Evidence-based Approach to Responsible AI

Microsoft Research Summit 2022
Responsible AI Workshop: An Interdisciplinary Approach
October 24, 2022

Jonathan Zhu 祝建華

Centre for Communication Research
CityU 香港城市大學
City University of Hong Kong

CityU School of Data Science
香港城市大學
City University of Hong Kong

Web Mining Lab
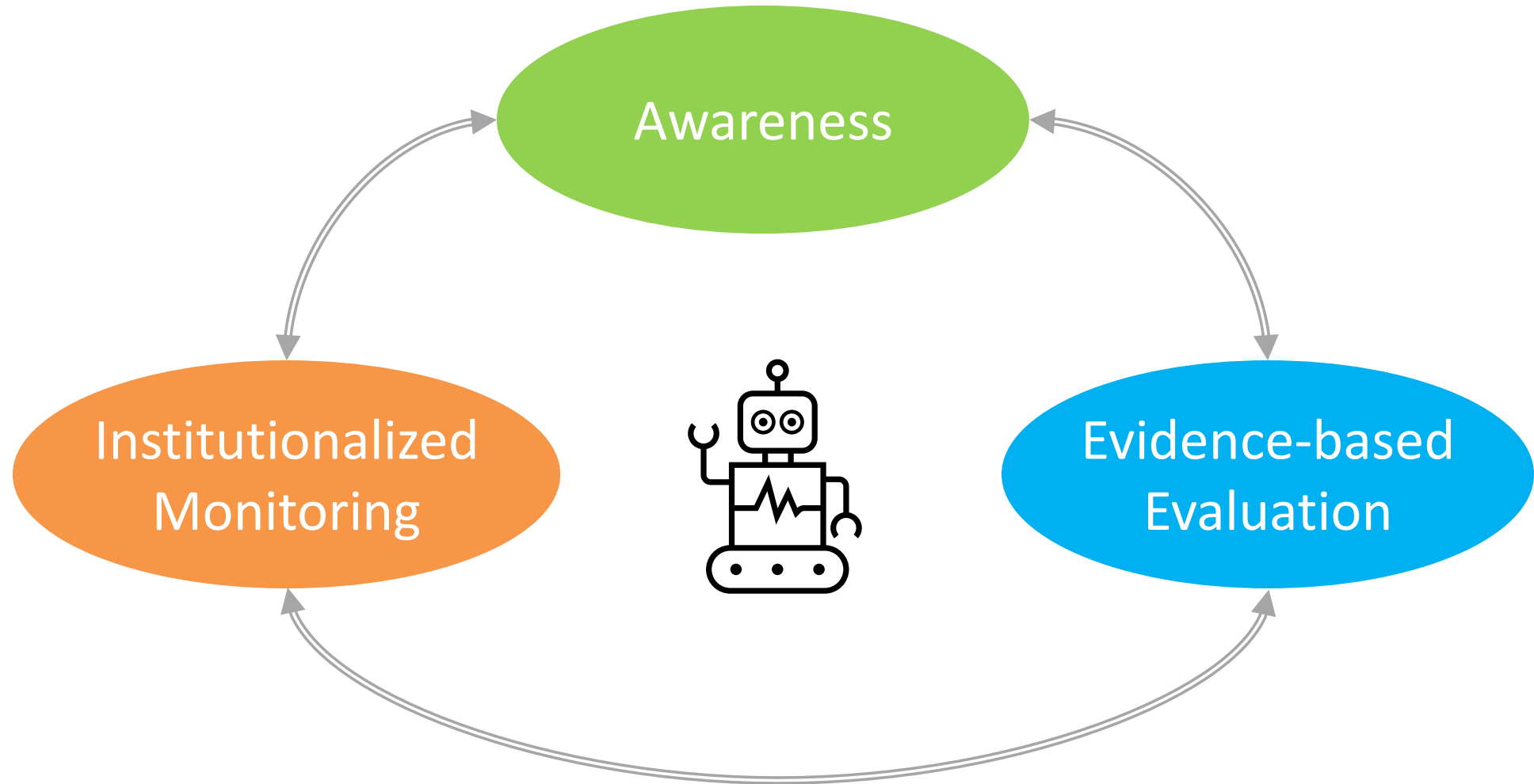互聯網挖掘實驗室

Four Ws in Responsible AI

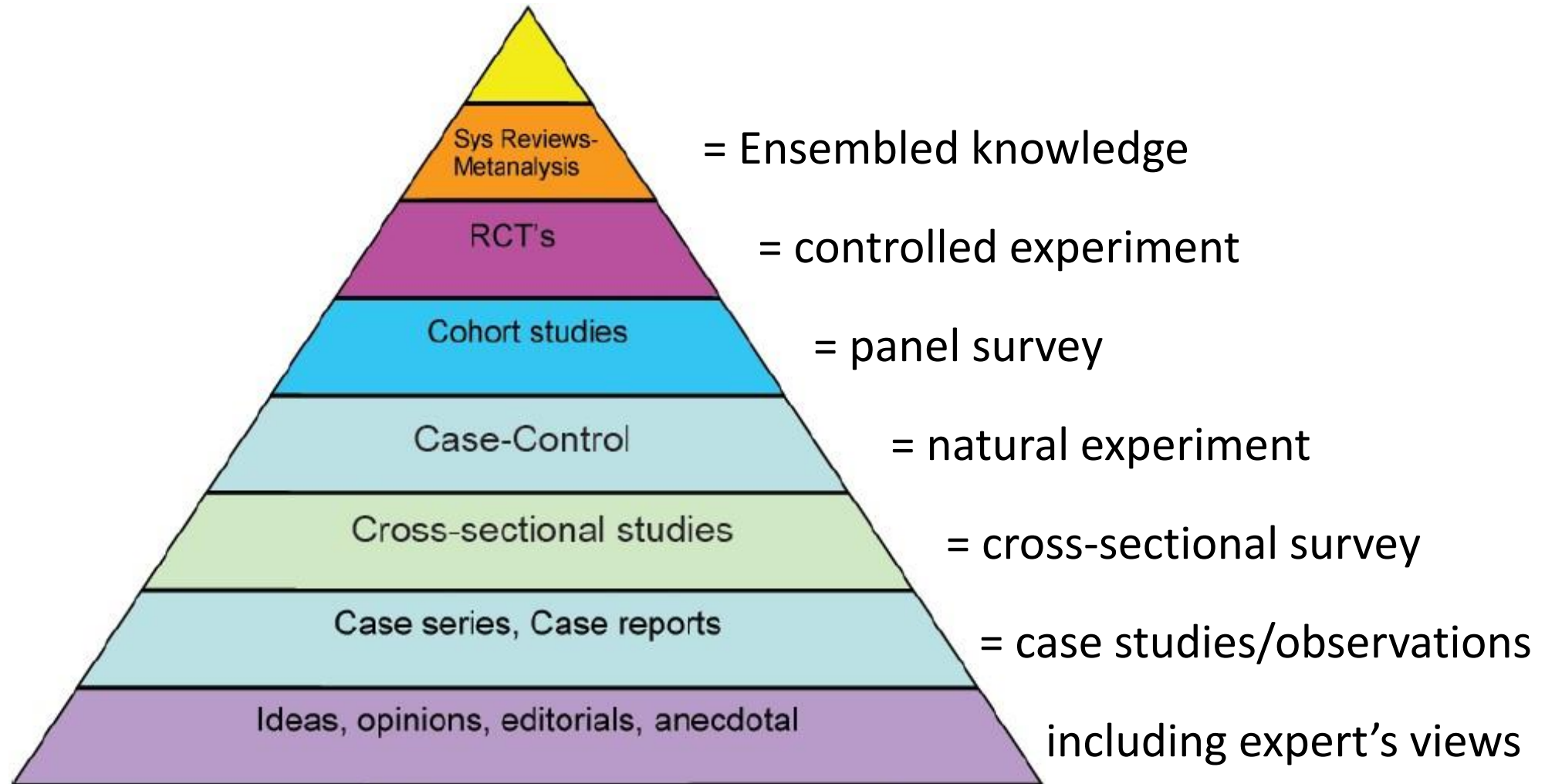WHAT is responsible AI?

WHY is responsible AI necessary?

HOW to produce responsible AI?

HOW to know the AI is responsible?

# Three-leg Approach for AI to be Responsible (Post-launch)

Evidence-based Evaluation

Sys Reviews-Metanalysis = Ensembled knowledge

RCT's = controlled experiment

Cohort studies = panel survey

Case-Control = natural experiment

Cross-sectional studies = cross-sectional survey

Case series, Case reports = case studies/observations

Ideas, opinions, editorials, anecdotal including expert's views

Web Mining Lab
互聯網挖掘實驗室

4

Evidence-based Production vs. Evidence-based Evaluation

Evidence-based Production

| Training | → | Test | → | Production | → | Deployed |

Evidence-based Evaluation
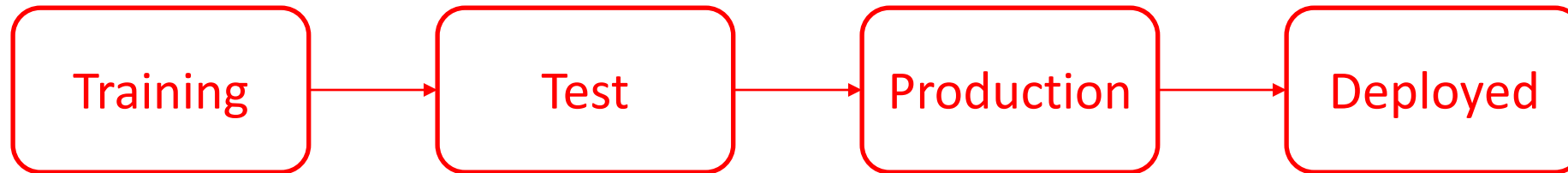
| Training | → | Test | → | Production | → | Deployed | → | **Evaluation** |

1. by independent party
2. based on independent data
3. to evaluate side effects

Web Mining Lab
互聯網挖掘實驗室

Test Effectiveness vs. Side Effects of Clinical Trials

I. Side effects

II. Effectiveness (+ side effects)

III. Effectiveness (+ side effects)

IV. Effectiveness + side effects
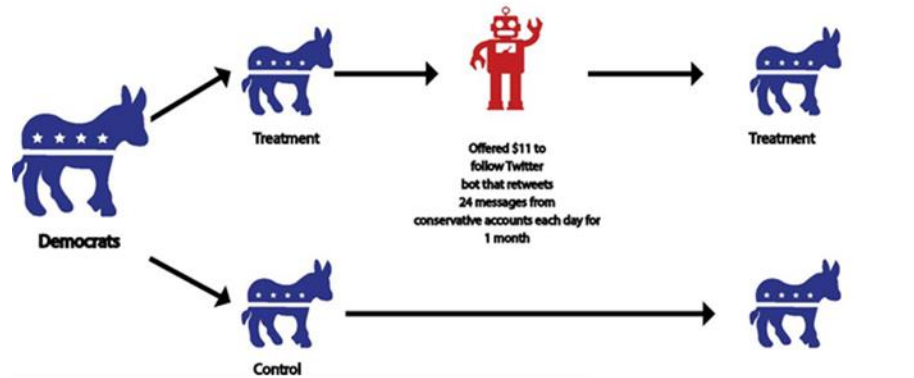
Web Mining Lab
互聯網挖掘實驗室
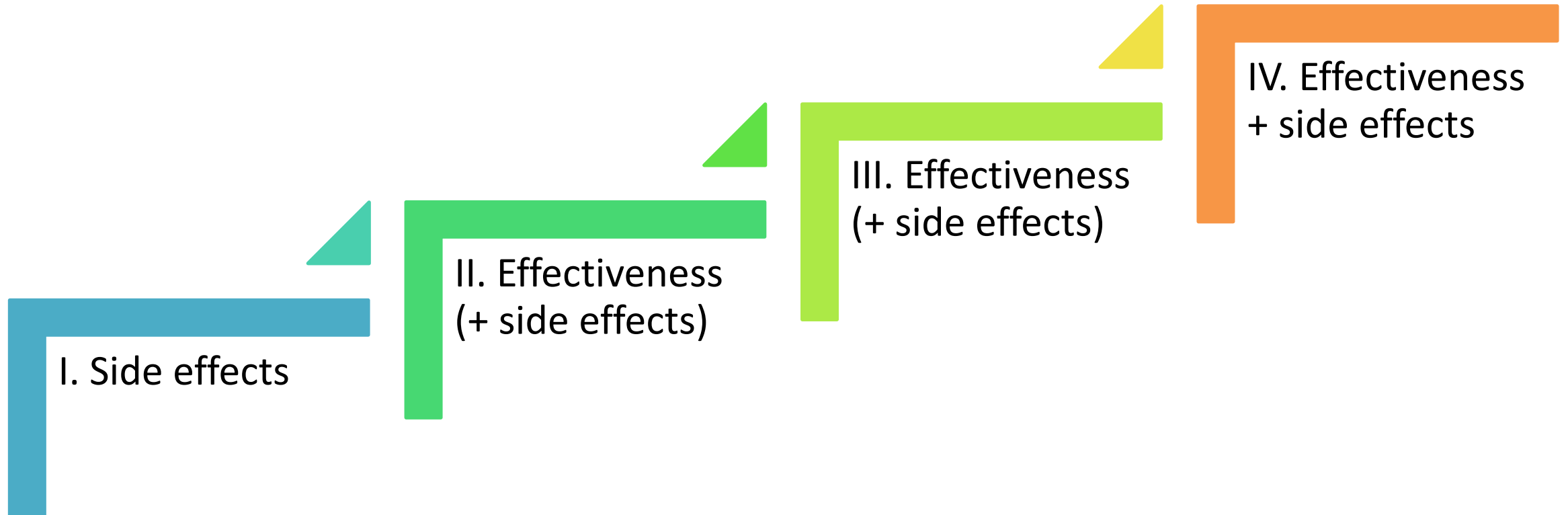
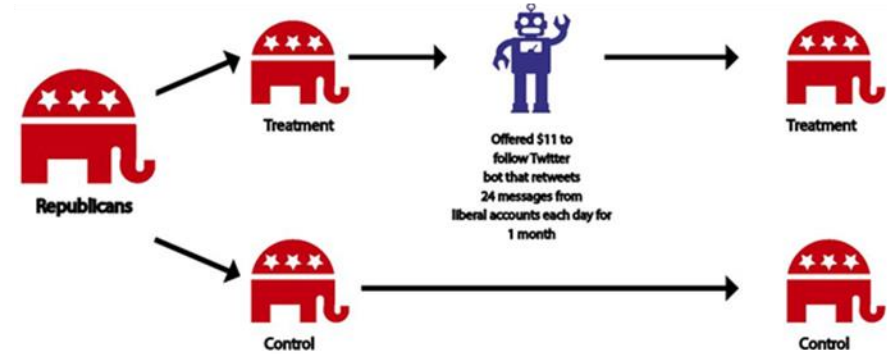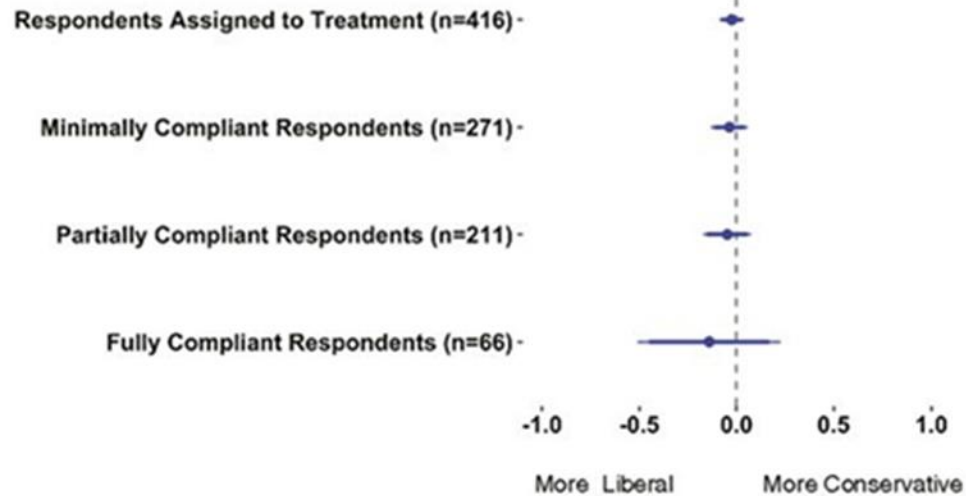# Evidence-based Evaluation Isn't Easy, Quick, or Cheap

- Evidence-based evaluation is required to have both internal and external validity
  - Internal validity: the AI product truly causes the side effects in "clinical" trial;
  - External validity: the causal relationship holds in real life.
- Causal explanation research requires
  - multiple/independent data (to test generalizability)
  - longitudinal data (to test long-term side effects)

Web Mining Lab
互聯網挖掘實驗室

Does Cross-cutting Exposure Reduce Polarization?

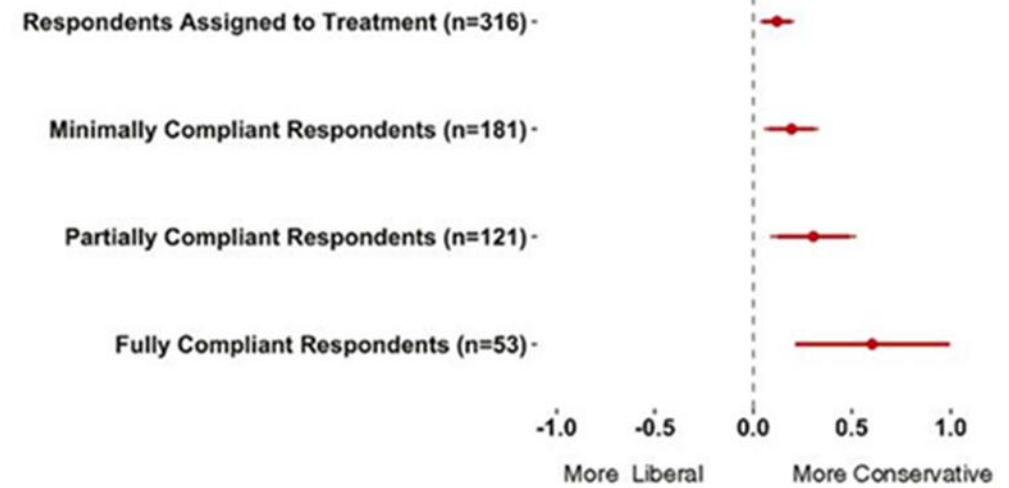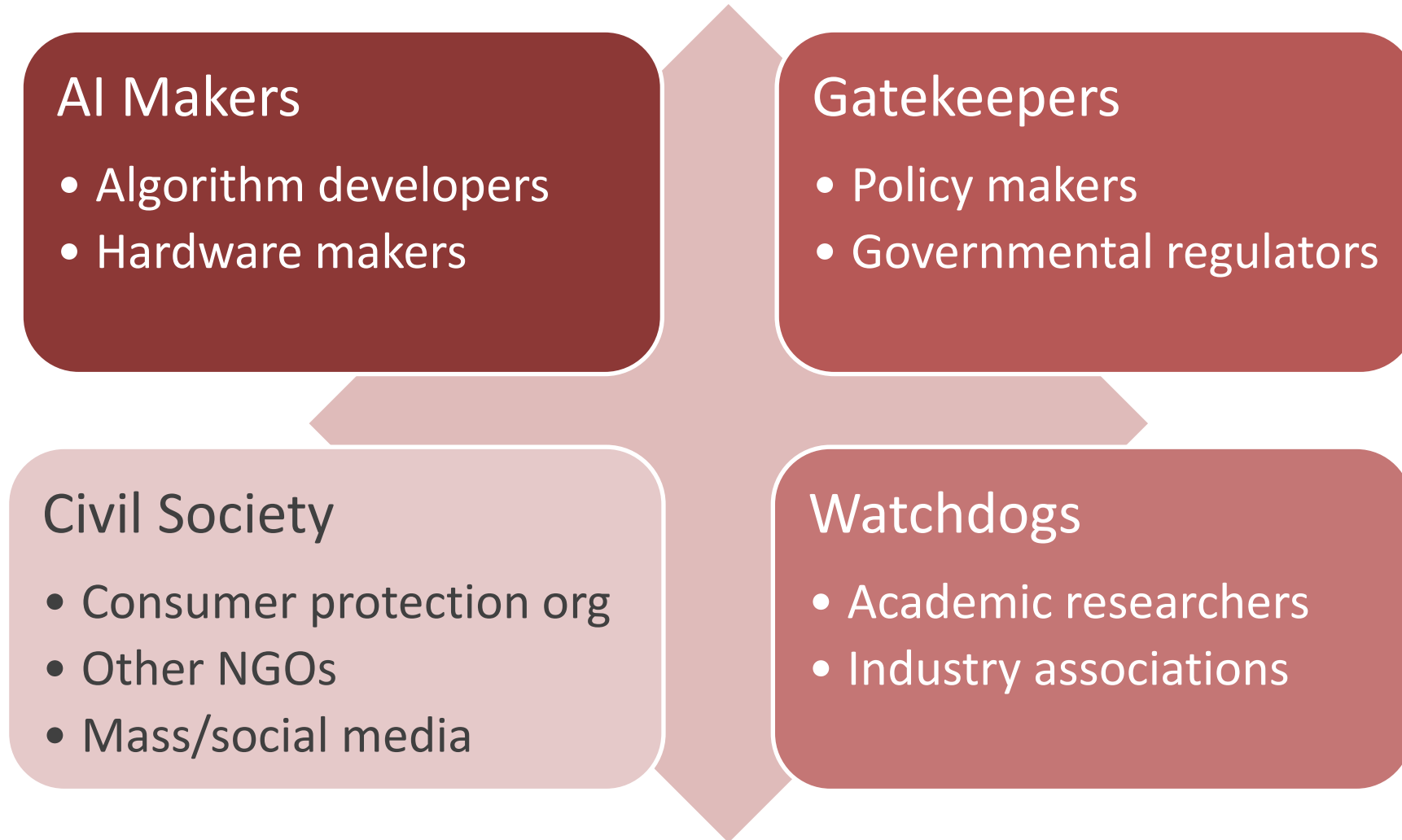Bail et al. (2018). Exposure to opposing views on social media.

Democrats

Republicans

Institutionalized Monitoring System

AI Makers
- Algorithm developers
- Hardware makers

Gatekeepers
- Policy makers
- Governmental regulators

Civil Society
- Consumer protection org
- Other NGOs
- Mass/social media

Watchdogs
- Academic researchers
- Industry associations

Web Mining Lab
互聯網挖掘實驗室

Funding Models for Independent Watchdogs

- Collaboration with AI makers

- Sponsorship from AI makers

- Funding from government/public agencies

- Funding from private foundations

- Self-financed with/without collaboration with AI makers

- etc.

The Role of the Government

- Governmental regulation is indispensable

- Governmental intervention should be kept at a modest rate

- An optimal institutionalized monitoring system is bottom-up, instead of top-down, with the government being reactive, rather than proactive or over-reactive

FDA Model vs. CPA Model

- Food Drag Administration (FDA)
  - Gatekeeping
  - Critical review of producer's data
  - Funded by involuntary taxpayers

- Consumer Protection Agency (CPA)
  - Quality monitoring
  - Assessment of independent data
  - Funded by voluntary consumers

Web Mining Lab
互聯網挖掘實驗室
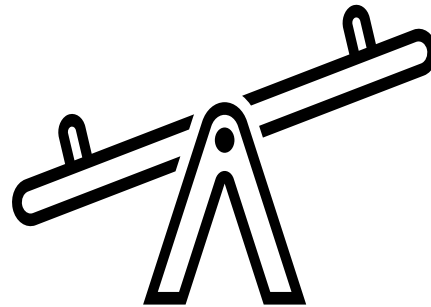
Takeaways

- Responsible AI is necessary, expensive, and ultimately beneficial for all stakeholders.

- Responsible AI requires attention, resources, and sustainable monitoring from AI makers, regulatory gatekeepers, academic/professional watchdogs, and NGOs.

- After all, evidence-based evaluation should serve the gold standard for entire cycle life of responsible AI.

Web Mining Lab
互聯網挖掘實驗室

THANK YOU

j.zhu@cityu.edu.hk

https://ccr.cityu.edu.hk

https:/weblab.com.cityu.edu.hk

Web Mining Lab
互聯網挖掘實驗室