

# Human-Centered Policy Frameworks for Autonomous Agent Technologies

By: Ramani Satsikumar<sup>1</sup>, Anastasia Kouvaras Ostrowski<sup>1</sup>, Cynthia Breazeal<sup>2</sup>

<sup>1</sup>Purdue University, <sup>2</sup>MIT Media Lab

## Introduction

- As autonomous agent (AA) technologies like robots, chatbots, and drones integrate into society, their design and deployment must prioritize human well-being and equity.
- These technologies offer great potential but pose social, ethical, and policy challenges [2].
- This research explores policies experts from academia, industry, and policy-making see as essential for human-centered AA integration.

### Research Goals:

- Identify key policy concerns in AA technology integration
- Explore the social and ethical implications of AA technology integration
- Provide insights for creating human-centered AA policies

## Methodology

Data was collected through two types of participatory design workshops and themes were identified using a thematic analysis approach [1]. Participants included academics, industry members, and people in the policy-making space.

**Equitable Design Workshop:** Focused on supporting equitable design in robot technology development processes.

- 5 participatory design workshops (44 participants) → 2 academic workshops, 2 industry workshops, 1 policy workshop

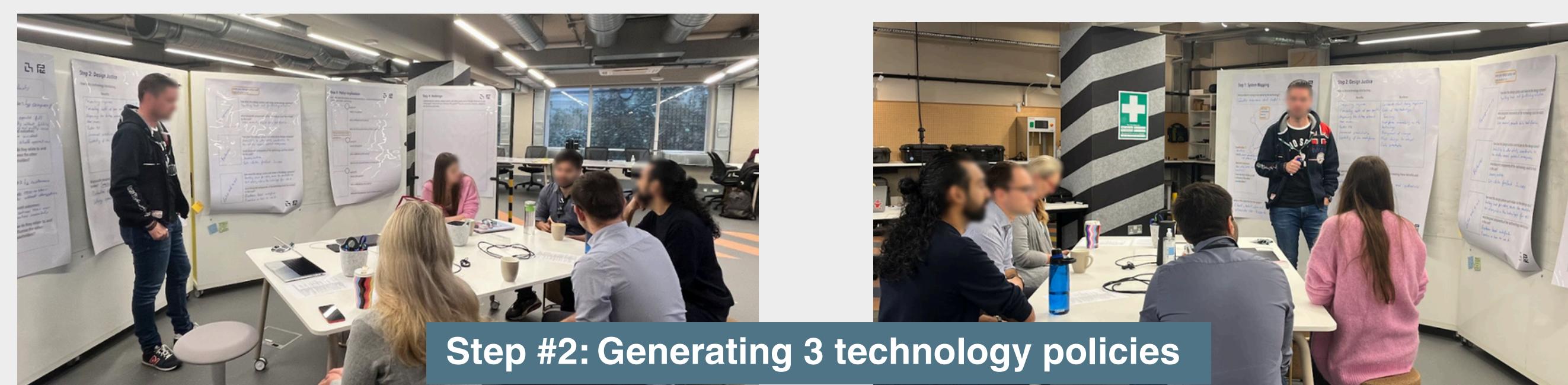
**Autonomous Agent Design Workshop:** Empowered employees involved in AA technology development.

- 3 participatory design workshops (28 participants) → 1 robot, 1 chatbot, 1 drone

**Data Collected:** Transcripts from discussions and written outputs



Step #1: Identifying benefits + harms of technology



Step #2: Generating 3 technology policies

## Identified Themes

### Data Sharing

"I think there should be some policy on restricting law enforcement ability to use individual's data" - PW1

### Transparency

"Companies should be restricted on what they can reveal about users to advertisers and let users know what data is collected and how it is used." - AW1

### Sensitive Data Access

"Data should be accessed on a need-to-know basis (medical data etc.)" - RW

### Data Storage

"The cloud becomes a big problem...what kind of data can be transmitted?" - PW1

### Chain of Accountability

"There should be a clear accountability framework for all robots and who's responsible along all stages of the AA development to deployment process" - RW

### Admin Overhead

"Clearly define the who operates and controls the robot...operating procedure, maintenance etc." - IWI workshop

### Trust

"I think we need to decide who is the endpoint in the trust, like with accountability... human or robot?" - CW

### Safety

"Duty of the company to ensure robot operates safely..." - RW

### Upskilling

"Since in the long-term automation is supposed to save us money...some of that money can be allocated back toward retraining people who are impacted by [job replacement]" - DW

### Explainability

"AA should be able to be interrogated... If something goes wrong, we can trace back why it made a certain decision." - DW

### Themes

### Ownership

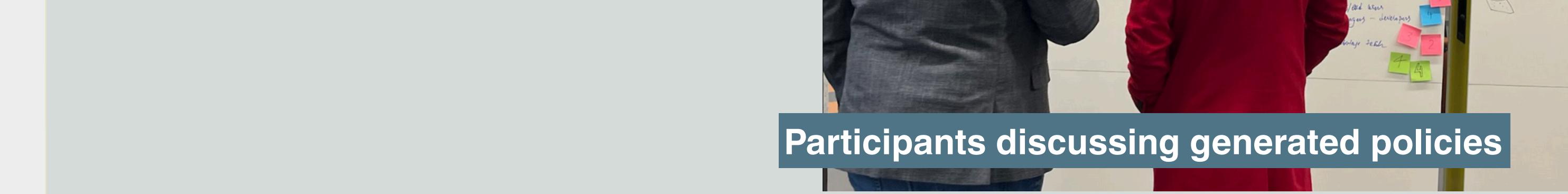
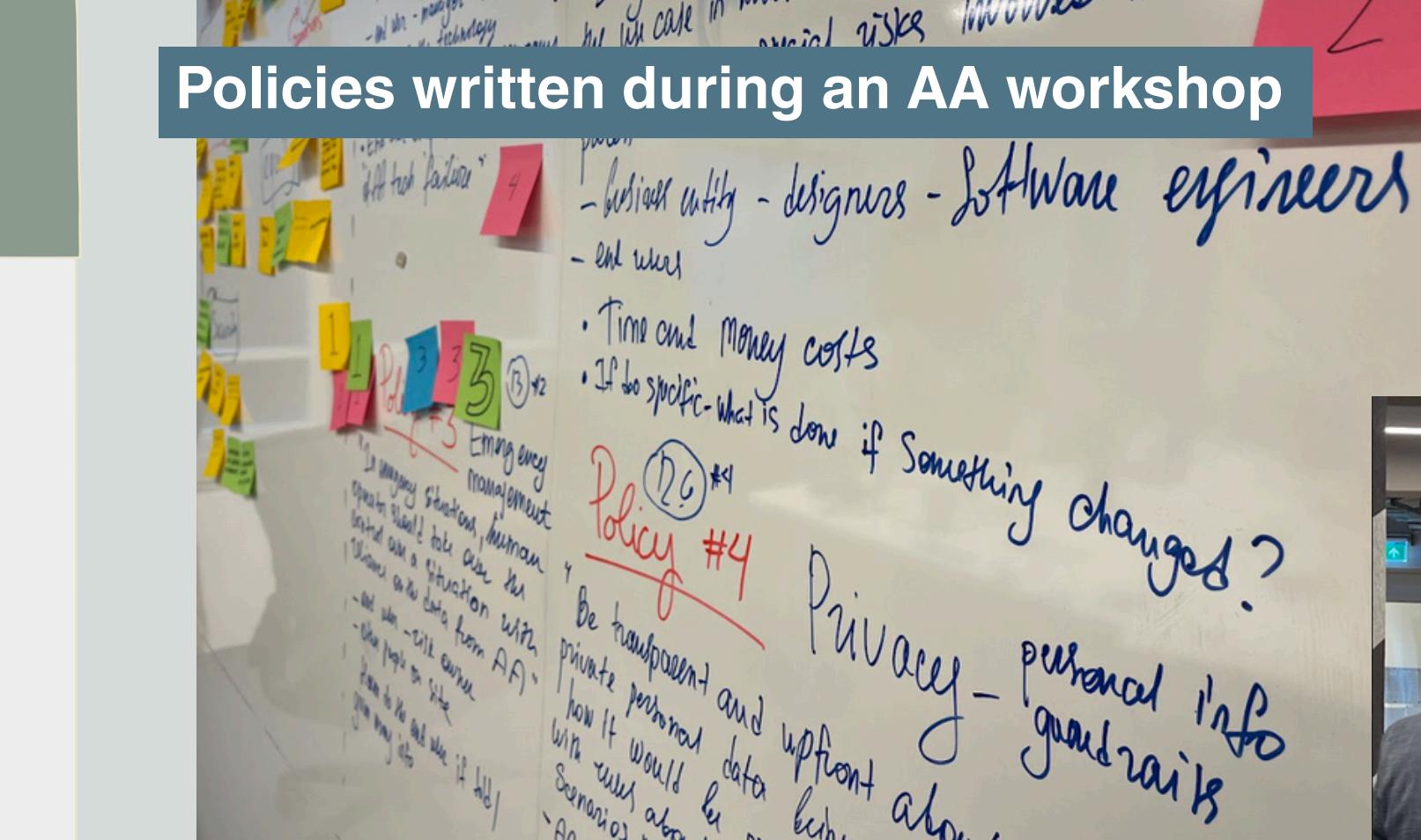
### Power Dynamics

"There is a power relationship between companies and employees....[companies] need to inform everything on a regular basis." - PW1

### User Rights

### Bias

"There must be testing requirements to demonstrate that technology is not biased against certain groups." - PW1



Participants discussing generated policies

## Conclusion

- This research showcases the social, ethical, and policy challenges of integrating AA technologies considering policy design, focusing on privacy, ownership, and user rights policy dimensions.
- Actionable policies prioritizing transparency, accountability, and user well-being as emphasized by participants can help ensure AA technologies have a positive societal impact.

## Policy Design Implications

- Translate data privacy concerns into actionable policy guidelines and establish clear frameworks for data ownership and control.
- Strengthen user rights through transparency, consent, and involvement of multiple stakeholders in AA technology integration.
- Develop policies that prioritize equity, human well-being, and ethical considerations in AA technology deployment.

## Acknowledgements

Thank you to the participants who collaborated with us in this work. This work was supported by a Tech Justice Lab's Faculty Mentor Award, MAC3 Society and Ethics in Computing Research Award, Amazon Robotics Fellowship, BP Design, and BP Digital Science.

## References

- [1] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.
- [2] M. A. Goodrich and D. R. Olsen, "Seven principles of efficient human robot interaction," SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483), Washington, DC, USA, 2003, pp. 3942-3948 vol.4, doi: 10.1109/ICSMC.2003.1244504.