Is "AI" Responsible for Its Mistakes?
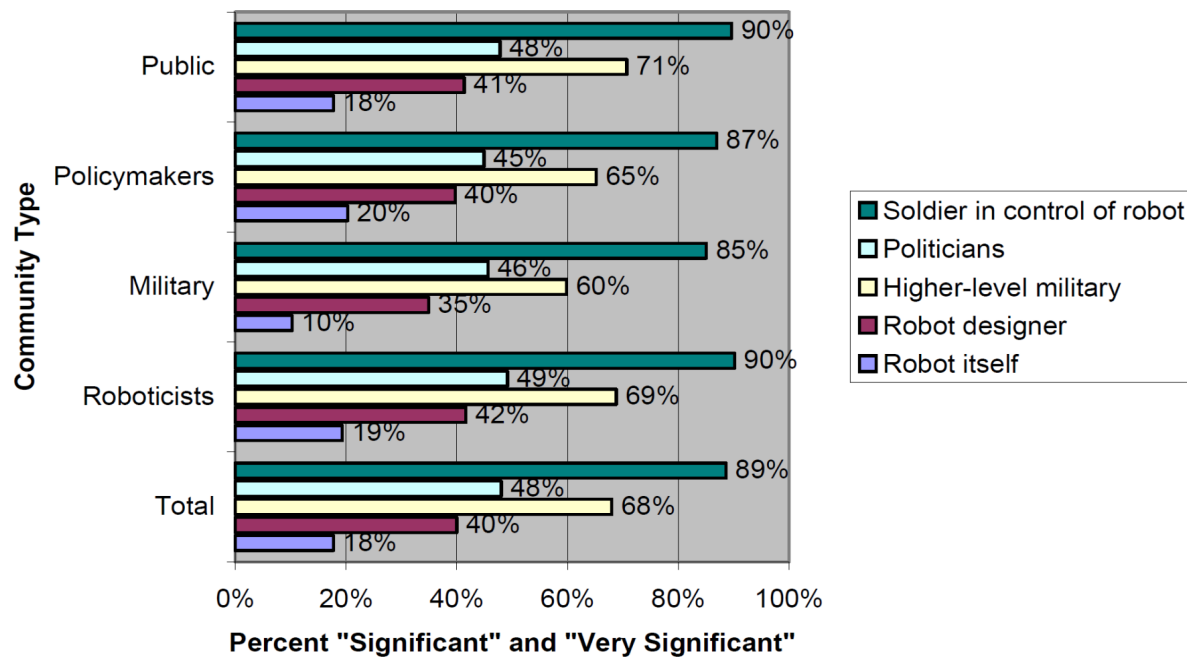
Lightning Talk at AAAI Bridge: AI and Law

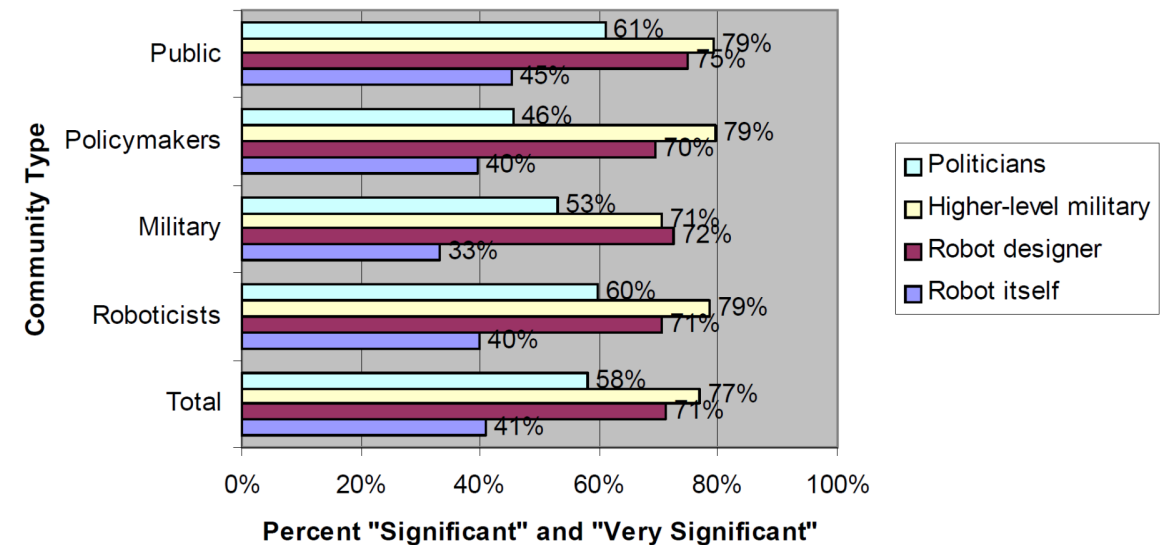Dr. Lilia Moshkina

February 8, 2023

Public Opinion 15+ Years ago

If, hypothetically, a robot makes a lethal mistake, many parties, including ROBOTS are found responsible

Responsibility for Lethal Mistakes of Robot as Extension

Responsibility for Lethal Mistakes of Autonomous Robot

Robot acting as an extension of human soldier: robot under direct authority of a human, including authority over use of lethal force

Autonomous robot: a robot that does not require human involvement except for high-level mission tasking; makes its own decisions, including lethal force

Large-scale survey on public opinion on lethal autonomous systems*

* Moshkina, L. & Arkin, R. 2007. Lethality and Autonomous Systems: Survey Design and Results, TR GIT-GVU-07-16, GaTech

In 2020, DoD adopts 5 principles of Ethical AI

1. **Responsible**. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.

2. **Equitable**. The department will take deliberate steps to minimize unintended bias in AI capabilities.

3. **Traceable**. The department's AI capabilities will be developed and deployed so that staffers have an appropriate understanding of the technology, development processes, and operational methods that apply to AI. This includes transparent and auditable methodologies, data sources, and design procedure and documentation.

4. **Reliable**. The department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing.

5. **Governable**. The department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

Address **"is AI to blame"** issue

Ambiguity Remains

- Public Opinion seems to be at odds with DoD guidelines:

    - Public is willing to hold AI responsible for its actions, especially as its scope and autonomy increase

    - DoD holds the person in charge (be it user, developer or stakeholder) responsible for AI decisions and actions, provided that the AI in question is governable (can be stopped) and traceable throughout

    - And what about Tesla?

- It is time for law to speak up:

    - Who is accountable if a deployed AI system makes errors?

    - How does the responsibility/accountability depend on the severity of errors (e.g., death of many vs not getting a job)?

    - How can the responsibility/accountability be proven in a court of law?

A Hypothetical Future Scenario: Active Shooter

- A system with many AI components detects what it **perceives** as a *sniper in the window* during an active shooter incident

- A SWAT team is notified and is dispatched to eliminate the threat

- As the doors are broken down, the system detects a previously known shooter with a weapon in a different location, who then continues to proceed on a rampage – there was no sniper in the window

Where were the mistakes made, and who OR what was responsible?

New Info: Legitimate Business

Known Shooter