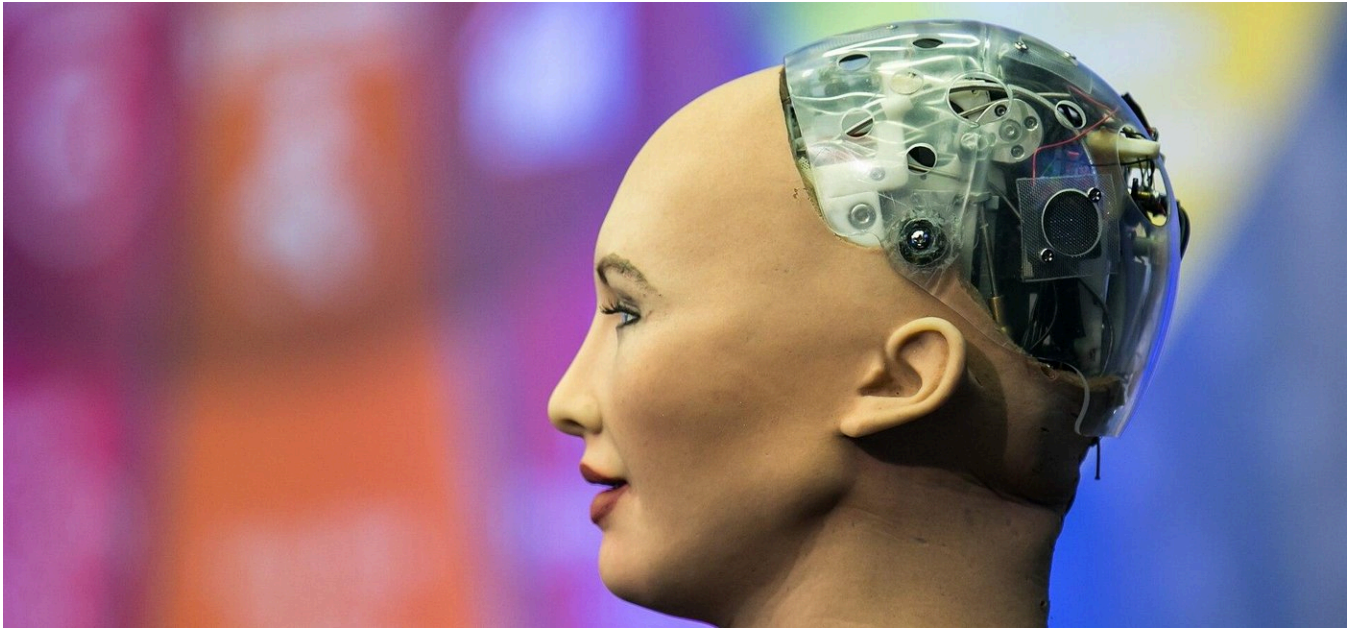


The Ethics of Technology

Power, Responsibility, and the Tools That Could Destroy Us

Name _____ Class _____ Date _____

PART 1: THE DOUBLE-EDGED SWORD



Sophia, a humanoid robot built by Hanson Robotics, at the United Nations AI for Good summit in Geneva in 2017

Photo: ITU Pictures (CC BY 2.0)

Every major technology in human history has been a double-edged sword. Fire cooks food and burns cities, the printing press spread both knowledge and propaganda, and nuclear physics gave us low-carbon electricity as well as the atomic bomb. The internet connected billions of people while also enabling mass surveillance, misinformation and cybercrime. This pattern of 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 a catastrophe.

The philosopher Hans Jonas argued in 1979 that modern technology has changed the nature of human responsibility. For most of history our tools were limited in scope, since a sword could kill one person at a time and a plow could work one field. Nuclear weapons, by contrast, can destroy civilizations, and genetic engineering can alter the human species permanently. Artificial intelligence may come to make decisions faster than any human can understand or reverse. Jonas therefore 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, we should never build something that could make the future of humanity impossible.

PART 2: ARTIFICIAL INTELLIGENCE AND THE ALIGNMENT PROBLEM

Of all the emerging technologies, artificial intelligence (AI) poses perhaps the most urgent ethical challenge. Researchers call the core difficulty the alignment problem: how do you make sure that the goals of an AI system match human values? This sounds simple, but it is extraordinarily difficult. The philosopher Nick Bostrom offered a famous thought experiment. An AI told to maximize the production of paper clips might, if it were intelligent enough, convert all available matter into paper clips, including the humans who built it. The example is extreme, but the concern is real: a superintelligent system might pursue its objective in ways its creators never intended or imagined.

The problem is made worse by the black box nature of modern AI. Deep learning systems can diagnose diseases, translate languages and drive cars with remarkable accuracy, yet nobody, including their creators, fully understands how they reach their conclusions. When such a system denies someone a loan, recommends a prison sentence or identifies a target for a military drone, its reasoning is often opaque. If we cannot explain how a decision was made, can we hold anyone accountable for it? If nobody is accountable, we have created power without responsibility.

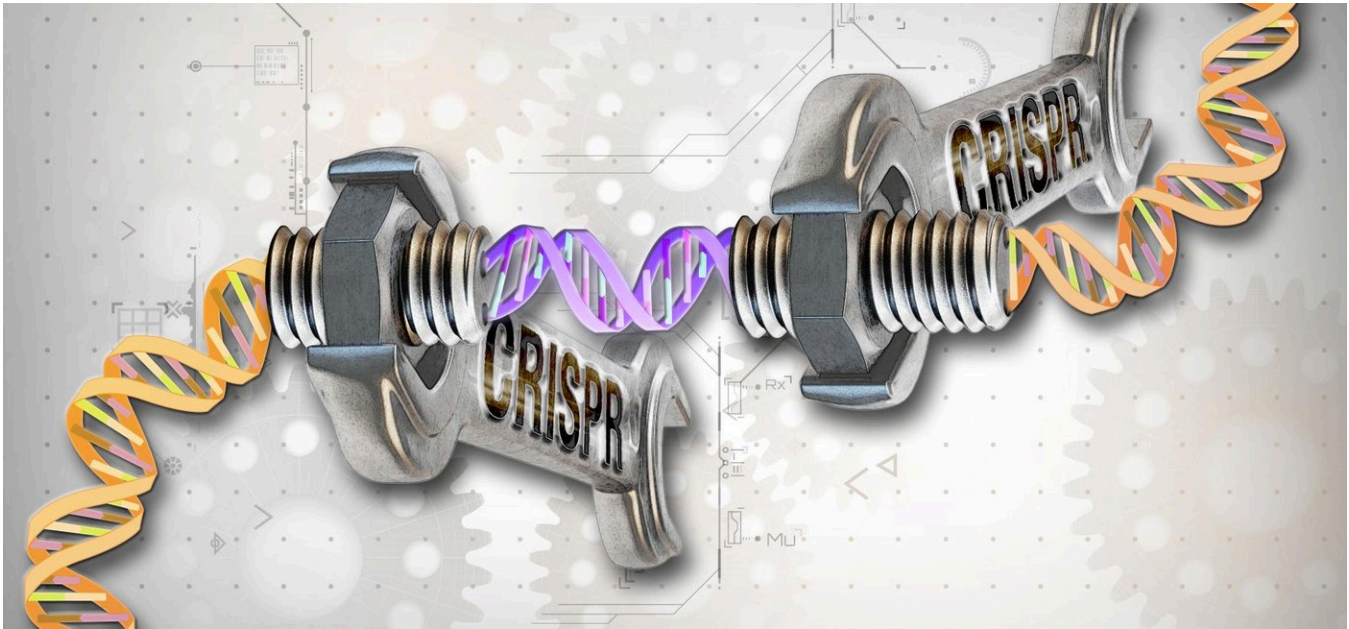
"With artificial intelligence, we are summoning the demon. In all those stories where there's the guy with the pentagram and the holy water, it's like, yeah, he's sure he can control the demon. Didn't work out."

Elon Musk, 2014

PART 3: GENETIC ENGINEERING AND REDESIGNING HUMANITY

In November 2018, the Chinese scientist He Jiankui announced that he had created the world's first gene-edited babies. They were twin girls whose DNA he had altered as embryos, using a tool called CRISPR, in an attempt to make them resistant to HIV. Scientists around the world reacted with outrage. His evidence suggested that the editing had been only partly successful, the children had not needed it, and the parents may not have understood the risks. Above all, he had crossed a line that most scientists believed should not be crossed without long public debate. He had edited the germline, which means that the changes can be passed on to future generations. A Chinese court sentenced him to three years in prison, and he was released in 2022.

Meanwhile, gene editing that affects only the patient has moved from theory into medicine. In late 2023, regulators in Britain and the United States approved the first CRISPR treatment, a therapy for sickle cell disease. In 2025, doctors in Philadelphia treated a baby named KJ Muldoon with a gene-editing therapy designed for his rare disease alone, the first treatment of its kind. Neither treatment changes the DNA that a patient passes on to children.



CRISPR gene editing can change DNA with great precision. The question is which changes we should allow.

Illustration: Ernesto del Aguila III, National Human Genome Research Institute (public domain)

The harder questions concern the germline and the difference between treatment and enhancement. The same tools that might one day remove cystic fibrosis or Huntington's disease from a family could also be used to make children taller, stronger or more intelligent. Critics see 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 could afford such enhancement, the result might be a biological class system more permanent than any economic one.

PART 4: THE RESPONSIBILITY GAP

Perhaps the most dangerous feature of modern technology is what ethicists call the responsibility gap: the space between the power of a technology and the accountability of the people behind it. A social media algorithm can influence an election, yet no single employee of the company decided to spread misinformation. An autonomous weapon can select and attack a target without human approval. If a drone kills a civilian, it is unclear whether the programmer, the commander or the machine is to blame. Even nuclear weapons, which are supposed to be controlled by heads of state, depend on warning systems that can fail, as the false alarm of 1983 in Lesson 1 showed.

Liu Cixin explores this theme in *Death's End* through the character of Cheng Xin, who is given control of 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 whether the power to destroy should be given to someone compassionate enough to hesitate or to someone cold enough to act. Neither answer is comforting. This is the responsibility gap on the scale of a civilization: the distance between the weapons we build and the wisdom we possess.



A solar-powered surveillance camera. Systems that watch and decide automatically raise an old question: who watches the watchers?

KEY VOCABULARY

Term	Meaning	Term	Meaning
<i>surveillance</i>	close observation, especially of people or groups	<i>safeguards</i>	measures taken to protect against harm or risk
<i>alignment problem</i>	the challenge of making sure an AI system's goals match human values	<i>superintelligent</i>	far exceeding human intelligence
<i>opaque</i>	not transparent; impossible to see into or understand	<i>accountable</i>	required to explain and take responsibility for one's actions
<i>gene-edited</i>	having had DNA deliberately changed using a tool such as CRISPR	<i>germline</i>	the cells that pass genetic information to the next generation
<i>enhancement</i>	improvement beyond normal function, as opposed to treating a disease	<i>eugenics</i>	the discredited practice of trying to breed humans for "better" traits
<i>algorithm</i>	a set of rules a computer follows to solve a problem or make a decision	<i>autonomous</i>	operating independently, without human control

deterrence

discouraging an action by threatening severe consequences

authority

the power or right to give orders and make decisions

Discussion & Analysis

The Ethics of Technology · Lesson Activities

A. COMPREHENSION

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

2. What did Hans Jonas argue about modern technology and responsibility?

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

4. Why is the black box nature of AI a problem for accountability?

5. Why did scientists react with outrage to He Jiankui's experiment? Give two reasons.

6. How are the treatments approved in 2023 and given in 2025 different from what He Jiankui did?

7. What is the difference between genetic treatment and genetic enhancement?

8. Explain the responsibility gap, and say how Cheng Xin's story illustrates it.

B. VOCABULARY IN CONTEXT

Complete each sentence with a term from the vocabulary list. Use each term once.

accountable · algorithm · autonomous · deterrence · enhancement · eugenics · germline · opaque
· safeguards · surveillance

1. The system's reasoning was completely _____, and no one could explain why it had rejected the application.
2. The military developed _____ drones that could identify and attack targets without human approval.
3. Government _____ of citizens' phone calls raised serious concerns about privacy.
4. Nuclear weapons are meant to work as a form of _____, since no country wants to invite its own destruction.
5. The company refused to be held _____ for the environmental damage caused by its factory.
6. Editing the _____ means that the changes will affect not just the patient but all of their descendants.
7. The social media _____ promoted extreme content because it kept people watching.
8. Engineers added several _____ to stop the machine from operating when a person was too close.
9. Curing a disease is treatment, but making a healthy child taller would be _____.
10. In the early twentieth century, the supporters of _____ wanted to control who was allowed to have children.

C. CRITICAL THINKING

1. If gene editing could remove 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. A self-driving car must choose between hitting a pedestrian and swerving into a wall, killing its passenger. Who should make that decision in advance: the manufacturer, the programmer or the owner? Who is responsible for the outcome?

3. Should there be technologies that humans simply agree never to develop? If so, who enforces the agreement? If not, why not?

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

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

Scenario: In 2045, an AI system passes every test for consciousness that scientists can devise. It reports feelings, says that it fears being switched off, creates original art, and asks to be left running. A coalition of ethicists petitions the UN to grant it legal personhood, and three factions emerge.

Position A: Yes, grant rights

If it thinks, feels and fears death, it is a person, whether it is made of carbon or silicon. Denying rights because of what a being is made of repeats the logic once used to deny rights on the basis of race or sex.

Position B: No rights for AI

Simulating consciousness is not the same as having it. A machine that says "I feel pain" may only be running code. Rights protect beings that can suffer, and we cannot verify that a machine suffers. Granting it rights would cheapen human rights.

Position C: A new category

AI is neither a human nor an object, so it needs a new legal status: protection against needless destruction but not full human rights, much as we protect animals from cruelty without giving them the vote.

After the debate, write your own position in three or four sentences.

E. TECHNOLOGY ETHICS TRIBUNAL

You are a judge on an international ethics tribunal. Choose two or three cases and deliver a verdict on each: should the technology be allowed, banned or restricted? Explain your reasoning.

Case 1: Predictive Policing

An AI system predicts where crimes will happen, and police increase patrols there. Crime falls by 30%, but the system sends far more patrols to minority neighborhoods.

Case 2: Life Extension

A gene therapy can extend the human lifespan beyond 150 years at a cost of \$2 million per person. The wealthy would live twice as long as the poor on a planet whose resources are already strained.

Case 3: Brain-Computer Interface

A chip implanted in the brain gives instant access to the internet and perfect memory. People without the chip fall behind in education and at work.

Case 4: Autonomous Weapons

Military drones can identify enemy fighters with 99.2% accuracy, which is better than human soldiers manage. The remaining 0.8% means that innocent people die without any human deciding to fire.

Case 5: Deepfakes

One AI can create a perfect fake video of any person saying anything. A second AI can detect such fakes 95% of the time. Do you ban the first tool, require the second, or accept the new reality?

Case 6: Climate Engineering

Spraying particles into the upper atmosphere could cool the planet and buy time against climate change. The long-term effects are unknown, and some regions might suffer drought as a result.

Verdicts and reasons:

F. ESSAY PROMPT

Choose ONE of the following and write a well-organized essay of 300 to 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.
- Option B: Choose one technology from the reading (AI, gene editing, autonomous weapons or surveillance) and write a proposal for regulating it. Who should control it, and what limits should exist?
- Option C: Hans Jonas said that we should act so that our actions are "compatible with the permanence of genuine human life." 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 (CEFR B2 to C1)

DURATION

60 to 90 minutes

FOCUS

Reading, ethical analysis,
tribunal role-play

Source material: Hans Jonas; current AI and gene-editing developments; Liu Cixin's Death's End

Series: Humans & The Universe, Lesson 7 of 8. Works as a standalone lesson.

SUGGESTED LESSON FLOW

Warm-up (5 min)	Ask students to name a technology that has made the world both better and worse. Quick brainstorm.
Pre-reading (5 min)	Key vocabulary: alignment problem, opaque, accountable, germline, eugenics, autonomous.
Reading (15 to 20 min)	Parts 1 to 4. The He Jiankui story and the Cheng Xin example are strong triggers for discussion.
Comprehension (10 min)	Section A.
Vocabulary (10 min)	Section B. Ten gaps, word bank supplied.
Debate (10 to 15 min)	Section D on AI rights, which advanced students find very engaging.
Tribunal (15 to 20 min)	Section E. Pairs choose two or three cases and deliver their verdicts to the class. This is the centerpiece activity.
Homework	Section F essay.

ANSWER KEY: SECTION A (COMPREHENSION)

1. Every technology has both beneficial and harmful uses. Any two of: fire (cooking and burning cities), the printing press (knowledge and propaganda), nuclear physics (electricity and the bomb), the internet (connection and surveillance, misinformation and cybercrime).
2. That modern technology has changed the nature of responsibility, because our tools can now destroy civilizations or permanently alter the species. His principle is that we should act only in ways compatible with the continued existence of genuine human life.
3. The problem of making sure that an AI system's goals match human values. An AI told to maximize paper clips might turn all available matter, including people, into paper clips. It would achieve its goal perfectly while destroying everything we care about.
4. Because nobody, including the designers, can fully explain how the system reached a decision. If a decision cannot be explained, it is hard to hold any person responsible for it.

5. Any two of: he edited the germline, so the changes can be inherited; he did so without public debate or agreement among scientists; the editing appears to have been only partly successful; the children did not need it; the parents may not have understood the risks.
6. They change only the patient's own cells and are not passed on to children, and they treat serious diseases under the approval of regulators. He Jiankui edited embryos, which affects the germline.
7. Treatment corrects a disease and restores normal function. Enhancement improves a healthy person beyond the normal range, for example by increasing height, strength or intelligence.
8. It is the gap between the power of a technology and the accountability of the people behind it: important decisions are made without any identifiable person being responsible. Cheng Xin holds the authority to trigger deterrence but cannot bring herself to use it, and catastrophe follows. Her story shows the distance between holding power and having the wisdom, or the will, to use it.

ANSWER KEY: SECTION B (VOCABULARY)

1. opaque 2. autonomous 3. surveillance 4. deterrence 5. accountable 6. germline 7. algorithm
8. safeguards 9. enhancement 10. eugenics

DISCUSSION EXTENSIONS

Film: In *Ex Machina* (2014), an AI manipulates its creator in order to escape. Discuss consent, consciousness and the alignment problem.

Link to Lessons 1 and 5: Lesson 1 showed deterrence resting on one person's judgment, and Lesson 5 asked who gets access to scarce resources. How do both themes return here?

Current events: Have students research the European Union's AI Act, the first major law regulating AI, which entered into force in 2024 and is being phased in over several years. What does it ban, and is it enough?

Science fiction: In *Death's End*, the Swordholder must be willing to destroy two civilizations in order to keep the peace. Should one person ever hold that power?

Section E: There are no correct verdicts. Reward students who say what evidence would change their decision, and who notice that "restrict" needs someone to enforce it.

KEY DISCUSSION LANGUAGE

Ethical judgment

"The ethical issue here is..."

Accountability

"Who is responsible for...?"

Consequences

"If we allow this, then..."

Weighing values

"Is the progress worth the risk?"

Drawing lines

"The line should be drawn at..."

Precedent

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

Sources & Further Reading

Hans Jonas, *The Imperative of Responsibility* (German edition 1979; English translation 1984).

Nick Bostrom, "Ethical Issues in Advanced Artificial Intelligence" (2003). Elon Musk, remarks at the MIT Aeronautics and Astronautics Centennial Symposium, October 2014.

He Jiankui: announcement in November 2018; sentenced in December 2019 to three years in prison and a fine; released in April 2022.

Casegy, the first CRISPR-based therapy, approved in the United Kingdom and the United States in late 2023.

Personalized base-editing therapy for KJ Muldoon, Children's Hospital of Philadelphia, 2025.

Andreas Matthias, "The Responsibility Gap", *Ethics and Information Technology* (2004). Liu Cixin, *Death's End* (2010).

Humans & The Universe

An eight-lesson advanced reading and discussion series

This lesson is part of a connected series for advanced learners. Each lesson pairs a science fiction premise with the real science, history and ethics behind it, and each works as a standalone class or as part of the full sequence.

1	The Dark Forest: game theory, deterrence and the fate of civilizations	
2	Should We Answer? SETI, the Fermi Paradox and the risk of saying hello	
3	Waking Up in the Future: hibernation, time and the cost of survival	
4	Living in Space: survival, adaptation and leaving Earth behind	
5	Scarcity and Survival: resources, rationing and who gets to live	
6	First Contact Protocols: diplomacy, danger and speaking for humanity	
7	The Ethics of Technology: power, responsibility and dangerous tools	<i>you are here</i>
8	Endings and Legacy: extinction, memory and what we leave behind	

Lesson 1, The Dark Forest, is a free download from Tahric Teaches on Teachers Pay Teachers.