

An Analysis of Kantian Ethics, Its Influence and Applicability in the Present Era of Artificial Intelligence

Dr. Gopa Bhattacharyya

Assistant Professor, Department of Philosophy
Santipur College, Santipur, Nadia (University of Kalyani)

Email-gopaphil@gmail.com

Abstract

German Philosopher Immanuel Kant provides a moral theory for evaluating the ethical actions of people and states. These theories may be used in advanced thinking discussions such as the use and advancement of artificial intelligence and robotics. It may prove to be an important topic today. Its ideology is inclusive. It includes moral grounds and valid reactions to unethical actions and extends to all individual actors, regardless of whether they are wrongdoers, illegitimate warriors or unfair rivals. Two independent contributions to the discourse make up Kantian ethics. Firstly, they provide a human-focused ethical structure in which human life and skill focus on a moral philosophy that establishes norms and guides our interpretation of moral behaviour. Second, practical theory specific and appropriate to accomplishing good behaviour is the end objective of Kantian ethics. This paper aims to examine Kantian ethical fundamental philosophies and its impact on the contemporary philosopher. I would also discuss the moral concerns discussed above by discussing its application in robotics and artificial intelligence. I would refer to the central elements of his ethical ideas on A. I, exploring completely autonomous methods that establish laws in humanity, moral thinking, and reverence for human dignity and the discrepancy between the human will and the machine will.

Keyword: Kantian ethics, goodwill and duty, Categorical Imperative, Universalizability, Formula of Autonomy, Influences, Medical ethics, Abortion, Sexual ethics, Lying, Artificial Intelligence

Article Received: 18 October 2020, Revised: 3 November 2020, Accepted: 24 December 2020

Introduction:

Kantian philosophy is based on the theory that "It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a good will". The theory was developed as a result [2- 4] of enlightenment rationalism, which claims that "an action can only be good if its maxim—the principle behind it—is a duty to the moral law, and arises from a sense of duty in the actor."

The categorical imperative that works on all individuals irrespective of their needs or desires is fundamental to Kant's moral law construction. Kant proposed the categorical imperative in many different ways. His universalization theory demands that action be carried out without contradiction to be permissible for all citizens. The second part of the categorical imperative of Kant's human formulation notes that as an end in itself, people never need to regard other people merely as means for a reason, but rather as ends in themselves. The concept of autonomy assumes that

rational actors are bound to moral rules by their own free will. In contrast, Kant's definition of the Kingdom of Ends demands that people behave as though they are forming a law for a hypothesized domain. Also, Kant differentiated between perfect and imperfect tasks. As an application of his ethics, Kant used the example of lying. [5 to 8]

Goodwill and duty

Kant developed the basis for an ethical rule in his joint works with the notion of duty [3]. Kant launched his ethical philosophy by arguing that goodwill is the only virtue that can be good without qualification. No other virtue has this status as the other virtues may be used for immoral reasons. For example, the virtue of loyalty is not good if one is loyal to an evil person. The goodwill is special because it is still good and holds its spiritual meaning even though its moral aims cannot be achieved[4]. Kant sees goodwill as one universal principle, which openly exploits other virtues for moral reasons [5].

For Kant, goodwill is a larger understanding than a duty. A will that works out of obligation can be seen as a will that overcomes moral law observance barriers. Goodwill is also a particular case of goodwill, and is seen under unfavourable situations. Kant claims that only duty-related actions have a positive meaning. This does not mean that actions done purely on the grounds of service are meaningless [6].

The definition of duty by Kant does not mean that people would perform their duties unwillingly. While duties frequently constrain people and lead them to act against their inclinations, they are still the product of an agent's will: they want to uphold the moral rule. So it is because the logical incentives matter more to an agent than to his opposing inclinations that he conducts an action from the assignment. In acknowledging the demands that rationality makes on us, Kant wished to step beyond morality as externally imposed obligations and offer an ethic of autonomy. [7]

Perfect and imperfect duties

Applying the categorical imperative, duties arise because failure to fulfil them would either result in a contradiction in conception or in a contradiction in the will. The former is classified as perfect duties, the latter as imperfect. A perfect duty always holds true. Eventually, Kant argues that there is only one perfect duty in reality—the Categorical Imperative. An imperfect duty provides flexibility because we are not obligated to be entirely beneficial at all times, but can select the times and places in which we are [8]. Kant claimed that perfect duties are more important than imperfect duties: if there is a discrepancy between duties, the perfect duty must be fulfilled [9].

Categorical Imperative

The categorical imperative, [10] from which he derived four more formulations, is the primary formulation of Kant's ethics [11]. Kant made a distinction between imperatives that are categorical and hypothetical. If we want to fulfil our needs, a hypothetical imperative is one that we must obey: 'go to the doctor' is a hypothetical imperative because we are only obligated to obey it if we want to get better. Regardless of our wishes, a categorical obligation binds us: everyone has a responsibility not to lie, regardless of circumstances and even though it is in our interest to do so. . These conditions are morally binding since they are focused on reason rather than on

contingent evidence regarding an agent [12]. We may not opt-out of the categorical imperative, because we can not opt-out of being reasonable agents, unlike hypothetical imperatives that attach us insofar as we are part of a community or society to which we owe duties. By virtue of being rational agents, we owe an obligation to rationality; therefore, rational moral standards still extend to all rational agents [13].

Universalizability

The first formulation of the Categorical Imperative by Kant is that of universalisability: act only according to the principle by which you can, at the same time, make it a universal law. It is according to a rule or maxim when someone acts. For Kant, an act is only appropriate if one is ready for the theory that makes it possible for the action to be a universal law under which everybody behaves. Maxims fail this test if, when universalized, they create either a contradiction in creation or a contradiction in the will. "A inconsistency in conception exists when it ceases to make sense if a maxim were to be universalized, since the "maxim would inevitably undermine itself as soon as a universal law was created." [16-20] For example, if the principle "It is acceptable to break promises" was universalized, no one would trust any promises made, because the idea of a promise would become meaningless. The theory is not moral since universalization is logically impossible. A maxim may also be unethical if it when universalized, produces inconsistency in the will. This does not imply a logical paradox, but it leads to a state of affairs that no reasonable being can wish for by universalizing the maxim. For instance, Julia Driver argues that when universalized, the maxim 'I will not give to charity' creates a contradiction in the will since a world in which no one gives to charity will be undesirable for the individual who acts according to that maxim.

Humanity as an end in itself

The second wording of the Categorical Imperative in Kant is to regard humankind as one's own goal: behave so that mankind is always at the same time as an end, whether in yourself or in another's person, and never as a means.

Kant believed that human beings should never be regarded as ends alone; they should always be treated as ends themselves, demanding fair regard

for their own logical motivations. This stems from Kant's argument [21-24] that the motivation of reason is moral: it allows one to value reason as a justification for all people, including others. A human being can't rationally accept that it is actually used as a method, so it must still be viewed as a reason. Kant argued that moral obligation is a logical necessity: what is genuinely sought is morally correct. Because all logical actors themselves will rationally become an objective and never only away, they will be regarded as a result of morality. This does not mean that we should never treat a human being as an end, it is indeed an end in itself.

Formula of Autonomy

Kant's autonomy formula reflects the notion that because of their rational will, rather than any outside force, an individual is obligated to obey the Categorical Imperative. Kant argued that the Categorical Imperative would be rejected by any moral law motivated by the desire to satisfy any other interest, leading him to conclude that moral law could only emerge from a reasonable will. This theory allows people to accept others' right to behave autonomously and suggests that what is required of one person is required of all, as moral laws must be universalizable.

Kingdom of Ends

The Kingdom of Ends is another formulation of Kant's Categorical [26] Imperative: A rational person must always regard himself as giving laws in a kingdom of ends that is made possible by freedom of will, either as a member or as sovereign. This formulation demands that acts be treated as though their principle is to provide a rule for a conceptual Kingdom of Ends. Accordingly, people have a duty to behave on values that would be recognized as rules by a society of reasonable agents. Each individual can only embrace maxims in such a society that can rule every member of the group without treating any member merely as a means to an end. While the Kingdom of Ends is an ideal, as legislators of this ideal kingdom, the acts of other people and events of nature mean that actions of good intentions often lead to harm, we are still expected to behave categorically.

Influences on Kantian ethics

1. Lutheran Pietism, to which Kant's family subscribed, stressed integrity and spiritual living, more concerned with emotion than rationality,

over doctrinal belief. Kant believed that reason is necessary, but that morality and good will should be concerned. As a version of the Lutheran doctrine of righteousness, Kant's description of moral development as the turning of inclinations towards the fulfilment of obligation has been identified.

2. Kant's view of human beings' basic meaning was influenced by the political philosopher Jean-Jacques Rousseau, whose social contract. Pojman also cites current ethical controversies as critical to the growth of the ethics of Kant. Kant preferred rationalism over empiricism, which meant that, rather than anything centered on human desire, he regarded morality as a form of experience.

3. Natural law, the assumption that, by nature, the moral law is decided.

4. Intuitionism, the idea that humans are conscious of objective moral realities in an unconscious way.

Influenced by Kantian ethics:

Karl Marx

While Karl Marx opposed many of the ideas and assumptions contained in the ethical writings of Kant, Philip J. Kain claims that his views on universalisation are much like Kant's views on the categorical imperative, and his notion of freedom is similar to Kant's notion of freedom. In his theory of communist society, which is set up by a historical agent who will make the realization of morality possible, Marx was also inspired by Kant.

Jürgen Habermas

A theory of discourse ethics has been proposed by the German philosopher Jürgen Habermas, who claims to be a descendant of Kantian ethics. He argues that behavior should be based on communication with those concerned, in which their desires and intentions are addressed so that everyone can understand them. Rejecting any sort of coercion or manipulation, Habermas believes that it is important for a moral decision to be made to agree between the parties. Like Kantian ethics, discourse ethics is a cognitive ethical theory, in that it supposes that truth and falsity can be attributed to ethical propositions. It also formulates a law by which ethical acts can be decided and suggests that ethical actions should be universalisable in a similar way to Kant's ethics.

Karl Popper

Karl Popper changed the ethics of Kant and concentrated on his moral theory's subjective aspects. Like Kant, Popper believed that morality can not be obtained from human nature and that

moral virtue is not the same as self-interest. He radicalized the autonomy conception of Kant, removing its components of naturalism and psychology. He argued that it is not possible to justify the categorical imperative by logical nature or pure motives.

John Rawls

In his work *A Philosophy of Justice*, Kant's ethics was inspired by the moral philosopher John Rawls's social contract theory. Rawls argued that it would be equal to a fair world.

Thomas Nagel

In the associated areas of moral and political philosopher Thomas Nagel was very influential. Nagel has long advocated Kantian and rationalist approaches to the moral philosophy, under John Rawls' guidance.

Contemporary Kantian ethicists

The philosopher Onora O'Neill, a modern Kantian ethicalist working under John Rawls at Harvard University, advocates a Kantian approach to social justice problems. O'Neill claims that Kantian's popular account of social justice does not focus on unjustified idealizations or stereotypes. She states that Kant has historically been accused by philosophers of idealising individuals as autonomic agents without social backgrounds or life goals but insists that without such an idealistic perspective, Kant's ethics can be read.

Marcia Baron

On this subject, Marcia Baron has attempted to defend Kantian ethics. She claims that these issues only occur when people misinterpret what their obligation is, after providing a variety of explanations that we might find acting out of duty unacceptable. It is not intrinsically wrong to act out of obligation, but immoral consequences will arise when individuals misinterpret what they are obligated to do. It is not appropriate to see duty as cold and impersonal: one may have a duty to develop their personality or to strengthen their personal relationships.

Different fields in the light of Kantian ethics

Autonomy

In its dual argument that people are co-legislators of morality and that morality is a priori, a number of philosophers have all argued that the Kantian interpretation of ethics rooted in autonomy is contradictory. They claim that if anything is

inherently a priori (i.e. unchangingly existing before experience), then it should not also be partially dependent on individuals who have not always existed. In the other hand, if people actually do enforce morality, they are not objectively bound by it, so they are still free to change it.

Applications

Medical ethics

Kant claimed that human beings' innate capacity to reason should be the pillar of morality, and that it is the capacity to reason that makes human beings morally powerful. Therefore, he claimed that all people should have the right to common integrity and respect. Margaret L. Eaton claims that, according to Kant's ethics, even though they were the patient themselves, a medical practitioner must be satisfied for their own methods to be used by and on others. In other words, Kant's autonomy criteria would mean that patients should be able to make an informed choice on medication, which would make testing for the unknowing patient unethical.

Abortion

Kant suggested, she suggests treating a woman as a dignified self-employed citizen with her body's control. In Kantian ethics, she claims that women's right to choose is paramount and that abortion is the choice of mother.

Sexual ethics

Kant saw humans as subject to the animalistic impulses of self-preservation, preservation of species, and preservation of pleasure. He argued that people are responsible for preventing negative or degrading maxims, including suicides, sexual damage and drunkenness. This led Kant to regard sex as degrading, because it reduces people to a pleasure subject. He confesses that sex is worse than suicide, reducing a human to below the status of an animal; he argued that the violation should be penalized with castration and that bestiality calls for the expulsion from society. He acknowledged that it was only during marriage, which he considered to be "a merely animal union."

Commercial sex

Feminist philosopher Catharine MacKinnon argued that Kant's standards view many modern practices as unethical because of women's

dehumanization. She argues that sexual abuse, prostitution, and pornography objectify women and do not fulfil Kant's human autonomy standard. Commercial sex has been criticized for making all partners into objects (and thereby using them as a means to an end); mutual consent is problematic when people decide to objectivise themselves while consenting. Alan Soble has observed that more liberal Kantian ethicists believe that women's consent will legitimize their involvement in pornography and prostitution, depending on other contextual factors.

Animal ethics

Since Kant regarded rationality as the basis for a moral patient, he thought there were no moral rights for animals. According to Kant, animals are not logical, so it is impossible to be complied with immorally. Although Kant claimed we had no duty towards animals, he thought it wrong to be cruel to them because our actions could affect our attitudes towards people: if we get used to harming animals, then causing harm to human beings are more likely to be acceptable.

Lying

Kant claimed that the Categorical Imperative gives us the principle that even though we want to have positive outcomes, such as lying to a killer to keep them from locating their intended victim, we should not lie under any circumstances. Kant argued that the outcome could be unexpectedly harmful because we cannot completely understand the implications of any action. We should then behave instead of preventing a probable error to prevent the known — lying —. If the effects are negative, we are irrefutable that we did our job. Driver claims that if we want to formulate our maxims differently, this may not be a problem: the maxim 'I would lie to save an innocent life' can be universalized. However, the murderer can always be viewed as a means to an end that we have an obligation to stop. So in the case of Kant, we would still have to tell the truth to the killer.

To summarise:

Advantages of Kantian ethics

- The morality of Kant is rather clear and reason-based, making it available to everyone.
- Duty is an aspect of human experience.
- Morality relies not on intentions, effects or moral laws.

- The categorical imperative gives us laws that apply to all and orders us to honour human life.
- It makes apparent that ethics is doing one's duty and not only doing emotions. We cannot presume that what's good for us is good for anyone else.
- It seeks to treat all equally and fairly and wants to correct the utilitarian idea that some can suffer as long as others are happy.
- As they are the logical high point of life, Kant sees people as being of intrinsic value. This ensures that they will not be enslaved or harmed (Basis of the Declaration of Human Rights)
- Fair treatment of people gets rid of biases that often impair decision-making.
- Categorical imperatives tell us precisely what is right and wrong, giving us a strong sense of moral guidelines.
- Kant makes a simple difference between obligation and preference.
- The moral value of an action stems from its inherent correctness, thereby avoiding teleological ethics issues
- Individuals usually have the same morality principles.
- Ethical behaviour should be based on reason rather than emotion.
- Many people recognize the meaning of obligation, which is part of what it means to be human.
- There is a distinction between responsibility and desire.

Disadvantages of Kantian ethics

- The theory of Kant is abstract and not always easy—it tells you what kind of actions are good, but not the right thing to do in specific circumstances.
- Alasdair MacIntyre claims that the concept of universality can be used to justify virtually anything.
- If his ethics are deontological or teleological, Kant seems confused. The basic definition is deontological, but the kingdom of ends has a potential objective.
- Some philosophers conclude that Kant is in favor of liberty and that moral agents must obey the values provided in the Categorical imperative that contradict each

other. As long as I follow these rules, am I free?

- Individuals rarely behave out of obligation, since they still have some expectation of what they will get in return.
- In Kant's philosophy, there is no room for love and personal relationships. Some philosophers claim that putting duty above emotion is cold and inhuman.
- To understand the rationally ordered universe, Kant's view relies on some conception of God, meaning atheists do not support this theory.
- Kant advises us in general terms to value others and not treat them as a means to an end, but doesn't tell us what to do in specific cases
- Philippa Foot, among others, has criticized the theory of Kant, which does not support the double effect condition.
- There are no exceptions to the use of individuals as means to ends that severely limit our conduct.
- Kant never tackles what a 'human' is (issues with abortion, euthanasia etc)
- Duties clash under some situations. W D Ross argued that we should take on Prima Facie duties (first sight) so pursue it unless a contrasting duty emerges to make a larger point.
- Universal sable maxims when applied to moral dilemmas are challenging.
- People are different and do not always have the same 'good will' feeling.
- Not everybody is willing to make rational moral choices.
- Love and compassion are part of every action we take because we are human beings.
- Before acting, it is human nature to weigh the consequences.

STRENGTHS AND WEAKNESSES OF KANT'S THEORY

STRENGTHS:

- Not a consequentialist, Kant realized that a bad behaviour may have positive consequences.
- Universal: Offers universal moral rules, irrespective of community.
- Clear: The theory of Kant is argued as straightforward. 'If anyone did that to you,

would you like it? "No?-No?" "Don't do it to someone else then"

- Autonomy-The greatest respect for human dignity and autonomy is granted to Kant.
- Rational: Kant is not motivated by emotion. Favouritism does not make his theory. It is a hypothesis that is purely rational.
- Human Rights- Provides a human rights base.
- Equality and Justice- Provides a framework for contemporary equality and justice conceptions.
- International Law provides the framework for a great deal of British and international law.
- Objective- Objective norms in cases, rather than subjective.
- Duty-Bad alternatives may arise from acting with love/compassion. It's always right to behave out of obligation.
- A system of rules works efficiently, and everybody understands the responsibilities there.
- Authority-the laws of Kant are rational and reasonable-as if someone broke promises, they would mean nothing anymore.
- Ends in themselves: Kant honors human life as ends rather than means, but current medical ethics contradicts this.

WEAKNESS

- Consequences-Consequences may often be so serious that violating rules might be appropriate.
- Inflexible- If the condition warrants it, violating an unhelpful rule should be appropriate.
- Lack of motivation-The knowledge that anything is unreasonable does not offer any incentive to do the right thing.
- Conflicting duty: Caring for your mother vs. caring for your dad. Which one should I follow?
- Absolute Duty-Ross says that we have an absolute obligation-we have a duty to break a promise often.
- Moral law- The nature of moral law is disputed by some philosophers. Why are

we expected to assume that objective morality exists?

- Anthropocentric- Kant sees no inherent worth for non-human species.
- Too Ambiguous- It is not clear how extensive our application of the CI should be. Uh. E.g. When my council every 2 weeks needs to collect garbage. But I assume it's too long, rationally speaking. Is that morally wrong, really?
- If the SS asked if you were hiding Jews, it would be difficult to form maxims. What is the maxim you're following? 'Don't lie' or 'Don't expose violence to others'?
- A priori: Some slander the a priori technique. In cases such as medical ethics, experience is not better.
- Unrealistic-Simply because we follow this direction does not indicate that everybody else too. For instance, by being a pacifist, I would leave myself open to a non-Kantian attack.
- Unforgiving-Kant believed "an eye for an eye" in retributive justice. Whereas someone like Bentham felt that it should be rehabilitative, things should be made better.
- Every situation is unique-In a world where every scenario is different, universal laws are not much use. If no case is the same, morality should not be absolutist, but rather relativist.

Kantian Ethics in the Age of Artificial Intelligence and Robotics

Artificial intelligence and robotics are prevalent in everyday life and are likely to extend to new levels, probably replacing human decision-making and action. Some examples are self-driving vehicles, robots for homes and healthcare, and autonomous weapons. There seems to be a difference between potentially benign civilian technology usage (e.g. unmanned aerial vehicles carrying drugs) and potentially malicious military usage (eg lethal autonomous weapons killing human combatants). The metaphysical foundations of human life and ethical behaviour were questioned by machine-mediated human interaction. There are philosophical concerns regarding the desirability of replacing human roles and the human mind with such technologies, aside from the technological difficulties of maintaining ethical behaviour in artificial intelligence and robotics. In order to act ethically, how can

artificial intelligence and robotics participate in moral reasoning? Is a new set of moral rules necessary? What happens to human contact when technology mediates it? Should technologies like that be used to bring an end to human life? Who bears responsibility for artificial intelligence and robotics misconduct or adverse behaviour? In regards to Kantian ethics, this paper addresses these matters.

Core elements of Kantian ethics

Kantian ethics offers a human-cantered ethical structure that places human life and ability at the core of a standard-creating philosophy that governs our understanding of moral behaviour [1]. The works of Kant can be criticized for being dense and opaque, but his ultimate objective was practical philosophy that could contribute to the development of practical implementation-capable principles or laws. Kantian ethics, which defines the human-cantered ethical structure, explores the following main elements: the categorical imperative; will independence; rational beings and rational thinking capacity; and human dignity and humanity as an end in itself [8].

Moral rules capable of universalisation in relation to artificial intelligence and robotics

How can Kant's key formulation of the categorical imperative be applied by artificial intelligence and robotics-'act only on the theory by which you can at the same time make it a universal law '? Artificial intelligence and robotics do not have the potential for human critical reasoning or a free will to grasp what constitutes a necessarily desirable rule, practical and useful to universalize. But in the design, development, testing and implementation of such technology, there is a human entity, so that human beings are responsible for enforcing the categorical imperative. In order to ensure ethical use and moral behaviour, humans decide the regulations are programmed into the technology. They must be 'accessible and shareable' for these laws to be worthy of universalisation. For instance, there is a lot of controversy in the civilian sphere regarding open access and the use of artificial intelligence to collect personal data, potentially endangering privacy. Discussions on lethal autonomous weapons in the military sphere under the auspices of the United Nations Convention on Certain Conventional Weapons reflect a mechanism for universalizing rules that can govern or prohibit such weapons. Indeed, there are emerging opinions among some states for a rule of

preventive prohibition, and most states recognise that ethical, legal, and humanitarian considerations must be taken into account by any rules regulating lethal autonomous weapons. The potentially broad objectives and uses of artificial intelligence and robotics technology can lead to the emergence of competing rules that may or may not arise here, but some preliminary issues are considered relevant to the design and form of laws.

How will legislation be designed to govern the ethical use and functioning of technology? This depends on whether the technology is intended to replace human functions and critical thought entirely or to complement and supplement those human features. Completely autonomous technology refers to the substitution of human critical thinking abilities and free will by artificial intelligence and robotics, such that laws come from the technology itself rather than from humans. On the other hand, human-machine integrated technology refers to technology that, in some cases, supports and assists humans in order to develop, influence, monitor and tailor rules through a combination of interaction and intervention between humans and machines. There are legal ramifications of both kinds of rule-generating methods.

Fully autonomous rule-generating approach

A truly autonomous approach to producing rules would suggest that technology creates its own rules and actions without human reference or interference. The software makes its own decisions after the initial design and programming by people. This is 'machine learning' or 'dynamic learning systems' whereby the machine relies on its own database and experiences to produce potential rules and behaviour. For example, fully autonomous weapons systems will have independent thinking ability in terms of obtaining, tracking, choosing, and attacking human targets in combat based on previous military scenario experiences. Such an apprentice The rules that a completely autonomous weapons system would create beyond what it was supposed to do are ambiguous and unpredictable, so that it would not be consistent with international humanitarian law or Kantian ethics. In the civilian domain, fully autonomous technology will produce rules that, by causing human redundancies, unemployment, and income insecurity and inequality, adversely affect human self-worth and development. In the rule-generating process, detrimental effects on human self-worth and development, and ambiguity and unpredictability are counter to what is essentially

beneficial to humanity; such a system does not generate rules that are necessarily desirable, practicable, valuable and universalizable. Without any restrictions, a perverse 'machine subjectivity' or 'machine free will' would exist, analogous to Kant's 'hypothetical imperatives' created by human subjective desires.

Human-machine rule-generating approach

In both the civilian and military worlds, a human-machine rule-generating method currently exists. For example, IBM prefers the word 'increased intelligence' rather than artificial intelligence because this better represents their aim of developing systems that improve and scale human knowledge and skills rather than replacing them. The technology focuses on realistic applications that help individuals perform well-defined functions (for example, robots that clean houses; robots working with humans in production chains; warehouse robots that take care of the tasks of an entire warehouse; companion robots that entertain, talk, and help elderly people maintain contact with friends, relatives, and doctors). Remotely controlled and semi-autonomous weapons in the military sphere merge human activity with technology for weapons. Human intervention is required to decide when it is appropriate to carry out an attack order or to trigger an abortion function. This form of rule-generating strategy holds the individual at the core of decision-making. But what happens if there are human-machine interface problems (e.g. errors; deficiencies in performance; communication breakdown; communication connection loss; mis-coordination)? This could prove to be fatal in communication and coordination-based integrated human-weapon systems, and a back-up mechanism will need to be in place to suspend or abort operations. What happens if alternate or random rules that cause failure, non-performance, or harmful effects are hacked into the technology? The same issue applies to completely autonomous technology and seems to be a valid reason for limiting the capacity to use and output to set tasks, managed scenarios or environments where any possible harm can be contained.

For the formulation of ethical behaviour in artificial intelligence and robots, the possible exclusion of non-human beings and inanimate objects from Kant's human-centric approach to the categorical imperative can explicitly apply. Human interference will set the categorical imperative if there is concern about computer unpredictability and ambiguity in generating its own laws, as

'technology must always prioritize human existence over property harm or non-human animal life.' With the German Government recently approving ethical guidelines for autonomous vehicles requiring that: 'the safety of human life takes top priority in balancing legally protected rights, this human-centered approach is already being tested in self-driving cars. Thus, the systems must be designed to tolerate harm to animals or property in a dispute within the limits of what is technologically feasible, if this means that personal injury can be avoided.

'Difference between 'human will' and 'machine will'

It is difficult to transpose Kant's autonomy of the will into technology because it depends on principles such as self-worth, integrity, independence, capacity to create laws, and interaction. A computer does not have a sense of or be able to apply meaning to these principles. The "human will" grows to instruct moral behaviour through character and experience. 'Machine learning' or 'dynamic learning systems' that produce rules and actions based on a database of previous experiences, may resemble a type of 'machine will' that makes ethical decisions based on behavioural rules that are internally learned. But human will is far more dynamic, elusive, and capable of dealing with spontaneity in reaction to new circumstances outside of rule-based behavior. To achieve a state of moral standing and be able to participate in moral actions, autonomy of the will requires the inner and outer development of the individual. An inherent sense of right and wrong is indicative of this. Can computers imitate this kind of 'will'? In autonomous weapons, artificial intelligence can allow machine logic to evolve over time to recognize correct and incorrect behaviour, demonstrating a limited sense of autonomy. But the computer does not have its own 'will' or understand what independence is and how to achieve it by following values that will establish the will's inner and outer autonomy. It does not have a self-determining ability to make decisions between various degrees of right and wrong. The person can decide to challenge or go against the rules, but even in cases of failure and mis-programming, the computer cannot. It has no understanding of liberty and how it can be improved for both people and themselves. Moral dilemmas would not burden the system, so the deliberative and analytical aspect of decision-making is totally absent (vital for recognizing the implications of actions and ensuring appropriate responses). There is a limited sense in which the

outer aspect of Kant's autonomy of the will can mimic artificial intelligence and robotics. To encourage cooperation and avoid conflict among themselves, robots will have a common code of interaction. In order to prevent collisions and failures, autonomous weapons operating in swarms should establish rules that govern how they communicate and coordinate behaviour. But these are examples of practical interaction between person and machine that do not apply to human interaction, and thus do not constitute a form of universalization-capable autonomy of the will.

We really mean reliability when we speak about trust in the sense of using artificial intelligence and robotics. Confidence relates to people's statements and acts and is not an abstract thing. Machines should not be attributed to trust without the sovereignty of the will, in the Kantian sense, and without the capacity to make claims. Algorithms cannot decide whether anything is trustworthy or not. Thus, trust is used metaphorically to mean functional reliability; that the system performs tasks without appropriate error or minimal error for the set purpose. But there is also an expansion of this notion of human agency-related faith in creating and using artificial intelligence and robotics. Can we trust people involved in the creation of such technology to do so with ethical considerations in mind, i.e. to minimize needless suffering and harm to people, not to violate basic human rights? Can we trust those who would use it to do so for benevolent rather than malevolent purposes once the technology is developed? These concerns also occur in discussions on data security and the right to privacy in relation to technology trafficking practices involving personal data. Again, this goes back to the principles that represent ethical behaviour that will be installed and allow the technology to discern right from wrong.

Kantian notion of rational beings and artificial intelligence

Kant's emphasis on people's critical thinking ability applies to potential rather than actual possession of rationality, taking into account defective rationality, unethical actions, and circumstances in which people can purposefully behave irrationally in order to gain some advantage over an opponent. If it participates in a pattern of logical thought from which it rationalizes and takes action, technology can be considered to have rational thinking capacity. Although the idea of Kant is primarily reserved for

people who can create a set of rules regulating moral actions (a strictly human endeavour and not one that can be generated mechanically), rather than actual logical thought, the ability element can be fulfilled by artificial intelligence and the potential of robotics. But this tends to be a low level that poses questions about the predictability and certainty of real-life technology scenarios. There would also need to be even greater consistency and assurance as to what kind of rationality the technology would have and how it would be used in human scenarios.

There is a clear contrast between a calculating machine's logic and the wisdom of human judgment when we equate machines with humans. Machines conduct cost-effective and rapid peripheral processing operations based on quantitative analysis, repetitive behaviour, and data sorting (e.g. mine clearance; and detection of improvised explosive devices). They are excellent at automatic reasoning and in certain tasks they will outperform humans. But in human contexts where artificial intelligence can be used, they lack the deliberative and sentient aspects of human thought required. They have no complex cognitive capacity to assess a specific situation, exercise judgment and refrain from taking action or restricting harm. Robots have no instinctive or innate ability to do the same, unlike people who can pull back at the last minute or select a workable alternative. For example, the use of discretion during warfare is vital for the enforcement of rules for the prevention of needless suffering, the adoption of precautionary measures and the evaluation of proportionality. In robotics, such discretion is missing.

Universal and particular moral reasoning in artificial intelligence and robotics

In order to act ethically, how can artificial intelligence and robotics participate in moral reasoning? Does technology have universal or basic moral reasoning in place? The universality of moral reasoning in ethical philosophy means that there is a justification for doing so in any case where an agent can morally do something. Kant's categorical imperative makes it clear that it is a particular form of reason; one based on a universalisation-capable law. In comparison, 'particular' moral reasoning, instead of searching for similar circumstances from which laws arise, does not rely on universal rules to explain moral responsibilities and reasons for actions. Does universal moral reasoning in artificial intelligence

reference all specific instances that involve specific moral reasoning?

In the civil and military realms, recent developments illustrate moral dilemmas and the relevance of human moral thinking to mediate between conflicting social interests and values. Companion robots may need to be aware of privacy and security concerns related to supporting their human companion and communicating with third parties (e.g. safety and disclosure of personal data; strangers who may pose a threat to the property, physical and mental integrity of the person) (eg hospitals; banks; public authorities). It may be important to build companion robots so that they do not have full control over their human companion's life, which violates human dignity, autonomy, and privacy. In general, robots will need to lack the ability to deceive and control humans in order to maintain human critical thought and free will. Then there is the question of whether, in the lethal force decision-making phase to destroy another human being, completely automated weapons should be built to replace human combatants. Is there a universal moral logic to overcome certain dilemmas that technology might possess? Or does a particular moral rationale, specific to the technology or situation, have to be available?

A mixture of knowledge, judgment, experience and feelings is involved in human moral reasoning. Societal, cultural, political, and religious factors can also rely on it. The Universal Declaration of Human Rights of 1948 offers, arguably, a common standard of universal moral reasoning in defining general human rights, which are considered universal, indivisible and inviolable. Specific moral reasoning can attempt to restrict reasoning-relevant factors based on the capabilities of the technology or the situation in which it is used. An autonomous weapon that is only capable of targeting and destroying buildings, for example, would not have to consider factors relevant to a human combatant's position, appearance, intentions or activities. In the other hand, if the weapon is used in uncomplicated and non-mixed environments and is capable of human targeting, moral justification that conforms to the concepts of differentiation, proportionality and undue suffering will have to be involved.

However, the relative meaning and importance of some human rights that may lead to arbitrary and contradictory implementation may or may not be interpreted by computer moral reasoning. One way

to resolve this is to design a value-neutral technology to classify individual lives such that it is not focused on ethnic, racial, gender, or religious prejudices. An example is the current ethical guidelines for autonomous vehicles of the German government that states that 'any distinction based on personal characteristics (age, gender, physical or mental constitution) is strictly prohibited in the case of inevitable accident situations\

Can artificial intelligence and robotics respect human dignity and humanity as an end in itself?

In accepting the reasonable potential and free will of individuals to be bound by moral law, as well as by principles of accountability and responsibility for wrongdoing, human dignity is granted. We agree that everyone needs to be held accountable and liable when crime is committed. How will artificial intelligence convey and satisfy this human integrity element (i.e. accountability for misconduct means treating moral agents as equal members of the moral community) with individual-to-person accountability? Could we ever recognize the fair membership of artificial intelligence? There is also the question of whether, in the Kantian context, artificial intelligence and robotics would be able to treat humanity as an end. The use of lethal autonomous weapons is arguably used for a relative reason in the military sphere (ie the desire to eliminate a human target in the hope of preventing harm to others). Relative ends, for Kant, are lesser values that can be replaced by an equal. It is not adequately morally grounded to circumvent human dignity to destroy a human being in the expectation that it will avoid more damage and can be irresponsible if alternatives and consequences are not considered. Utilitarians can counter that balancing interests requires taking into account the common good that is intended to avoid harm to others in this case.

The theory of proportionality under Article 51 of the API illustrates consequentiality reasoning and utilitarian calculus, which involves an evaluation of whether an attack is likely to cause an unnecessary incidental loss of civilian life in relation to the concrete and direct military benefit predicted. But the issue of applying a quantitative evaluation of life for a prospective common good, which considers human beings sacrificed as mere objects and establishes a hierarchy of human integrity, cannot be overcome by utilitarianism. They will extinguish a priceless and irreplaceable goal end held by all rational beings; human integrity, unless autonomous weapons can only be

used to track and classify rather than destroy a human target.

In the first place, the use of automated arms to extinguish life destroys the rationale for possessing morals; rational beings' human dignity with autonomy of will. A relative end over an objective end is given priority in doing so. The absence of face-to-face killing establishes a hierarchy of human dignity. In order to destroy another human being, military officers, remote pilots, commanders, programmers, and engineers are immune from fair and ethical decision-making and do not witness the consequences. The combatant's human integrity is not only maintained but lifted above the human target by replacing the human combatant with a machine. This can also be seen as a relative end in that, as an objective end, it selfishly protects your own warriors from harm at all costs, even breaching the universal concept of morality.

Conclusion

Kantian ethics provides a human-centric approach to formulating moral principles. Kantian ethics's core elements contribute to a concentration on self-determining human capacity for rule-making and obedience to rules. These elements demonstrate the fundamental ways in which human characteristics and skills, such as practical reasoning, decision exercise, self-reflection and deliberation, enable the formulation of moral rules capable of universalisation to be established. In artificial intelligence and robotics, certain human qualities and capacities are non-existent, so that human entities must be at the forefront of developing and taking responsibility for their ultimate actions and action. A limited sense of critical thinking ability can be programmed in the machine, but as established under the Kantian notion of rational beings, it will not have the self-reflective and deliberative human capacity, so that the machine will not be able to analyze a given circumstance and exercise judgment in choosing a specific action or not. This restricted critical thinking ability can be sufficient in closed situations where technology is used for specified tasks, as seen in the civilian sphere, where it would not be necessary to exercise discretion.

It depends on whether there is a completely autonomous rule-generating or human-machine rule-generating approach to whether rules can be generated to satisfy the Kantian categorical

imperative norm. Both pose ethical questions as to who actually decides on the rules governing ethical behaviour and whether in case of malfunction or adverse damage this is adequately controllable and alterable. More complex scenarios involving open-ended machine learning tasks or dynamic learning systems used to build rules pose complexity and unpredictability issues. Such a system would not be of fundamental benefit to humanity because it is unable to generate laws that are necessarily desirable, feasible, useful and capable of universalisation. There is also a small context in which technology can actually be considered to have its own 'will'; certainly not in the Kantian sense of the will's sovereignty, but maybe a 'computer will' that has the ability to set and adhere to laws. This limits the machine-to-machine interaction rule-making ability to the exclusion of human ethical considerations.

References:

1. https://en.wikipedia.org/wiki/Kantian_ethics
2. Johnson R. (2004) Kant's Moral Philosophy;
3. Gregor M., (1996) Practical Philosophy, Cambridge University Press, Online ISBN:9780511813306,
4. Wood A., Giovanni G., (1996) Religion and Rational Theology, Cambridge University Press, Online ISBN:9780511814433
5. Louden, Robert, 2000, *Kant's Impure Ethics*, New York: Oxford University Press.
6. Heath P & Schneewind J. B., (2001) Lectures on Ethics .
7. Guyer p. & Wood A, (1998) Critique of Pure Reason, Cambridge University Press, Online ISBN:9780511804649, DOI:<https://doi.org/10.1017/CBO9780511804649>
8. Ulgen O., (2017) Kantian Ethics in the Age of Artificial Intelligence and Robotics;
9. <https://getrevising.co.uk/grids/kantian-ethics#:~:text=Another%20reason%20Kant's%20theory%20is,cold%20when%20you%20don't.>
10. Kantian Ethics in the Age of Artificial Intelligence and Robotics by OZLEMULGEN on Oct 31, 2017
11. Engelhardt, Hugo Tristram (2011). Bioethics Critically Reconsidered: Having Second Thoughts. Springer. ISBN 9789400722446.
12. Freeman, Samuel (2019). "Original Position". Stanford Encyclopedia of Philosophy. Retrieved 7 March 2020.
13. Guyer, Paul (2011). "Chapter 8: Kantian Perfectionism". In Jost, Lawrence; Wuerth, Julian (eds.). Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics. Cambridge University Press. ISBN 9781139494359.
14. Hacoheh, Malachi Haim (2002). Karl Popper - The Formative Years, 1902-1945: Politics and Philosophy in Interwar Vienna. Cambridge University Press. ISBN 9780521890557.
15. Hare, John (1997). The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance. Oxford University Press. ISBN 9780198269571.
16. Hare, John (2011). "Kant, The Passions, And The Structure Of Moral Motivation". Faith and Philosophy: Journal of the Society of Christian Philosophers. 28 (1): 54–70.
17. Harris, Dean (2011). Ethics in Health Services and Policy: A Global Approach. John Wiley & Sons. ISBN 9780470531068.
18. Hill, Thomas (2009). The Blackwell Guide to Kant's Ethics. John Wiley & Sons. ISBN 9781405125819.
19. Hirst, E. W. (1934). "The Categorical Imperative and the Golden Rule". Philosophy. 9 (35): 328–335. doi:10.1017/S0031819100029442. ISSN 0031-8191. JSTOR 3746418.
20. Janaway, Christopher (2002). Schopenhauer: A Very Short Introduction. Oxford University Press. ISBN 0-19-280259-3.
21. Janaway, Christopher; Robertson, Simon (2013). Nietzsche, Naturalism, and Normativity. Oxford University Press. ISBN 9780199583676.
22. Johnson, Robert (2008). "Kant's Moral Philosophy". Stanford Encyclopedia of Philosophy. Retrieved 11 September 2013.
23. Johnson, Robert N. (2009). "1: Good Will and the Moral Worth of Acting from

- Duty". In Hill Jr, Thomas E. (ed.). *The Blackwell Guide to Kant's Ethics*. Wiley-Blackwell. ISBN 9781405125826.
24. Kain, Philip J. (1986). "The Young Marx and Kantian Ethics". *Studies in Soviet Thought*. 31 (4): 277–301. doi:10.1007/BF01368079. ISSN 0039-3797. JSTOR 20100119.
25. Kant, Immanuel (1785). *Groundwork of the Metaphysics of Morals* – via Wikisource.
26. Kant, Immanuel (1788). *Critique of Practical Reason* – via Wikisource.
27. Knight, Kevin (2009). "Catholic Encyclopedia: Categorical Imperative". *Catholic Encyclopedia*. Retrieved 21 June 2012.
28. Korsgaard, Christine (1996). *Creating the Kingdom of Ends*. Cambridge University Press. ISBN 978-0-52149-962-0.