AI IN ASIA 2016

Safety of AI, or AI for Safety?

December 16, 2016

Sung-Bae Cho
Dept. of Computer Science, Yonsei University
http://sclab.yonsei.ac.kr

Fiction vs Reality

- Fantastic predictions for AI in the popular press

- Reality
  - No cause for concern that AI is an imminent threat to humankind
  - No machines with self-sustaining long-term goals and intent have been developed, nor are they likely to be developed in the near future
  - Increasingly useful applications of AI, with potentially profound positive impacts on our society and economy are likely to emerge

What is AI?

- Study or technology to understand the substance of human intelligence, and realize it artificially

- Strong AI
  - Study or technology to implement human intelligence
  - Technology to make machine think like human
    - Creativity / thought / emotion

- Weak AI
  - Study or technology to solve a particular problem by imitating human intelligence
  - Technology to solve a specific problem like human
    - Large-scale data processing tirelessly and unbiasedly

Difficulty and Approaches of AI

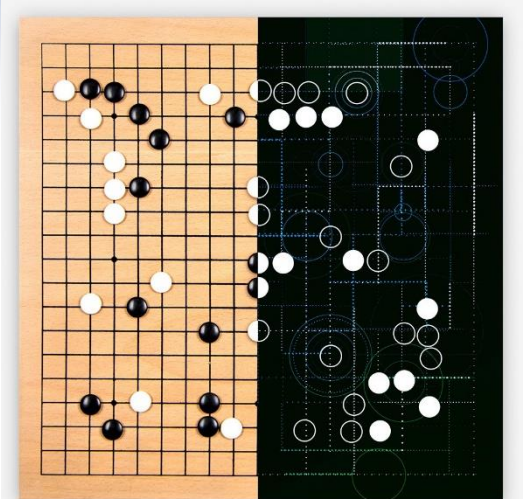
- Difficulty
  - "What" is clear, but "How"?

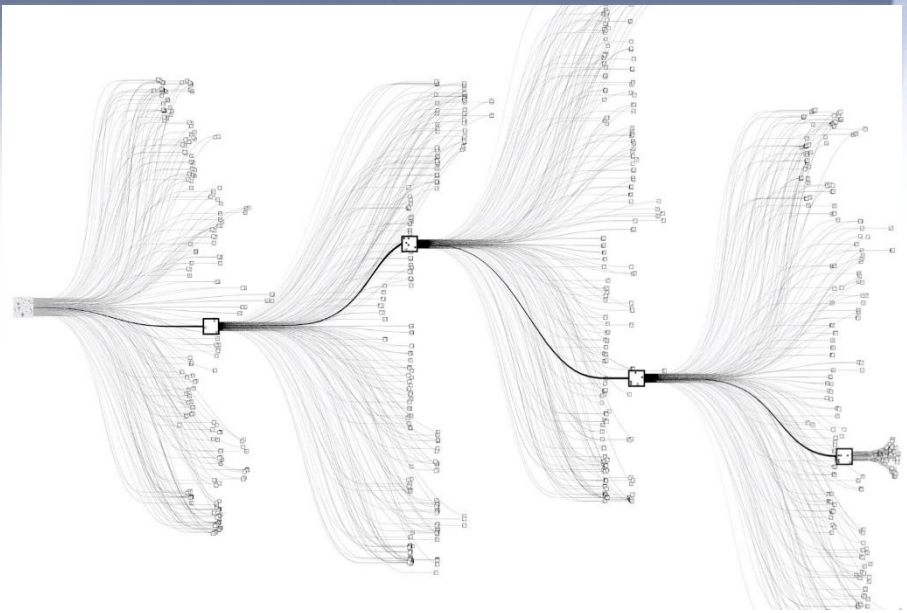
- History of AI technology: Continuous fade in and out of new technology for 60+ years since the invention of computer
  - Logic, optimization theory, probabilistic model, search theory, knowledge-based systems, expert systems, fuzzy logic, neural networks, genetic algorithm, chaos theory, ...

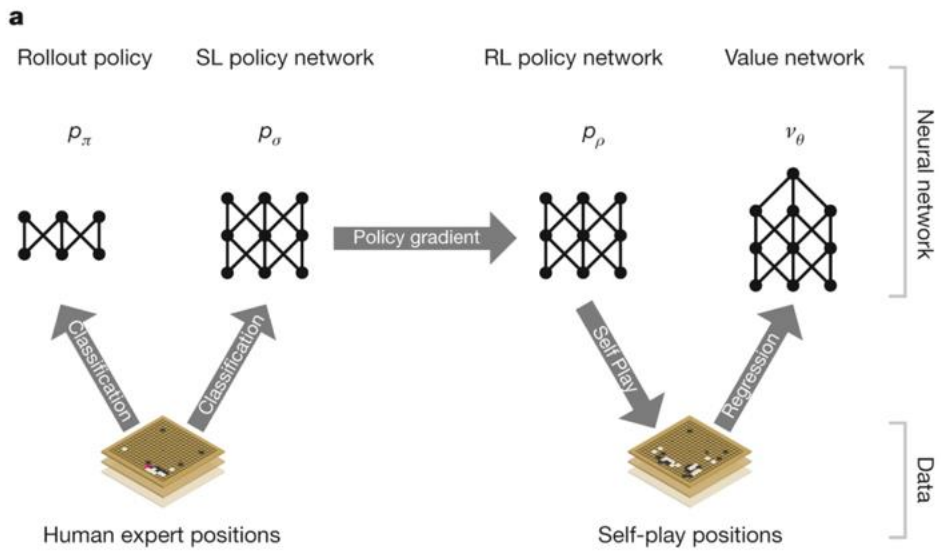
- Two approaches to developing AI
  - Knowledge-based AI: Decision making with stored knowledge
  - Data-driven AI: Decision making with knowledge extracted from data

Knowledge-based AI: IBM Watson

Data-driven AI: Google AlphaGo







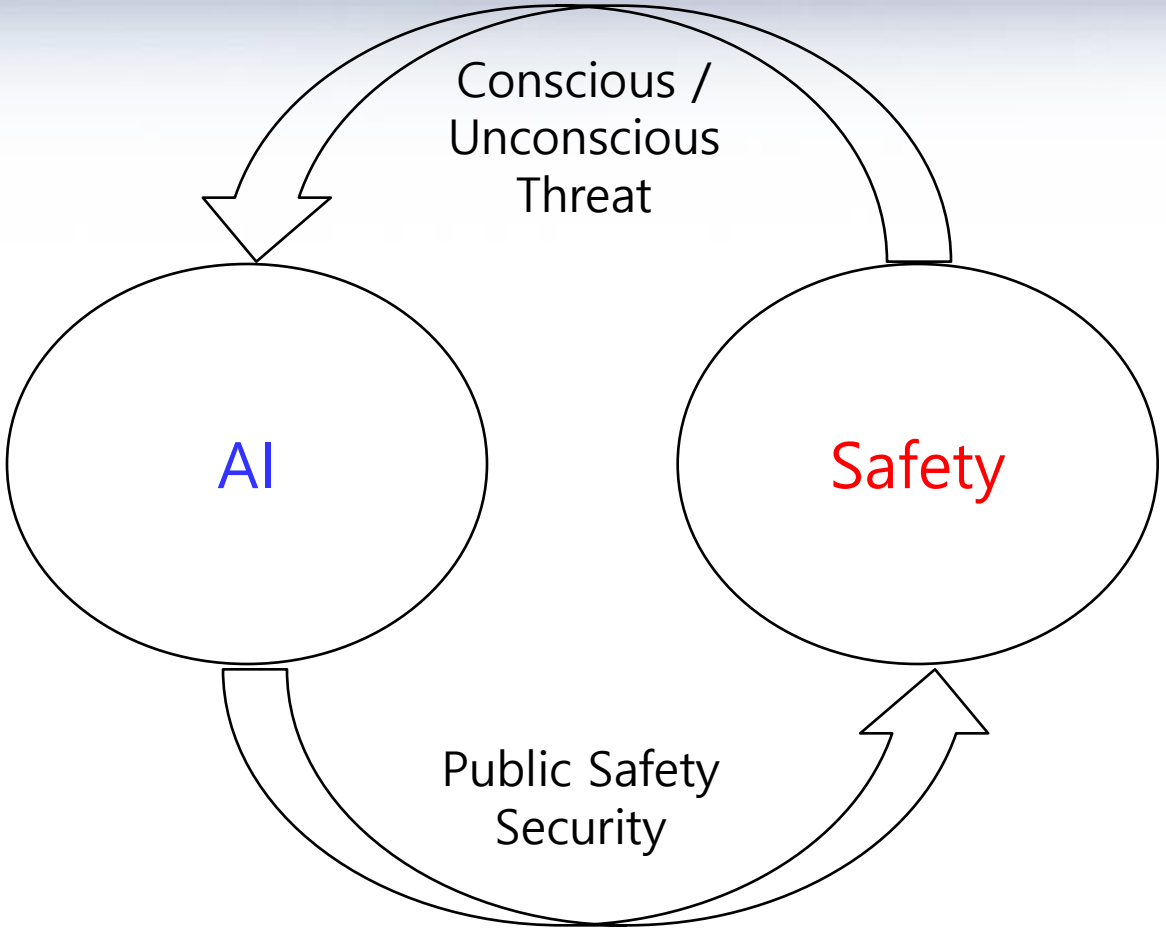
Challenges and Solutions of AI

- Challenges
  - Various techniques are independently developed in various areas
  - Serious overestimation and misuse / abuse of techniques due to name confusion

- Solution 1: Integrative Intelligence Technology
  - Require the cooperation between high-level intelligence of conventional AI and low-level intelligence of various approaches
  - Connectionism and symbolism
  - Artificial life and AI
  - Reactive (behaviorism) and deliberative

- Solution 2: AI Architecture for Problem Solving
  - Require the collaboration between symbolic representation and connectionist representation
  - Need of consideration on social ability, emotion, sensibility, etc.

AI and Safety



Safety of AI

- Conscious threat
  - Criminal threat
    - killer drones
    - LAWS (lethal autonomous weapons systems)
  - Infringement of privacy and personal information
  - Need careful consideration of AI safety for critical applications

- Unconscious threat
  - Bug / malfunction of AI system
    - Malfunction of AI system for subway and airplanes might cause mass casualties
  - Excessive dependence on AI
    - Need to control the dependence on AI
    - Need new ethics, laws, regulations, technologies, etc.

→ The safety problem is caused not by AI but by people!

AI for Safety

- Prospects
  - Cameras, drones and software to analyze crime patterns should use AI in ways that reduce human bias and enhance safety without loss of liberty or dignity
- Risks
  - AI may become overbearing or pervasive in some contexts
- Benefits
  - AI may enable policing to become more targeted and used only when needed
  - AI may also help remove some of the bias inherent in human decision-making

AI for Public Safety and Security

- One of the more successful uses of AI analytics is in detecting white collar crime, such as credit card fraud
- AI tools may also prove useful in helping police manage crime scenes or search and rescue events by helping commanders prioritize tasks and allocate resources
- AI will better assist crime prevention and prosecution through greater accuracy of event classification and efficient automatic processing of video to detect anomalies
- Machine learning significantly enhances the ability to predict where and when crimes are more likely to happen and who may commit them

AI Policy, Now and in the Future

- Some existing regulatory regimes for software safety (for example, the FDA's regulation of high consequence medical software) require specific software engineering practices at the developer level

- Modern software systems are often assembled from library components which may be supplied by multiple vendors, and are relatively application-independent

- It doesn't seem desirable to subject all such developers to the standards required for the most critical, rare applications. Nor does it seem advisable to allow unregulated use of such components in safety critical applications

- Tradeoffs between promoting innovation and regulating for safety are difficult ones, both conceptually and in practice. At a minimum, regulatory entities will require greater expertise going forward in order to understand the implications of standards and measures put in place by researchers, government, and industry

Concluding Remarks

- If society approaches AI technologies primarily with fear and suspicion, missteps that slow AI's development or drive it underground will result, impeding important work on ensuring the safety and reliability of AI technologies

- If society approaches AI with a more open mind, the technologies emerging from the field could profoundly transform society for the better in the coming decades

Q & A