



# CYBER POLITICS AND POLICIES

## CHAPTER 12: ARTIFICIAL INTELLIGENCE AND AUTONOMY

# AT THE END OF THIS CHAPTER, STUDENTS WILL BE ABLE TO:

Describe political and ethical problems related to dependence on data and algorithms – including algorithmic governance, bias

I. Define key terms associated with artificial intelligence – including Meaningful Human Control (MHC) and Levels of Autonomy

I. Apply 3 paradigms (Realism, Liberal Internationalism and Constructivism) to describing and analyzing the dynamics of an “AI Arms Race”

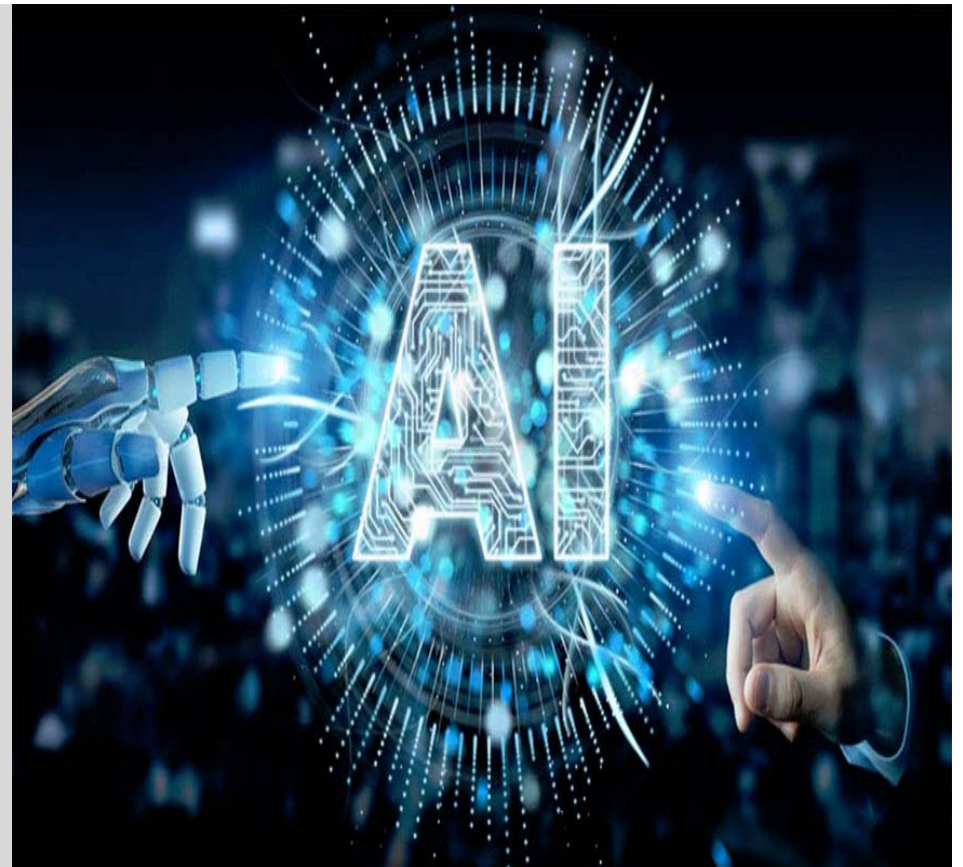
# TERMS AND CONCEPTS FOR THINKING ABOUT FUTURE TECHNOLOGICAL DEVELOPMENTS





# EMERGING TECHNOLOGIES

- **Technology** which is “radically novel and relatively fast growing . . . characterized by a certain degree of coherence persisting over time and with the potential to exert a consider impact on the socioeconomic domain



# FOURTH REVOLUTION

- **Fourth Revolution:** the advent of technologies, like 3D printing of human prosthetics, voice control over technologies (like a home heating or entertainment system), the use of technology in medical diagnostics, and the growing influence of the internet of things and big data.

- **First Industrial Revolution** (1784): steam power, mechanical equipment
- **Second Industrial Revolution** (1870): specialization in labor, mass production, electricity
- **Third Industrial Revolution** (1969): Increased dependence on information technology and automated production

# PRECAUTIONARY PRINCIPLE

- Used when thinking about the introduction of new technologies whose long-term effects are unknown.
- Decision-makers should assume a technology is harmful until it can be proven safe – rather than assuming it is safe until it is proven harmful.
- When thinking about possible harms from a technology, decision makers should consider not only those who are currently present, but also those who may be present several generations from now



The Precautionary Principle

# PART TWO: NEW DEVELOPMENTS // BIG DATA AND AI



# BIG DATA: CHARACTERISTICS

- A methodology in which existing data collections of all types are aggregated together, and made to communicate with one another through sharing data

- Exists in high volumes
- Accumulated at a high velocity
- Created in real time
- Extensive variety of data
- Exhaustive in scope
- 'Fine grained'
- Can be indexed and classified
- Is relational.
- Can be expanded rapidly.



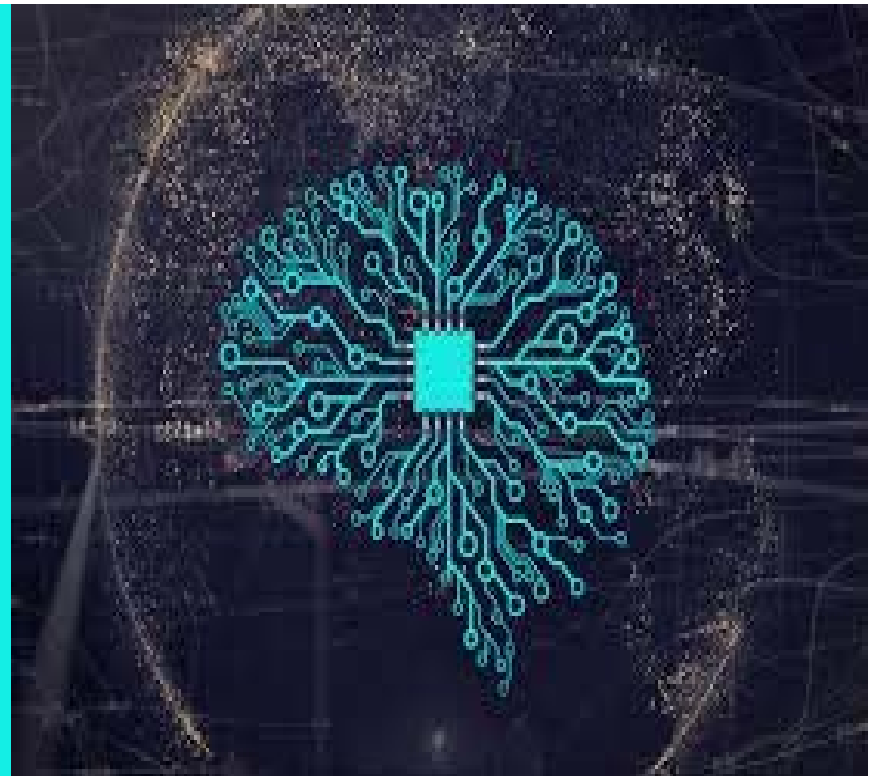
## 3 ISSUES WITH BIG DATA

- 1. Governments may be tempted to utilize data in order to engage in surveillance of citizens.

- 2. Your information is being stored preemptively, in response to the likelihood that you might in the future commit a crime and the government might require evidence in order to prosecute you.
- 3. Big Data has an ideology. Does not serve individual citizen interests, but rather serves the interests of major corporations and the national security interests of government.

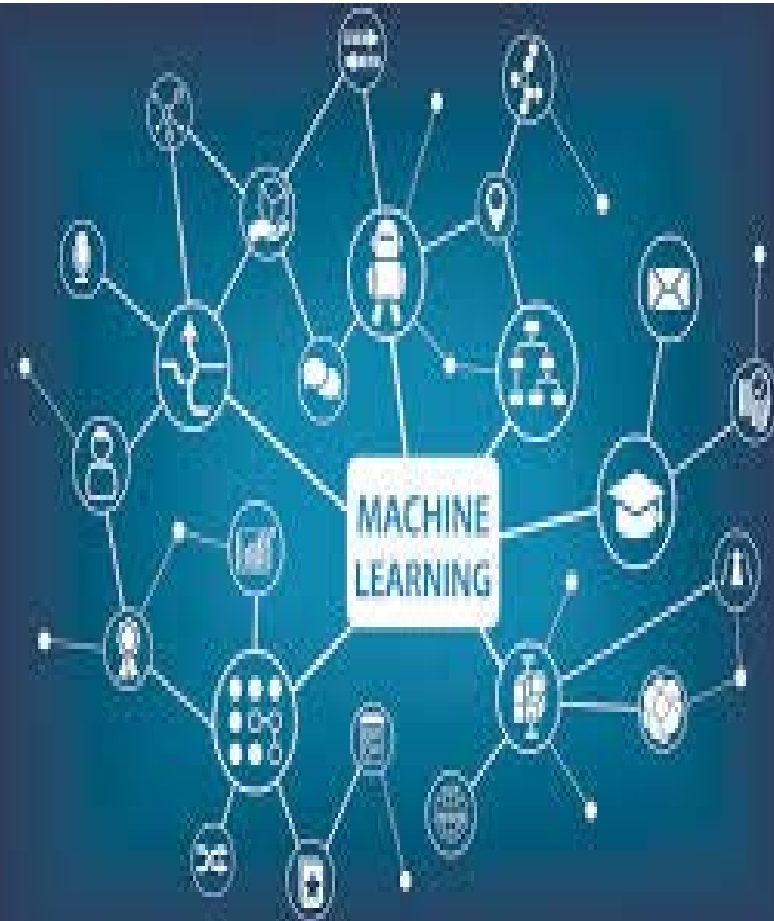
# DEFINING ARTIFICIAL INTELLIGENCE

- “The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.



# MACHINE LEARNING

- An algorithmic computer approach to making a determination or prediction in which the computer may act without being explicitly programmed, automatically improving itself.”
- Exists along a spectrum – systems can be fully autonomous in their machine learning functions, or they can be semi-supervised, in a scenario where they are ‘fed’ labelled data.
- **Deep learning:** A system whereby artificially intelligent agents utilize multi-layer artificial neural networks that mimic human neural architecture.



# PART THREE: WHAT IS AUTONOMY





# US DEPARTMENT OF DEFENSE

- Fully autonomous weapons system: a weapon system that, once activated, can select and engage targets without further intervention by a human operator.”



# LEVELS OF AUTONOMY

- **semi-autonomous;** a semi-autonomous robot possesses some level of independence but ultimately is not self-directed. That is, such a machine does not possess decision-making capabilities and may require human participation
- **‘human in the loop’**

- **Human on the loop:** Weapons capable of independently targeting and delivering force while under the supervision of a human operator
- **human out of the loop:** Weapons can select targets and deliver force without any human input or interaction.

# DISTRIBUTED RESPONSIBILITY

- A system where human and autonomous agents might share responsibility (ethically, legally, and morally) for outcomes

Could an AI system which makes medical diagnostic or treatment decisions be sued for malpractice if a problem occurs?

Could an AI system be held criminally liable for certain actions (like vehicular manslaughter) and could it be held liable for breach of warranty?

## PART FOUR: CRITIQUING AUTONOMY AND AI





# LIMITATIONS OF ARTIFICIAL INTELLIGENCE

- Chinese box experiment
- Can machines learn to 'behave ethically'?
- The Hidden Layer



# RULE BY ALGORITHM

- Worry that citizens will cede too much personal control to inanimate entities without understanding the long-run implications of this decision.
- By the time users realize what they have done, it may be too late for humans to wrest back this control from the machines



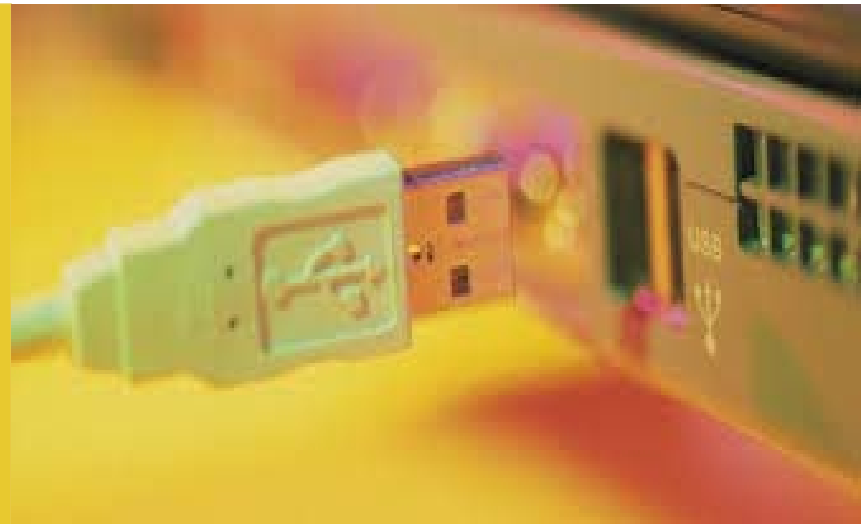
# SHOULD WE TRUST AI?

- Problem of Undertrust
- Problem of Overtrust



# SPEAKING BACK TO TECHNOLOGY

- Relinquishment
- Limitation and regulation
- Augmentation





# FOR DISCUSSION

- Should robots be treated as independent agents, or 'citizens' with the right to own property?
- To testify (perhaps against you) in court?
- To vote in national elections?
- Should robots be treated as wholly responsible for their actions in, for example, being tried for manslaughter if they cause an auto accident or the death of a pedestrian – or does their need to be an individual who exerts **Meaningful Human Control (MHC)** over the robot?

