

## Assignment 1 – Research into AI Ethics & Responsible AI

- **Due** 11:59pm Sunday 13 April 2025 (Week 6)
- **Contributes** 20% to your final subject result, subject to moderation if required.
- **Individual**.

*This assignment requires you to work individually to conduct research into the issues around AI Ethics and Responsible AI.*

This assignment requires you to conduct research into the issues around AI Ethics and Responsible AI. Students will take full responsibility for performing the research while the teaching staff will be their facilitator and weekly check the student progress on the assignment. Students are encouraged to use a Github repo or OneDrive to provide progress updates to their respective tutor.

### Background

The starting point for your research will be the following two documents (D1 and D2):

D1. Burton, E., Goldsmith, J., Koenig, S., Kuipers, B., Mattei, N. and Walsh, T., 2017. Ethical considerations in artificial intelligence courses. *AI magazine*, 38(2), pp.22-34.

<https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/download/2731/2632>

D2. Australia's government proposal paper on "Introducing mandatory guardrails for AI in high-risk settings: proposals paper" (2024): <https://consult.industry.gov.au/ai-mandatory-guardrails>  
(Note also that the Australia's government has made substantial initiative on this topic and further info/documentation can be found: [https://consult.industry.gov.au/find-consultations?labels=mct\\_categories.mlb1a14785fef72b741661fb](https://consult.industry.gov.au/find-consultations?labels=mct_categories.mlb1a14785fef72b741661fb))

After reading the above two documents, you should form an understanding on the topics as well as more general issues related to AI Ethics and Responsible AI. In addition to these two documents, you are encouraged to broaden your knowledge by using reliable sources such as books and peer-reviewed journal and magazine articles. You can also use suitable Internet sources but you must ensure that they are reliable sources of information.

### The Research Topics

Once you have gained a solid knowledge and understanding of the area, you need to focus on a specific topic in this area. A list of topics based on the two base documents D1 & D2 includes:

1. How Should AIs Behave in Our Society?
2. What Should We Do If Jobs Are in Short Supply?
3. Should AI Systems Be Allowed to Kill?
4. Should We Worry About Superintelligence and the Singularity?
5. How Should We Treat AIs?

Document D2 introduces a list of proposed principles (see pages 19-25 of D2) defined for high-risk AI systems (called PPs herein).

6. Do the PPs adequately capture high-risk AI? Would you recommend adding or removing any principles?
7. Are there ways to define the principles so that they better capture harms to First Nations people, communities and Country or other minority groups?
8. Are there other alternatives for defining and capturing high-risk AI systems? What are the pros and cons of such alternatives, especially when comparing them with the approach in D2?
9. Are the PPs flexible enough to capture new and emerging forms of high-risk AI, including general-purpose AI (GPAI)/Artificial General Intelligence (AGI)?

The above list of topics is not exhaustive and there are many topics that are very relevant and important. For instance, document D2 lists 16 questions for consultation. Each question can be a topic for your research. Moreover, if you identify a relevant topic that you are interested in, please feel free to discuss with your tutor to gain their permission for the topic.

## Objectives of the Assignment

On successful completion of this assignment, you will be able to demonstrate your ability to conduct research into a complex topic and present your research outcomes in a well-written report. More specifically, your report will:

1. *Identify the key issues in the question and explain them clearly and logically*; that is, you are able to understand the area, the topic and the associated problem(s) to be researched to the point that you can explain them clearly.
2. *Demonstrate the ability to find and collect the relevant information, analyse and synthesize information and put them together to support your argument and conclusion*; that is, you are able to conduct systematic research to find the relevant information from reliable sources to provide comprehensive answers and analysis on the issues.
3. *Select and present information clearly and coherently to your target audience*; that is, you can present your findings and the outcomes of your research in a convincing manner to ensure that the issues have been addressed satisfactorily with clear conclusions.

## Tasks and Methodology

1. **Week 2-3:** Read the base documents D1 and D2; Get a good understanding of the area and identify the topic/question of your choice; Identify the key issues in the question; Start writing the report, especially the Introduction (approx. 500-800 words) to capture this understanding & analysis.
2. **Week 3-4:** Conduct systematic research to find the relevant information from reliable sources to address the key issues in the question; Analyse and synthesize information and form your arguments with a logical structure for your findings: Prepare a draft for these sections of your report (approx. 1,800-2,500 words)
3. **Week 5:** Consolidate your report to ensure that the information is presented clearly and coherently (including writing the Conclusion for the report). Make sure that you properly cite all your sources. Your report must have between 2,500-3,000 words (excluding the Cover page and the references). Submit your assignment to Canvas.
4. You need to make constant progress for the above 3 milestones. Your tutor will check your progress to ensure that you follow the above timeline (it's OK to get ahead but it's not OK to fall behind). **Failure to make satisfactory progress will be penalised.** Thus, you cannot wait until Week 5 to start your Assignment. The easiest way to demonstrate your progress is to put the content you created for each milestone on Github (e.g., on the Github Wiki) or OneDrive.

### Tip:

It's very likely that the topic or question you choose has already been investigated by someone. Please feel free to learn from those sources (as long as they are reliable and make sense). However, it's not good to select a single source (e.g., a research article or a blog post) and reproduce it in your writing. It's important that you research from multiple sources and get different views (even when they are conflicting) to form your own arguments and conclusions.

## Submission instructions

Your report has to be either in Microsoft Word or in PDF whose name is your Student ID (for example, 1234567.PDF). You must submit your work via Canvas.

Standard late penalties apply - 10% for each day late, more than 5 days late is 0%.  
You can follow the following structure for the final report:

- **Cover page:** including your student details (i.e., your full name and student ID).
- **Table of contents (TOC).**
- **Executive Summary:** (2 paragraphs) First paragraph provides a summary of the topic/question and the key issues associated with this question. Second paragraph summarises your findings and conclusions in regard to the key issues and the question as a whole.
- **Main report: (2,500-3,000 words)**
  - **Introduction:** Provide a background on the area, especially on the aspects related to your chosen topic. Discuss and explain the key issues in the topic/question. Discuss the methodology you use for the research. You can also discuss the scope of your research.
  - **Findings:** Analyse and synthesise the information you gather from reliable sources to form arguments to support your findings and conclusions. It's a good idea to structure this section into multiple subsections to improve the readability and presentation of your report. For example, you

- can structure it according to different key issues in the question. Make sure that you cite all the sources and references you used (including the use of AI tools) when generating your report.
- **Conclusion and recommendations:** Give a summary of the conclusions and findings presented earlier in your report. Make sure that these conclusions and findings address all the key issues you discuss in the Introduction of the report. If there are recommendations you want to make regarding these key issues, please include them here as well.
  - **Acknowledgements/Resources:** Include in your report a list of the resources you have used to create your work. A simple list of URL's is not enough. Include with each entry a basic description of how the person/AI tool or website assisted you in your work.
  - **References:** Cite the sources you referred to in your report.

**Tips:**

- If you choose to include a picture/diagram, one figure is equivalent to 50 words.
- Report presentation should include header/footer information (pages numbers, etc.)

## Marking Scheme

Requirements (We adopt the <b>Research Paper Rubric</b> from <b>Cornell College</b> for a more specific rubric for this report. See next page.)	Mark
<b>Content</b>	<b>75</b>
<b>Presentation</b>	<b>25</b>
<b>Total</b>	<b>100</b>
Failure to provide satisfactory updates of the assignment progress to your tutor	<b>Up to -60</b>

**Cornell College's Research Paper Rubric**

(<https://www.cornellcollege.edu/library/faculty/focusing-on-assignments/tools-for-assessment/ResearchPaperRubric.pdf>)

	EXPERT	PROFICIENT	APPRENTICE	NOVICE
INTEGRATION OF KNOWLEDGE	The paper demonstrates that the author fully understands and has applied concepts learned in the course. Concepts are integrated into the writer's own insights. The writer provides concluding remarks that show analysis and synthesis of ideas.	The paper demonstrates that the author, for the most part, understands and has applied concepts learned in the course. Some of the conclusions, however, are not supported in the body of the paper.	The paper demonstrates that the author, to a certain extent, understands and has applied concepts learned in the course.	The paper does not demonstrate that the author has fully understood and applied concepts learned in the course.
TOPIC FOCUS	The topic is focused narrowly enough for the scope of this assignment. A thesis statement provides direction for the paper, either by statement of a position or hypothesis.	The topic is focused but lacks direction. The paper is about a specific topic but the writer has not established a position.	The topic is too broad for the scope of this assignment.	The topic is not clearly defined.
DEPTH OF DISCUSSION	In-depth discussion & elaboration in all sections of the paper.	In-depth discussion & elaboration in most sections of the paper.	The writer has omitted pertinent content or content runs-on excessively. Quotations from others outweigh the writer's own ideas excessively.	Cursory discussion in all the sections of the paper or brief discussion in only a few sections.
COHESIVENESS	Ties together information from all sources. Paper flows from one issue to the next without the need for headings. Author's writing demonstrates an understanding of the relationship among material obtained from all sources.	For the most part, ties together information from all sources. Paper flows with only some disjointedness. Author's writing demonstrates an understanding of the relationship among material obtained from all sources.	Sometimes ties together information from all sources. Paper does not flow - disjointedness is apparent. Author's writing does not demonstrate an understanding of the relationship among material obtained from all sources.	Does not tie together information. Paper does not flow and appears to be created from disparate issues. Headings are necessary to link concepts. Writing does not demonstrate understanding any relationships
SPELLING & GRAMMAR	No spelling &/or grammar mistakes.	Minimal spelling &/or grammar mistakes.	Noticeable spelling & grammar mistakes.	Unacceptable number of spelling and/or grammar mistakes.
SOURCES	More than 5 current sources, of which at least 3 are peer-review journal articles or scholarly books. Sources include both general background sources and specialized sources. Special-interest sources and popular literature are acknowledged as such if they are cited. All web sites utilized are authoritative.	5 current sources, of which at least 2 are peer-review journal articles or scholarly books. All web sites utilized are authoritative.	Fewer than 5 current sources, or fewer than 2 of 5 are peer-reviewed journal articles or scholarly books. All web sites utilized are credible.	Fewer than 5 current sources, or fewer than 2 of 5 are peer-reviewed journal articles or scholarly books. Not all web sites utilized are credible, and/or sources are not current.
CITATIONS	Cites all data obtained from other sources. APA citation style is used in both text and bibliography.	Cites most data obtained from other sources. APA citation style is used in both text and bibliography.	Cites some data obtained from other sources. Citation style is either inconsistent or incorrect.	Does not cite sources.

**Content** = INTEGRATION OF KNOWLEDGE + TOPIC FOCUS + DEPTH OF DISCUSSION + COHESIVENESS

**Presentation** = SPELLING & GRAMMAR + SOURCES + CITATIONS