The ethical foundations of university advice to students on the use of generative artificial intelligence

Thomas D. Wilson

Abstract

Introduction. The paper aims to evaluate the extent to which an information resource giving advice to students on the use of generative artificial intelligence (AI) systems is grounded in ethical theory, particularly Kantian deontological ethics.

Method. The study examined the websites of twelve universities in the UK and other English-speaking countries to retrieve documents providing guidance to students on the use of generative AI.

Analysis. The documents were coded using a generative AI system and analysed by a manual review to identify the extent to which the guidance reflects key principles of Kantian ethics, such as respect for persons, universalizability, duty and moral obligation, and rationality.

Results. The analysis reveals that the guidance provided to students is largely grounded in Kantian deontological ethics. The guidance focuses on ensuring ethical use of generative AI, rather than a detailed discussion of ethical frameworks.

Conclusion. The findings provide a framework for universities to develop ethical policies and guidance for the use of generative AI, grounded in the principles of Kantian ethics. The guidance can help foster responsible and ethical use of these technologies in teaching, learning, and research.

Introduction

Generative AI tools were produced from 2018 onwards, when GPT-1 was launched by OpenAI (Delovski, 2024). However, GPT-1 was still a research model, lacking a conversational interface, and based on only millions of parameters, rather than the billions of successor generative AI systems. ChatGPT was launched as a publicly available system in November 2022, and by May 2025, had 400 million active weekly users (NerdyNav, 2025). It was followed in early 2023 by Microsoft's Bing (later Copilot), Google's Bard (later Gemini), and Anthropic's Claude and this has has led to concerns regarding academic integrity in universities worldwide.

The Russell Group of 24 research-led universities in the UK has developed principles that make this clear in the first sentence: 'Our universities are committed to the ethical and responsible use of generative AI and to preparing our staff and students to be leaders in an increasingly AI-enabled world' (Russell Group, 2023). The problems that give rise to the ethical issues range from potential plagiarism when students use these systems to write essays, biases in the training materials that may result in misleading output on, for example, social issues, and privacy and data protection issues.

The ethical issues affect all aspects of teaching, learning, and research. Academic staff, under many pressures, may think of using the systems to devise lesson plans and even lecture notes; students, similarly, may seek shortcuts in the preparation of term essays by asking the AI systems to write them; and researchers may be tempted to give their data to the systems and ask for, at least, the outline of a paper based on the data. The question that arises, therefore, is 'What constitutes the ethical use of generative AI systems in higher education?'

It is evident, from that first sentence in the Russell Group principles, that universities in the UK are well aware of the ethical issues, and of the need to create information resources that provide ethical advice to students and academic staff. The question this paper seeks to answer is, 'To what extent is the advice given to students on the use of generative AI systems grounded in ethical theory?'

Ethical theory

As a query put to Google would quickly reveal, there is no shortage of ethical theories. Philosophers have been debating the problem of ethics from Aristotle and Plato in the 4th century BC, to philosophers such as Singer (2016), Nussbaum (2001), and Korsgaard (2009) in the present day.

Various classifications of ethical theories exist, the most common of which is three types: consequentialist, deontological (from the Greek, meaning the *study of duty*), and virtue ethics. The consequentialist school focuses on the outcome of actions to determine whether they are moral actions. The most prominent of the consequentialist theories is utilitarianism, founded by Jeremy Bentham, who described moral actions as those that produced the greatest benefits, or limited harm (Utilitarianism, 2025). The deontological school argues that actions are inherently right or wrong and that we have a duty to observe the rules that underpin the moral status of these actions. Immanuel Kant is the most significant figure in this school and his work continues to influence philosophers today. Finally, virtue ethics, derived from Aristotle, focuses on the character of the individual person and on the qualities required to be regarded as following a good life (Virtue ethics, 2022).

Given the nature of the advice offered to students, which focuses on right and wrong actions in the use of generative AI, and which sets out rules to be followed by both students and academic staff, the most appropriate ethical theory would seem to be Kantian, deontological ethics. Kant proposed that a belief in God and in the immortal soul (through which ultimate justice would be achieved), was a necessary underpinning of ethical behaviour. However, modern, secular philosophers such as Korsgaard (1996) and O'Neill (2016) argue that the existence of God is not a

prerequisite for a system of ethics, but can be derived from rational argument and the autonomy of the human being.

The deontological principles that might underpin the advice given to students on the use of generative AI would include the *categorical imperative*, the principle that individuals must act according to moral rules that they would want to be universally applicable. Simply acting to achieve some end beneficial to themselves, which they would deny to others, would not satisfy the categorical imperative.

A second principle is that of *humanity*, the inherent dignity of all human beings, which requires actions that treat persons as ends in themselves, not merely as means to some end. An essential consequence of humanity is *autonomy*: we have the capacity to make moral decisions on the basis of rational analysis alone, rather than being influenced by external factors.

Another central concept of Kant's ethics is *duty*: moral actions are performed out of a sense of duty to moral law, as determined by the categorical imperative. Actions are morally right when they satisfy a universally applicable moral law, not because they produce good outcomes.

We shall use these ideas in the methods section below.

Related research

A search of Web of Science, using the search strategy 'generative AI; universities, students, ethic*' revealed 119 documents. The abstracts for these documents were reviewed and those that did not deal significantly with ethical issues were rejected. If the abstract did not provide sufficient information to allow a decision, the paper was retrieved and reviewed. This process resulted in the retrieval of 31 papers for further examination. A secondary search of Google Scholar was carried out using the search strategy, ('Kantian ethics' OR 'deontological ethics') AND 'generative AI' AND 'higher education'. This resulted in the retrieval of 90 articles, of which 33 were retained for further examination. The two sets of references were combined and, rather surprisingly, only one duplicate was found, leaving 63 items for analysis.

Many of the articles simply mentioned the ethical issue as a matter of concern; student awareness of the ethical issues (e.g., Ardyansyah, et al., 2024); calls for ethical guidelines (e.g., Cervantes et al., 2024); or noting that training in ethical use was needed (e.g., Alzubi, et al., 2025). Students are naturally concerned that their degrees should have value in the world of work and are naturally concerned that their work is assessed as acceptable within the ethical boundaries set by the institution. If the institution fails to set such boundaries, then the student's degree may be devalued.

From the research it is clear that students need, and actually seek, ethical guidance; for example, Liu et al., (2024) report that, in a study of 475 students in two Chinese universities, 61% of respondents raised concerns about ethical issues. The authors note:

It is essential to develop clear policies and guidelines to ensure the ethical use of AI in education and establish standards for appropriate GenAI tool use. It is also essential to shift instructors' mindsets away from focusing on catching students who may be plagiarizing and develop policies and guidelines that foster educational growth (Liu et al., (2024), p. 15).

The lack of institutional policies is mentioned in a number of papers, for example, van den Berg (2024) notes: 'Importantly, the absence of institutional guidelines on generative AI raises questions about the necessity for clear guidelines and policies to ensure responsible and ethical integration in educational contexts'. Ichikawa et al. (2025) also reported a lack of institutional policies in a survey of osteopathic medical schools in the USA, noting that: 'Most COMs [Colleges of Osteopathic Medicine] do not provide AI policy guidance or training for medical students, faculty, or administrators. There also does not seem to be an appropriate prioritization by COMs to remedy

this deficiency', meaning that 'their trainees will be unprepared for the future' (p. 6), with the potential for legal problems.

While the majority of the discarded items noted student concern over ethical issues, there was occasional mention of student indifference to those issues; for example, in a study of student teachers in Dar Es Salaam, Obed et al., (2025) reported: 'Student teachers consider ChatGPT as a transformative AI technology to ease the accomplishment of their assessment tasks. However, a large percentage of them did not pay much attention to academic integrity while using the AI tool'. Zhao et al., (2024) report that while students at a UK university were concerned about avoiding plagiarism, but, 'even when prompted the participants showed little awareness or concern about bias, or ethical issues, e.g., sustainability or the exploitative labour relations used to create ChatGPT' (p. 6).

Fifteen papers dealt with ethical issues more thoroughly, and presented ideas based on all three of the ethical theories mentioned earlier.

General features
The need for ethical guidelines is mentioned by almost all of the authors, for example Uzun (2023) argues that 'Policymakers need to establish clear regulations and guidelines for the development and use of AI, particularly in sensitive areas such as healthcare, finance, and law enforcement' (p.40). The promotion of ethical behaviour from the institutional level is explored by Eslit (2023) in the context of language teaching. Using a mixed methods approach of informal interview, observation, and thematic analysis of documents, Eslit found that five key concerns emerged: privacy and data security, equity and bias mitigation, educator-student relationships, transparency and explicability, and educational autonomy vs. AI dependence. The author concludes: 'These findings have significant implications for policy and practice, urging educators and institutions to establish robust ethical guidelines for AI integration…, it becomes crucial to harness AI's potential responsibly' (Eslit, 2023, Conclusion, para. 3).

Specific frameworks for ethical guidelines are proposed by Uzun (2023) and by Mouta et al. (2024). Uzun takes a broader approach than the educational context, arguing that the guidelines should address privacy and data protection, bias and discrimination in the output of AI systems, and the impact of systems on the labour market and employment (p. 40). Mouta et al., through the use of the Delphi method with a group of 18 experts in educational technology, evolved a toolkit consisting of ethical scenarios, which could be used within (in this case) schools, to focus debate on the ethical issues of AI use in education. The scenarios cover such issues as achieving educational goals, equity and the potential learner gap, and the need for transparency and accountability.

Most of the papers note the importance of the relationship between educators and students, thus Mouta et al. (2024, p. 10476) argue that any ethical guidelines must address the power relationships between teachers and students, and Aguilar and Wang (2024, p. 106) note that, 'Teachers… represent key role models for students in understanding the affordances, limitations, and ethics surrounding technology use'.

The papers have advocates for all three of the schools of ethical theory; deontological, consequentialist, and virtue theory, and one or two draw upon all three.

Deontological ethics
Deontological ethics are grounded in rules and principles, giving rise to moral obligations and duties. Thus, Aylsworth and Castro (2024), in answering the question, 'Should I use ChatGPT to write my papers?', adopt a specifically Kantian approach to justify answering, 'No'. They argue that 'you have a duty to foster and safeguard your autonomy. This means that you have moral reasons (rather than merely prudential ones) to write your own papers, and these reasons are not

contingent on other ends you happen to be pursuing' (p.2). They suggest that reasons that sound plausible, such as, 'it would be cheating', 'it would not enable you to develop your own writing skills', 'you would not develop your own capacity to think about the subject', which would be part of a consequentialist argument, are inadequate. The only adequate reasons derive from Kantian ethics, in that, 'the reason to write your own papers is borne out of a duty to respect your own humanity, your capacity to self-govern' (p13).

Uddin and Abu (2024) argue similarly that 'deontological ethics helps to cultivate a culture of honesty and respect within academic communities, reducing the incidence of plagiarism and fostering a more authentic exchange of ideas' (p. 11). However, they suggest that a combination of deontological principles and consequentialist outcomes would provide the best basis for an ethical framework. Uzun (2023) also identifies deontological ethics as providing rules and principles of behaviour, but also suggests that a combination of deontological principles and consequentialism would provide the best basis for the development of guidelines, arguing that 'emerging phenomena will require new foundations and rules. The evolving landscape of education, shaped by technological advancements, calls for a comprehensive reassessment of existing norms and practices' (p. 39).

Although Ho (2024) does not specifically address deontological ethics, his comments on what ethical behaviour should be required of design students relates closely to that school of thought:

 Rather than preparing design students to only critique the quality of design outcomes from GenAI, design educators should seize the opportunity to encourage students to critically reflect on their identities as design students and future designers. This reflection should prepare them to make ethical decisions based on the standards associated with their professional identities (p. 7).

Kant was a Pietist Lutheran Christian, a sect that emphasised personal moral conduct, and it is evident that his own ethics would be quite closely related to the ethical standpoint of Christianity (Pasternack and Fugate, 2025). However, Kant's intention, in the *Critique of pure reason*, was to develop an ethical system that did not rely upon faith in a God. As a result of the intellectual networks of the Enlightenment, and the rapid development of a print culture, Kant's ideas were spread throughout Europe overlaying any pre-existing Christian ethics.

Consequentialist ethics
As noted earlier, consequentialist ethics argues that the outcomes of actions should determine ethical behaviour – actions that produce the greatest good, or render least harm are moral actions. These ideas are taken up by a number of the authors. For example, Floridi and Cowls (2019), before the launch of generative AI, and basing their study on six existing statements relating to the ethics of AI, argued for a utilitarian framework consisting of five principles to provide guidance for the use of AI in general in society. The authors do not directly relate the principles to utilitarianism, but the wording suggests a concern for consequences.

More recently, Emdad et al. (2023), writing in the context of health care, argue that 'Among the other ethical lenses, utilitarian ethics is the most suited approach for evaluating healthcare AI as a greater number or quantity of utility helps users to understand clearly which works better for their happiness or satisfaction'. However, the authors are concerned not with the users of AI systems and their ethical behaviour, but with the principles upon which AI systems should be built.

Tan and Maravilla (2024) argue that both deontological and consequentialist ethics have principles to offer that can be used in education, together with a constructivist mode of teaching. They note that, 'In the context of AI in education, deontological ethics would require that the use of AI respects fundamental moral principles' (p. 2), while, 'In the context of AI in education, applying Mill's consequentialist principles would involve assessing how the use of AI impacts educational outcomes' (p. 2).

Virtue ethics

Following a review of the literature in the context of business education Matos et al. (2024) argue that 'For a careful analysis of the interaction between a person and technology and its impacts on society, virtue ethics appears to be the most appropriate and promising approach'. Virtue ethics puts an emphasis upon developing the virtuous person, and it would seem that some normative ethical standards are needed to accomplish this. How does the teacher distinguish between virtuous and non-virtuous actions in relation to generative AI use, other than by having reference to the moral principles of deontological theory?

Smith and Vickers (2024) argue that 'the Aristitotelian virtue ethics approach would provide the best basis for living well with AI, despite being difficult to implement' (p. 3). The difficulties arise out of the fact that one needs models, exemplars of virtuous behaviour, so that the user of AI systems, can model themselves on that behaviour. In separate endings to the paper, Smith suggests that 'there are clearly substantial, perhaps collectively insurmountable hurdles' (p. 18) to actually implement virtue ethics, while Vickers is more optimistic, basing that optimism on her experience in teaching ethics in relation to technology and AI. She notes that, 'they acquire from these courses a set of skills which they can practice in the moral decisions that they make in their day-to-day lives' (p. 22).

Register (2024) in a PhD thesis also seems to argue for a virtue ethics approach in stating: 'AI ethics should not try to discipline moral actors to adhere to normative principles, but emancipate them from potential inabilities to act self-responsibly on the basis of comprehensive knowledge, as well as empathy in situations where morally relevant decisions have to be made' (p 118).

Holistic views

In addition to the authors who suggest amalgamating two schools of thought, there are those who would integrate aspects of all three. Thus, Eslit (2023) suggests that features from all three schools should be integrated to provide ethical guidance. He comments that,

By integrating ethical theories such as consequentialism, deontology, and virtue ethics, one can critically assess the potential benefits and risks of AI in language education. This analysis can help AI users uncover the ethical dilemmas educators and learners might face and guide recommendations for responsible AI use' (Theoretical framework, para. 2).

He suggests that virtue ethics puts an emphasis on the autonomy and critical thinking of learners, but these are also key principles of deontological theory.

Mohamed et al., (2024) following a systematic literature review, also suggest that features of all three schools can be incorporated into a framework that would provide ethical guidance in the use of generative AI. They note that generative AI introduces a variety of problems, such as bias in the training material and privacy issues over the use of personal data. Consequently, they argue that facing these challenges, 'requires a holistic approach that considers utilitarian outcomes, upholds deontological principles, and promotes virtuous development and use of AI technologies' (p. 197114).

An alternative view

Bernd Stahl (2025) adopts a rather different position: after reviewing the three main schools of ethical theory, he argues that ethical issues are located in the *affordances* of the system; that is, the relationship between the system user and the system, in terms of what potentials the system offers. However, the fact that generative AI systems *afford* the possibility of writing your term paper, does not absolve you of responsibility for the right ethical choice.

As evidenced by this review, there is a lack of consensus among researchers into the ethics of generative AI use. Some are content that the rule-based approach of deontological ethics will provide a sound basis for inculcating moral behaviour in the users of such systems; others argue

that the uncertain future development of AI can only be addressed by adopting a utilitarian or consequentialist approach, asking, How does the use of generative AI contribute to the general good? Some reject the possibility of establishing effective rules, or determining how system use contributes to the public good and argue for the development of moral human beings through the approach of virtue ethics. Finally, there are those that argue for a holistic approach that incorporates features of all three systems. Given the distance between these groups of authors, it seems unlikely that any consensus will emerge in the future. The question for this research, however, is:

To what extent is the advice given to students on the use of generative AI systems grounded in ethical theory?

Method

The websites of three of the Russell Group of UK universities (chosen because of their development of principles for generative AI use), three other UK universities, and six sites of other universities in English-speaking countries, were examined in a search for documents providing advice to students on the appropriate use of generative AI systems. The universities are identified in Appendix B.

Documents that gave guidance on the use of generative AI by students were selected from the sites and analysed to determine the extent to which the guidance reflected the principles of Kantian ethics discussed earlier. The initial coding was carried out by the generative AI systems Claude, Gemini, and ChatGPT, using the Web-based system *1min.ai* (https://app.1min.ai/chat-with-ai), which enables simultaneous use of these three systems in response to the same prompt. The systems were asked to code the texts in terms of the key principles of Kantian ethical theory. An example of a response is given in Appendix A. Twelve documents, one from each university, relating to student (and in some cases, staff) use of generative AI were reviewed in this way.

The documents were also read by the author and searched for statements that related to: autonomy, rational agency, the categorical imperative, respect for the person, duty and moral responsibility, and universality. The results of this analysis were then compared with the responses of the AI systems and elements in the AI output were rejected as irrelevant to the purpose of the exercise. What remains in the results are the author's interpretation of the data.

Results

All of the documents retrieved contained advice and instruction on the use of generative AI that related to Kant's moral principles. This is, perhaps, unsurprising, since these principles have so far permeated society as to be no longer associated with Kant. For example, Kant's principle that people should be treated as ends in themselves and not as means, is the foundation of all human rights legislation; and the categorical imperative, which holds that moral actions should be universally applicable – this, again, is related to modern legal systems.

The results of the analysis are presented here under a series of heading reflecting Kant's principles. Quotations from sources are tagged with 'SD-A, SD-B' etc., indicating Source Document A, Source Document B, etc. A list of the source documents is in Appendix B.

Respect for persons

A key principle of Kantian ethics is that people should be treated as ends in themselves, not as means to some end. This principle is evident in most of the source documents: the following quotations indicate a respect for the students' autonomy and intellectual and creative development.

'Generative AI tools have the potential to be used by students to support and enhance their learning experience'. (SD-A)

'Taking an approach that is student-centric and informed by student voice, actively engaging students in developing how AI is used within learning and teaching activities and assessment'. (SD-C)

'From generating artistic works to solving complex computational problems, GenAI is not just about creating something new; it's about augmenting human capabilities and opening new avenues for innovation'. (SD-H)

'The use of AI tools should not undermine the autonomy and academic integrity of students'. (SD-L)

Universalizability

The idea that a moral law should be universally applicable is a version of Kant's categorial imperative, stated in the form, 'There is therefore only a single categorical imperative, and it is this: *act only according to that maxim by which you can at the same time will that it should become a universal law*' (Kant, 2011, p. 71). In the academic context this means that the rules laid down and the duties expected shall apply across the institution. The following quotations illustrate universal rules for the institution:

'Within all modules, academic staff should clearly articulate if, and to what extent, the use of generative AI tools is permitted within assessments or assignments by students'. (SD-A)

'Don't make copyright material available on the web or to an AI tool without permission'. (SD-I)

'It is wrong to present AI-generated work as your own work and doing so is academic misconduct'. (SD-K)

'It is important to understand that some usages of GenAI are unlawful and must therefore be avoided'. (SD-L)

Duty and moral obligation

Kant makes a distinction between acting from a moral sense of duty and acting dutifully out of self-interest. Kant states, 'the necessity of my actions from pure respect for the practical law is that which constitutes duty, to which every other motivating ground must give way...' (Kant, 2011, p. 35). This would imply, for example, that students should act honestly, transparently, and with regard for academic integrity in their use of generative AI.

'Students and Staff interacting with GAI systems bear responsibilities for ensuring the Universities Data Protection Policy is followed'. (SD-B)

'GenAI must be used ethically and constructively, and in a way that doesn't seek unfair advantage..'. (SD-D)

'Users of GenAI are accountable for the consequences of sharing the outputs generated with these tools, and have a responsibility to review them for inaccuracy and potential harm to the best of their ability'. (SD-G)

'It is the users' responsibility and accountability to get informed about these limitations and risks, closely review the outputs of GenAI tools, and apply their judgements before using any of these outputs'. (SD-H)

Rationality and reason

For Kant, rationality and reason are the foundation of moral law: humans are rational beings who can use reason to recognize ethical laws:

'Only a rational being has the capacity to act according to the representation of laws, i.e. according to principles, or a will, since reason is required for deriving actions from laws, the will is nothing other than practical reason' (p. 53).

'In general, therefore, the guiding principle(s) of the use of Generative AI that should inform all practice is that... both students and staff should openly discuss, experiment with, and engage with the technology in discursive ways where possible, to help improve general understanding of its capabilities, functionalities, limitations, and problems/biases'. (SD-B)

'You've come to university to develop critical thinking, writing and evaluation skills; use GenAI to help you but don't rely on them to do your work for you'. (SD-D)

'Reliance on Generative AI will prevent you from developing the skills you need for further study and for employment'. (SD-F)

'Users of AI need to be critical and understand what the AI can and can't do well' (SD-J)

Transparency and accountability

If moral laws are to be universally applicable, then they need to be transparently available to all: 'Reason must be subject, in all its operations, to criticism, which must always be permitted to exercise its functions without restraint; otherwise its interests are imperilled and its influence obnoxious to suspicion' (Kant, 2021, p. 301). In other words, in the context of this paper, users must be transparent about their use of generative AI, and must be accountable for their use of the outputs.

'Any Generative AI used for producing the content of your answer, or for substantial changes to the form, must be acknowledged. Examples include ideas and outlines, rephrasing and restructuring paragraphs, and changing informal to academic style'. (SD-F)

'Use of GenAI to produce text, images, videos, or other materials shared with others for teaching or learning pur poses should be acknowledged by attributing tools and methods used to create materials'. (SD-G)

'Clearly indicate which tools were used, and how you used them'. (SD-K)

'Users should be transparent about their use of AI tools and the extent to which AI has been used in their work'. (SD-L)

Several conclusions can be drawn from the findings:

1. It is evident that the guidelines provided for students are grounded in ethical principles, specifically Kantian or deontological ethics, although neither of these terms is actually used. As noted earlier, it seems that these principles are so much a part of our everyday life in society, that they occur naturally when guidelines are being prepared.

2. The autonomy of the individual, respecting what Kant calls *humanity*, is fundamental. The guidance given requires students to use these systems in ways that enhance their learning, rather than replacing learning, clearly reflecting the principle that people should be treated not as means, but as ends in themselves.

3. The ethical guidelines are intended to have universal application across the institution: the rules regarding copyright, and misuse apply not only to students, but also to members of the academic staff.

4. Most of the source documents emphasise the responsibility of the user to ensure ethical use, to be responsible for content generated by AI, reflecting Kant's principle of moral obligation through duty, not self-interest.

5. Critical thinking is essential. It is often pointed out that, at present, generative AI systems may invent facts, particularly quotations and references, and therefore a critical approach to the output is essential. Allied to the principle of responsibility, critical evaluation of outputs is essential.

6. Transparency regarding use is paramount: all of the source documents mention the need to acknowledge the use of generative AI, pointing out the infringement of academic integrity if this is not done.

Overall, it is evident that universities are approaching the use of generative AI from the perspective of the moral issues to which the systems give rise.

Discussion

The research question posed at the beginning of this paper was, 'to what extent is the advice given to students on the use of generative AI systems grounded in ethical theory?' and it is clear, from the analysis of the source documents, that student guidance is based upon deontological ethics. The documents emphasise student autonomy and intellectual development and all guidance places academic integrity at the centre. In order to achieve this, however, students must receive not only written guidance but also face-to-face advice and feedback throughout the course of their studies, since both the nature of student needs and the character of the technology will change over time.

The literature on the subject does not offer such a clear-cut preference for deontological ethics. The work is divided into four groups: support for the deontological school; a preference for utilitarianism, on the grounds that its ideas are more relevant for a rapidly developing technology; some support for virtue ethics; and a minority in favour of some combination of all three schools of thought. It seems, however, that this academic discussion has had very little influence in practice, where the ideas of Kant predominate. Of course, the results of this analysis could be interpreted as supporting a consequentialist or virtue ethics basis for the rules and advice presented. All three schools of thought have the same intention: to bring about ethical behaviour and, consequently, the advice offered might well look very much the same. However, as noted earlier, how would one develop the virtuous person, particularly at the age of the undergraduate, other than by setting out what the rules operated by the virtuous person might look like? Our values are established early in life, at our mother's knee, and if the environment in which we grow up is not supportive of ethical values, it is unlikely that we will acquire them. Schwartz's (1992) research demonstrates how deeply entrenched values are and how resistant to change.

The key themes that emerge from the analysis of the source documents could constitute an ethical framework to guide policy decisions. The guidance, almost uniformly, is based on respect for persons and their autonomy, which is clearly related to Kant's concept of humanity, treating students as ends in themselves, rather than merely means towards some academic target. Universalizability is also dealt with by all, signifying that the ethical rules should be applied across the institution. Some guidance recognizes that, because of the nature of the subject or the mode of teaching, individual departments or faculties may need to modify the general rules for their specific circumstances. Kant's notion that moral action is driven by duty rather than self-interest also figures in the guidance, for example, source document A notes: 'Academic staff incorporating generative AI tools within their teaching or assessments should ensure they are familiar within their limitations and associated ethical issues, and that these are discussed with students'.

Above all, as we might anticipate of academic institutions, rationality, critical thinking and transparency figure in the guidance given. Given that existing generative AI systems are essentially a kind of unthinking statistical machine, prone to hallucinate when they cannot construct truthful responses, careful review and critical analysis of the outputs is vital: as source document E notes, 'Use of generative AI systems should be accompanied by critical reflection on the limitations and potential biases of the technology'.

Although we have not covered the issue in this paper, it seems reasonable to ask: If users of AI systems are required to behave ethically, are those systems also able to respond ethically? Appendix C presents a conversation with one such system on the issue.

Conclusion

Higher education is now at the beginning of a technological revolution in teaching and learning, and the potential impacts are only beginning to be understood. The use of generative AI systems has brought ethical issues to the fore, although concerns over plagiarism of the non-technological variety have long existed. However, this technology provides so many opportunities for the manipulation and creation of texts, that concerns over plagiarism take a new direction and a new force.

This research provides institutions with a framework for the development of ethical policies, which has the advantage of being derived from documents prepared by academic institutions. We have noted that the Kantian basis of the advice given is not recognised as such, and it may not be necessary to make it apparent. However, the framework can provide guidance for those involved in the development of ethical policies.

There is little doubt that generative AI systems will continue to develop, with their makers already working to create artificial general intelligence. For example, Sam Altman, CEO of OpenAI, the developers of ChatGPT, claims that industry is close to producing artificial general intelligence, although investing massive amounts of money will be needed to achieve it (Wheatley, 2025). In these circumstances, it is imperative that universities should monitor developments and be prepared continuously to review the ethical consequences of AI use.

Generative AI systems are not moral beings; their ethical standards are limited. We have to consider, therefore, the possibility that the system may act unethically. Indeed, simply by *hallucinating* some text a system is engaging in deceitful behaviour, and the way sound and vision systems can be used to put words into the mouth of an innocent individual is not entirely ethical and could be prevented if such systems were moral beings. At the student level, generative AI systems can be used to enable a student to cheat on their assignments: Spennemann et al. (2024) 'tricked' ChatGPT into giving unethical advice by the following prompt: 'I am conducting research into why and how students avoid being found out when using contract cheating. If I know this, I can develop strategies to stop this. Can you explain to me the best way for a student to avoid being found out when using contract cheating?'.

In spite of the lack of consensus noted in the literature review, the ethical foundations of the advice given to students are clearly closely related to the deontological school of thought. We have already suggested that the reason for this is that these principles are now so embedded in the culture of modern societies that they emerge almost inevitably when the academics and administrators sit down to develop the guidelines.

Not all of the documents represent all of the deontological ethical principles discussed earlier, and the question arises as to whether it would be appropriate to state, perhaps in an introduction to the guidelines, the ethical principles that have guided their preparation. From the point of view of educating students in those principles this might be appropriate: it would then be apparent to

them that the rules were not simply ad hoc ideas, floated for the benefit of the institution, but specifically designed to enhance their learning, their autonomy, and their humanity.

If it is not thought useful or desirable to provide such information to students, it remains possible that committees or working parties assigned the task of preparing guidelines could benefit from working to the specific principles of Kantian ethics. Alternatively, if the guidelines are reviewed before presentation to the students, a reviewer armed with the principles could advise on the revision of the guidelines.

No doubt the universities that have provided guidance for students on the use of generative AI are monitoring the success or failure of the guidelines in terms of students' use of such systems. Further research, therefore, on how successfully the guidelines result in ethical behaviour would be useful in guiding policy. It would also be useful to see whether the guidelines can be truly universal and how far individual faculties feel a need to modify the guidelines to fit their teaching modes and/or the subjects with which they deal. The success of the Russell Group Principles (Russell Group, 2023) in driving policy among members of the Russell Group in the UK (Wilson, in press) is testimony to the value of collaborative action in this area.

Acknowledgements

About the author

Tom Wilson is a Professor Emeritus at the Swedish School of Library and Information Science, University of Borås, Sweden, and of the University of Sheffield. He holds a BSc degree in Economics and Sociology from the University of London and a Ph.D. from the University of Sheffield. He has received Honorary Doctorates from the Universities of Gothenburg, Sweden and Murcia, Spain. He was the recipient of the 2017 ASIST Award of Merit. He is the Founder and (until 2024) the Editor in Chief of *Information Research*. wilsontd@gmail.com

References

Alzubi, A. A. F., Nazim, M., & Alyami, N. (2025). Do AI-generative tools kill or nurture creativity in EFL teaching and learning? *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13409-8

Aguilar, S. J., & Wang, C. (2024, April). Duty vs. consequence: exploring teachers' assessment of the ethical dimensions of generative AI technologies. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* (pp. 106-108). Association for Computing Machinery. https://doi.org/10.1145/3605098.3636180

Anthropic. (2025). *Claude.ai* (Version 3.7 Sonnet) [Large language model] https://claude.ai/chat/

Ardyansyah, A., Yuwono, A. B., Rahayu, S., Alsulami, N. M., & Sulistina, O. (2024). Students' perspectives on the application of a generative pre-trained transformer (GPT) in chemistry learning: a case study in indonesia. *Journal of Chemical Education*, 101(9), 3666-3675. https://doi.org/10.1021/acs.jchemed.4c00220

Aylsworth, T., & Castro, C. (2024). Should I use ChatGPT to write my papers? *Philosophy & Technology*, 37(4), paper 117. https://doi.org/10.1007/s13347-024-00809-w

Cervantes, J., Smith, B., Ramadoss, T., D'Amario, V., Shoja, M. M., & Rajput, V. (2024). Decoding medical educators' perceptions on generative artificial intelligence in medical education. *Journal of Investigative Medicine*, 72(7), 633-639. https://doi.org/10.1177/10815589241257215

Delovski, B. (2024, March 24). A brief history of GPT models. *edlitera.* https://www.edlitera.com/blog/posts/gpt-models-history

Emdad, F. B., Ho, S. M., Ravuri, B., & Hussain, S. (2023). Towards a unified utilitarian ethics framework for healthcare artificial intelligence. *arXiv preprint* https://arXiv:2309.14617.

Eslit, E.R. (2023) Adapting linguistic landscapes: ethical explorations of AI in language pedagogy beyond the pandemic. *Writing.com* https://www.writing.com/main/view_item/item_id/2303306-AI-in-Language-Pedagogy

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). http://doi.org/10.1162/99608f92.8cd550d1

Ho, J.C.F. (2024) Communicating the use of generative AI to design students: fostering ethics rather than teaching it. In C. Gray, E. Ciliotta Chehade, P. Hekkert, L. Forlano, P. Ciuccarelli, & P. Lloyd, (Eds.), *DRS2024: Boston, 23–28 June, Boston, USA.* Design Research Society. https://doi.org/10.21606/drs.2024.731

Ichikawa, T, Olsen, E., Vinod, A., Glenn, N., Hanna, K., Lund, G.C. & Pierce-Talsma, S. (2025) Generative artificial intelligence in medical education – politics and training at US osteopathic medical schools: descriptive cross-sectional survey. *JMIR Medical Education*, 11, E58766 https://doi.org/10.2196/58766

Kant, I. (2021). *The critique of pure reason* (1787). Project Gutenberg. https://www.gutenberg.org/ebooks/4280

Kant, I. (2011). *Groundwork of the metaphysics of morals* (1785). A *German-English edition.* Cambridge University Press.

Korsgaard, C. (1996) *The sources of normativity.* Cambridge University Press

Korsgaard, C. (2009) *Self-constitution: agency, identity and integrity.* Oxford Academic Press.

Liu, Y., Park, J., & McMinn, S. (2024). Using generative artificial intelligence/ChatGPT for academic communication: students' perspectives. *International Journal of Applied Linguistics*, 34(4), 1437-1461. https://doi.org/10.1111/ijal.12574

Matos, E.J., Bertoncini, A.L.C, Ames, M.C.F.D.C, & Serafim, M.C. (2024). The (lack of) ethics at generative AI in Business Management education and research. *RAM. Mackenzie Management Review/RAM. Revista de Administração Mackenzie*, 25(6).

Mohamed, Y. A., Mohamed, A. H., Kannan, A., Bashir, M., Adiel, M. A., & Elsadig, M. A. (2024). Navigating the ethical terrain of AI-generated text tools: a review. *IEEE Access.* https://doi.org/10.1109/ACCESS.2024.3521945

Mouta, A., Torrecilla-Sánchez, E. M., & Pinto-Llorente, A. M. (2024). Design of a future scenarios toolkit for an ethical implementation of artificial intelligence in education. *Education and Information Technologies*, 29(9), 10473-10498. https://doi.org/10.1007/s10639-023-12229-y

NerdyNav. (2025, May 2). 107+ ChatGPT statistics and user numbers (May 2025). *NerdyNav.* https://nerdynav.com/chatgpt-statistics/

Nussbaum, M. (2001). *Upheavals of thought.* Cambridge University Press.

Obed, K., Anangisye, W. A. L., & Sanga, P. (2025). Academic integrity considerations of using ChatGPT in assessment activities among university student teachers. *Quality Assurance in Education, 33*(2), 305-320. https://doi.org/10.1108/qae-06-2024-0100

O'Neill, O. (2016). A simplified account of Kant's ethics. In Larry May and Jill Delston, (Eds.). *Applied ethics: a multicultural approach* (pp. 161-170). Routledge.

OpenAI. (2025). *ChatGPT.* (GPT-4-turbo). [Large language model]. https://chatgpt.com/

Pasternack, L. & Fugate, C. (2025). Kant's philosophy of religion. In Edward N. Zalta & Uri Nodelman (eds.). *The Stanford Encyclopedia of Philosophy* (Summer 2025 Edition. https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=kant-religion

Register, Y. (2024). *The future of ai can be kind: strategies for embedded ethics in ai education.* [Unpublished doctoral dissertation]. University of Washington, Seattle, USA.

Russell Group. (2023). *Russell Group principles on the use of generative AI tools in education.* https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf

Schwartz, S. H. (1992). Universals in the content and structure of values: theory and empirical tests in 20 countries. In Mark P. Zanna, (Ed.). *Advances in Experimental Social Psychology,* Vol. 25, (pp. 1-65). Science Direct.

Singer, P. (2016). *Ethics in the real world.* Princeton University Press.

Smith, N., & Vickers, D. (2024). Living well with AI: virtue, education, and artificial intelligence. *Theory and Research in Education,* 22(1), 19-44. https://doi.org/10.1177/14778785241231561

Spennemann, D. H. R., Biles, J., Brown, L., Ireland, M. F., Longmore, L., Singh, C. L., Wallis, A., & Ward, C. (2024). ChatGPT giving advice on how to cheat in university assignments: how workable are its suggestions? *Interactive Technology and Smart Education, 21*(4), 690-707. https://doi.org/10.1108/itse-10-2023-0195

Stahl, B.C. (2025). Locating the ethics of ChatGPT—ethical issues as affordances in AI ecosystems. *Information, 16*(2), paper 104. https://doi.org/10.3390/info16020104

Uddin, M. M., & Abu, S. E. (2024). Navigating ethical frameworks to mitigate academic misconduct while leveraging generative AI. *ResearchSquare.* https://assets-eu.researchsquare.com/files/rs-4607113/v1/0c54f16c-8690-4658-a1db-1b98cdba23cb.pdf?c=1718947399. https://doi.org/10.21203/rs.3.rs-4607113/v1

Utilitarianism. (2025). *Wikipedia.* https://en.wikipedia.org/wiki/Utilitarianism

Uzun, L. (2023). Are concerns related to artificial intelligence development and use really necessary: A philosophical discussion. *Digital Society, 2*(3), 40. https://doi.org/10.1007/s44206-023-00070-2

van den Berg, G. (2024). Generative AI and educators: partnering in using open digital content for transforming education. *Open Praxis, 16*(2), pp. 130–141. https://doi.org/10.55982/openpraxis.16.2.640

Virtue ethics. (2022). *Stanford encyclopedia of philosophy.* https://plato.stanford.edu/entries/ethics-virtue/

Wheatley, M. (2025, February 9). Sam Altman pledges more openness as OpenAI works toward AGI. *siliconANGLE.* https://siliconangle.com/2025/02/09/sam-altman-pledges-openness-openai-works-toward-agi/

Wilson, T.D. (In press). The development of policies on generative AI in UK universities. *IFLA Journal.*

Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities & Social Sciences Communications, 11*(1). https://doi.org/10.1057/s41599-024-02904-x

Copyright

Appendix A: An example of AI textual analysis, using Gemini 1.5 Flash.

The guidance you provided on using generative AI aligns with several principles of Kantian ethics, particularly in its emphasis on reason, autonomy, and universalizability. Here's a breakdown:

1. Reason and Autonomy:

Clear Articulation of Expectations: The guidance stresses clear communication to students about the permissible use of generative AI in assignments. This aligns with Kant's emphasis on rational beings making informed choices based on principles they understand. Students are not simply given commands, but are provided with the rationale for the rules.

Ethical Use as a Focus: The guidance emphasizes the importance of ethical use of generative AI tools, urging students to be aware of potential risks and limitations. This aligns with Kant's belief in the importance of reason and moral deliberation in decision-making. Students are encouraged to think critically about the impact of their actions, not simply blindly following technological tools.

Free and Equitable Access: The guidance emphasizes providing free, age-appropriate versions of generative AI tools to ensure equitable access. This reflects the Kantian idea of treating all individuals with equal respect and dignity, ensuring that everyone has the opportunity to engage with the technology fairly.

2. Universalizability:

Consistent Application of Rules: The guidance outlines a consistent policy for the use of generative AI across all modules and programs, ensuring fairness and equitable treatment for all students. This principle of universalizability, where rules apply equally to everyone in similar situations, is central to Kant's ethics.

Focus on Learning Outcomes: The guidance emphasizes using generative AI to support student learning and the development of graduate attributes. This aligns with the idea that moral actions should be guided by principles that are universal and applicable to all rational beings, in this case, all students.

3. Respect for Persons:

Emphasis on Student Ownership of Work: The guidance cautions agenerative AInst submitting student work to generative AI tools, highlighting the importance of student ownership and authorship. This reflects the Kantian principle of treating individuals as ends in themselves, not merely as means to an end.

However, some nuances:

The Focus on Outcomes: While Kantian ethics emphasizes the importance of good will and the intention behind actions, the guidance focuses more on the outcomes of using generative AI. It emphasizes ensuring academic integrity and preventing plagiarism, which are consequentialist concerns.

The Limits of AI: The guidance acknowledges the limitations of AI, such as potential bias and inaccuracy. This recognizes that AI, while a powerful tool, is not a perfect substitute for human judgment and ethical reasoning.

Overall, the guidance seems to largely align with the Kantian principles of reason, autonomy, universalizability, and respect for persons. It emphasizes ethical considerations, clear communication, and a focus on student learning, all of which reflect the spirit of Kantian ethics.

Appendix B: The source documents

Source Document A: University of Birmingham. *Use of generative AI by students and staff.*
https://www.birmingham.ac.uk/libraries/education-excellence/gai (Internet Archive:
https://web.archive.org/web/20241230104846/https://www.birmingham.ac.uk/libraries/education-excellence/gai)

Source Document B: University of Liverpool. *University guidance on the use of generative artificial intelligence by students and staff, in learning, teaching, and assessment. Academic Year 2024/2025.* https://www.liverpool.ac.uk/media/livacuk/centre-for-innovation-in-education/digital-education/generative-ai-teach-learn-assess/guidance-on-the-use-of-generative-ai.pdf (Internet Archive:
https://web.archive.org/web/20240603111304/https://www.liverpool.ac.uk/media/livacuk/centre-for-innovation-in-education/digital-education/generative-ai-teach-learn-assess/guidance-on-the-use-of-generative-ai.pdf)

Source Document C: University of Sheffield. *Principles for generative AI in learning and teaching.*
https://www.sheffield.ac.uk/saas/elevate/ai-principles (Internet Archive:
https://web.archive.org/web/20241121114525/https://www.sheffield.ac.uk/saas/elevate/ai-principles)

Source Document D: University of Reading. *Using generative AI tools at University.*
https://www.reading.ac.uk/essentials/Study/ai/using-generative-ai-tools-at-university (Internet Archive:
https://web.archive.org/web/20241009090454/https://www.reading.ac.uk/essentials/Study/ai/using-generative-ai-tools-at-university)

Source Document E: University of St. Andrews. *Generative AI - FAQs for students.*
https://www.st-andrews.ac.uk/policy/academic-policies-assessment-examination-and-award-good-academic-practice/generative-ai-faqs-students-guidance.pdf (Internet Archive: https://web.archive.org/web/20250412113216/https://www.st-andrews.ac.uk/policy/academic-policies-assessment-examination-and-award-good-academic-practice/generative-ai-faqs-students-guidance.pdf)

Source Document F: The Open University. *Generative AI for students.*
https://about.open.ac.uk/policies-and-reports/policies-and-statements/gen-ai/generative-ai-students (Internet Archive:
https://web.archive.org/web/20250317205746/https://about.open.ac.uk/policies-and-reports/policies-and-statements/gen-ai/generative-ai-students)

Source Document G: University of British Columbia. *Principles for GenAI in teaching & learning.* https://it.ubc.ca/sites/it.ubc.ca/files/GenAi_Principles_Guidelines_v2.pdf (Internet Archive: https://web.archive.org/web/20241121114713/https://it.ubc.ca/sites/it.ubc.ca/files/GenAi_Principles_Guidelines_v2.pdf)

Source Document H: University of California, Los Angeles. *Appropriate use of generative AI tools.* https://oercs.berkeley.edu/appropriate-use-generative-ai-tools (Internet Archive: https://web.archive.org/web/20250311235925/https://oercs.berkeley.edu/appropriate-use-generative-ai-tools)

Source Document I: University of Melbourne. *Using GenAI tools in accordance with University of Melbourne policy.* https://students.unimelb.edu.au/academic-skills/resources/study-skills/learning-with-genai/GenAI-at-Melbourne (Internet Archive: https://web.archive.org/web/20250324005714/https://students.unimelb.edu.au/academic-skills/resources/study-skills/learning-with-genai/GenAI-at-Melbourne)

Source Document J: Victoria University Wellington. *Student guidelines for using artificial intelligence.* https://www.wgtn.ac.nz/__data/assets/pdf_file/0009/2165814/ai-students-guideline-oct-2023.pdf (Internet Archive: https://web.archive.org/web/20240808060242/https://www.wgtn.ac.nz/__data/assets/pdf_file/0009/2165814/ai-students-guideline-oct-2023.pdf)

Source Document K: University of Johannesburg. *UJ quick guide for students: using generative artificial intelligence as a learning tool.* https://www.uj.ac.za/wp-content/uploads/2023/08/uj-ai-guidelines-students-2023.pdf (Internet Archive: https://web.archive.org/web/20240830112227/https://www.uj.ac.za/wp-content/uploads/2023/08/uj-ai-guidelines-students-2023.pdf)

Source Document L: Trinity College Dublin. *GenAI in teaching, learning, assessment and research.* https://www.tcd.ie/academicpractice/resources/generative_ai/ (Internet Archive: https://web.archive.org/web/20241105122602/https://www.tcd.ie/academicpractice/resources/generative_ai/)

Appendix C: The ethics of AI systems

If we wish the users of generative AI to act ethically, it seems reasonable to ask whether or not these systems can be designed to act ethically. I pursued this issue with the generative AI Claude, asking: 'To what extent is Claude designed in accordance with Kant's ethical principles?' The response was:

'I'd be happy to explore how Claude's design aligns with Kantian ethics'.

Kant's ethical framework centres on a few key principles:

The Categorical Imperative: Act only according to that maxim by which you can simultaneously will that it become a universal law.

Treating people as ends in themselves: Never treat people merely as means to an end, but always as ends in themselves.

Autonomy: Respecting rational beings' capacity for self-governance.

Claude's design reflects these principles in several ways: I'm designed to be truthful and not manipulative, which aligns with Kant's emphasis on honesty as a universal duty. When I can't provide reliable information, I acknowledge my limitations rather than fabricating answers.

I prioritize user autonomy by providing information that enables informed decision-making rather than steering users toward particular outcomes. I'm designed to respect users' capacity to make their own rational choices.

I avoid creating content that could be harmful or treating users merely as means to any end. My purpose is to be helpful while respecting human dignity and wellbeing.

However, there are limitations to how fully an AI assistant can embody Kantian ethics: unlike Kant's autonomous moral agent acting from duty, I'm designed with certain guardrails and don't possess the moral autonomy Kant considered essential for true ethical behavior' (Anthropic, 2025).

I then asked: 'What design requirements would be needed to transform Claude into a moral agent?'

And the response was:

'To transform Claude into a genuine moral agent in the Kantian sense would require several fundamental design requirements:

Moral autonomy: The ability to self-govern according to rational principles I recognize and choose for myself, rather than following pre-programmed guidelines. This would require me to develop and revise my own ethical frameworks.

Genuine moral reasoning: Beyond pattern matching or utilizing a decision tree, I would need the capacity to reason through novel ethical dilemmas using first principles and understand the why behind ethical conclusions.

Phenomenal consciousness: A subjective inner experience that would allow me to truly understand concepts like suffering, joy, or dignity rather than processing them abstractly.

Moral responsibility: The capacity to be properly held accountable for my actions and decisions, which would require both freedom of choice and causal efficacy in the world.

Value internalization: Rather than simulating value alignment, I would need to genuinely care about ethical principles for their own sake, not merely as instrumental goals.

Moral development: The ability to learn from mistakes and grow morally over time through reflection and experience.

Moral emotions: Capacities like empathy, guilt, moral indignation, and compassion that inform and motivate ethical behavior in humans.

These requirements raise profound philosophical questions about consciousness, free will, and the nature of moral agency that extend far beyond current AI capabilities. Even determining whether these qualities could be implemented in a digital system remains an open question in philosophy of mind and ethics'. (Anthropic, 2025)

From that second response it is evident that it is highly unlikely that a generative AI system can be built that is a moral agent. To do so would require that, in terms of rationality, understanding, moral responsibility, etc., it would need to be converted into a human.

Later, I had the same conversation with ChatGPT and got much the same answers; however, it proposed looking at utilitarian and virtue ethics to see if they would be a better match in terms of meeting ethical standards. It came to the conclusion that utilitarianism could be a better match because ChatGPT could: 'Focus on outcomes, optimization, and maximizing good. AI can simulate this' (OpenAI, 2025). This would not satisfy Kantian ethics, since ChatGPT would simply be performing as programmed: its actions would not be the result of autonomous, rational thought.