Chad Jones

On Friday, May 26, Chad Jones joined Markus' Academy for a lecture on The A.I. Dilemma: Growth versus Existential Risk. Jones is The STANCO 25 Professor of Economics at Stanford Graduate School of Business and a research associate of the National Bureau of Economic Research. Introductory remarks by Markus Brunnermeier.

A few highlights from the discussion:

- **A summary in four bullets**
    - Prof. Jones' working paper tries to study the unique dual nature of AI: it can deliver incredible economic gains while also presenting an existential threat.
    - We covered two models where the social planner chooses whether to use A.I.
    - The first model makes the tradeoff between risk and reward clear. With a log utility and a high-level parametrization, we obtain that the social planner would use A.I. for 40 years to have consumption grow by a factor of 55 (roughly the amount of growth seen in the last 2000 years), at the cost of a 1/3 chance of extinction.
    - The second model focuses on the possibility of infinite economic growth (singularities) and the possibility that the A.I. could improve the life expectancy of the population. Interestingly, the planner is much more willing to use A.I. and risk extinction to improve life expectancy.

- **[00:00] Introduction**
    - An existential risk is the opposite concept to resilience. They are shocks that you cannot come back from.
    - Resilience can be a matter of society's speed of transition to mitigate risks. There is an A.I. policy debate on whether we should slow down the inventions, assuming that by slowing it down we will reduce risks. However the way we respond is also important, since our response can also amplify shocks
    - A.I. risk is similar to climate risk in that they are both fat-tail risks (hence changing discount rates Weitzman 2014). When compared to nuclear risks, it is harder to control the proliferation of A.I.
- **[8:50] The dual nature of A.I.**
    - In Aghion et al 2017, we showed that A.I. could raise the growth rate of the economy above 2%, which is very difficult to do in standard models (in the real world, electricity, the internet or semiconductors could not do it). If A.I. can take over some tasks, it can allow people to focus on coming up with new ideas.
    - Because of this, A.I. is fundamentally different from nuclear technology: there is an existential risk (some probability of human extinction), but there are also incredible potential benefits.
- **[14:26] The simple model**
    - Intuitive solution. Requires calibrating the existential risk.

  - The model is static. Planner chooses the intensity of the use of A.I. (think of it as “years of use”). This intensity boosts consumption growth, while it also increases the probability of extinction.
  - We maximize expected social welfare: (Prob of survival) * (Utility from consumption).
  - Optimal choice is when the value of a life (measured in units of consumption) equals the ratio of the growth parameter and the survival probability parameter (which is multiplied by the intensity) – this is the A.I. Cost Benefit ratio.
  - Notice the solution does not depend on population size or the discount rate. The intuition is that we use A.I. until the marginal value of the lost lives exceeds the growth gain.
  - With a basic parametrization and log utility we obtain: you run the A.I. for 40 years to have consumption grow by a factor of 55 (roughly the ratio of growth seen in the last 2000 years), at the cost of a 1/3 chance of extinction. This result is highly sensitive to the coefficient of relative risk aversion.
  - If the agent’s consumption is high enough without A.I., or if the existential risk parameter is too high, the value of life may be too high to use the A.I. and bear any risk at all.
- [41:02] **The richer model**
  - Add two aspects: (1) using A.I. can improve the mortality rate (e.g. cure cancer), and (2) the possibility that the A.I. could lead to a singularity: infinite consumption in finite time.
  - Rather than choosing the intensity, we now have a binary decision of using the A.I., which yields different growth rates and mortality rates. Existential risk probability now is linear in its parameter (before it was exponential), so 10% of people dying is equivalent to a 10% chance of everyone dying. If the planner is a total utilitarian, the mortality and existential probabilities are treated in the same way
  - We obtain a threshold for the existential risk parameter above which the A.I. is shut down: one minus the ratio of utilities in the two scenarios
  - With CRRA>1, infinite consumption delivers finite utility. This gives us a threshold that is decreasing in: (1) the value of life initially, (2) risk aversion, (3) the growth rate without A.I.; and is increasing in the discount and mortality rates (as if you enjoyed the infinite consumption for less time).
  - Adding the mortality reduction drastically changes results. The thresholds become much higher. The planner is willing to risk extinction to improve people’s life expectancy. You are trading off “lives vs lives”, instead of “lives vs consumption”. With no mortality reduction, the planner wanted to ensure avoiding extinction so that people could enjoy the infinite consumption in the singularity. Now, we the planner tries to figure out the best way to live a long time.

Timestamps: