

Hi everyone!

Below is a list of suggested reading and research tips for each of the essay questions.

At the end is some general essay writing advice. I know that for many of you this will be your first time writing a longish research essay. Don't panic! Your essays will all be great I am sure.

Okay, so again, here are the general essay details:

=====
=====

Word length – 2000w in total (not including bibliography or footnotes).
Due date – Friday October 28 (Week 7), 23:59.

Please submit your essay via turnitin (details to follow).

Answer one (1) of the following questions:

=====
=====

QUESTION 1. Compare and contrast act vs

rule utilitarianism. How might their use to motivate collision-avoidance protocols for autonomous vehicles result in different protocols? Which of act and rule utilitarianism do you prefer for this purpose? Why? Justify your answer.

This is a fun question. Like the others, it really is a /question/ – well actually it is a bunch of questions – and this means that your essay should be an /answer/ to it. Then, and this is super important, you need to give /reasons/ to your reader for why it is that you think your answer(s) is the correct one. Your goal is to convince your reader that you are right! So, let's move through Question 1 carefully, one sub-question at a time.

The first thing that Question 1 asks you to do is to compare and contrast act vs rule utilitarianism. To this properly, you need to describe both act and rule utilitarianism, and then explain /how/ it is that they are distinct. In other words, describe what they have in common, and what they do not. You are very free here to state

which of the two, if either, that you prefer. If you do so, then it is important that you give reasons for your choice.

The second thing that Question 1 asks you to do describe how it is that you think that collision-avoidance protocols for autonomous vehicles could be designed on the basis of act utilitarianism. It then asks you how it is that you think that collision-avoidance protocols for autonomous vehicles could be designed on the basis of rule utilitarianism. The best thing to do here is to make sure that you are already very clear on the distinction between act and rule utilitarianism, then it will be much easier :)

The third thing that the question asks you is this - do you prefer either act or rule utilitarianism for the purpose of designing collision-avoidance protocols? It is open for you to say that you think that both act and rule utilitarianism are so terrible at this that you dislike each of them equally. The important thing is that you state / why/ it is that you think this, and give detailed reasons for your belief. Similarly, it is open for you to say

that, on balance, you think that both act and rule utilitarianism are good, and equally good at the task. Here too, the important thing is that you provide /reasons/. Now, hopefully it is pretty obvious that if you prefer act over rule utilitarianism for this task, or rule over act utilitarianism, then of course either of these are fine, BECAUSE THE MOST IMPORTANT THING IS THAT YOU GIVE REASONS FOR YOUR BELIEF!

Giving reasons for your answer, and then arguing in favour of that answer, are the fourth and fifth things asked in Question 1 respectively. But we have just covered these above :) The moral here, is that the reasons that you give for your answers might be more important than the answers themselves.

Here are some other things that you might read with regard to writing your answer to Question 1. Note that all of them can be found on Google Scholar (you will need to be inside the UNSW network to access some of them, so VPN is essential if you are not on campus):

Contissa et. al. (2017): The Ethical Knob: ethically-customisable automated vehicles and the law, "Artificial

Intelligence Law", 25:365–378.

Goodall, N. J. (2) Machine Ethics and Automated Vehicles, in G. Meye and S. Beiker (eds.): "Road Vehicle Automation" 2014 (corrected 2018). Springer.

Liu, P. and Liu, J. (2021): Selfish or Utilitarian Automated Vehicles? Deontological Evaluation and Public Acceptance, INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION 2021, VOL. 37, NO. 13, 1231–1242.

Mirnig, A. G., and Meschtscherjakov, A. (2019): Trolled by the Trolley Problem: On What Matters for Ethical Decision Making in Automated Vehicles, CHI 2019, May 4–9, 2019.

Fournier, T. (2016): Will My Next Car Be a Libertarian or a Utilitarian? Who Will Decide? IEEE Technology and Society Magazine / j u n e 2 0 1 6

=====
=====

QUESTION 2. Expound and assess rule–

based/Kantian ethics. Analyse the extent to which such an ethics might be used to design an automated ethics. What do you think that the risks and opportunities of such an automated ethics might be? Why? Justify your answer.

This is a fun question too! All of the general advice for Question 1 is good advice for Question 2 also.

The first thing that Question 2 is asking you to do is to expound Kantian ethics. "Expound" is just a fancy way of asking you describe and explain something in detail.

The second thing that Question 2 is asking you to do is to assess Kantian ethics. So, this is a great opportunity for you to state what it is that you think about Kantian ethics, why you think it, and what reasons you have for your reader to agree with you that you are right!

The third thing that Question 2 asks you is just how useful you think a Kantian ethics might be for automated ethics. Another name for automated

ethics is "machine ethics".

The fourth and fifth things that Question 2 asks you respectively are to give reasons for your answer to the third thing, and to then argue for and justify these reasons.

I am sure that you can see a pattern here by now :) Here are some readings that will help you with your answer to Question 2. Again, note that all of them can be found on Google Scholar (you will need to be inside the UNSW network to access some of them, so VPN is essential if you are not on campus):

Manna, R. and Nath, R. (2021): Kantian Moral Agency and the Ethics of Artificial Intelligence, *Problemos* vol. 100, pp. 139–151

Nath, R., and Sahu, V. (2021): The problem of machine ethics in artificial intelligence, *AI & Society*, 35:103–111

Tonkens, R. (2009): A Challenge for Machine Ethics, *Minds & Machines* (2009) 19:421–438

Singh, L. (2022): Automated Kantian

Ethics: A Faithful Implementation.
Online at <https://github.com/lsingh123/automatedkantianethics>

=====
=====

QUESTION 3. Is kicking a robot dog morally wrong, or morally permissible? In your answer, make explicit detailed reference to virtue ethics. Which answer or answers might virtue ethics give us? Are any of these answers correct? Why? Justify your answer.

The first thing that Question 3 asks you is what you think about the moral status is of kicking a robot dog. A good answer here would give some space to the considerations that motivate your answer.

The second thing that Question 3 ask you is to engage in detail with virtue ethics! This is fun, as there is a lot of great literature out there on this very same topic. Think about it now for

just a moment. Neither utilitarianism nor Kantian ethics give as any obvious reason for why it is that kicking a robot dog is the wrong thing to do. In spite of this, some people have felt that there is /something/ wrong with kicking a robot dog.

There is some great literature on this topic. The videos of the Boston Dynamics robot dog "Spot" being kicked led to a very interesting debate, and virtue ethics played a central role in it.

The other things that Question 3 asks you require you to engage with this literature. Here is the literature. Again, note that all of them can be found on Google Scholar (you will need to be inside the UNSW network to access some of them, so VPN is essential if you are not on campus):

Coeckelbergh, M. (2021): How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance, *International Journal of Social Robotics*, 13:31–40.

Sparrow, S. (2021): Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might it be Explained? *International Journal of Social Robotics* (2021) 13:23–29.

Coeckelbergh, M. (2021): Does kindness towards robots lead to virtue? A reply to Sparrow's asymmetry argument, *Ethics and Information Technology* (2021) 23:649–656.

Coeckelbergh, M. (2021): Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots, with Implications for Thinking About Animals and Humans, *Minds and Machines* (2021) 31:337–360.

=====
=====

QUESTION 4. In his "The uselessness of AI ethics" (2022), Luke Munn criticises much practice in AI ethics. Do you think that Munn's criticism is justified, and that his proposed solutions are sufficient? Why? Justify your argument with explicit detailed

reference to Elettra Bietti's "From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy" (2020).

Question 4 makes explicit reference to the two readings that are on webcms for Week 4. I have combined them there into a single pdf. There is no need for you to look for extra readings for Question 4. However, if you want to do so, then there is certainly no prohibition against it! One thing to do in this case would be to look at the citations in Munn and Bietti's articles and follow your interests.

Again, the most important thing to do here is to provide an answer to the question!

=====
=====

QUESTION 5. Expound cognitivism and non-cognitivism in meta-ethics. Which do you find most plausible? Why? Of the one that you find most plausible, which variety do you find most plausible?

Why?

Question 5 is a wild adventure into meta-ethics. Question 5 is for those of you who I know have a strong interest in meta-ethics already. Note, very importantly, that if you choose to answer this question then you will need to do a large amount of independent research.

This is because although we will be covering meta-ethics in this course, we will not be doing so until /after/ your essay is due. Wild, no? I am available for consultation of course, and I shall be on Ed, but still, know that this question will keep you very busy.

A good place to start with regard to readings is here:

<https://plato.stanford.edu/entries/moral-cognitivism/>

This entry contains a lot of references, and a good essay that answers Question 5 would be expected to deal with many if them.

=====

=====

GENERAL ESSAY ADVICE:

- Start early and revise drafts. The more drafts you revise, the better your essay will be. Always. Every time.

- To revise, read your essay out loud to yourself. I do not know why this works but I promise you that it does. Better still, try reading it to a friend! If they do not understand what you mean, then you have more work to do.

- Remember that your reader has only what you have written on the page. They do not have magical mind-reading abilities. They are practicing magical mind reading abilities, but still.... :) Be clear. Be very clear. And then be clearer still.

- English hates long sentences. English is not French. The longer the English sentence, the less clear it is. The shorter the English sentence, the more clear it is. Short sentences are clear sentences. Clear sentences are good

sentences. Good sentences make good essays. Write in short sentences. Short is good. Good? Good.

– You are allowed to use the word "I".
Example – "I love writing essays".

– Use proper citations and have a bibliography. Any reference style is okay, as long as it is done properly and consistently.

– Your tutors want to know what you think. I want to know what you think. Have reasons for what you think. Argue for those reasons in your essay. A good tip is to try to imagine an objection to your argument, and then provide a response. Do this over and over again. This will make your argument "tighter", or "more thorough".

– No one's ideas are more important than YOUR IDEAS! I really do mean this. Do not be afraid to criticise anyone. Not ever. Defend your ideas bravely. They are your ideas and they need your help to make them strong.

See you in the lectures and on Ed,

Seb

