AI ethics –What's possible, probable, and preferred?

*Please refer to the online lesson plan on the DT Hub to access all website links and additional resources.*

**Years 9-10:** The development and ubiquity of Artificial Intelligence raise a number of social and ethical matters that students can explore in the Digital Technologies classroom.

This lesson idea outlines a project to help students frame such discussions using the curriculum Key Idea of **Creating preferred futures**, tying into Critical and Creative Thinking.

‘As students progress through the Technologies curriculum, they will begin to identify possible and probable futures, and their preferences for the future. They develop solutions to meet needs considering impacts on liveability, economic prosperity and environmental sustainability. Students will learn to recognise that views about the priority of the benefits and risks will vary and that preferred futures are contested.’

In the Digital Technologies curriculum, these discussions address content descriptors dealing with **impact** (see Curriculum links for more detail).

*This lesson idea is distilled from the Case Study What would my preferred AI future look like?, describing a project undertaken by a Year 10 class at Roseville College.*

# Curriculum links

Links with Digital Technologies Curriculum Area

| **Strand** | **Year** | **Content Description** |
| --- | --- | --- |
| Processes and Production Skills | Year 9–10 | evaluate existing and student solutions against the design criteria, user stories, possible future impact and opportunities for enterprise (AC9TDI10P10) |
| select and use emerging digital tools and advanced features to create and communicate interactive content for a diverse audience (AC9TDI10P11) |

# Assessment

In addition to an assessment of the quality and fullness of the student-made case study and interactive plan, the SOLO taxonomy may be useful for assessing the depth of student understanding demonstrated. Dimensions for a rubric may include:

* possible/proposed purpose and function of AI application
* possible social and ethical implications of AI application
* preferred future plan for AI application
* justification of development for AI application.

Students can also undertake a self-reflection, journal or peer review for the project.

# Learning hook: the hard numbers

Authorities and experts must sometimes make hard choices in which human lives become part of a mathematical equation.

* Engineers must decide on a bridge design where a lower percentage chance of a collapse is weighed against unreasonable cost.
* Politicians and epidemiologists must enact policies that weigh economic factors against deaths during a pandemic.

Now, designers and programmers of self-driving cars must make choices too.

Try out the Moral Machine at MIT's website. Starting by watching the video linked on the online lesson plan, then choose **Start Judging**.

DISCUSS: How did you choose your judgements in each scenario? Did you have a common set of principles? Did you change or update them with each scenario?

The Moral Machine is not dissimilar to the famous ethical thought experiment called the trolley problem (Wikipedia). **But it is more than a thought experiment.** Should self-driving cars become mainstream, it is inevitable that people will continue to die in road incidents, even if their overall frequency is reduced.

DISCUSS: Who is legally/ethically to blame when the self-driving car makes its ‘moral’ judgement during a road incident? The programmer? The pedestrian? The driver?

# Project outline

The project includes the following steps:

1. **Define** and **explore** AI.
2. Determine an AI **area of interest** for researching.
3. Do a deep dive in the form of a **case study** into a specific, emerging AI application or system.
4. Using the persona of a fictional universe character, make a **plausible plan** of a preferred AI future with the AI application or system you researched, and **justify** whether or not to further develop this technology.

You may wish to use or modify this student handout for the project.

# Step 1: Define and explore AI

Using the DT Hub resources on the Artificial Intelligence topic page, and the AI lesson plans, begin by exploring how we define AI and some of its applications. Classes that have already done this may choose to skip this step.

# Step 2: Determine an AI **area of interest** for researching

Student choice should not only be informed by personal interests, but also by the availability of research and ease of comprehension, in terms of concepts as well as language used.

It is important to find resources that show what is on the cusp of the public domain to facilitate extrapolation into what is probable in the future.

*TIP: For this step and Step 3, research may benefit from the involvement of a teacher-librarian, as was done in the Case Study What would my preferred AI future look like? undertaken by a Year 10 class at Roseville College.*

Areas of interest may range from current real-world problems to philosophical theory, including but not limited to:

* facial recognition technology and its applications
* self-driving cars and moral decision making
* ‘algorithmic bias’ and other problems associated with machine learning applications
* recommendation algorithms (for media, products and other applications) and their impact
* military and geopolitical applications of AI
* AI as a tool for addressing positive environmental and social causes
* the effect of artificial intelligences on human relationships
* consciousness and the legal/moral status of artificial intelligences.

# Step 3: Do a **case study** of a specific AI application or system

The case study should include:

* purpose and target audience of the AI application or system
* components: hardware, software, people and procedures
* what makes the application or system innovative or contributing to future innovations
* an analysis of *possible* and *probable* social and ethical implications, including sustainability.

The purpose of the case study is to hone into particular applications of AI to facilitate analysis of the technology’s impact on individuals and society. The plan is not to generalise but rather, in specificity, to get a better understanding of the issues at play when using or developing emerging technologies. Focusing on what is already or almost happening gives a starting point to more easily imagine what could happen.

# Step 4: Make a **plausible plan** of a preferred AI future

In this step, students choose a fictional Marvel Universe character who may be the developer and/or user of the AI application or system they researched. Pretend this AI future is real, in the Marvel universe.

In an interactive, multimedia form, create and explain a plausible plan for the preferred future for the use of this AI.

In the persona of your chosen Marvel character, include a justified conclusion on whether or not to further develop or use the technology.

## A note on multimedia

The following might be used to create interactive solutions:

* A **website** with multiple pages and/or interactive elements. Pages can include video or audio clips for a multimedia experience. *Google Sites and other free website tools are available, but consider your school's privacy requirements.*
* An **interactive slideshow** allows users to click to navigate to specific slides by clicking a button or link in the presentation. This can be combined with voiced narration or movie clips for a multimedia experience. *Common presentation software like Microsoft PowerPoint or Google Slides allows for user navigation via links or buttons.*
* **Dynamic data visualisations** allow users to change or filter the display of visual information from data by clicking on buttons or other interface elements. *This can be achieved in common spreadsheet software like Microsoft Excel or Google Sites, or with specialised online tools like Tableau.*
* Applications like Microsoft Sway combine some of the elements of the interactivity in the categories above, while removing some of the time-consuming layout work.

For examples of non-interactive multimedia works as well as traditional written reports for a similar project, see the case study What would my preferred AI future look like? undertaken by a Year 10 class at Roseville College.

# Resources

* AI ethics principles from the Department of Industry, Science, Energy and Resources
* Case study What would my preferred AI future look like?, undertaken by a Year 10 class at Roseville College
* MIT Moral Machine
* Digital Technologies Hub Artificial Intelligence topic page, and AI lesson plans