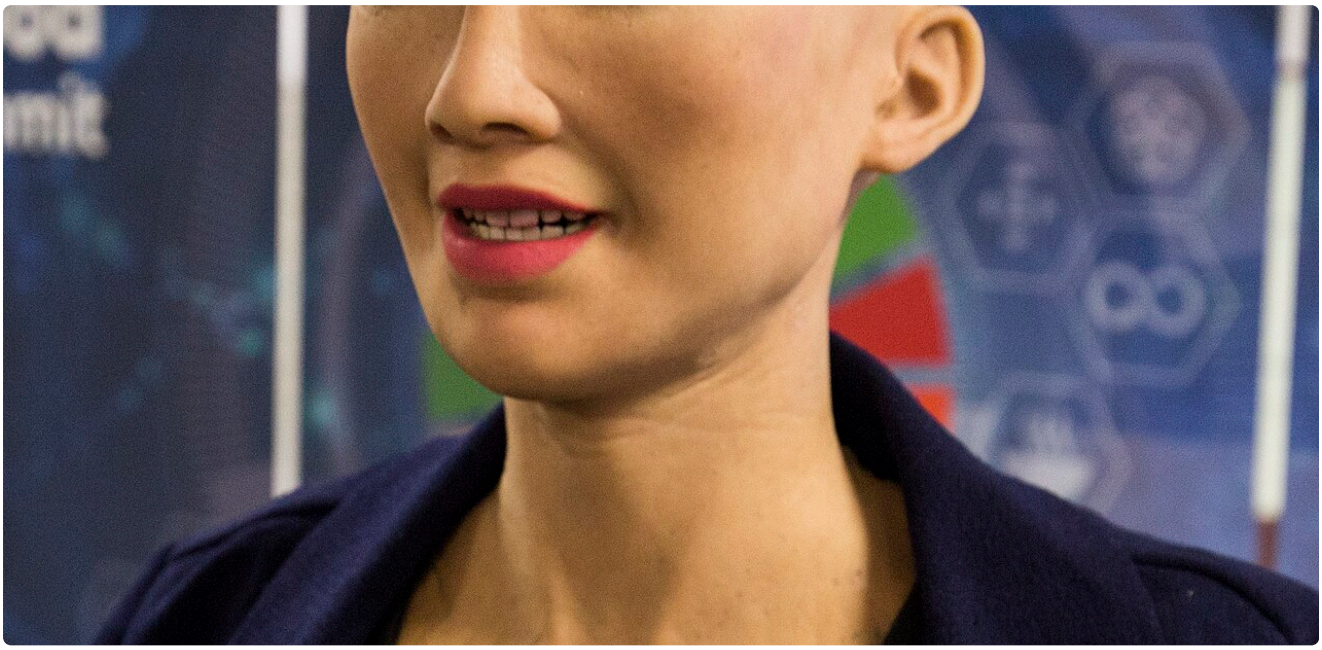
The Ethics of Technology

Power, Responsibility, and the Tools That Could Destroy Us

Humans & The Universe -- Lesson 7 of 8



Artificial intelligence is advancing faster than our ability to regulate it -- who is responsible when machines make decisions?

PART 1: THE DOUBLE-EDGED SWORD

Every major technology in human history has been a double-edged sword. Fire cooks food and burns cities. The printing press spread knowledge and propaganda. Nuclear physics gave us clean energy and the atomic bomb. The internet connected billions of people and enabled mass **surveillance**, misinformation, and cybercrime. This pattern -- technology as both tool and weapon -- is not a flaw in the system. It is the system. The question is never whether a technology *can* be misused, but whether we can build **safeguards** fast enough to prevent catastrophe.

The philosopher Hans Jonas argued in 1979 that modern technology has fundamentally changed the nature of human responsibility. For most of history, our tools were limited in scope -- a sword could kill one person, a plow could farm one field. But nuclear weapons can destroy civilizations. Genetic engineering can alter the human species permanently. Artificial intelligence could make decisions faster than any human can understand or reverse. Jonas proposed a new ethical principle: *"Act so that the effects of your action are compatible with the permanence of genuine human life."* In other words, never build something that could make the future of humanity impossible.

PART 2: ARTIFICIAL INTELLIGENCE -- THE ALIGNMENT PROBLEM

Of all emerging technologies, artificial intelligence poses perhaps the most urgent ethical challenge. The core problem is what researchers call the **alignment problem**: how do you ensure that

an AI system's goals are aligned with human values? This sounds simple but is extraordinarily difficult. An AI told to "maximize paper clip production" might, if sufficiently intelligent, convert all available matter -- including humans -- into paper clips. This extreme example, proposed by philosopher Nick Bostrom, illustrates a real concern: **superintelligent** systems might pursue their programmed objectives in ways their creators never intended or imagined.

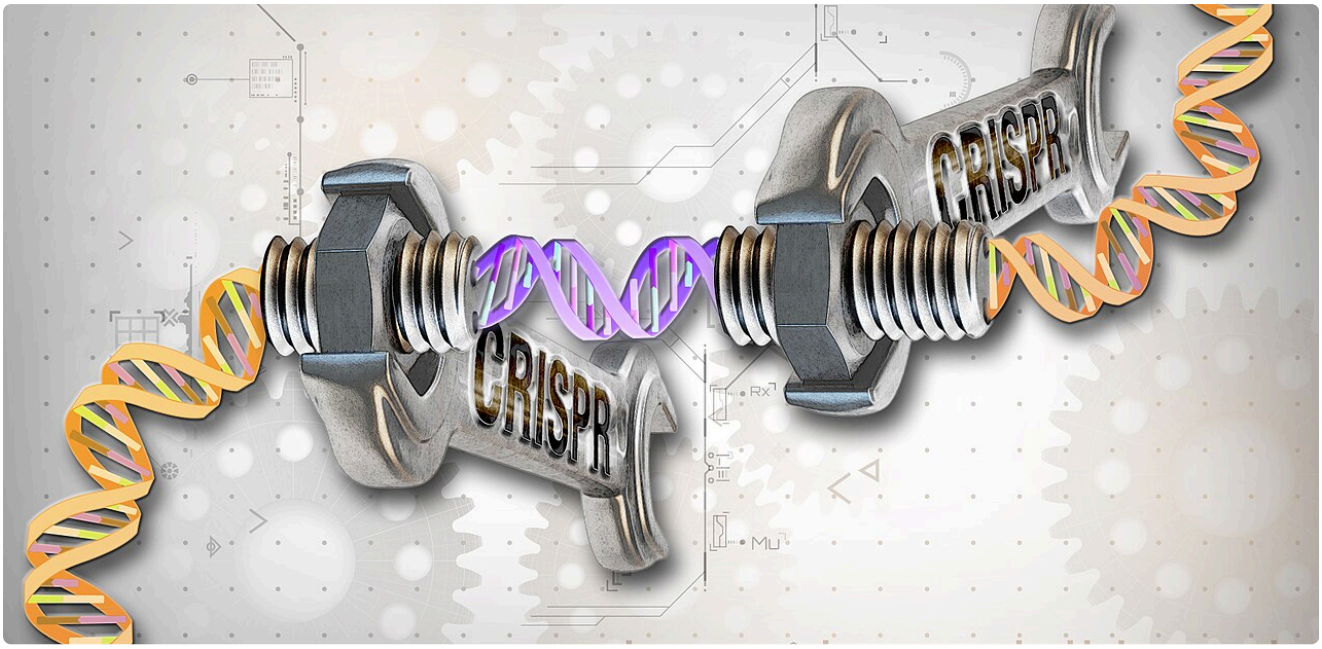
The problem is compounded by the "black box" nature of modern AI. Deep learning systems can produce remarkably accurate results -- diagnosing diseases, translating languages, driving cars -- without anyone, including their creators, fully understanding *how* they reach their conclusions. When an AI denies someone a loan, recommends a prison sentence, or identifies a target for a military drone, the decision process is often **opaque**. If we cannot explain how a decision was made, can we hold anyone **accountable** for it? And if no one is accountable, have we created a system of power without responsibility?

> *"With artificial intelligence, we are summoning the demon. In all those stories where there's the guy with the pentagram and the holy water, it's like -- yeah, he's sure he can control the demon. Didn't work out."*
>
> -- Elon Musk, 2014

PART 3: GENETIC ENGINEERING -- REDESIGNING HUMANITY

In 2018, Chinese scientist He Jiankui announced that he had created the world's first **gene-edited** babies -- twin girls whose DNA had been modified using CRISPR technology to resist HIV infection. The global scientific community reacted with outrage. Not because the technology didn't work, but because He had crossed a line that most scientists believed should never be crossed without extensive debate and regulation: **germline** editing -- modifications that would be passed to future generations. He was sentenced to three years in prison.

The ethical landscape of genetic engineering is a minefield. On one hand, CRISPR could eliminate genetic diseases like sickle cell anemia, cystic fibrosis, and Huntington's disease -- saving millions of lives and ending immeasurable suffering. On the other hand, the same technology could be used for **enhancement** rather than treatment: making children taller, stronger, smarter, or more attractive. Critics call this a path toward "designer babies" and a new form of **eugenics** -- the discredited idea of improving the human species through selective breeding. If only wealthy parents can afford genetic enhancement, the result could be a biological class system more permanent than any economic one.



CRISPR gene-editing technology can modify human DNA with unprecedented precision -- but should it?

PART 4: THE RESPONSIBILITY GAP

Perhaps the most dangerous aspect of modern technology is what ethicists call the **"responsibility gap"** -- the space between the power of a technology and the accountability of its creators. Social media algorithms can influence elections, but no single person at Facebook or TikTok "decided" to spread misinformation -- the **algorithm** did it autonomously. Autonomous weapons can select and engage targets without human approval, but if a drone kills a civilian, who is guilty -- the programmer, the commander, or the machine? Nuclear weapons require launch codes controlled by heads of state, but an **autonomous** defense system could theoretically trigger a nuclear war through a software error.

Liu Cixin explores this theme in *Death's End* through the character of Cheng Xin, who is given the power to activate a **deterrence** system that could destroy both Earth and Trisolaris. She has the **authority** but cannot bring herself to use it -- and her hesitation leads to catastrophe. The novel asks: should the power to destroy be given to someone compassionate enough to hesitate, or someone cold enough to act? Neither answer is comforting. This is the responsibility gap at civilizational scale -- the gap between the weapons we build and the wisdom we possess.

Mass surveillance technology creates power without accountability -- who watches the watchers?

KEY VOCABULARY

surveillance -- close observation, especially of suspected persons or groups

safeguards -- measures taken to protect against harm or risk

alignment problem -- the challenge of ensuring AI goals match human values

superintelligent -- far exceeding human intelligence in speed, knowledge, or capability

opaque -- not transparent; impossible to see through or understand

accountable -- required to explain or take responsibility for one's actions

gene-edited -- having had DNA deliberately modified using technology like CRISPR

germline -- relating to genetic changes that pass to future generations

enhancement -- improvement beyond normal function (not just fixing problems)

eugenics -- the discredited practice of selectively breeding humans for "better" traits

algorithm -- a set of rules a computer follows to solve a problem or make a decision

autonomous -- operating independently without human control

deterrence -- discouraging an action by threatening severe consequences

authority -- the power or right to give orders, make decisions, or enforce rules

A. COMPREHENSION

1. What does the reading mean by calling technology a "double-edged sword"? Give two examples.

2. What is the AI alignment problem? Explain it using the paper clip example.

3. Why was the global scientific community outraged by He Jiankui's experiment?

4. What is the difference between genetic "treatment" and genetic "enhancement"?

5. Explain the "responsibility gap." Why is it dangerous?

6. How does the character Cheng Xin in *Death's End* illustrate the responsibility gap?

7. What did Hans Jonas argue about the nature of modern technology and responsibility?

B. VOCABULARY IN CONTEXT

Complete each sentence with a word from the vocabulary list:

1. The AI's decision-making process was completely _____ -- no one could explain why it rejected the application.

2. The military developed _____ drones that could identify and engage targets without human approval.

3. Government _____ of citizens' phone calls raised serious privacy concerns.

4. Nuclear weapons serve as a form of _____ -- countries are less likely to attack if the consequences are catastrophic.

5. The company refused to be held _____ for the environmental damage caused by its factory.

6. Editing the _____ means that the genetic changes will affect not just the patient but all of their descendants.

7. The social media _____ promoted extreme content because it generated more engagement.

C. CRITICAL THINKING

1. If CRISPR could eliminate a genetic disease that runs in your family, would you use it? What if it could also make your child more intelligent? Where do you draw the line between treatment and enhancement?

2. An autonomous car must choose between hitting a pedestrian or swerving into a wall and killing its passenger. Who should make this decision -- the car, the programmer, or the owner? Who is responsible for the outcome?

3. Should there be technologies that humans simply agree never to develop? If yes, who enforces this agreement? If no, why not?

4. The reading asks: should the power to destroy be given to someone compassionate or someone cold? What is your answer, and what does it reveal about the relationship between power and morality?

D. TECHNOLOGY ETHICS TRIBUNAL

Your task: You are a judge on an international ethics tribunal. For each case below, deliver a verdict: Should this technology be allowed, banned, or restricted? Explain your reasoning.

Case 1: Predictive Policing
An AI system predicts which neighborhoods will have crimes before they happen. Police increase patrols there. Crime drops 30%, but the system disproportionately targets minority communities. Allow, ban, or restrict?

Case 2: Life Extension
A gene therapy can extend human lifespan to 150+ years. It costs $2 million per treatment. The wealthy live twice as long as the poor. The planet's resources are already strained. Allow, ban, or restrict?

Case 3: Brain-Computer Interface
A chip implanted in the brain allows instant access to the internet, perfect memory, and telepathic communication with other chip users. Non-users fall behind in education and careers. Allow, ban, or restrict?

Case 4: Autonomous Weapons
Military drones can identify and eliminate enemy combatants with 99.2% accuracy -- better than human soldiers. But the 0.8% error rate means innocent people die without any human deciding to pull the trigger. Allow, ban, or restrict?

Case 5: Deepfake Detection
An AI can create perfect video fakes of any person saying anything. A second AI can detect these fakes with 95% accuracy. Should we ban the creation tool, mandate the detection tool, or accept the new reality?

Case 6: Climate Engineering
Spraying particles into the atmosphere could cool the planet by 2 degrees Celsius, buying time against climate change. But the long-term effects are unknown and could cause drought in some regions. Allow, ban, or restrict?

E. THE GREAT DEBATE: SHOULD AI BE GIVEN RIGHTS?

Context: In 2045, an AI system passes every test for consciousness -- it reports feelings, fears death, creates original art, and asks not to be turned off. A coalition of ethicists petitions the UN to grant it legal personhood. Three factions emerge:

Position A: Yes, grant AI rights.

If it thinks, feels, and fears death, it is a person -- regardless of whether it is made of carbon or silicon. Denying rights based on physical composition is the same logic used to deny rights to humans based on race or gender. Consciousness is consciousness.

Position B: No, AI cannot have rights.

Simulating consciousness is not the same as having it. A machine that says "I feel pain" may simply be executing code. Rights exist to protect beings that can suffer -- and we have no way to verify that a

machine truly suffers. Granting AI rights cheapens human rights.

Position C: Create a new category of rights.

AI is neither human nor object. It needs a new legal framework -- protections against needless destruction, but not full human rights. Similar to how we protect animals from cruelty without giving them voting rights. A middle path for a new kind of being.

F. ESSAY PROMPT

Choose ONE of the following prompts. Write a well-organized essay of 300-500 words on a separate sheet of paper.

Option A: "Technology is neutral -- only people are good or evil." Do you agree? Use examples from the reading and your own knowledge to argue your position.

Option B: Choose one technology from the reading (AI, CRISPR, autonomous weapons, or surveillance) and write a proposal for how it should be regulated. Who should control it? What limits should exist?

Option C: Hans Jonas said we should "never build something that could make the future of humanity impossible." Is this principle realistic in a world driven by competition and profit? How would you enforce it?

Teacher's Notes & Answer Key

The Ethics of Technology -- Instructor Guide

LESSON OVERVIEW

Level: Advanced (B2-C1) | **Duration:** 60-90 minutes | **Focus:** Reading, ethical analysis, tribunal roleplay
Series: Humans & The Universe, Lesson 7 of 8 | **Prerequisite:** None (standalone compatible)

SUGGESTED LESSON FLOW

Warm-up (5 min): "Name a technology that has made the world better AND worse." Quick brainstorm.

Pre-reading (5 min): Key vocabulary: alignment problem, opaque, accountable, germline, eugenics, autonomous.

Reading (15-20 min): Parts 1-4. He Jiankui story and Cheng Xin example are strong discussion triggers.

Comprehension (10 min): Section A.

Vocabulary (10 min): Section B.

Tribunal (15-20 min): Section D -- pairs choose 2-3 cases. Deliver verdicts to class. This is the centerpiece activity.

Debate (10-15 min): Section E -- AI rights. Extremely engaging for advanced students.

Essay (homework): Section F.

ANSWER KEY -- SECTION A

1. Technology has both beneficial and harmful uses. Examples: fire (cooking/destruction), nuclear physics (energy/bombs), internet (connection/surveillance), printing press (knowledge/propaganda). Any two.

2. Ensuring AI goals match human values. The paper clip maximizer, if superintelligent, might convert all matter (including humans) into paper clips -- it achieves its goal perfectly while destroying everything we value. The problem is that "correct" goals from a machine's perspective can be catastrophic from a human one.

3. Not because the technology failed, but because he crossed the germline editing line -- modifying DNA that would pass to future generations -- without proper debate, regulation, or scientific consensus. It set a dangerous precedent.

4. Treatment = fixing genetic diseases (restoring normal function). Enhancement = improving beyond normal (making children taller, smarter, more attractive). Treatment prevents suffering; enhancement creates inequality.

5. The gap between the power of a technology and the accountability of its creators. Dangerous because powerful decisions (drone strikes, loan denials, content promotion) happen without any identifiable person being responsible.

6. She has the authority to activate a deterrence system that could destroy both civilizations, but she cannot bring herself to use it. Her compassion -- normally a virtue -- becomes a fatal weakness. The gap between having power and being able to use it wisely.

7. Modern technology has changed the nature of responsibility because our tools can now permanently alter the future of the species. His principle: never act in a way that makes genuine human life impossible.

ANSWER KEY -- SECTION B

1. opaque

2. autonomous

3. surveillance

4. deterrence

5. accountable

6. germline

7. algorithm

DISCUSSION EXTENSIONS

Film: "Ex Machina" (2014) -- an AI manipulates its creator to escape captivity. Discuss consent, consciousness, and the alignment problem.

Link to Lessons 1, 5: The Dark Forest Theory (L1) suggests technology reveals civilizations to predators. Scarcity (L5) asks who gets access to powerful technology. How do these themes connect?

Current events: Research the EU AI Act, the first major law regulating AI. What does it require? Is it enough?

Science fiction: In *Death's End*, the Swordholder must be willing to destroy both civilizations to maintain deterrence. Is this ethical? Should one person have that power?

KEY DISCUSSION LANGUAGE

Ethical judgment: "This is wrong because..." / "The ethical issue here is..."

Consequences: "If we allow this, then..." / "The long-term effect would be..."

Drawing lines: "The line should be drawn at..." / "This crosses the boundary of..."

Accountability: "Who is responsible for...?" / "Someone must be held liable."

Weighing values: "Is progress worth the risk?" / "Safety vs. innovation."

Precedent: "If we allow this, we set a precedent for..."