

Guiding Questions and Overarching Principles for Conducting Socially Responsible and Ethical Research to Counter Mis/Dis/Malinformation

TEN GUIDING QUESTIONS

- 1 Does your organization have institutional guidance, de facto norms, or legal requirements for human subject research?
- 2 Can you distinguish between rule-based and principles-based approaches to scientific experimentation?
- 3 How does the ethics and values mindset of your research team, organization or sponsor differ from your own?
- 4 What external factors could come into conflict with the ethics of your experiment?
- 5 How do you and your organization view data ownership; what privacy rights does an individual have once they've submitted their data and who is responsible for enforcing/guaranteeing those rights?
- 6 How comfortable are you in using PAI datasets even if the acquisition of the set was not acquired in ethically normative and acceptable ways? (Note: this includes secondary use)
- 7 Have you considered how your experiment may be perceived externally? What are the implicit and explicit risks to your organization, "tribe", and yourself personally?
- 8 How can your results be misused against (or, groups or individuals?
- 9 Can you articulate what constitutes "informed consent" for your experiment?
- 10 What actions, above and beyond IRB approval, have you implemented to ensure the safety and wellbeing of your participants and audience?

TEN OVERARCHING PRINCIPLES

...Plus one

- 1 Will not use unethical means to combat disinformation
- 2 Will not intentionally promote harmful stereotypes
- 3 Will adhere to institutional core values and IRB guidelines
- 4 Will exercise care when using public information which could characterize or identify individuals
- 5 Will maintain diligence in identifying and mitigating bias in algorithms and implementation
- 6 Will perform risk analysis to understand unintended consequences; including how the capabilities we develop could be weaponized or misused
- 7 Will not violate EULAs or User Agreements
- 8 Will conduct R&D with neutrality and without bias, and strive to remain apolitical
- 9 Will provide means for transparency and accountability
- 10 Will ensure that stakeholders, including researchers, are aware of their personal risks
- 11 Plus One: Will take reasonable precautions to prevent harm (as listed above) and, in the event that inadvertent harm occurs, will own and learn from it



Professor Andrew Maynard

Director of the Risk Innovation Lab at the School for the Future of Innovation in Society (SFIS) at Arizona State University (ASU)



amaynar2@asu.edu



Amna Greaves

Assistant Group Leader

AI Software Architectures and Algorithms



amna@ll.mit.edu

Risk Innovation is an approach to understanding Ethics and Social Responsibility for Counter MDM R&D, it is a way of approaching risk that leads to new knowledge, understanding, and capabilities, and translates these into approaches, tools, or practices that protect and grow societal, technical, and other value



**RISK INNOVATION
NEXUS**



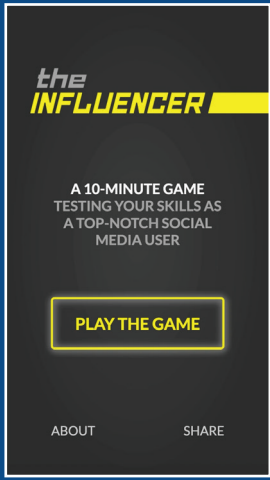
MITLL Rapid-Play Serious Game Practices for Capturing Social/Behavioral Data

Data Quality	Engaging <ul style="list-style-type: none"> • Players volunteer to play through many games • Players are invested in the outcome of each scenario 	Automated Data Collection <ul style="list-style-type: none"> • Games record every action taken by player • Actions are fused with game states to provide rich data set 	Experimental Design <ul style="list-style-type: none"> • Carefully chosen variable and scenario design • Data supports metrics baselining and hypothesis testing 	Model-Based <ul style="list-style-type: none"> • Domain-specific models • Required accuracy dictated by use cases
Data Quantity	Short and Iterative <ul style="list-style-type: none"> • Games executed from minutes to hours • Allows for multiple control variable per gaming session 	Accessible <ul style="list-style-type: none"> • No special equipment • Enables remote play 	Flexible <ul style="list-style-type: none"> • Easy to generate scenarios • Easy to modify objectives 	Rewind & Replay <ul style="list-style-type: none"> • Ability to save and replay from known states • Easy to 'what if'

The Influencer: A Serious Game to Emulate Misinformation

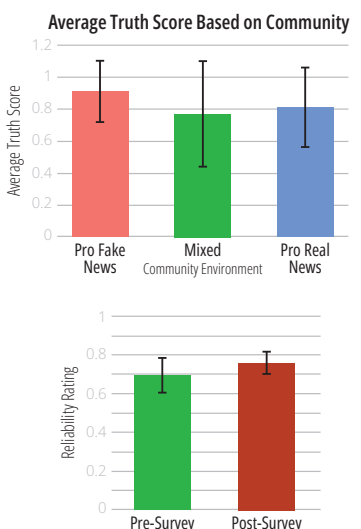
Create a controlled social media environment to run experiments that measure misinformation-related decision-making behavior

- What influencers people to post true or false information online?
- Will encouragement or discouragement from others in a group affect what individuals choose to post?
- Can bots help reduce the spread of misinformation?



Backfire Effect: Exposure to contrary views can increase polarization

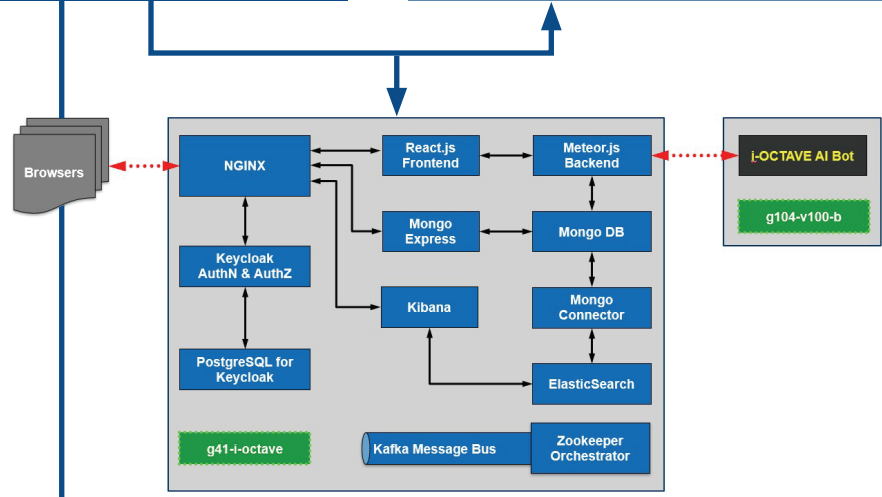
Sensitivity to MDM increases after gameplay



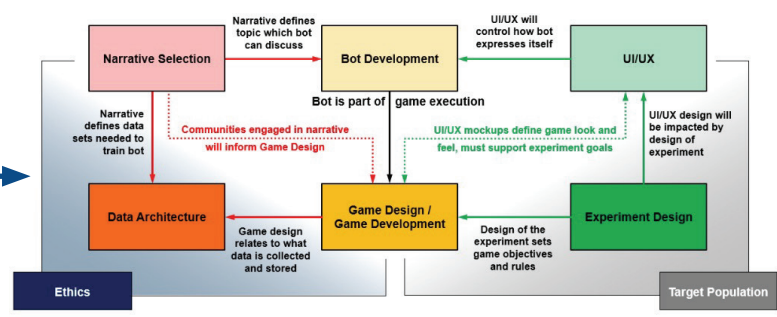
Community Environment	Average Truth Score
Pro Fake News	~0.9
Mixed	~0.75
Pro Real News	~0.8

Survey Stage	Reliability Rating
Pre-Survey	~0.7
Post-Survey	~0.75

Game Architecture



Functional Block Diagram



Design and demonstrate a reusable framework which can observe / measure the impact of bots within an online community