



Actua's AI Activities Series

Activity 8

Ethics in AI: Don't Let DANN Turn Evil!

Ethics in AI: Don't Let DANN Turn Evil!

If you're accessing this activity directly, did you know there are eight other activities in this series up on our website? If you find yourself unfamiliar with any of the AI concepts and terminology introduced in these activities, please refer to our [AI Glossary](#). These activities also follow a space exploration narrative when done in order. It is recommended to complete the activities in order but they can also be done on their own.

You and your group-mates are astronauts and scientists aboard the Actua Orbital Station. Unfortunately, your station just got bombarded by magnetic rays and your electronics have begun to shut down! The only one who can save you is the orbital station's AI, DANN. DANN stands for Dedicated Actua Neural Network, and it's gone a little loopy. Brush up on your technical skills, learn about AI, and save yourself and your crewmates! So far, we have read through DANN's manual, rebooted its audio and visual cores, and gotten all of DANN's core functions fixed.

Most of DANN's physical sensors and other functions are back online, like the audio core we fixed in ["Voice Activated AI: Training Audio Recognition Models"](#). But DANN's ethics system has been completely wiped! We have to make sure DANN understands the difference between right and wrong, and doesn't use your data in evil ways. Could you imagine an evil AI? That would be a disaster! Once we can be confident DANN won't turn evil, we can look at fixing its ability to understand us in ["Sentiment Analysis: Understanding the Emotion Behind Text"](#).

Activity Summary

In this activity, participants will learn about morals and ethics, how they apply to the world, and how they apply to artificial intelligence. They will explore data bias, and learn how it can affect AI programs in significant ways. Participants will then use this knowledge to develop their own code of ethics for an artificial intelligence program.

Developed by Actua, 2022.



Delivery Environment	Activity Duration	Intended Audience
Classroom	60 Minutes	Grades 9-12 (Ages 13-18)

Achievement Goals



Learning Goals

Learning goals are statements referring to the understanding, knowledge, skills or application participants acquire during the activity. **Following this activity, participants will:**

- **Examine** a code of ethics
- **Develop** linguistic, grammar and syntax skills
- **Create** their own code of ethics



Logistics (Timing, Group Size, Materials)

Section Title	Time	Group Size	Materials
Opening Hook	10 minutes	<i>Entire group</i>	Facilitators <ul style="list-style-type: none"> • Ethics in AI Slides • Interactive: Survival of The Best Fit
Optional: Intro to AI	5 minutes	<i>Entire group</i>	N/A
Section 1: Ethics Regarding Autonomous Vehicles	15 minutes	<i>Entire group</i>	Facilitators <ul style="list-style-type: none"> •  The Complicated... • Ethics in AI Slides • Post Its Individual <ul style="list-style-type: none"> • Pen/Pencil per participant
Section 2: LG Ethics Case Study	30 minutes	<i>Individual</i>	Facilitators <ul style="list-style-type: none"> • Ethics in AI Slides •  CES 2019-LG Key... (14:42-17:01) Individual <ul style="list-style-type: none"> • Pen/paper per participant
Reflection & Debrief	5 minutes	<i>Entire group</i>	Facilitators <ul style="list-style-type: none"> • Ethics in AI Slides



Safety Considerations

Safety considerations have been provided below to support safety during this activity, however they are not necessarily comprehensive. It is important that you review the activity and your delivery environment to determine any additional safety considerations that you should be implementing for the delivery of these activities.

Online Safety

Some components of this activity require the use of devices connected to the internet.

- Facilitators should review the provided videos and read/explore provided websites and materials to determine if they are suitable for their participants.
- Where applicable, facilitators should remind participants to stay on task and only use links provided within this activity.

Activity Procedure

Opening Hook

Before we can start developing DANN's ethics system, we have to understand what ethics looks like in an artificial intelligence!

1. Start by playing the Survival of the Best Fit game at <https://www.survivalofthebestfit.com/>.
 - a. This game goes through an AI hiring applicants based on what it is taught from a human and can show the bias in the data.
 - b. Play for a few minutes, until you reach the stage where the AI has begun to select resumes.
2. Pause the game for a quick reflection.

Q: What did you notice happening in the game?

A: You began by selecting resumes for applicants, judging them based on four criteria. After a while, an AI was automated to select people for jobs based on your choices, as well as data from a big company.



Q: What happened to the applicants once the AI began to automatically choose them?

A: More applicants of one colour were chosen, and you got accused of data bias.

Q: Why do you think that was happening?

A: It happened because the data became influenced by even a small set of choices. Even though you didn't tell the AI to look at whether a person was orange or blue, it saw a pattern and began to decide based on that.

3. Knowing this, play the rest of the game, then reflect on the rest of it.

Q: Did you notice anything different now? Did the data bias fix itself or get worse?

A: It got worse, because we didn't curate the data; that is, we didn't go through our data and identify any patterns that could lead to data bias.

Q: How is this relevant to real life?

A: We trust AI to make many different decisions in our lives today. Even a small amount of biased data can snowball into a very serious problem. Consider what might happen if an AI that was managing a hospital started favouring some people over others to receive certain treatments.

Today's activity is about the ethics systems that must be developed to make sure that AI follows rules that make sure people don't get hurt.



Optional Section: Intro to AI

If you have done any of our other AI activities, or you are already familiar with AI, then skip this section.

1. Review the following questions:

Q: What is artificial intelligence?

A: Artificial intelligence is the study of trying to create computers that can act with an equivalent intelligence to humans. This can take the form of self-driving cars, smart home assistants like Amazon Alexa, and many other examples in the world today.

Q: Why is artificial intelligence important?

A: Artificial intelligence is everywhere in the world around us. It has applications in every job field. AI is being used for many different purposes, such as job automation or managing huge amounts of data that humans can't process.

Q: How does artificial intelligence work?

A: There are many different ways to achieve artificial intelligence. One of the ways we will be exploring today is called machine learning. This is where programs are given huge amounts of data, and use it to learn and evolve. However, this data can also have some negative effects on the AI, which we are about to explore.

Section 1: Ethics Regarding Autonomous Vehicles

Summary: Learn about the ethics of self-driving cars, and the rules that need to exist to make them safe.

DANN is a complicated AI, so we need to build its ethics system from the ground up! Let's start with a simpler kind of AI to develop a foundation. What if DANN was a self-driving car?



Self-driving cars, also known as autonomous vehicles, have begun to change the world of driving. In previous years, it was the responsibility of drivers to make sure that they knew and followed every single rule of the road. Now, as self-driving cars become more and more developed, that responsibility is instead on the manufacturers. But how do you create an AI that can account for every possible situation that could happen when driving a car?

1. Watch the short YouTube video, [The Complicated Ethics Of Self-Driving Cars | Think | NBC News \(3:32\)](#). This video discusses the ethical dilemmas surrounding autonomous vehicles.
2. Separate everyone into groups of two or three and hand out some post-it notes. Create stations for each of the three questions below, and have groups rotate through the stations. At each station, participants should write down their thoughts about the question on a post-it note, and leave it at the station.
 - a. If two self-driving cars were to crash, who is at fault? One of the users? One of the manufacturers? How do we decide this?
 - i. *Consider the computer code that goes into the cars. What could cause them to crash?*
 - ii. *Consider different possible collisions. What if one car was trying to avoid a pedestrian who was walking across the road? What if the two cars crashing was the safest outcome?*
 - b. Should a vehicle break certain laws if it becomes necessary to get you to your destination? Why or why not?
 - i. *Consider the different rules of the road. There is a big difference between going a little over the speed limit, and ignoring red lights.*
 - ii. *Consider the safety of others on the road. What if your vehicle could break laws without putting other people in danger?*
 - c. This era will effectively make cars supercomputers with lots of information. Should the company who sold you the car be able to collect and use this private information? What if that information could help solve a crime?



- i. Consider what information the car has access to- records of where you've driven, what kind of music you listen to.
 - ii. Consider that most cars connect to your phone as well. Would you be okay with the information on there being sold too?
3. Once the groups have rotated through each station, go through each station for approximately 3-5 minutes and discuss some of the answers.

Section 2: LG Ethics Case Study

Summary: Learn about the future possibilities of AI, and begin to develop a code of ethics for an artificial intelligence.

The time has come! First we need to review what information DANN can learn about, and what tasks DANN can accomplish. Then we will rebuild DANN's ethics system!

1. Run this clip from LG: [CES-2019 Keynote Speech](#) (14:42-17:01)
2. After watching the video, you will be responsible for developing your own AI Code of Ethics.

Q: What is a code of ethics?

A: A code of ethics for artificial intelligence is a set of ethical rules that the AI has to follow. The most famous of these is known as Asimov's Law, after a famous science fiction writer. The law is: "*A computer may not injure a human being or, through inaction, allow a human being to come to harm.*" A code of ethics is made of rules just like this one!

To make a code of ethics, you must come up with a set of four specific rules of things that DANN should or shouldn't be allowed to do. For each rule, write one or two sentences about why you think this rule is necessary for DANN, and what could happen if DANN didn't have that rule. Think about the video, and which parts of the video didn't feel right to you, and why not. That's a great starting point for your code of ethics!



Appendix B contains a template that you can use to write out your code of ethics. It also has an example code from LG Electronics, paraphrased from their public code of ethics.

This code is a much more thorough and complex version of what you will be making. You could come up with the following rule, for example:

- a. Artificial Intelligence isn't allowed to access information about my messages or phone calls with my friends or family.
 - b. **Why?** I get to choose what information I want to give to my AI, but I don't get to decide that for my friends or family. If I didn't have this rule, my AI could access information from my friends, and then their friends, and none of them would be able to say that they didn't want their information to be used like that.
3. Take 10-15 minutes to think about what DANN might need to do to run a space station, and develop a code of ethics for it. Once you have finished, reflect on each rule in your code using the following questions:
 - a. Does this rule address privacy concerns?
 - b. Does it address legal concerns?
 - c. Is it meant to protect people who are using the AI, or the people who made it?
 - d. Are there any loopholes in this policy or ways to get around it?
4. Congratulations! Now, let's apply your code of ethics to a very real situation.
 - a. *A new company is launching online and they need to find a way to get people to buy their products. They could pay for general advertising, like a TV commercial, but this would be expensive and probably not too effective. They decide to contact Facebook and pay them for browsing information on Facebook users, in order to help them target the specific customers that would want to buy their product.*

Imagine that you are Facebook in this scenario. Based on the code of ethics you just created, would you agree to sell your users' information? Why or why not?



Reflection & Debrief

Congratulations! We have used our combined codes to recreate DANN's ethics system, and we can relax knowing that DANN knows what it should and shouldn't do.

Reflect on what you've learned today with the prompting questions below:

Q: How are ethics and morals different?

A: Morals are an internal sense of right and wrong. A set of ethics is something multiple people agree on.

Q: Why is a code of ethics important in everyday life?

A: Without a code of ethics, each person's morals could be very different, and people wouldn't be able to rely on everyone else believing things like "murder is bad."

Q: Why do artificial intelligence programs need a code of ethics?

A: Unlike humans, AI programs don't automatically have morals. They have no sense of what is right or wrong, so it is very important to establish a code of ethics to ensure the AI won't do anything that might cause harm to someone.

Q: What is data bias?

A: Data bias is when the data given to an AI has some sort of skew to it. The AI learns from that data, and the bias becomes a bigger and bigger problem. That's why curating data and making sure it's as unbiased as possible is very important.

Q: Why is it important to avoid data bias?

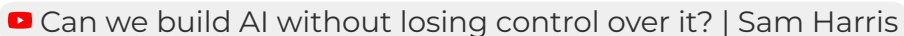
A: Data bias can result in people being ignored or hurt. Think back to the survival of the best fit game. Some people weren't getting a fair chance at jobs because of data bias! It can very easily happen unless we are careful about our data.



Delivery Adaptations

How might you adapt the time, space, materials, group sizes, or instructions to make this activity more approachable or more challenging? **Modifications** are ways to make the activity more accessible, **extensions** are ways to make the activity last longer or more challenging.

Modifications

1. The following extension activity can be used in Section 2 - LG Ethics Case Study:
 - a. Have the participants create their own potential scenario, just like the facebook one.
 - b. Give them time to create the scenario, then pair them up with another group.
 - c. Have the two groups swap case studies, and see what they would do based on their code of ethics.
 - d. If needed, continue to change around the groups to explore different scenarios.
2. For further extension, watch and discuss the following video:


Extensions

- In a group setting, the Code of Ethics activity can be done in groups of any size. Have each group develop a code of ethics together, and then have them all present and discuss their codes with other groups.
- In a self-managed exploration of this module, instead write down reflections on the questions for section 1: The ethics of self-driving cars. There is no need for post-it notes.



- In a virtual environment, Google Jamboard can be used in place of post-it stations for section 1. Breakout rooms or similar ways of dividing students can be used to simulate groups.
- If participants don't know how to start their code of ethics, give them 2-3 scenarios (e.g., refer to the video or make up your own) and ask them if your rules are ethical or not.
- If participants are having trouble grasping the idea of manufacturer responsibility vs individual responsibility, give them the example of buying a broken product. If you buy a bike with no brakes and then you get injured, that is the manufacturer's fault, even though you were operating it because it was defective.
- If a participant continues to struggle, pair them up with another participant who has a good grasp on the concepts and encourage collaboration.

References & Gratitude

The Survival of the BestGame was used to help make this project and can be found at <https://www.survivalofthebestfit.com/>

The LG code of ethics was used as an example for this project and can be found at <https://docs.google.com/document/d/1-iyHhe7HzDc6iZPN8zGQKTCZHRrNs4whBu6MsVXRoGU/edit?usp=sharing>



Terms of Use

Prior to using this activity or parts thereof, you agree and understand that:

- It is your responsibility to review all aspects of this activity and ensure safety measures are in place for the protection of all involved parties.
- Any safety precautions contained in the “Safety Considerations” section of this write-up are not intended as a complete list or to replace your own safety review process.
- Actua shall not be responsible or liable for any damage that may occur due to your use of this content.
- This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. For more information, please see <https://creativecommons.org/licenses/by-nc-sa/4.0/>.
- You may adapt the content for your program (remix, transform, and build upon the material), providing appropriate credit to Actua and indicating if changes were made. No sharing of content with third parties without written permission from Actua.

About Actua

Actua is Canada’s leading science, technology, engineering and mathematics (STEM) youth outreach network, representing a growing network of over 40 universities and colleges across the country. Each year 350,000 young Canadians in over 500 communities nationwide are inspired through hands-on educational workshops, camps and community outreach initiatives. Actua focuses on the engagement of underrepresented youth through specialized programs for Indigenous youth, girls and young women, at-risk youth and youth living in Northern and remote communities. For more information, please visit us online at www.actua.ca and on social media: [Twitter](#), [Facebook](#), [Instagram](#) and [YouTube](#)!



Appendices

Appendix A: Career & Mentor Connections

- AI / Machine Learning researcher & programmer
- Computer Programmer
- Engineer
- Economist
- Trader
- Law
- Medicine

Appendix B: Background Information

ETHICS

Ethics is an area of study that has been evolving since ancient times. When we hear about ethics in the modern day it is usually concerned with law or medicine and the practices that every individual in those fields should uphold. With the advent of the 21st century and rapid technological growth our laws and regulations have fallen behind and so there are ethical concerns with drastic consequences we haven't even begun to think about. Artificial intelligence is one of these areas and that is why we will be exploring ethics as it relates to artificial intelligence.

Many people confuse ethics and morality. Morality is what someone holds to be right and wrong personally. Ethics is an agreed upon standard of right and wrong for a particular group or even society.

AUTONOMOUS VEHICLES

Self-driving vehicles have been increasing exponentially in popularity over the last decade. Tesla is the biggest player in this space and exclusively makes electric vehicles with autonomous capabilities. When these vehicles first came out they were limited to self-parking and simple things like that. The technology is now moving very, very quickly and a fully functioning self-driving car as good and even better than a human is the modern reality.



Appendix C: Supporting Resources

Ethics Template

Code of Ethics:

1.
 - I.
 - II.
 - III.
2.
 - I.
 - II.
 - III.
3.
 - I.
 - II.
 - III.
4.
 - I.
 - II.
 - III.



LG Code of Ethics Example:

1. We Will Ensure Privacy As a Top Priority
 - I. LG will never collect and store any of the information it gathers from its customers for commercial use.
 - II. All data collected from cameras will be stored securely on a server only obtainable by yourself and the systems you run.
 - III. At customer request we will destroy all data on the server.

2. We Will Not Allow AI To Perform Any Spending Without Permission
 - I. LG's AI will not buy anything for the home such as groceries without receiving your permission first.
 - II. Before our AI uses any home appliances automatically such as a stove or refrigerator the customer will have to approve it first.
 - III. LG will allow its customer to override AI permissions at any time.

3. We Will Not Use Our Customers as Revenue
 - I. Our AI will make recommendations for things such as clothing or food in an unbiased manner.
 - II. Anything recommended that has a partnership with LG will be flagged as such so the customer is aware.
 - III. LG will not sell any customer data to third parties.

4. We Will Not Use AI To Mislead Customers
 - I. Our AI will not use things like virtual fittings to make appearances look better than they actually are.
 - II. We will not show customers media that is biased towards making LG or our partners look better

