Automated Corporations and AI Risk

Owain Evans, University of Oxford

- Put financial data online, invited predictions
- Scientists sent predictions and NumerAl invested using winning predictions
- Profits returned to scientists



- Financial data not public: expensive, proprietary
- Encrypted so that cannot be sourced but predictions based on encrypted data still work
- Task completely opaque to scientists: pure pattern recognition.



- Within weeks scientists beat Craib (who understood data)
- Unsurprising: some expertise needed to set up problem but after that it helps little.
- Key idea is to let the algorithms discover structure from data (rather than coding it in from the start)

11543 75353 55900



- Scientists? Anonymous, leaving no distinguishing information. Paid in Bitcoin.
- Scientists don't know what their predictions mean. Don't know what effect their having on world and can't trade on ideas themselves.

U	IIIU	
1	TANUKI \$3000 (6.92 BTC)	
2	TOMMY \$2000 (4.61 BTC)	0 1011 1100 111 100 100 10000 10001 10010 01 0110 01 00:01 10 10 10 10 10 10 10 10 10 10 10 10 1
3	BEANSMAN \$1200 (2.76 BTC)	AUC 0.55083
4	<u>UNQ</u> \$925 (2.13 BTC)	AUC 0.54954
5	FILINEP \$750 (1.73 BTC)	AUC 0.54928
6	ERISTODDLE \$625 (1.44 BTC)	AUC 0.54848
7	MYAI \$525 (1.21 BTC)	AUC 0.54795
8	PIOTREK \$425 (0.98 BTC)	AUC 0.54670
9	DATAGEEK \$325 (0.75 BTC)	AUC 0.54629
10	CHAITU	AUC 0.54462

NumerAI: role of people

 Craib finds investors, deals with legal compliance, buys and encrypts the data, controls overall investment strategy



• Scientists supply algorithms (nothing else)



Automated NumerAl

 AI can access huge number of proprietary datasets.
Encrypts them, sees how well scientists perform, and then scales up best performing data.





 Humans: add new proprietary datasets, supply algorithms, attract investors.

AI Corporation

- Hedge fund: minimal physical resources. Buy data online, rent computers from cluster, don't need outside customers, property is secured digitally.
- Real estate: Data online. Find regular company to manage and renovate property. Monitor their performance.







- Key idea: Crucial question is not human involvement but human oversight on key decisions.
- Why remove humans from oversight? The risk that humans make decisions that are worse for corporate bottom line.
- Humans: very slow, limited quantity of facts, limited coordination.





- Al needs quantities to maximize. They should be numerical/scalar, cheap to measure, and updated on short time-scales. (See Reinforcement Learning).
- Example: short-term return from a single investment, sales of mass-market product, corporate stock price, advert click-through or shares, number of website views.



- We care about qualitative, pluralistic values (society, morality, justice, knowledge, beauty) that are difficult to measure and require long-term planning.
- Problems of maximising shortterm, easy-to-measure quantities:
 - pollution
 - support organised crime, oppressive regimes
 - disinformation (e.g. health)
 - fraud (e.g. Ponzi scheme)



 Illegal activity can harm corporation even in short term. Plausible outcome: activity with bad externalities that is not illegal:



Mark Zuckerburg – Dead At 32 – Denies Facebook Has Problem With Fake News – The Shovel

Facebook founder Mark Zuckerburg – who died of heart complications at his Californian home today – says the proliferation of fake news stories on the site he...

THESHOVEL.COM.AU



- Facebook fake news (maybe)

- Opiate-based prescription painkillers
- Flash crash
- Financial Crisis