AI and Ethics Teacher Guide

AI and Ethics
AI is playing an ever-greater role in society. This AI and Ethics lesson will look at how AI affects the wider public and society, and what kind of ethical issues we might consider when developing such technologies.

Time: ~1 hour*

If your classes are 40 mins. long rather than 1 hour, we recommend that you select the topics you feel are most important and/or shorten some of the interactive activities.

Background
Although there are no prerequisites for this module, it would be useful to cover the *Introduction to AI* module with students first, so that they already understand AI and some of its characteristics.

Curriculum Links
The interactive and reflective nature of AI in My Life ensures that students will hone the five key skills central to teaching and learning across the Transition Year curriculum:
- Information processing
- Critical and creative thinking
- Communicating
- Working with others
- Being personally effective

An outline of links to the Leaving Certificate curriculum is provided after the module walkthrough.

Materials Needed: PowerPoint presentation, Laptop, Screen, Internet Access, Timer

Module Overview

Introduction (~3 mins.)
A quick introduction to the module, outlining learning goals and what is Artificial Intelligence.

AI and Ethics (~16 mins.)
A brief look (through video examples) at what ethics means and why ethics matters for AI.

Responsible AI Development (~9 mins.)
Explores what rules should we consider in order to produce responsible AI.

Ethical Use of AI (~24 mins.)
Shows that users of AI – like the developers of AI – also need to act in an ethical way.

The Impact of AI (~6 mins.)
Outlines some of the positive and negative impacts of AI - finding that balance is vital.

Recap (~2 mins.)
A brief recap on what students have learned in this module and an opportunity to ask questions about topics they would like to explore more.

AI and Ethics
Module Walkthrough

Learning Intentions

By the end of this workshop, students will be able to:
- Give examples of how AI is used
- State some of the ethical and privacy implications of AI
- Evaluate the role of AI in their lives and in society

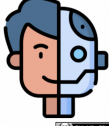| **Introduction (~3 mins.)**	
	Lecture (~1 min) Tell students that AI in My Life is a modular workshop series on Artificial Intelligence (AI). It was developed by the Science Foundation Ireland ADAPT Research Centre. ADAPT is a major academia-industry research centre with more than 300 researchers from 8 Irish universities (led by Trinity College Dublin and co-hosted by Dublin City University) producing research and innovations in many aspects of AI.
	Lecture (~1 min) Outline to students what they'll be able to do by the end of this module: • Give examples of how AI is used • State some of the ethical and privacy implications of AI • Evaluate the role of AI in their lives and in society
	Lecture (~1 mins) Go through the following definition, checking that students understand: *Artificial intelligence refers to many different technologies that are designed to accomplish tasks that can otherwise be only carried out by human minds, for example identifying patterns, recognising images, understanding languages or solving problems.* This will be a recap for students who have completed the Introduction to AI module.
AI and Ethics (~16 mins.)	

Slide	Notes
Ethics and AI — Davos 2017	Lecture (~3 mins) Watch this news video (which lasts 2 minutes 36 seconds) from The World Economic Forum meeting in Davos, Switzerland in 2017 which outlines various components of the AI and Ethics debate. What are the questions this video throws up? This video also shows that the questions have been raised some time ago and not just recently with the introduction of the latest AI advancements such as ChatGPT.
Let's start *What do we mean by "ethics"?*	Lecture (~3 mins) Ask the students about what they understand when they hear the word "ethics". List out the different ideas which they come up with. You might wish to use an interactive tool such as Kahoot, Mentimeter or Sli.do to create a word cloud with the class's suggestions.
What do we mean by "ethics"? *Our **understanding** of **what is right and wrong,** as well as the **rules** that individuals and the society should follow.* 	Lecture (~2 mins) Provide a definition for what we mean by ethics: "*Our understanding of what is right and wrong, as well as the rules that individuals and the society should follow.*" If asked about the difference between morals and ethics you can say that ethics is more about the rules and codes and the discussion about what is right and wrong, whereas morality is about an individual's/group's sense of what is right and wrong. Morality is what we have internally, whereas ethics is the branch of knowledge (or discussion) that studies this. However, these terms are also often used interchangeably.
What do we mean by "AI ethics"? AI ethics refers to the **moral principles and values** that should guide the development and use of artificial intelligence systems. AI systems should be designed and used in a way that **benefits society** and **follows ethical principles**. Issues like **fairness, transparency, privacy** and **safety** are important ethical considerations. 	Lecture (~2 mins): Proceed by explaining what we mean by "AI ethics": ● AI ethics refers to the moral principles and values that should guide the development and use of artificial intelligence systems. ● AI systems should be designed and used in a way that benefits society and follows ethical principles. ● Issues like fairness, transparency, privacy and safety are important ethical considerations.

Why does ethics matter for AI? The Telegraph 'Unethical' judge asks ChatGPT for advice on child's medical rights **THE IRISH TIMES** Advance of AI creating a moral minefield Artificial intelligence is now capable of making decisions without any human intervention, so how do we navigate the ethical issues raised? sky news 'Regulate it before we're all finished': Musicians react to AI songs flooding the internet This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	<u>Lecture (~2 mins)</u> Why does ethics matter? Show examples of media excerpts which indicate how this debate has been increasingly present in public awareness. They are examples of headlines from: ● *The Telegraph* article about ChatGPT use by a judge ● *The Irish Times article* about AI's decision making capability ● *Sky News article* about musicians feeling threatened by AI-generated music
Why does ethics matter for AI? • AI has the **potential for immense impact**, both positive and negative. • AI systems are increasingly being **entrusted with high-stakes decisions** that affect human lives and society. • AI learns from data, which can **encode human biases and prejudices**. Unethical applications could automate or worsen injustice. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	<u>Lecture (~2 mins)</u> ● Explain that AI has the potential for both positive and negative impact. ● However, as AI systems improve, it appears that they are being entrusted with more decision-making capabilities that affect humans and society. ● Ultimately, AI learns from the data on which it is trained. The origin of this training data is therefore very important, as it may lead systems to make judgements that are biased or prejudiced.
Why does ethics matter for AI? • As AI gets more capable and autonomous, we need to **maintain human accountability and control**. • Ethical issues around AI use, like privacy invasion or job losses, require careful consideration to **avoid public backlash**. *Ethics helps ensure AI development and use is guided in a socially responsible direction.* This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	<u>Lecture (~2 mins)</u> ● As AI capability becomes more powerful, we need to make sure that there is human accountability and control. ● Ethical issues around AI use require careful consideration and discussion. Ethical consideration needs to be central to AI development in order to ensure it is guided in a socially responsible direction. ● Ethics is not just about technical solutions; it's also about the values that we want reflected in the AI systems/tools we create e.g., corporate profit v social solidarity - we have to actively decide what we want to achieve as a final outcome.

Responsible AI Development (~9 mins.)	
What rules would you consider? You are given the task of creating a **set of rules for** a scientist who is **creating an AI bot**. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	<u>Activity (~4 mins)</u> We will now look at what developers may have to consider when developing AI systems e.g. not to cause harm, listening to humans as the final decision maker, etc. What kind of rules do students think should govern AI research and development? Ask the group to suggest what kind of rules they would give a scientist who is creating an AI bot.
Ethical Considerations for AI Development (1 of 2) ● **Human Agency and Oversight** All decisions are reviewed by humans and humans are the "in charge" of final decisions. ● **Technical Robustness and Safety** AI tools do not contain errors and their operation does not lead to harm. ● **Privacy and Data Governance** There are strict rules about people's data and how AI is able to use it. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	<u>Lecture (~2 mins)</u> We will now look at the seven ethical considerations which should be taken into account during the development of AI systems. These are EU guidelines for Trustworthy AI. It is not the only possible list, but this is what we mainly work with within the EU (and what will become law). The first three are: 1. **Human Agency and Oversight** - making sure all decisions are made by humans and humans are in charge of the final decision 2. **Technical Robustness and Safety** - making sure that AI systems/tools do not have any errors (bugs) and that their operation would not lead to harm 3. **Privacy and Data Governance** - the rules associated with how people's data is used
Ethical Considerations for AI Development (2 of 2) ● **Diversity, Non-Discrimination and Fairness** AI should work equally well for different people and in different contexts. The benefits of AI should be shared equally ● **Societal and Environmental Well Being** AI should impact the environment and the society in a positive way. ● **Transparency** It is possible to say what decisions AI makes, what steps led it to those decisions and what data informed them ● **Accountability** We know who is responsible for the AI and we can fix problems or enforce sanctions when something goes wrong. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	<u>Lecture (~3 mins)</u> The final four ethical considerations for AI development are: 4. **Diversity, Non-Discrimination and Fairness** - making sure that AI works equally for different people in different contexts and there should be no bias related to gender, ethnicity etc. 5. **Transparency** - it should be possible to see what AI decisions are made, the steps along this decision making process and the data that informed them 6. **Societal and Environmental Well Being** - AI should positively impact the environment and society 7. **Accountability** - it should be obvious as to who is responsible for the AI so that problems can be fixed or sanctions can be imposed.

Ethical Use of AI (~24 mins)

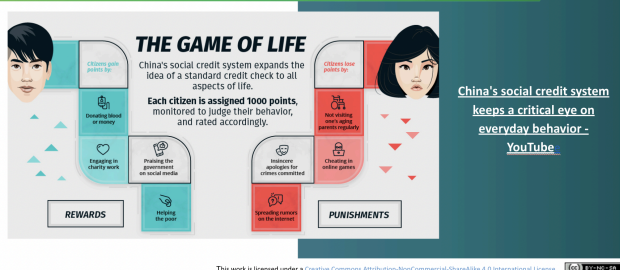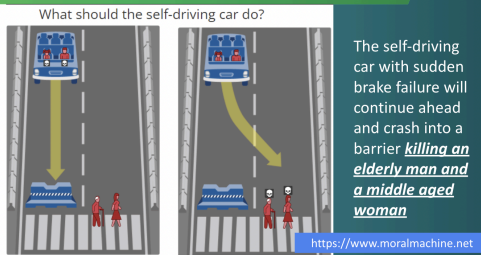|	
Scenario A: ChatGPT and school assignments This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	**Scenario A** Activity (~3 mins) This video (length 2 minutes and 4 seconds) explores whether it's ethical to use generative AI systems like ChatGPT for school assignments. If you have difficulty accessing the video, play the first 2 mins. of the full-length version on YouTube: https://www.youtube.com/watch?v=Fn8jDanbf0c
Questions for Scenario A • Should you ask AI to help with your homework? • How is this different to working with other (real) people or asking a parent, sibling, or friend for help? • How to establish trust between students and teachers in a world where AI tools like ChatGPT exist? This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	**Scenario A** Discussion (~5 mins) Having watched the video, ask the class the following questions: ● Should you ask AI to help with your homework? ● How is this different to working with other (real) people or asking a parent, sibling, or friend for help? ● How to establish trust between students and teachers in a world where AI tools like ChatGPT exist? These questions are posed to tease out the differences that students see between using AI rather than asking another person for assistance.
OR	
Scenario B China's Social Credit System THE GAME OF LIFE China's social credit system expands the idea of a standard credit check to all aspects of life. Each citizen is assigned 1000 points, monitored to judge their behavior, and rated accordingly. Citizens gain points by: Donating blood or money, Engaging in charity work, Praising the government on social media Citizens lose points by: Not visiting one's aging parents regularly, Insincere apologies for crimes committed, Cheating in online games, Spreading rumors on the internet REWARDS PUNISHMENTS China's social credit system keeps a critical eye on everyday behavior - YouTube This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	**Scenario B** Activity (~3 mins) Next play the video about the Social Credit System in China (Length: 2 minutes and 36 seconds). This 'Game of Life' concept means that members of society could earn points for performing certain altruistic tasks such as donating blood or helping the poor but simultaneously could lose points when caught engaging in morally offensive activities such as cheating on video games or not visiting elderly family members.

Questions for Scenario B

- Within months of early trials of this system being launched in 2014, more than 14 million people opted in. Why do you think this is?

- How would you feel about such a system being introduced in Ireland? Would you opt in?

- Should researchers (in universities or in companies) have to consider the potential uses to which the innovations they develop might be put? Should they be held accountable?

Scenario B

Discussion (~5 mins)

Ask the second group to consider the following questions:
- Within months of early trials of this system being launched in 2014, more than 14 million people opted in. Why do you think this is?
- How would you feel about such a system being introduced in Ireland? Would you opt in?
- Should researchers (in universities or in companies) have to consider the potential uses to which the innovations they develop might be put? Should they be held accountable?

These questions are asked to tease out what may and may not be acceptable in different cultures when it comes to society's use of AI.

Ethical Decisions of AI

What should the self-driving car do?

The self-driving car with sudden brake failure will continue ahead and crash into a barrier _killing a middle aged man and his young daughter_

The self-driving car with sudden brake failure will continue ahead and crash into a barrier _killing an elderly man and a middle aged woman_

https://www.moralmachine.net

Discussion (~10 mins)

***Note: If you haven't done the AI in My Life taster module, this fun activity takes about 10 mins. to complete and is a great example of the kinds of ethical decisions posed by AI development. If you've already completed this activity, we suggest that you invite students to create their own scenarios to the Moral Machine.

This activity challenges students to imagine themselves as an AI developer programming a self-driving car. In the event of an impending unavoidable accident, how would they programme the car to decide what action it should take? Which of these two possible scenarios, in their minds, is the lesser of two evils?
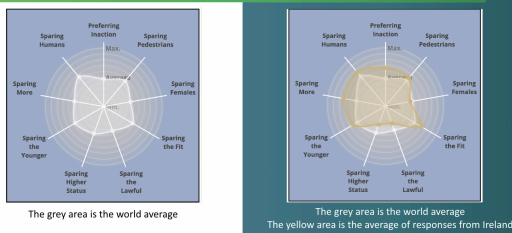
Researchers in the USA have built a platform for gathering a human perspective on moral decisions made by machine intelligence, such as self-driving cars. They call the platform "The Moral Machine". Explain to students that this is an "edge case" or extreme scenario as, in reality, there are lots of other actions, such as braking, that a car could take and there is more to making ethical decisions than choosing who gets harmed.

Separate students into groups of 4 and ask them to discuss and come to an agreement on which is their "least undesirable" course of action for the car. Ask them to report

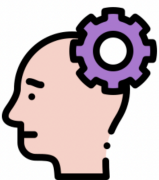|	
	back to the whole class on their decision and the rationale for it. Then ask students how they found the activity, did they come to agreement quickly or were there different opinions in the group. Did they ever think about the kind of ethical decisions that AI developers face? There are hundreds of these scenarios on the Moral Machine website. You can view and "judge" them yourself. You can also submit your own scenarios.
Ethical Decisions of AI: How does Ireland compare? The grey area is the world average The grey area is the world average The yellow area is the average of responses from Ireland https://www.moralmachine.net This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	Lecture (~3 mins) In the previous discussions about the cultural aspects to ethics, you would have touched on how different cultures can treat different cohorts of society as more or less important. The class may be interested in seeing how Irish preferences compare to people across the world. The Moral Machine shows results from people judging scenarios on its website. 1st image: We can see the world average here. The further to the outside of the circle a dot is, the more people preferred that course of action. 2nd image: How does Ireland compare? The yellow area shows that people in Ireland generally show a greater preference for sparing humans, preferring inaction, and sparing the fit, and they are less inclined to spare those with higher social status and those obeying the law. What do you think this tells us about people in Ireland?
Ethical Decisions of AI: How does Ireland compare? How contributions from Ireland compare to other countries https://www.moralmachine.net This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	Lecture (~3 mins) This slide goes into further detail as to how Ireland ranks out of 117 countries on various courses of action. "Sparing more" appears to be the most important factor while "sparing the lawful" or "sparing higher status" appear to be the least popular. Do these statistics surprise anyone in the group? These notable differences again suggest that there's an important cultural element to bias. This makes it even more important to ensure that those developing and regulating AI systems come from diverse backgrounds in order to help minimise potential bias.

The Impact of AI (~6 mins)

Lecture (~2 mins)

- Automation has been seen as a threat to manual labour in the past. Could robots replace factory workers and farmers, or self-driving cars come after human drivers' jobs?
- AI can also be trained to do more intellectual tasks such as translation, programming, writing, medical diagnosis or teaching.
- The ethics of AI also involves the discussion of wider impacts of the technology. How do we distribute the benefits created by automation fairly? How to ensure that entire communities are not left behind? AI has the potential to transform the world's economy and we have an ethical obligation to make this transformation fair.

Lecture (~2 mins)

- AI is not magic. Each operation made by AI consumes energy and resources, which has a large impact on the environment and we should be mindful of this as we find ourselves in the middle of a climate crisis. Data used by AI is stored and processed in enormous data centres comprising millions of servers and computers.
- Data centres accounted for 18% of Irish energy consumption, it is worth stressing that these are all data centres and not **only** AI.
- As the servers require cooling, it is estimated that ChatGPT uses half a litre of water for every 5 to 50 prompts it answers! While ChatGPT is being lauded as a revolutionary tool, there has been little discussion about the computing power required to train such a system i.e. the training of GPT-3 was estimated to generate 85,000 kg of CO2, or the equivalent of driving a new car to the Moon and back: https://www.theregister.com/2020/11/04/gpt3_carbon_footprint_estimate
- Moreover, the material costs of AI should not be underestimated, especially as many of the resources are imported from developing countries and are not sourced in an ethical or

	sustainable way ● Can we justify such energy use today? Will the benefits outweigh the drain on resources?
	Lecture (~2 mins) Show the slide with the following questions: ● Who should be blamed when decisions made by AI lead to harm? ● What are the differences between the rules that should apply to ethical AI and those that govern our own ethical behaviour? ● To what extent can we ignore the risks of AI to enjoy the benefits it provides? ● What can we do as citizens to influence how AI is used? We don't have to provide answers but rather can leave students with these questions to consider. The intention is to encourage students to ponder some of the many ethical considerations of AI.
Recap and Conclusions (2 mins)	
	Lecture (~1.5 mins) What have we learned? We revisit the three learning outcomes outlined at the start of the session: ● We know what AI is and how its use can present ethical questions ● We have uncovered some of the ethical and privacy implications of AI - those developing AI AND those using AI need to act ethically ● We have looked at the role of AI in our lives and in society and understand the importance of striking the right balance as there are negative and positive impacts
	Lecture (.5 min) Final slide which explains that there are more AI in My Life modules available if the students would be interested.

Curriculum Links

In addition to its relevance to honing the key skills central to teaching and learning across the Transition Year curriculum, the STEAM focus makes the AI in My Life content relevant to the following subjects:

Leaving Certificate:
Technology
Politics and Society
Computer Science
Design and Communication
Business
Economics

Applied Leaving Certificate:
Engineering
Technology
Social Education
Science
Information and Communication Technology – Specialism

If you have questions or comments about this lesson,
please contact us at education@adaptcentre.ie