Future of Decentralization, AI, and Computing Summit: Cooperative AI with Decentralized Trust

Gordon Liao*

Circle Internet Financial

August 27, 2023

The views expressed in this presentation are those of the author and not those of affiliated organizations.

Why is cooperative AI research important?

- Most AI research and development so far have focused on a single-agent setting or adversarial setting in multi-agent context (e.g. zero-sum games)
- Cooperation is often needed to achieve pareto optimal in games, e.g. prisoner's dilemma

	Cooperate	Defect
Cooperate	2, 2	-1, 3
Defect	3, -1	0, 0

- Real-world examples of AI:
  - self-driving cars interacting with other autonomous or human drivers
  - algo pricing models reacting to other merchants' pricing
  - trading bots that generate pecuniary externalities, e.g. flash crashes and liquidation cycles

What are the main challenges to cooperative AI?

- Cooperative AIs face typical challenges in contract theory:
  - **Information asymmetry** in the form of *adverse selection* (hidden information) and *moral hazard* (hidden action)
  - **Incomplete contracting** (not all states of the world are knowable in advance)

What are the main challenges to cooperative AI?

- Cooperative AIs face typical challenges in contract theory:
  - **Information asymmetry** in the form of *adverse selection* (hidden information) and *moral hazard* (hidden action)
  - **Incomplete contracting** (not all states of the world are knowable in advance)

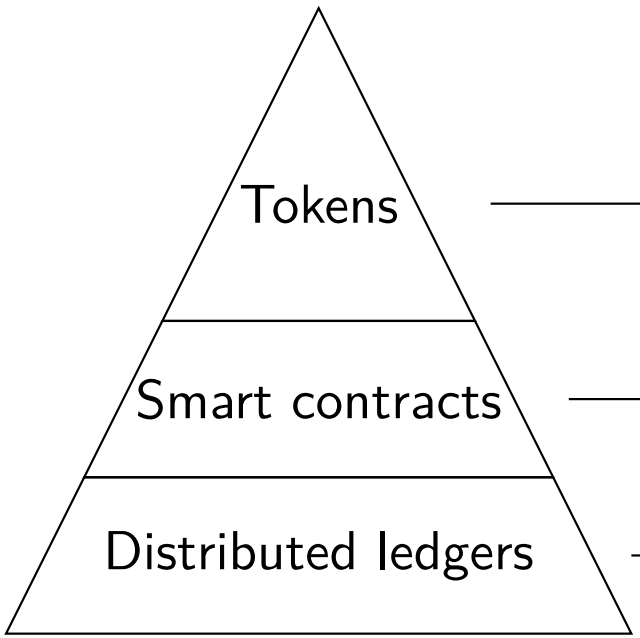
- AIs might lack the "human touch" in cooperation:
  - Tacit knowledge and understanding of norms
  - Communication and commitment device
  - Recall of history in repeated games

What are the main challenges to cooperative AI?

- Cooperative AIs face typical challenges in contract theory:
  - **Information asymmetry** in the form of *adverse selection* (hidden information) and *moral hazard* (hidden action)
  - **Incomplete contracting** (not all states of the world are knowable in advance)
- AIs might lack the "human touch" in cooperation:
  - Tacit knowledge and understanding of norms
  - Communication and commitment device
  - Recall of history in repeated games
- The "human touch" learned in model training might be problematic
  - Turing experiments with GPT models on games such as ultimatum game, Milgram shock experiment have produced outcomes that resemble human tendencies [1]

How can decentralization aid cooperative AI development?

Decentralization of data, contracting, and finance can address key informational frictions for cooperative AI



→ align incentives on uncertain, contingent payoffs

→ programmed for signaling and commitments

→ reduce information asymmetry among agents

Centralization in cooperative AI can pose serious downsides

- Centralization in AIs can lead to too much information sharing and excessive cooperation in a societally harmful way — e.g. price collusion [2]
- Bias and exclusion are harder to correct with too much centralization in AI models
- Excessive coercive behavior can arise with too much centralization- e.g. gang up on new entrants

Reference

Aher, G. V., Arriaga, R. I., and Kalai, A. T.
Using large language models to simulate multiple humans and replicate human subject studies.
In *International Conference on Machine Learning* (2023), PMLR, pp. 337–371.

Assad, S., Clark, R., Ershov, D., and Xu, L.
Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market.

Dafoe, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K. R., Leibo, J. Z., Larson, K., and Graepel, T.
Open problems in cooperative ai.
arXiv preprint arXiv:2012.08630 (2020).