AI – Danger or Blessing?

A) Three Laws of Robotics (by Isaac Asimov, as explained in short story *Runaround*, 1942):

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

- What is the purpose of these laws?
- Do you think these laws are useful?
- Do you see any problems that can arise within these three laws?

B) *Ted Talks*: Sam Harris (2016) & Nick Bostrom (2015)

Sam Harris

Nick Bostrom

1. Why does Sam Harris talk about ants?

He shows up how we treat ants and showed us how we could be treatet from intelligent machines, so if the ants are in our way we just kill them, not we hate them but the are in our way and that could be the same with the super intelligent machines

2. Why, according to Nick Bostrom, could it be problematic to give a machine with AI tasks?

AI will optimize the world to fullfill the task and could ignore what we want

3. What do <u>you</u> think could be done to avoid the risks mentioned by Bostrom?

more specific task
implement rules with clear definitions
kill switch
AI proposes way to solve task in advance -> humans give permissions
monitoring system

4. What does Bostrom suggest we should do?

We have to teach the AI our values, so it will b on our side

C) What speaks for AI, what against AI? Fill in the table. Think about:

- *Ted Talks* by Harris and Bostrom
- movie excerpt from *Space Odyssey*
- Video “What is Artificial Intelligence?” (last session)
- AI in our daily lives
- Impacts of AI on your job / What groups of jobs should be worried about AI
- etc.

Advantages	**Disadvantages**
- faster work process - easier life for humans - automation process better quality - long term cheaper and no salary - space explorations - dangerous tasks	- may go rogue - unpredictable - less jobs - hackers could compromise machines - bugs