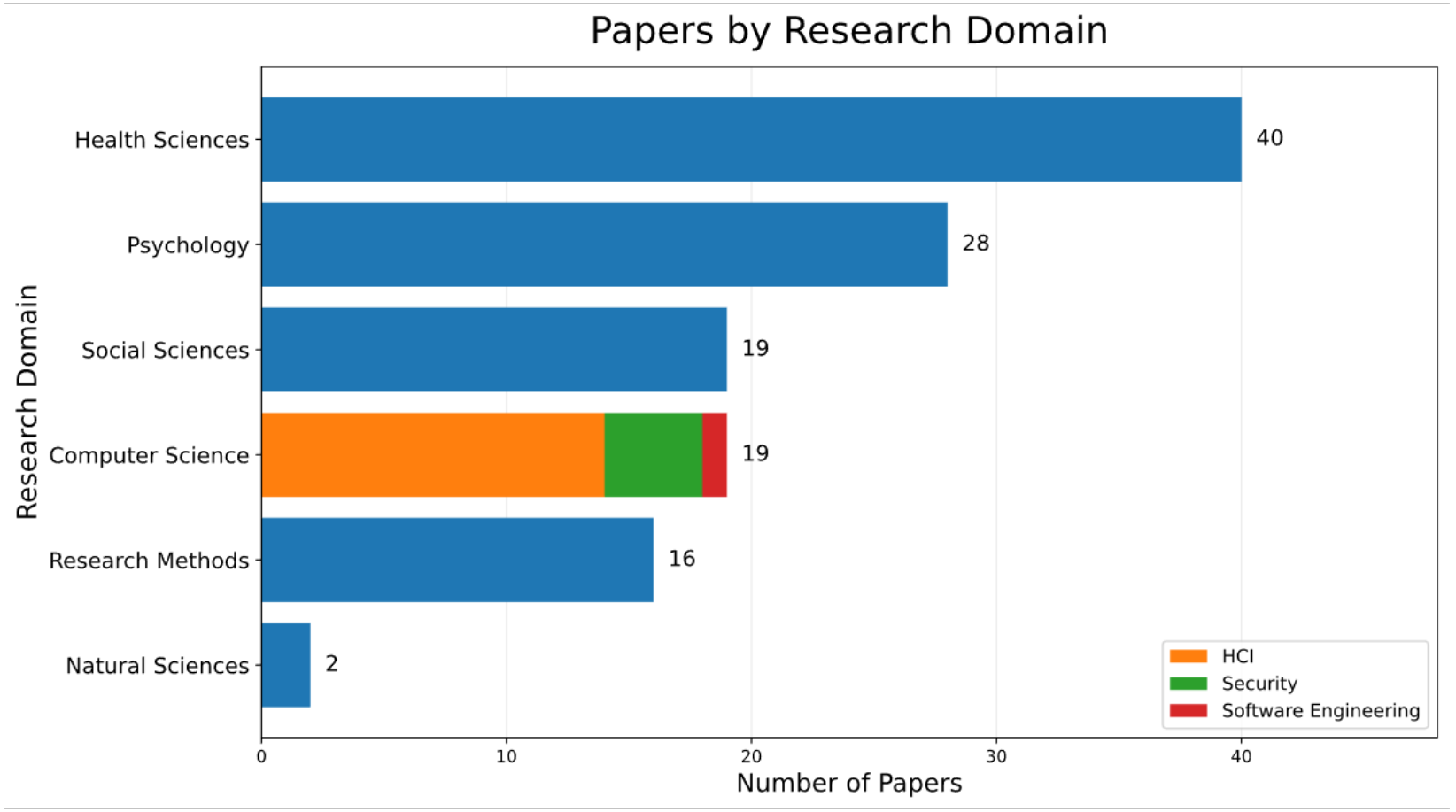
1. Corpus built following the PRISMA 2020 protocol

2. Inclusion criteria
   - Papers published between **2010 - 2025**
   - *Plus* at least one of the following:
     - Conceptualization of Fraud
     - Detection or Mitigation
     - Guidance or Implications

Google Scholar dblp computer science bibliography

Initial search	**107**
Snowballing	**+ 166**
Subtotal	**= 273**
Out of scope	**- 74**
Not available	**- 11**
Subtotal	**= 188**
Data quality *only**	**- 64**
Total	**= 124**

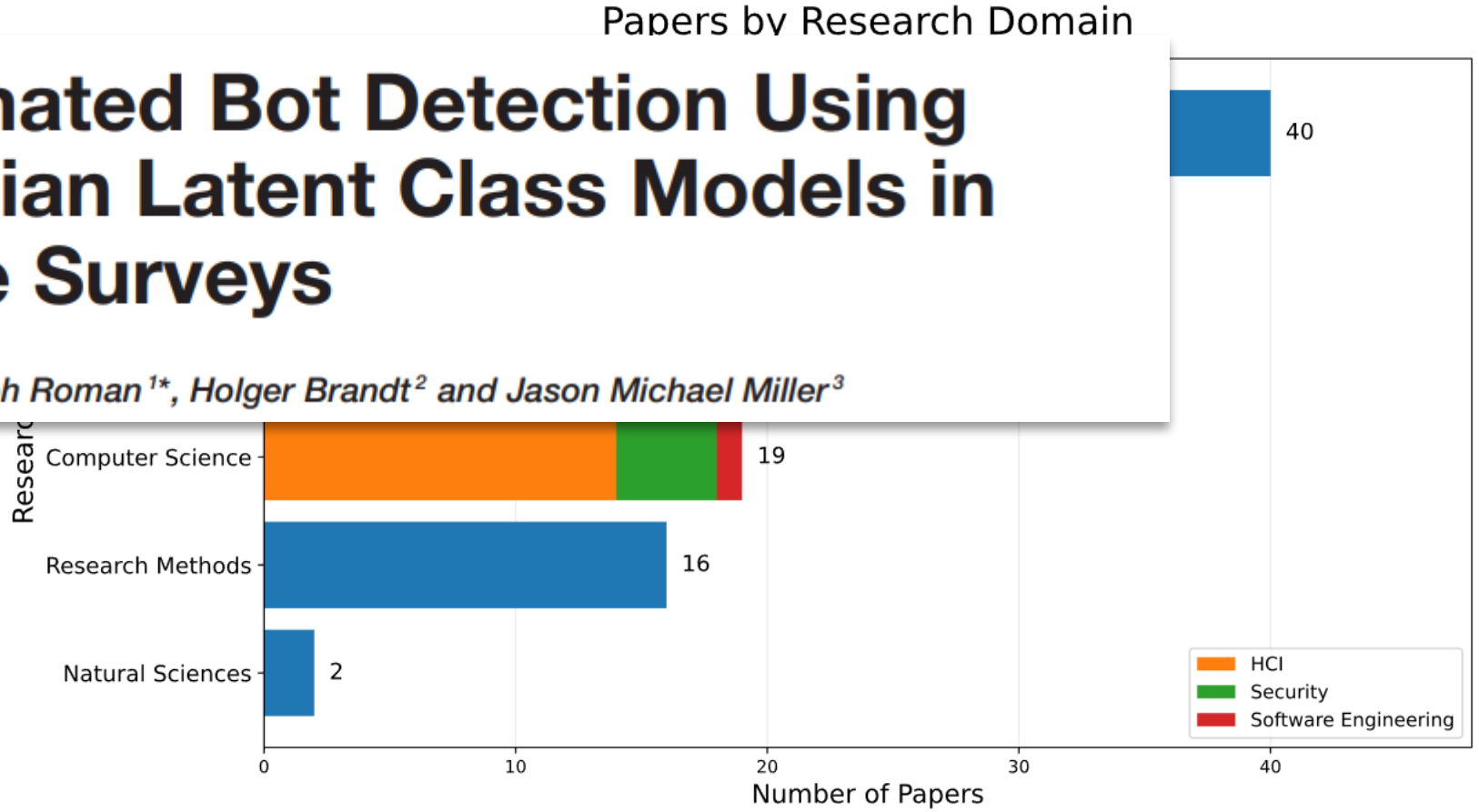
Overview of the studies



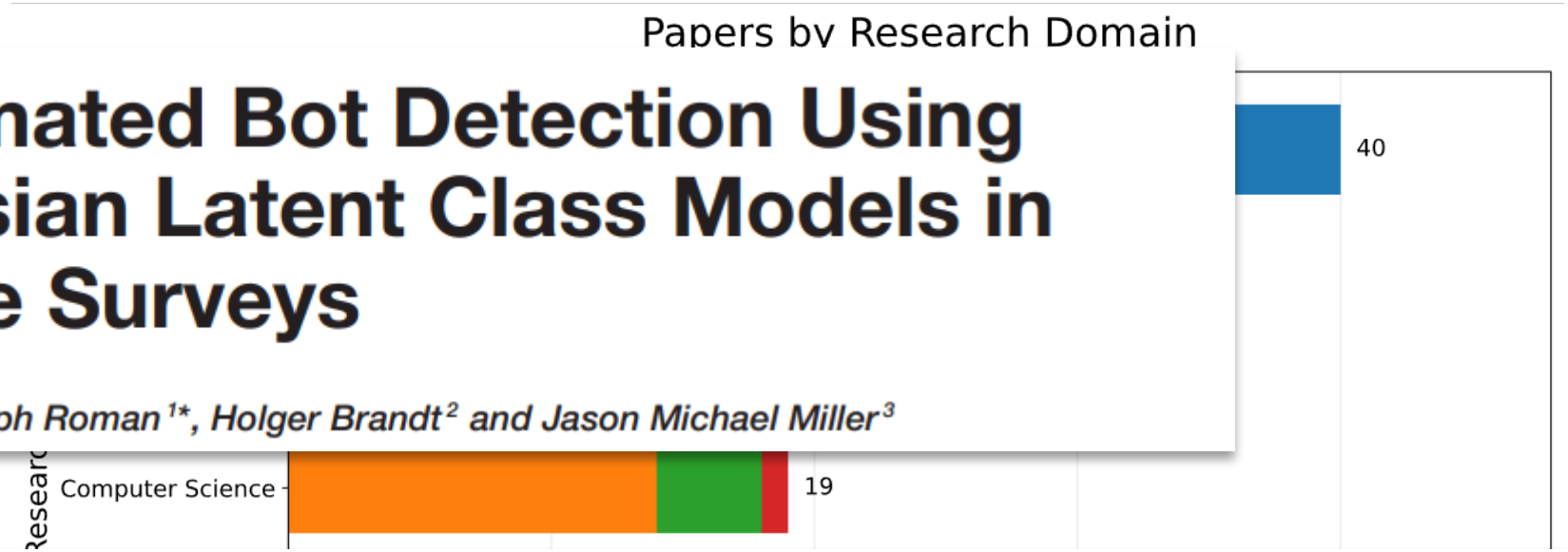
Overview of the studies

Automated Bot Detection Using Bayesian Latent Class Models in Online Surveys

Zachary Joseph Roman[1], Holger Brandt[2] and Jason Michael Miller[3]*



Overview of the studies



Automated Bot Detection Using Bayesian Latent Class Models in Online Surveys

Zachary Joseph Roman[1], Holger Brandt[2] and Jason Michael Miller[3]*

Fraudulent participation in psychological research using virtual synchronous interviews: ethical challenges and potential solutions

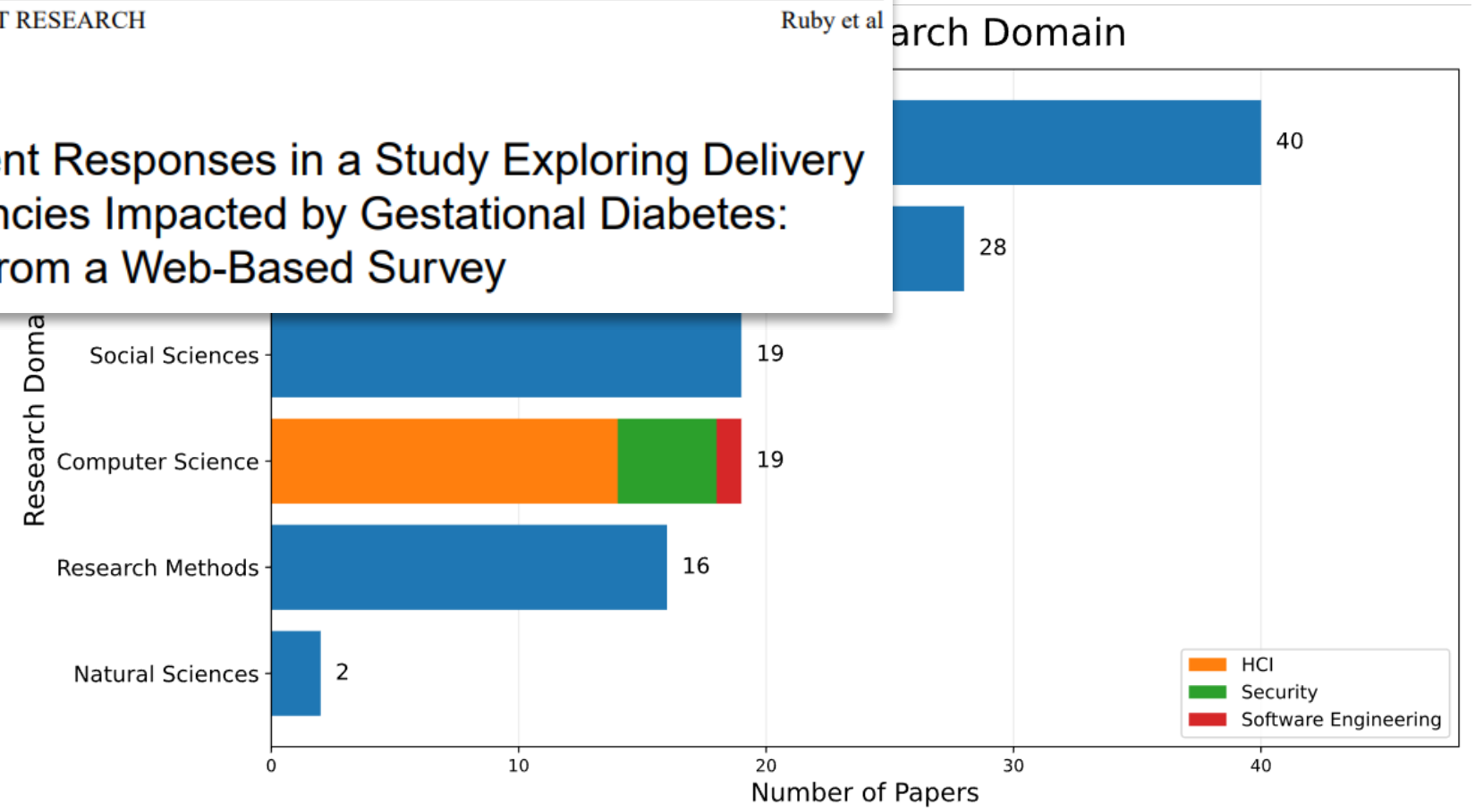
Kaitlyn McLachlan[a], Emma E. Truffyn[a], Bianka Dunleavy[a], Delane Linkiewich[a], Deborah Powell[a], Anna Taddio[b], and C. Meghan McMurtry[c]

Overview of the studies

JOURNAL OF MEDICAL INTERNET RESEARCH Ruby et al

Viewpoint

Identifying Fraudulent Responses in a Study Exploring Delivery Options for Pregnancies Impacted by Gestational Diabetes: Lessons Learned From a Web-Based Survey



Overview of the studies

JOURNAL OF MEDICAL INTERNET RESEARCH Ruby et al

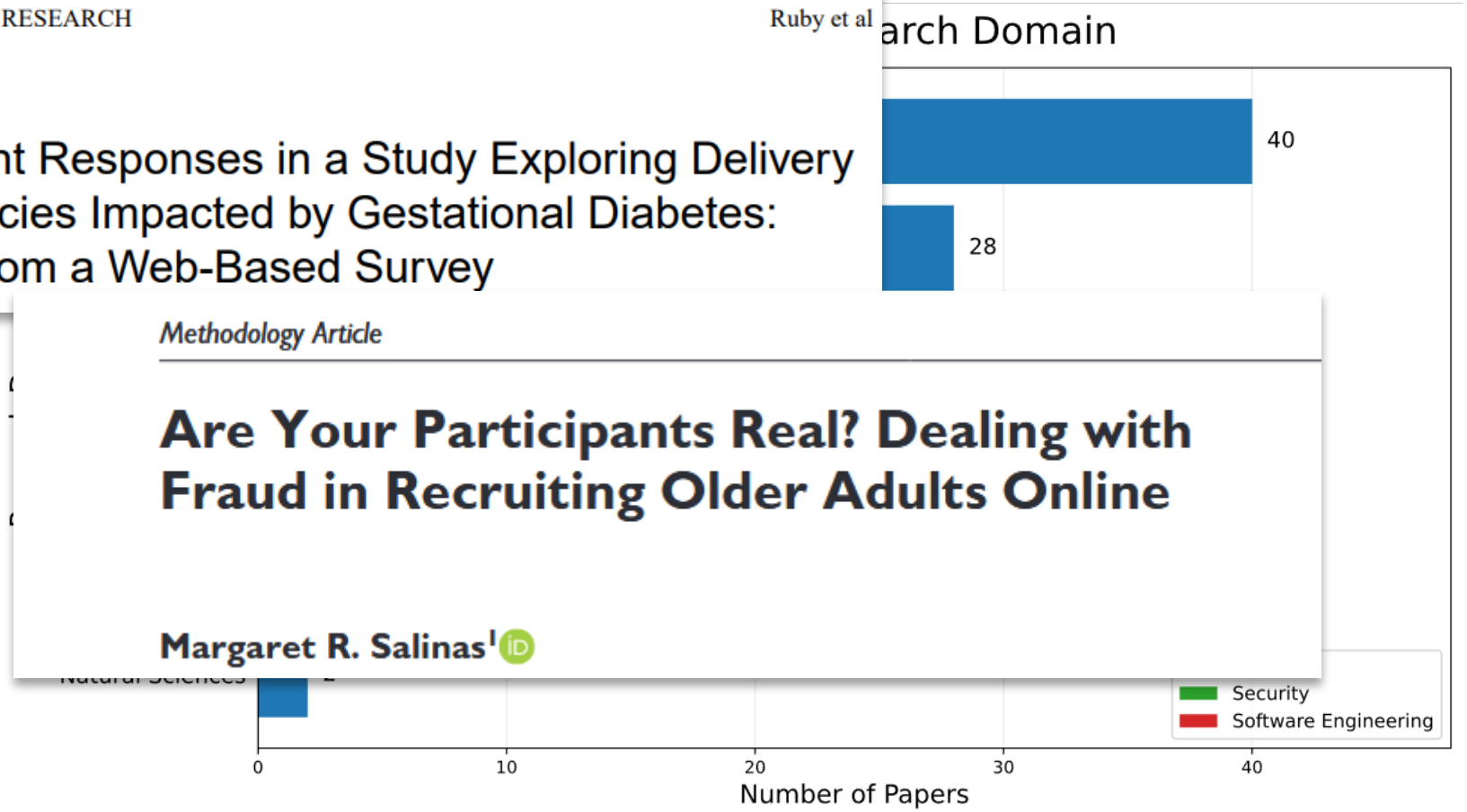
Viewpoint

Identifying Fraudulent Responses in a Study Exploring Delivery Options for Pregnancies Impacted by Gestational Diabetes: Lessons Learned From a Web-Based Survey

Methodology Article

Are Your Participants Real? Dealing with Fraud in Recruiting Older Adults Online

Margaret R. Salinas[1]



Overview of the studies

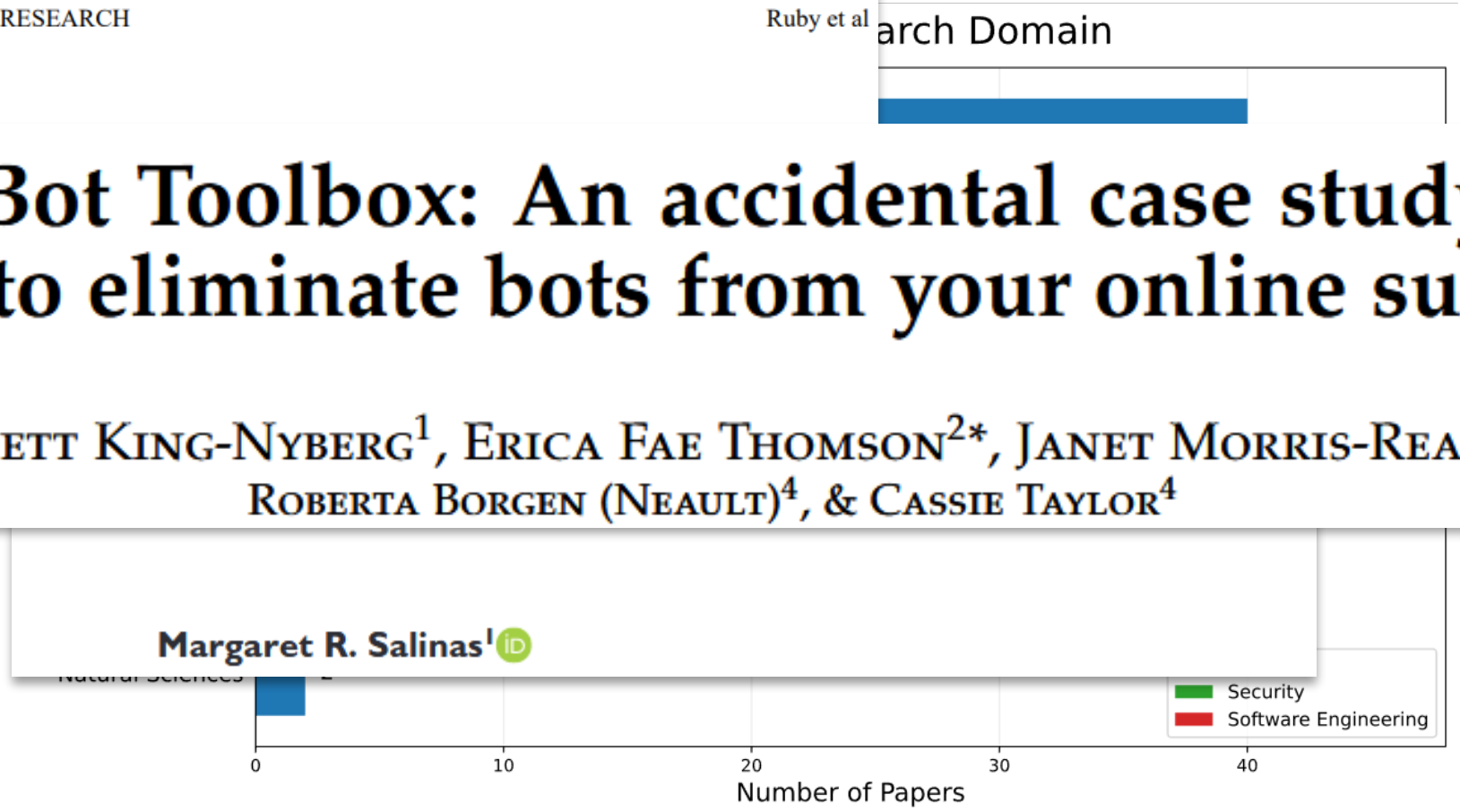
JOURNAL OF MEDICAL INTERNET RESEARCH Ruby et al

Viewpoint

The Bot Toolbox: An accidental case study on how to eliminate bots from your online survey

Bennett King-Nyberg[1], Erica Fae Thomson[2*], Janet Morris-Reade[3], Roberta Borgen (Neault)[4], & Cassie Taylor[4]

Margaret R. Salinas[1]



RQ1: What is considered fraud?

No definition

No explicit definitions

Implicit/Ad-hoc

RQ1: What is considered fraud?

No definition

No explicit definitions

Implicit/Ad-hoc

Intentional data fabrication

Mischievous participants

Interviewer data fabrication

RQ1: What is considered fraud?

No definition

No explicit definitions

Implicit/Ad-hoc

Intentional data fabrication

Mischievous participants

Interviewer data fabrication

Careless and Insufficient-Effort

Effortless, Inattentive

Random

Low or little motivation

RQ1: What is considered fraud?

No definition

No explicit definitions

Implicit/Ad-hoc

Intentional data fabrication

Mischievous participants

Interviewer data fabrication

Careless and Insufficient-Effort

Effortless, Inattentive

Random

Low or little motivation

Automated, AI-mediated

Scripted bots

Coordinated automation

“LLM pollution”

RQ1: What is considered fraud?

No definition	Intentional data fabrication	Careless and Insufficient-Effort	Automated, AI-mediated
No explicit definitions	Mischievous participants	Effortless, Inattentive	Scripted bots
Implicit/Ad-hoc	Interviewer data fabrication	Random	Coordinated automation
		Low or little motivation	“LLM pollution”

Common fraudulent behaviors:

RQ1: What is considered fraud?

No definition	Intentional data fabrication	Careless and Insufficient-Effort	Automated, AI-mediated
No explicit definitions	Mischievous participants	Effortless, Inattentive	Scripted bots
Implicit/Ad-hoc	Interviewer data fabrication	Random	Coordinated automation
		Low or little motivation	“LLM pollution”

Common fraudulent behaviors:

False eligibility claims

Location masking

Random responses

RQ1: What is considered fraud?

No definition	Intentional data fabrication	Careless and Insufficient-Effort	Automated, AI-mediated
No explicit definitions	Mischievous participants	Effortless, Inattentive	Scripted bots
Implicit/Ad-hoc	Interviewer data fabrication	Random	Coordinated automation
		Low or little motivation	"LLM pollution"

Common fraudulent behaviors:

False eligibility claims

Duplicate submissions

Location masking

Bots/AI-generated

Random responses

Coordinated use of eligible accounts

RQ2: Mitigation Strategies Often Come Too Late

SURVEY LIFECYCLE



RQ2: Mitigation Strategies Often Come Too Late

Lifecycle stage at which fraud vectors are addressed in papers from our corpus*

Fraud vector	Recruitment	In-survey	Post-hoc
Identity spoofing / false claims	23	28	31
Multi-accounting	17	11	12
Careless responding	7	37	24
Bot-generated	15	32	23

*See the paper's **Table 4**, p.9, for more fraud vectors

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE

Pre-Survey

Recruitment, Eligibility

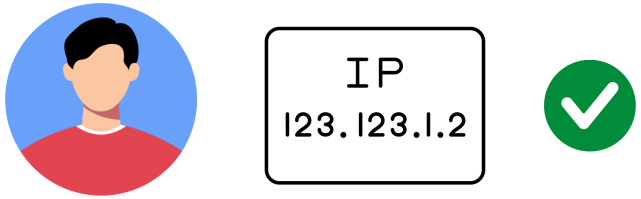
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

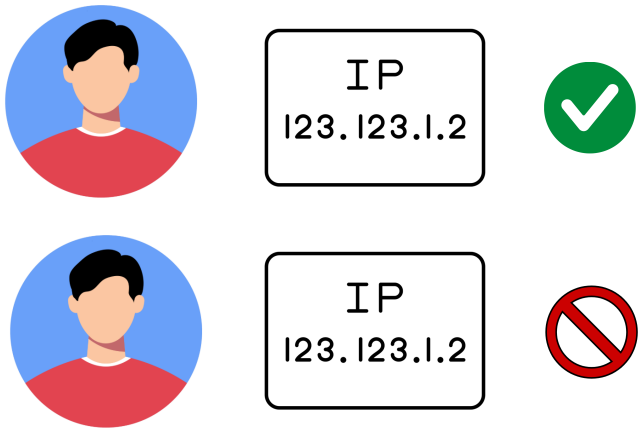
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

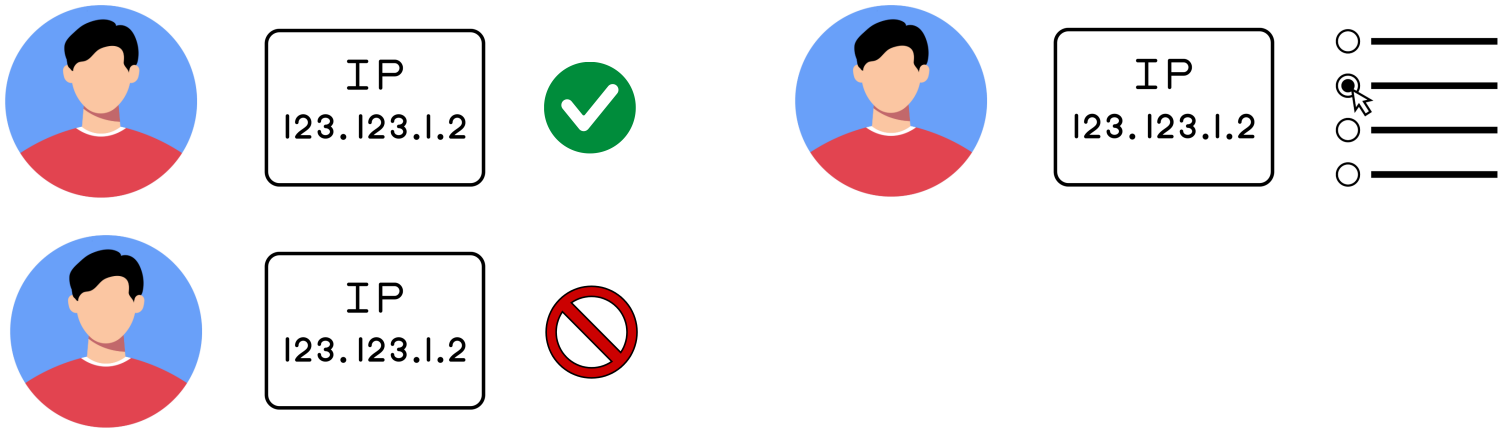
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

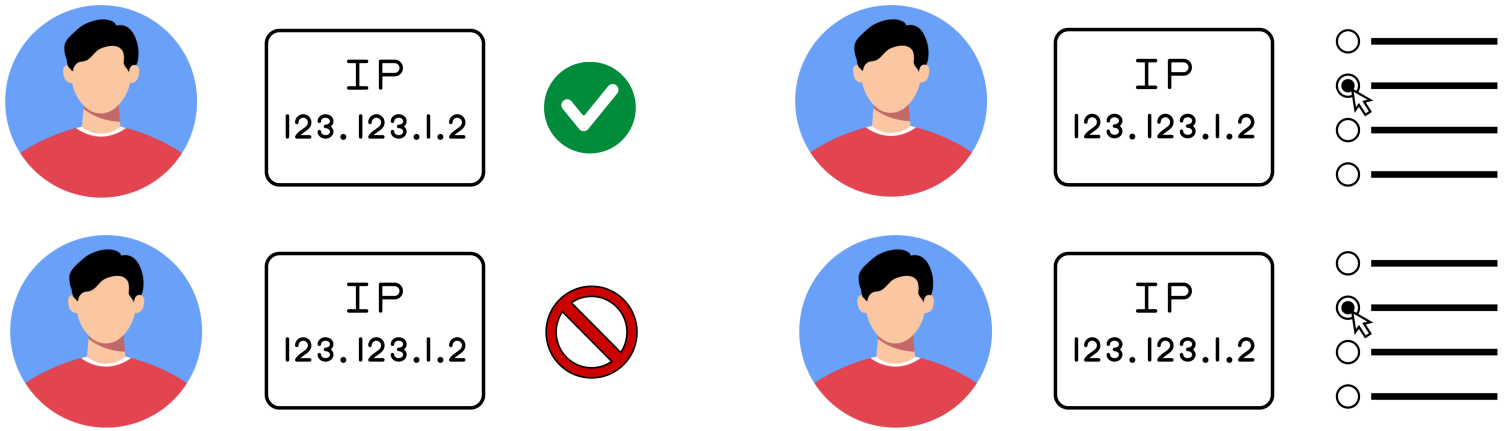
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

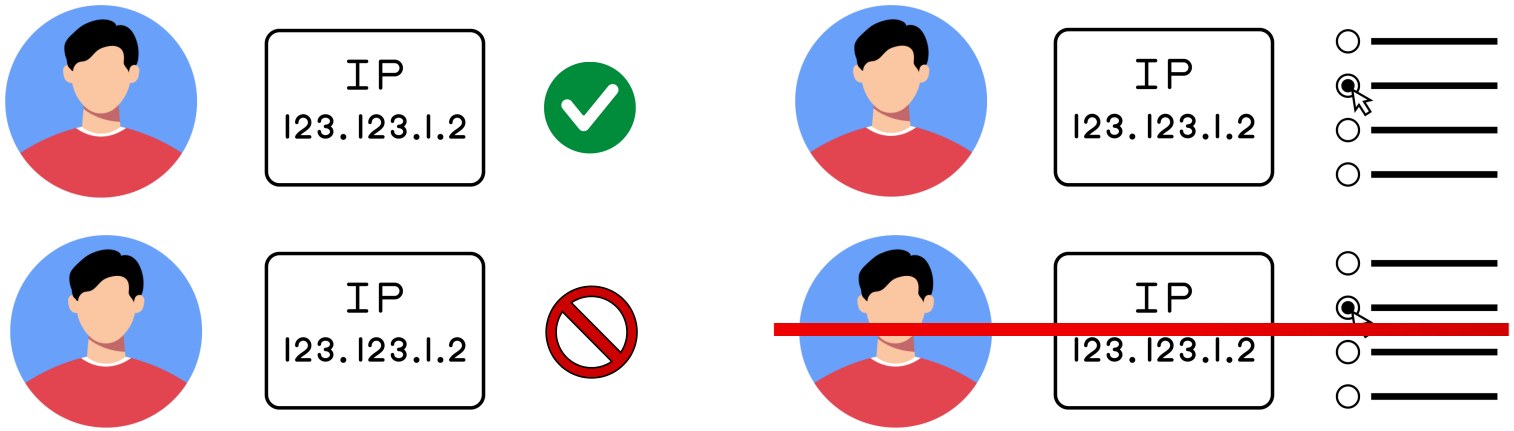
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

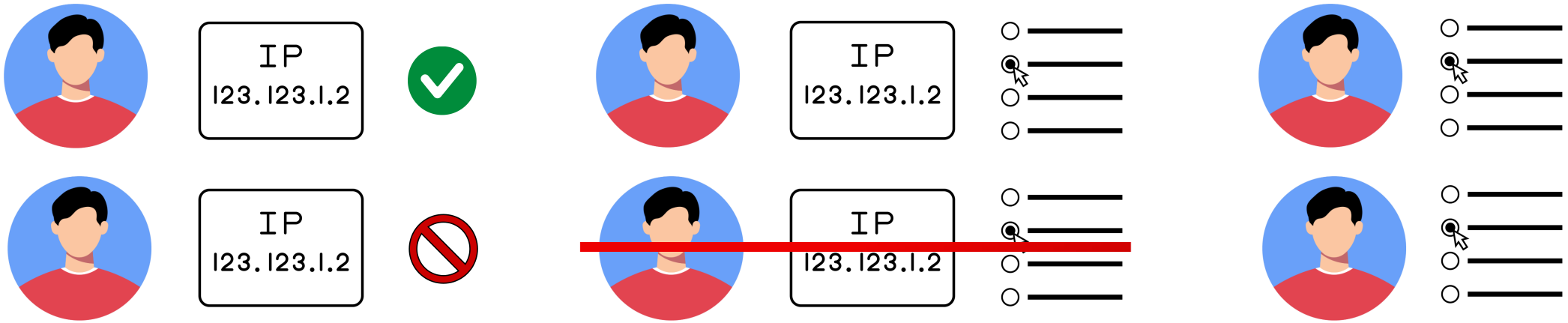
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

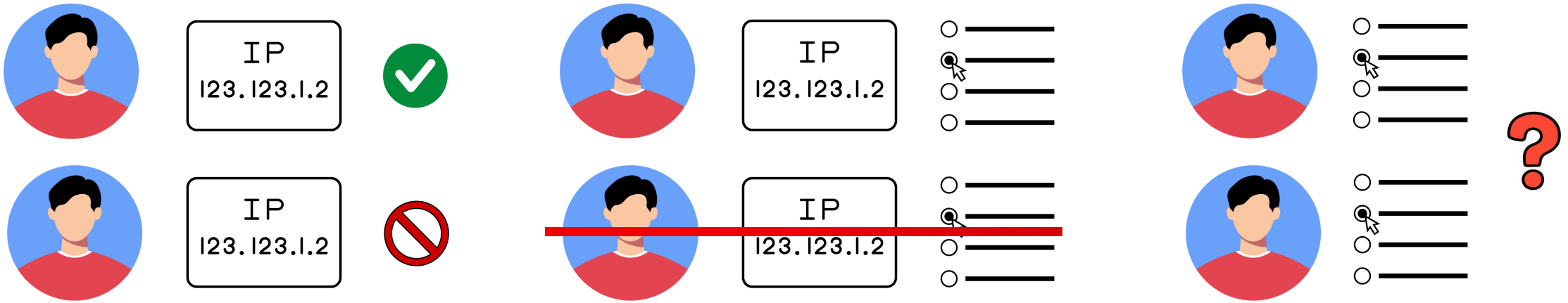
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



Pre-Survey

Recruitment, Eligibility

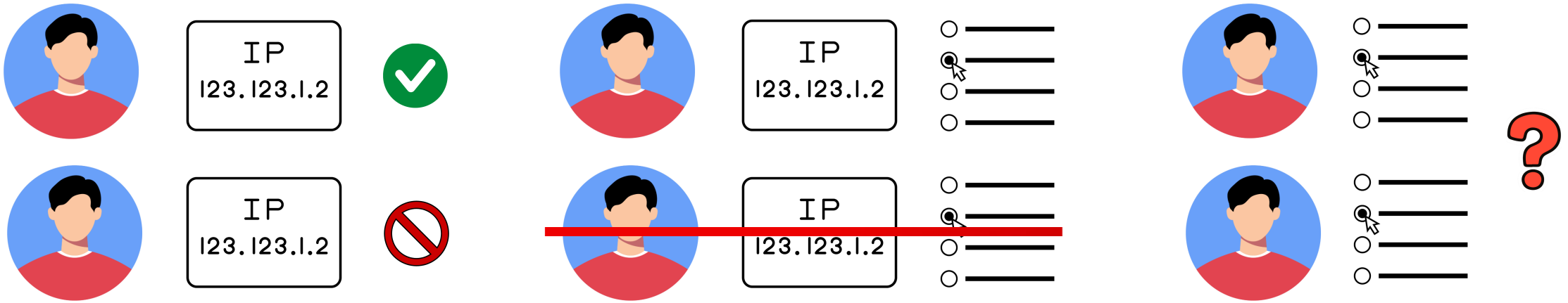
In-Survey

Post-Survey

Data analysis

Mitigation Strategies: Multiple Submissions

SURVEY LIFECYCLE



The infrastructure and data you decide to collect **before the study begins** can determine what fraud you are able to detect later.

RQ2: Alignment between threats and defenses

SURVEY LIFECYCLE

Pre-Survey

Recruitment, Eligibility

In-Survey

Post-Survey

Data analysis

RQ2: Alignment between threats and defenses

SURVEY LIFECYCLE



False claims

Location masking



Knowledge checks

VPN/proxy detection

Pre-Survey

Recruitment, Eligibility

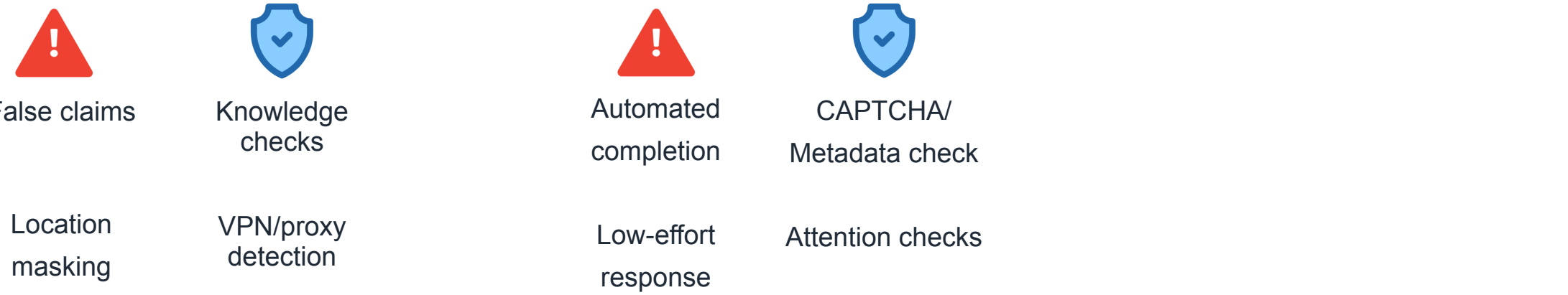
In-Survey

Post-Survey

Data analysis

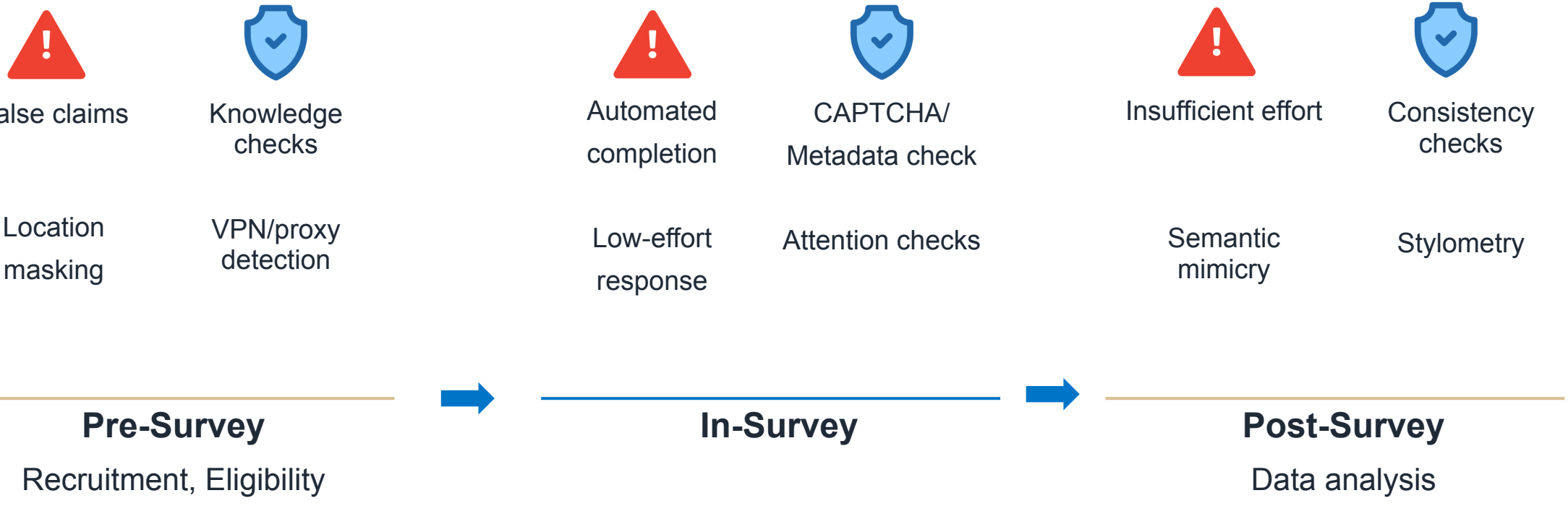
RQ2: Alignment between threats and defenses

SURVEY LIFECYCLE



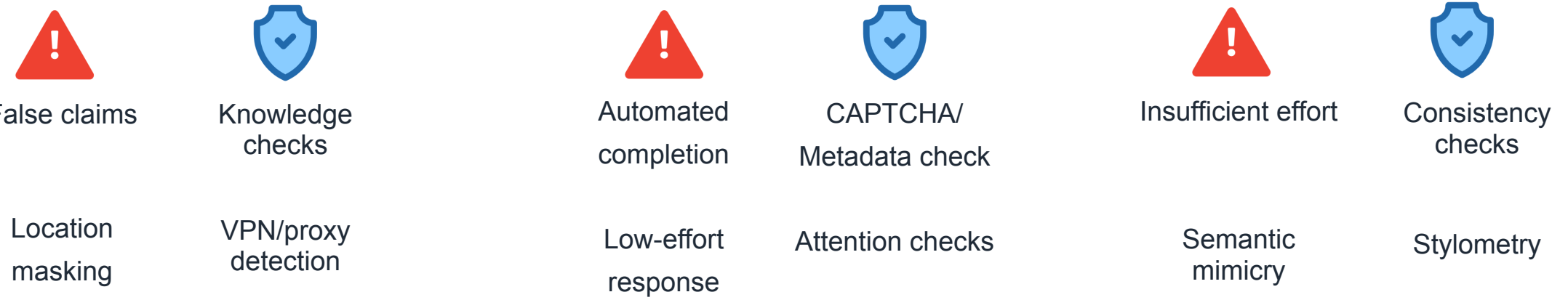
RQ2: Alignment between threats and defenses

SURVEY LIFECYCLE



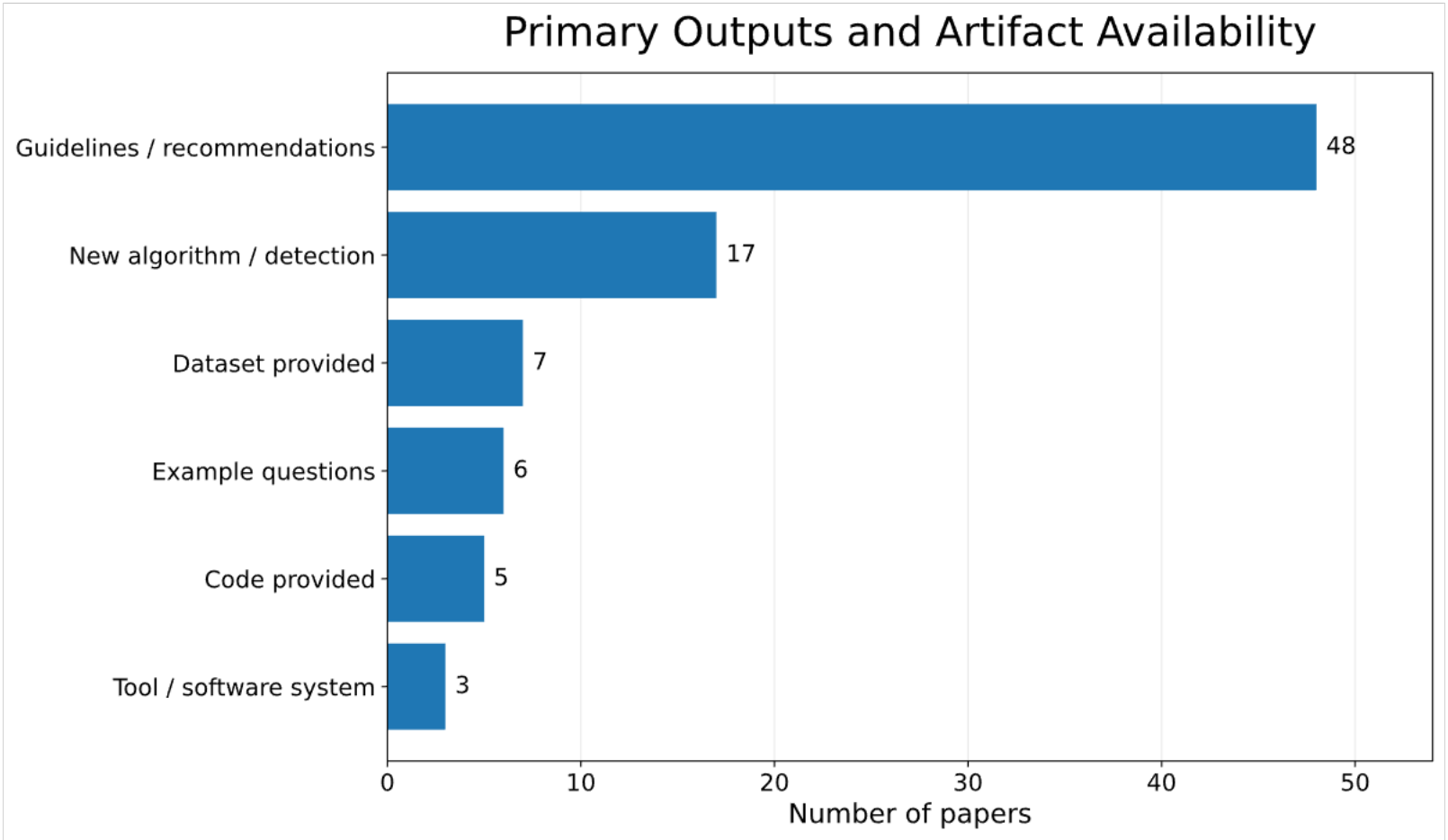
RQ2: Alignment between threats and defenses

SURVEY LIFECYCLE

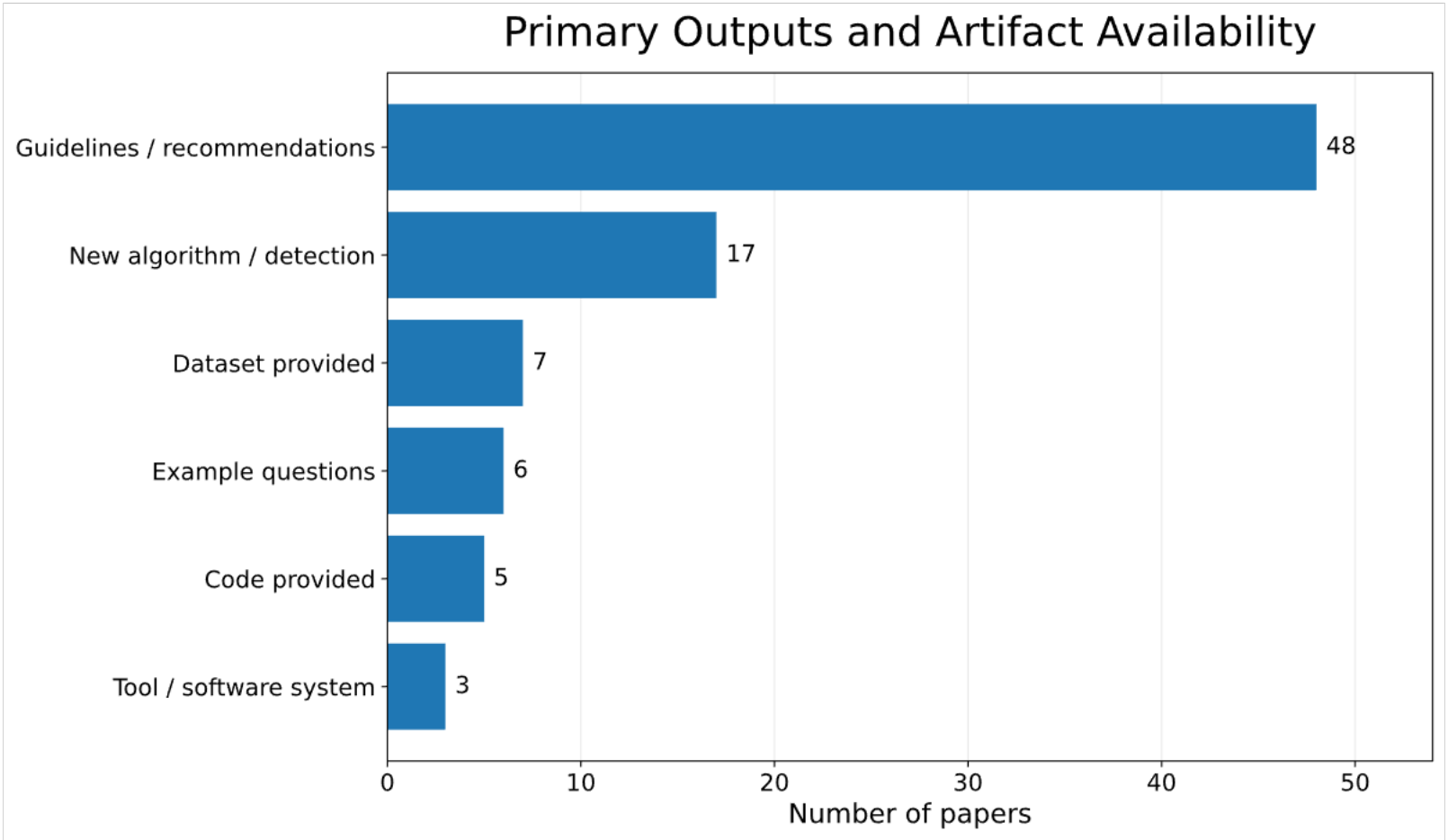


*See the paper's **Tables 3** (p.8) and **5** (p.9) for more fraud vectors and detection methods

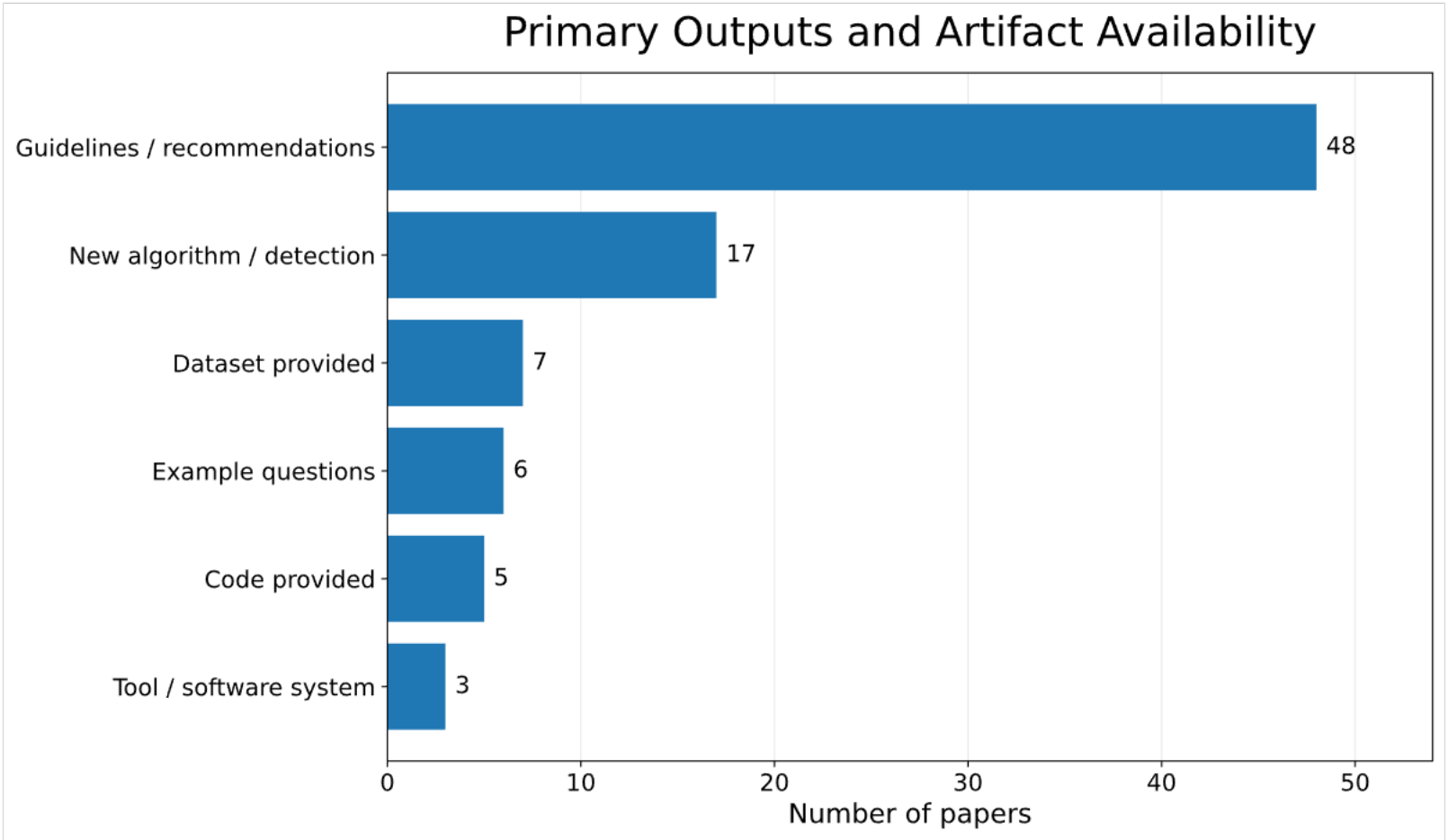
RQ3: What Researchers Share (and What is Left Out)



RQ3: What Researchers Share (and What is Left Out)

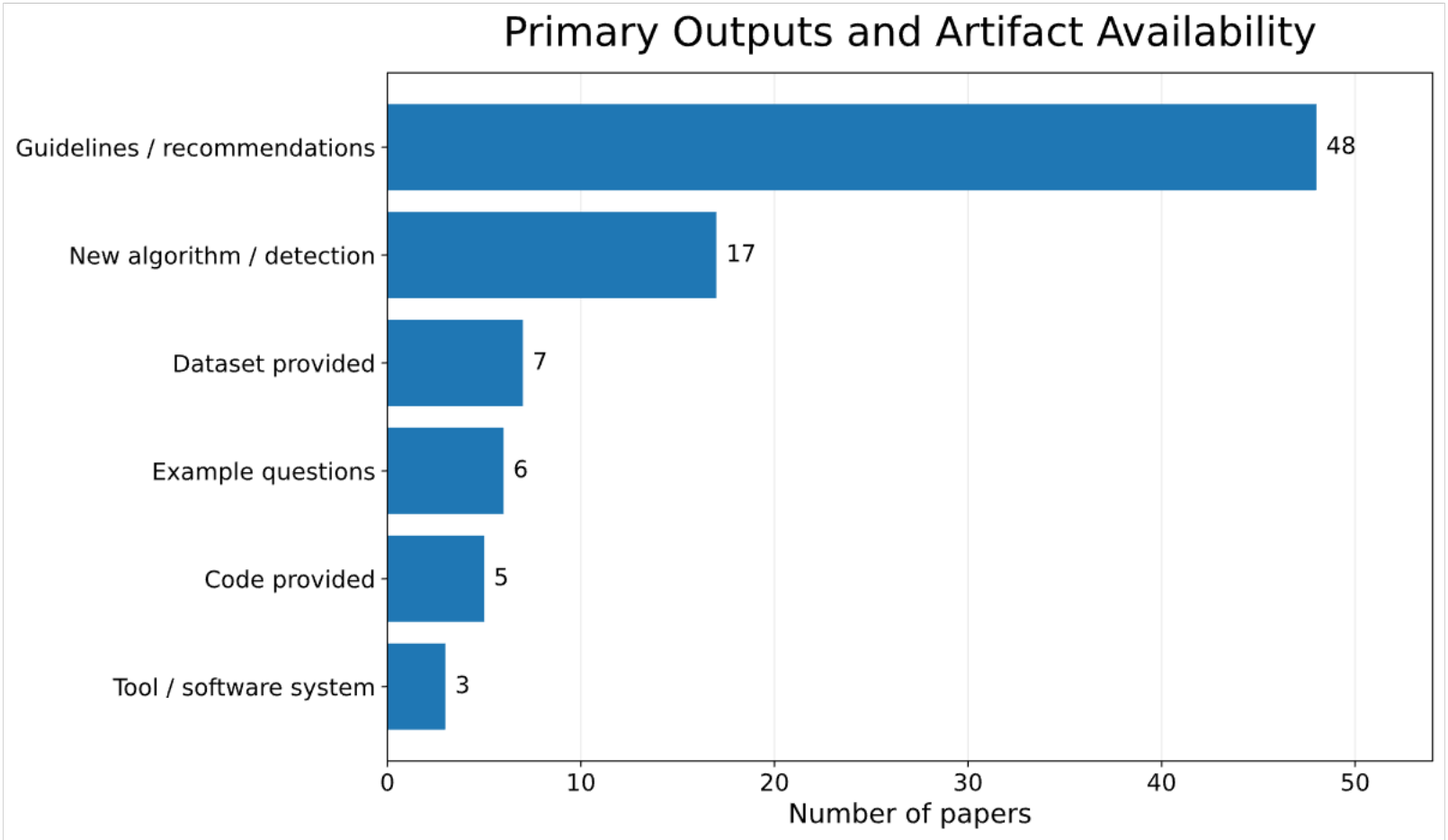


RQ3: What Researchers Share (and What is Left Out)



Frequently missing:

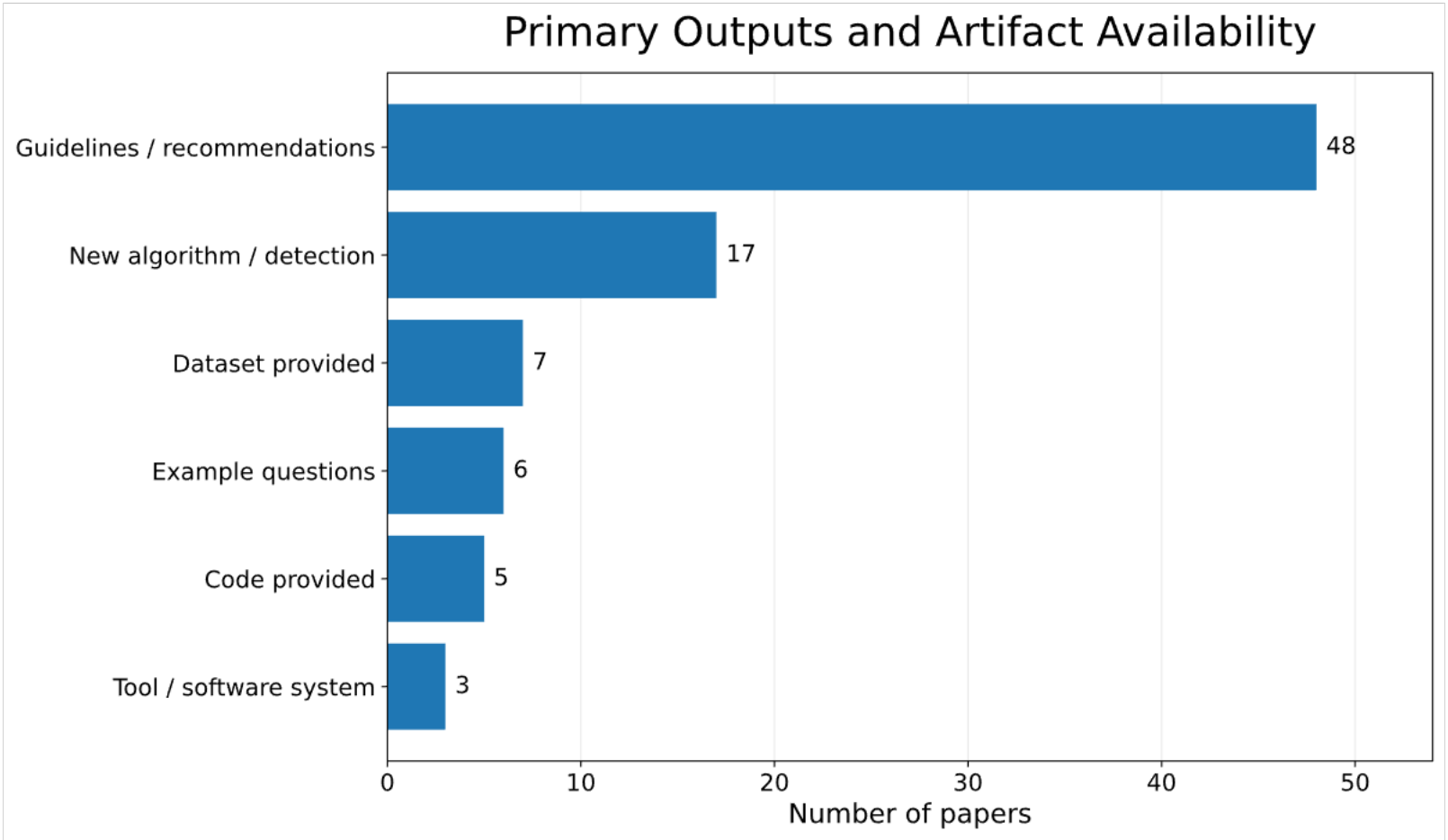
RQ3: What Researchers Share (and What is Left Out)



Frequently missing:

1. Decision thresholds

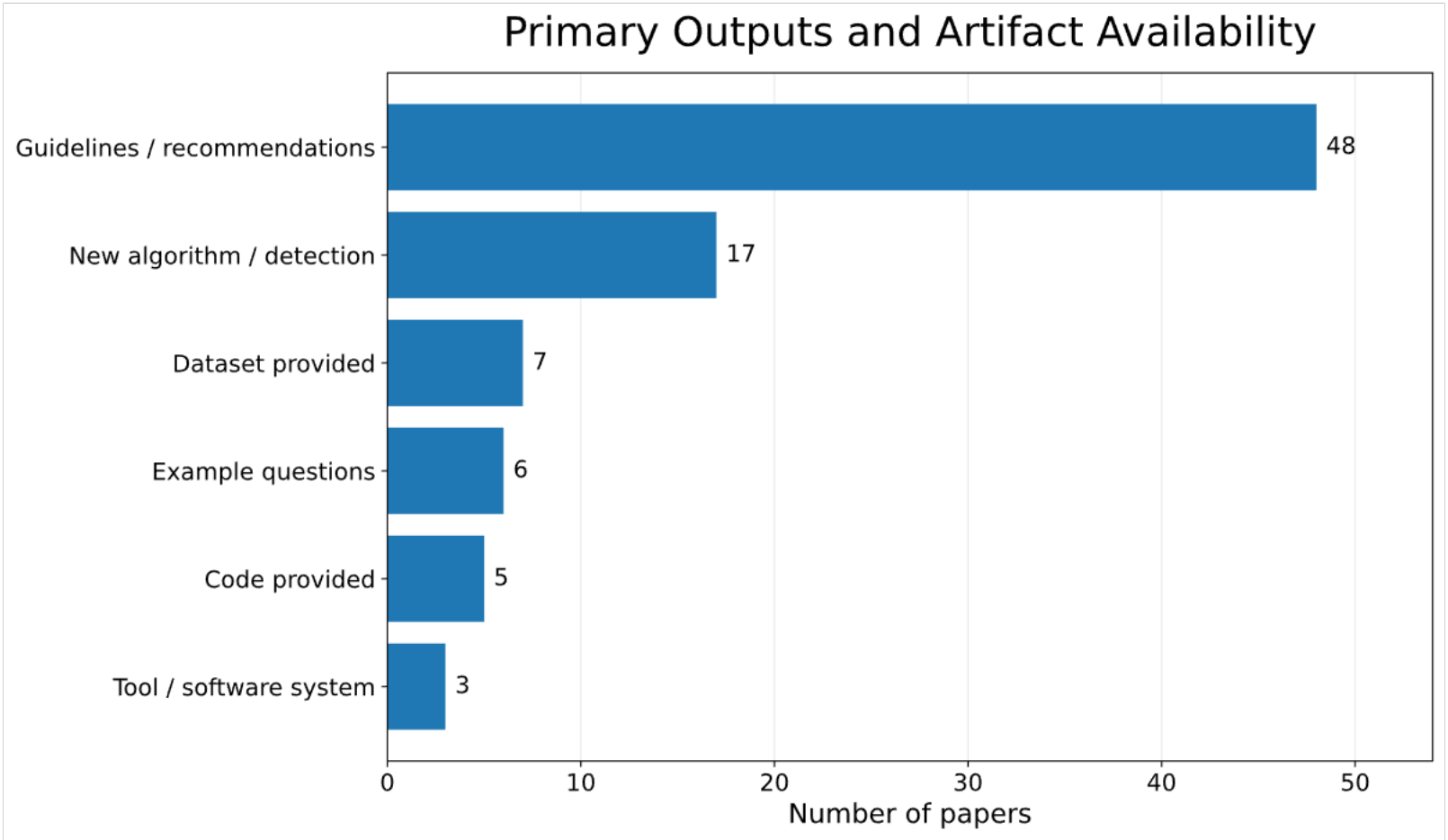
RQ3: What Researchers Share (and What is Left Out)



Frequently missing:

1. Decision thresholds
2. Exclusion criteria

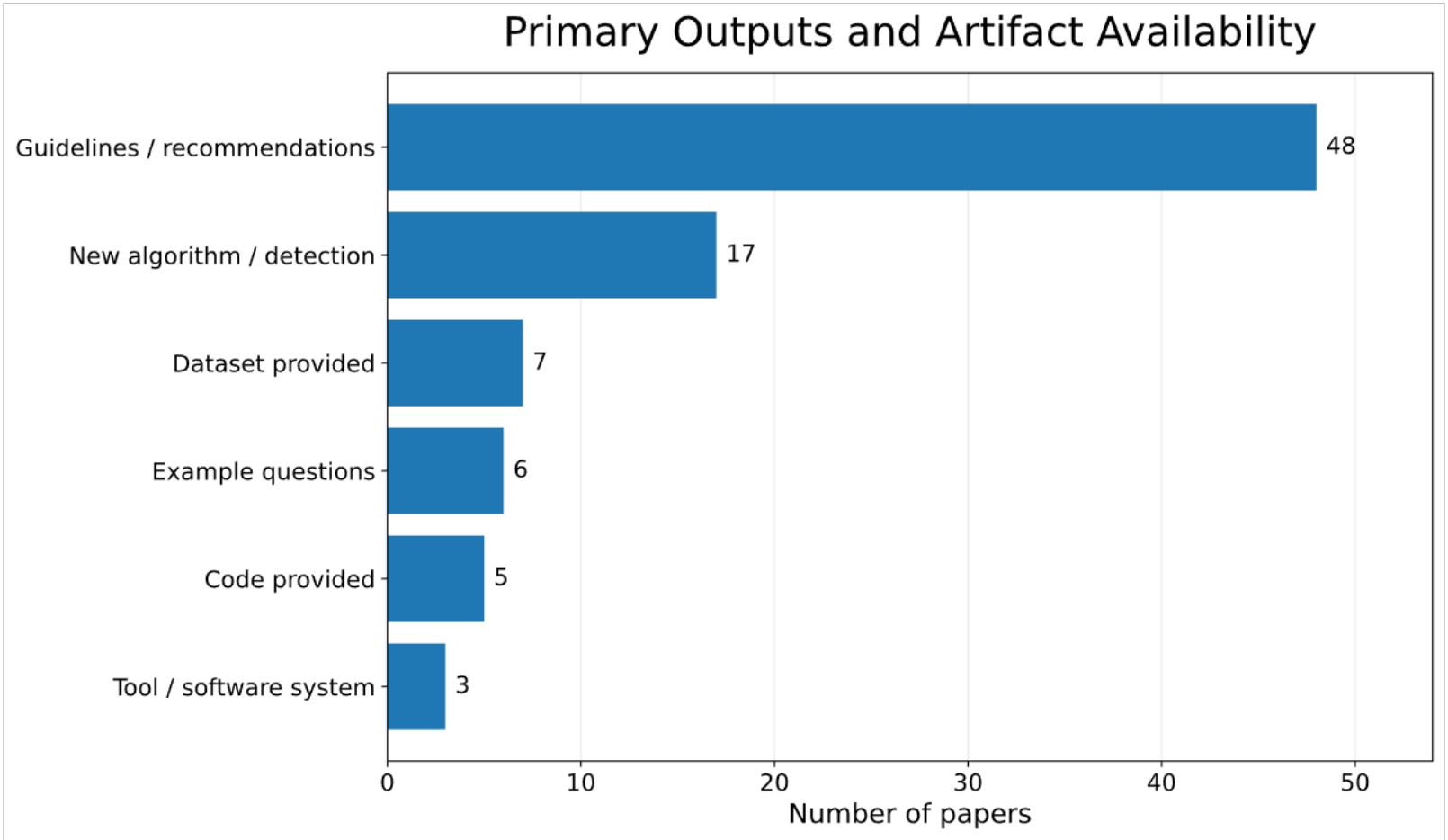
RQ3: What Researchers Share (and What is Left Out)



Frequently missing:

1. Decision thresholds
2. Exclusion criteria
3. % of responses removed

RQ3: What Researchers Share (and What is Left Out)



Frequently missing:

1. Decision thresholds
2. Exclusion criteria
3. % of responses removed
4. Detection methods used and their effectiveness
5. Consequences to participants (denying compensation, reporting participants to platform)

Survey fraud mitigation is difficult to evaluate when threats are poorly specified, defenses are not aligned with them, and mitigation decisions are not reported.

So, what should researchers do differently?

GUIDELINES

- **G1.** Explicitly specify the ***threat model*** underlying fraud mitigation choices.

So, what should researchers do differently?

GUIDELINES

- **G1.** Explicitly specify the ***threat model*** underlying fraud mitigation choices.

- **G2.** ***Align defenses*** with fraud vectors across the survey lifecycle.

So, what should researchers do differently?

GUIDELINES

- **G1.** Explicitly specify the ***threat model*** underlying fraud mitigation choices.
- **G2.** ***Align defenses*** with fraud vectors across the survey lifecycle.
- **G4.** Treat inattentiveness checks as ***limited signals*** rather than general-purpose defenses.

So, what should researchers do differently?

GUIDELINES

- **G1.** Explicitly specify the ***threat model*** underlying fraud mitigation choices.
- **G2.** ***Align defenses*** with fraud vectors across the survey lifecycle.
- **G4.** Treat inattentiveness checks as ***limited signals*** rather than general-purpose defenses.
- **G6.** ***Report*** mitigation decisions and their downstream effects on data.

So, what should researchers do differently?

GUIDELINES

- **G1.** Explicitly specify the ***threat model*** underlying fraud mitigation choices.
- **G2.** ***Align defenses*** with fraud vectors across the survey lifecycle.
- **G4.** Treat inattentiveness checks as ***limited signals*** rather than general-purpose defenses.
- **G6.** ***Report*** mitigation decisions and their downstream effects on data.
- **G10.** Consider the ***ethical concerns*** of fraud mitigation strategies.

Key takeaways

1. Survey fraud is a shared socio-technical problem, not an isolated methodological issue.

Key takeaways

1. Survey fraud is a shared socio-technical problem, not an isolated methodological issue.
2. Current defenses are often reactive and mismatched to evolving threats.

Key takeaways

1. Survey fraud is a shared socio-technical problem, not an isolated methodological issue.
2. Current defenses are often reactive and mismatched to evolving threats.
3. Prevention requires lifecycle-aware defenses, shared responsibility, and transparent reporting.

Key takeaways

1. Survey fraud is a shared socio-technical problem, not an isolated methodological issue.
2. Current defenses are often reactive and mismatched to evolving threats.
3. Prevention requires lifecycle-aware defenses, shared responsibility, and transparent reporting.

The usable security community can help by applying **threat modeling, lifecycle thinking, and usability principles** to design online research systems that protect both **research integrity** and **participants**.

Acknowledgments

HIRE HER!



Shiza Ali
Postdoctoral Associate | GWU

ON THE JOB MARKET

shiza.ali@gwu.edu
linkedin.com/in/theshizaali
shizaali.com



Wellington Barbosa
GWU



Matthias Fassl
GWU



Aditi Ganapathi
ASU



Jaron Mink
ASU



Adam Aviv
GWU

Full Paper