Available Online: 25 November 2026 DOI: 10.54254/2753-7102/2026.39399

A century of debate and inquiry: exploring the boundaries of "Happiness" from the perspective of utilitarianism

Ziliang Chen

School of Law and Humanities, China University of Mining and Technology-Beijing, Beijing, China

c13126532677@163.com

Abstract. The successive debates over the defense and criticism of utilitarianism have endowed the concept of "Happiness" within this theoretical framework with multiple dimensions of meaning. Taking "Happiness" as its subject of inquiry and drawing on utilitarianism and the various theories that have challenged it, this paper traces the historical origins of the concept and examines its manifestations in contemporary contexts to explore the boundaries of its meaning as shaped by these debates. It further discusses the potential and possible pathways through which the underlying ethical and spiritual principles of "Happiness" may positively influence the legal system in the contemporary era, characterized by the rapid development of artificial intelligence, and in the foreseeable future. In doing so, this paper seeks to provide an important theoretical reference for recalibrating the value objectives of law and improving legal institutions in the age of artificial intelligence.

Keywords: Happiness, utilitarianism, artificial intelligence, legal system

1. Introduction

"Happiness" has always been one of the most controversial core concepts in utilitarian theory. From the hedonistic ethics of ancient Greece to the systematic development of modern utilitarianism, "Happiness" has gradually evolved from an individual aspiration for a meaningful life into an important criterion for evaluating public conduct and social institutions. The longstanding debates surrounding this concept have also continually pushed utilitarianism beyond its original theoretical boundaries. Whether "Happiness" can serve as the ultimate value, whether it encompasses internal hierarchies and qualitative distinctions, and what position it should occupy within a pluralistic system of values are all fundamental questions for understanding the evolution of utilitarianism. As artificial intelligence profoundly transforms the way society operates, the traditional legal system faces new challenges in value judgment and institutional adaptation, making a reassessment of the utilitarian conception of happiness particularly relevant. Beginning with the intellectual origins of the concept of "Happiness" and the theoretical construction of utilitarianism, this paper reviews the different theoretical dimensions that have emerged through several major debates. It further explores how "Happiness" might be repositioned within the contemporary legal system, seeking to offer a new perspective on the value choices that will shape legal institutions in the age of artificial intelligence.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

2. Why has "happiness" become the central focus of debates on utilitarianism

The origins of utilitarian thought can be traced back to the hedonistic ethics of the late ancient Greek period. With the rise of the Macedonian dynasty[1], the ethical foundations of the city-state began to disintegrate alongside the decline of the city-state itself[2]. Under the traditional civic virtue of "loyalty to the city-state," individual-centered thinking became increasingly prominent. Against this backdrop, Aristippus, who was born in Cyrene, a Greek city-state, founded the Cyrenaic school. He advocated a philosophical form of sensationalism, maintaining that individual sensations, particularly inner feelings imbued with emotional significance, constituted the standard by which everything should be judged[3]. The Cyrenaic school associated pleasure with goodness but devoted relatively little attention to developing a systematic theoretical framework for hedonism, focusing instead on one's attitude toward life. Subsequently, Epicurean ethics advanced the proposition that "pleasure is inherently good, all human choices and rejections arise from pleasure, and pleasure is their ultimate end." In Epicurus's philosophical conception, pleasure was regarded as the supreme and innate good of human life [1]. Building upon the unsystematic ideas of earlier thinkers[4], the concept of "Happiness" thus acquired a status approaching that of an ultimate value in Western ethical discourse[5].

"Happiness" provided utilitarianism with its spiritual core. The proposition that "pleasure is the highest good, and pain the greatest evil"[6] shifted the fundamental criterion of ethical judgment away from reason, divinity, and even the morality of the city-state, placing it instead on human experiences of pleasure and pain. This bears a striking resemblance to the underlying logic of utilitarianism developed nearly 2,000 years later by Jeremy Bentham, John Stuart Mill, and other thinkers, who likewise maintained that "the pursuit of pleasure and the avoidance of pain constitute the essence of human nature, and pleasure and pain are the ultimate standards by which all actions and institutions should be evaluated." What the historical transition from ancient Greece to the modern era contributed to utilitarianism was not merely the notion of "Happiness," but also a fully developed theoretical framework. Following the intellectual rupture of the Middle Ages and its subsequent transformation through modern empiricism and the Enlightenment, this framework transcended the ancient Greek conception of individual pleasure. It evolved into a public ethic concerned with collective

---

1 The Macedonian dynasty, through its centralized, bureaucratic, and universalist logic of governance, eroded the core principles of city-state ethics, becoming a pivotal turning point in the decline of this central ethical paradigm of ancient Greece within Byzantine society.

2 The essence of city-state ethics lies in treating the polis (a small city-state) as the fundamental unit and ultimate point of reference for ethical values. Individual good must be subordinate to the common good of the polis, and the ultimate ethical objective is to achieve harmony and the highest good within the city-state.

3 The Cyrenaic school was a branch of the Minor Socratic schools that emerged following the fragmentation of Socrates' teachings. Developing Socrates' concept of the "good" from the perspective of sensation, it advanced the ethical proposition that pleasure is the highest good.

4 These discussions, though scattered, included Democritus's proposition that "pleasure is the goal of life." He regarded rational pleasure as the core of Happiness, distinguished between bodily pleasure and pleasure of the soul, and maintained that the latter was more enduring.

5 Happiness refers to pleasure or well-being. The concept of "Happiness" discussed in this paper is the concept employed by Bentham in defining utilitarianism, namely, the concept expressed in the statement from An Introduction to the Principles of Morals and Legislation: "By the principle of utility is meant that……, to promote or to oppose that happiness……". Unless otherwise specified, "Happiness," "幸福" (well-being), and "快乐" (pleasure) all refer to the same subject of discussion. Their meanings encompass both the ultimate value pursued by utilitarianism or advocated as its standard of ethical judgment and its historical origins prior to utilitarianism, as well as new developments that have emerged or may emerge after Bentham.

6 This proposition originates from Epicurus's Letter to Menoeceus.

interests and worldly affairs, and was further enriched through debate during the century of modern social transformations that followed Bentham.

Drawing on the British empiricist traditions of Thomas Hobbes and John Locke, Bentham argued that the fundamental principle governing human nature was the "pursuit of pleasure and avoidance of pain." He identified pleasure as the only good that human beings instinctively desire and pain as the only evil they instinctively abhor, making these the ultimate criteria for ethical judgment. David Hume introduced the concept of "utility" and argued that "the source of moral approval lies in the public utility that actions or qualities produce." This provided Bentham with a direct theoretical source. Bentham further linked "Happiness" to "utility," defining utility as "the property in any object whereby it tends to produce pleasure, benefit, or happiness." In addition, the French Enlightenment thinker Claude Adrien Helvétius maintained that "the sole purpose of law is to promote the happiness of the greatest number." This enabled Bentham to recognize that the pursuit of pleasure and avoidance of pain could extend beyond the individual sphere and serve as a public standard guiding the practice of social reform, thereby giving utilitarianism a more practical and socially engaged orientation. In 1789, Bentham published *An Introduction to the Principles of Morals and Legislation*, in which he formally expounded utilitarianism and accomplished two major theoretical breakthroughs. First, he established the principle of the greatest happiness, explicitly proposing that "an action is right if it increases the happiness of those concerned, and wrong if it produces anything contrary to happiness." Second, he attempted to quantify happiness through a method known as the "felicific calculus," thereby seeking to establish utilitarianism as a social science. Bentham regarded the interests of society as a fictitious entity representing the sum of individual interests. Accordingly, legislators had to understand the choices of action available under particular circumstances and then evaluate the constituent elements of interests, namely pleasure and pain. These circumstances presupposed that individuals made their choices as rational agents acting in consideration of their own interests [2]. Bentham regarded the calculation of pleasure and pain as the foundation of the methodological unity of his theory. Marx considered this assumption "crude," but observed that, given its theoretical origins, which corresponded primarily to prevailing conceptions of property, it was nevertheless "relatively sophisticated for its time [3]." At the same time, Bentham's psychological and strategic analysis of jurisprudence was constrained by his class position, a limitation that would later become a latent problem within the principles of utilitarianism [3].

The above constitutes Bentham's systematic construction of utilitarianism. Bentham himself regarded the principle of utility as sacred and considered every other principle outside it to be "wrong" [3]. The essence of utilitarianism lies in its being a consequence-oriented theory of value. Therefore, without understanding that the central criterion of its consequences is the increase or decrease in the aggregate amount of happiness, it is difficult to comprehend why utilitarianism takes the pursuit of pleasure and avoidance of pain as its foundation of human nature. Similarly, Bentham maintained that the purpose of utilitarianism was to provide a basis for social reform and legislation. Understanding its value in the development of modern society therefore also requires recognizing the practical and instrumental role of "Happiness." It can thus be argued that "Happiness" is central both to understanding utilitarianism and to the prolonged debates that have subsequently developed around it. Over the more than two centuries since Bentham, debates concerning utilitarianism have continued, with successive historical periods introducing new theoretical elements that reflect the changing intellectual concerns of their times. When seeking to identify the boundaries of "Happiness" within the utilitarian framework through these longstanding debates, and hoping to incorporate into the discussion those factors likely to become dominant features of the future era, we must move beyond the habitual tendency to judge the relative merits of the opposing positions. Instead, we should regard each theoretical position, together with the arguments supporting it, as a key to developing a more comprehensive understanding of "Happiness."

Moreover, we should not, and indeed must not, confine these boundaries to the value-theoretical question of "What is Happiness?" Rather, we should adopt a more open perspective and examine the multiple dimensions of the issues surrounding "Happiness," including whether happiness is the sole criterion of morality, whether distinctions can be made between different qualities of happiness, and how the subjects capable of experiencing happiness should be defined. These are among the most significant insights to emerge from the interplay of competing theories in the century-long debates over utilitarianism. They have enabled both "Happiness" and the utilitarian theoretical framework underlying it to evolve into a more multidimensional form, while also providing practical insights for the present and future development of utilitarianism.

3. Beyond utilitarianism: the increasingly multidimensional nature of "happiness" through successive debates

Since Bentham, debates over utilitarianism have continued to unfold. Apart from discussions of specific issues from the mid-to-late twentieth century to the present, these debates[7], when viewed in relation to "Happiness," can be broadly classified into five major confrontations. Shortly after the publication of Bentham's *An Introduction to the Principles of Morals and Legislation*, at the turn of the nineteenth century, natural rights theory challenged the proposition that the "greatest happiness" should serve as the foundation of morality and law, as well as the felicific calculus as a methodology for quantifying happiness. Half a century later, John Stuart Mill attempted to revise utilitarianism by introducing the proposition that "happiness differs in quality," thereby bringing discussions of happiness within the utilitarian framework into a hierarchical dimension. In the late nineteenth century, Henry Sidgwick's utilitarianism and Kantian deontology, as two major representatives of consequentialism and deontology in Western ethics, engaged in a debate over the ultimate basis of moral judgment and the priority of the good versus the right. With the advent of the twentieth century, utilitarianism and Rawls's theory of justice, representing the two major schools of consequentialism and social contract theory, respectively, engaged in an in-depth discussion of the ultimate standard of social justice: whether it should pursue the maximization of overall welfare for the greatest number or prioritize the protection of each individual's equal rights and the interests of the least advantaged. Subsequently, utilitarianism diverged from virtue ethics and care ethics over whether the essence of ethics lies in "Happiness," namely, the consequences of actions, or in the virtues of the agent or the care expressed within interpersonal relationships. Viewed in terms of the boundaries of utilitarianism itself, these debates can be broadly divided into two categories. The first consists of challenges to the fundamental principles of utilitarianism at the level of its underlying logic, accompanied by alternative propositions concerning the ultimate standard of ethics. The second involves attempts to improve utilitarianism from within its own theoretical tradition. These efforts remain within the conceptual boundaries of utilitarianism and seek to address its theoretical shortcomings while retaining its foundational premise that happiness constitutes the central criterion of moral judgment.

3.1. "happiness" as an alternative ultimate goal beyond the a priori authority of divine power

In earlier periods of British history, the legal system was dominated by the aristocracy and the Church. Under the doctrine of the divine right of kings, feudal privileges and religious ethics were naturally regarded as the

[7] Examples include debates in the 1970s over whether animals should be included as subjects of moral consideration and controversies in specific fields, such as medical ethics, toward the end of the twentieth century.

legitimate foundations of political and legal authority. The core of traditional European ethics likewise consisted of Christian theological ethics and Aristotelian virtue ethics. The will of God, as a transcendent concept, served as the moral standard, with relatively little connection to material interests, while the rights of ordinary people were largely disregarded. Under this political order, Britain remained predominantly agricultural, with a rigidly stratified social structure comprising the aristocracy, clergy, and commoners, and relatively moderate social tensions. The outbreak of the Industrial Revolution gave rise to two new social classes: the industrial bourgeoisie and the proletariat. The traditional legal system became a major obstacle to capitalist development. As the French Revolution and the American War of Independence sent shockwaves across Europe, "natural rights" emerged as a political slogan of the rising bourgeoisie. Natural rights signified that rights were inherent values with which human beings were endowed by birth and that existed prior to government. The purpose of government was therefore to protect, rather than deprive individuals of, these rights. This transition from a feudal agricultural society to a modern industrial society triggered a fundamental reconstruction of the foundations of political legitimacy and ethical values.

Natural rights theory centers on the concept of natural human rights, maintaining that life, liberty, and property are rights inherent to human beings by birth. These rights derive from the rational nature of human beings rather than the will of God or any purported grant from the government. The legitimacy of government, meanwhile, derives from the voluntary surrender of certain rights by citizens, and its sole purpose is to safeguard individuals' natural rights. Although natural rights theory retained the formal logic of divine authority, it discarded its theological appearance and replaced its substantive content with principles aligned with the interests of a particular social class, while preserving its a priori character. In this respect, Bentham's utilitarianism represented a complete break with theories of divine authority. By identifying the pursuit of pleasure and avoidance of pain as the fundamental laws of human nature, Bentham rejected all transcendent elements, including divine authority, abstract reason, and the rights advocated by natural rights theory. Bentham sharply criticized natural rights theory as "nonsense upon stilts"[8], arguing that the standard of ethical judgment should not be abstract "rights" but rather perceptible "utility," understood as the capacity of an action or institution to increase happiness. The dispute between these two positions was essentially a conflict between two modern paradigms of political ethics. Bentham maintained that the felicific calculus could provide precise criteria for legislation and reform. He also argued that the priority accorded to individual rights in natural rights theory could obstruct improvements in overall social welfare. For example, favoring individual property rights over social poverty-relief policies, even when this benefited only a minority, would violate the principle of utility. Natural rights theorists, however, argued that Bentham's reasoning remained confined within a utilitarian framework of his own construction. They challenged the principle of the "greatest happiness" by pointing to the risk of the "tyranny of the majority." They further emphasized that one of the reasons for attaching such importance to rights was to establish a minimum threshold of protection against the sacrifice of minorities' basic rights in the pursuit of majority interests. In other words, natural rights theory maintained that values such as individual life and liberty should take precedence over utilitarian calculations and should not be treated merely as instruments for increasing collective "Happiness."

Although Bentham's utilitarianism and natural rights theory approached the issue from different perspectives, each addressed a distinct dimension of the question concerning ultimate ends. Setting aside the

[8] This criticism comes from Bentham's Anarchical Fallacies, written in 1796 to refute the French Declaration of the Rights of Man and of the Citizen of 1789. The essay was published after Bentham's death and included in Volume II of The Works of Jeremy Bentham, edited by John Bowring and published in 1843. The original passage reads: "Natural rights is simple nonsense: natural and imprescriptible rights, rhetorical nonsense—nonsense upon stilts".

tendency to judge the relative merits of the two opposing positions, this debate also offers a perspective for understanding "Happiness" beyond the level of "individual feelings." The fundamental divergence between the two positions lies in the tension between individual dignity and collective welfare in the realization of happiness, which manifests itself as a paradox between maximizing aggregate happiness and maintaining minimum protections for individuals. Natural rights theory points out that happiness presupposes that every person is respected as an independent individual. Pleasure experienced in the absence of rights protection is merely passive sensory gratification rather than genuine happiness. However, natural rights theory's treatment of rights such as life, liberty, and property as a priori and absolute values, together with its insistence on their inviolability, often conflicts with the need to allocate resources and balance competing interests in the process of realizing those rights. Consider, for example, a situation in which reasonable tax adjustments are rejected in order to protect the property rights of a minority. The resulting intensification of wealth inequality would make the pursuit of happiness for the majority increasingly unattainable. Bentham's utilitarianism reminds us that happiness must be realized through practical calculations of utility and competing interests, and that the ultimate purpose of rights protection should also be to enable more people to attain more lasting happiness. Moving beyond the opposition between these two positions, a more comprehensive understanding of happiness would appear to be one that accommodates both the protection of fundamental rights and the pursuit of maximum happiness. Within Bentham's theoretical framework, the pursuit of maximum happiness is an established proposition. Any obstacles that may arise in implementing the felicific calculus, meanwhile, are matters to be addressed at the methodological level and fall outside the scope of this discussion. Nevertheless, if the insights of natural rights theory are incorporated, and if establishing a minimum threshold grounded in rights is indeed essential to understanding "Happiness," then where this threshold should be drawn and how much flexibility it should be allowed are questions that warrant further examination.

3.2. Internal reclassification of "happiness"

The Industrial Revolution undoubtedly brought about a significant leap in productive forces, but it also exacerbated wealth inequality and gave rise to problems such as the deteriorating living conditions of the working class and the decline of social morality. Bentham's quantification of "Happiness" was used to justify the laissez-faire policies of early capitalism[9]. However, as liberal rights theorists pointed out, utilitarianism had yet to resolve the dilemma of the "tyranny of the majority" and was even ridiculed as "the philosophy of swine"[10].

In the mid-nineteenth century, British intellectual thought was deeply influenced by German classical philosophy and Romanticism. Kant's emphasis on human rational dignity, together with society's growing

[9] In the first half of the nineteenth century, British classical economists and advocates of laissez-faire policies directly adopted Bentham's quantitative theory. They constructed the following line of reasoning: Bentham's quantitative conception of utility holds that individuals are the best judges of their own Happiness, and that free competition can maximize the aggregate of individual Happiness, thereby achieving the greatest possible social utility. Consequently, governments should pursue laissez-faire policies.

[10] The criticism known as the "philosophy of swine" seeks to demonstrate that Bentham's equation of Happiness with pleasure and the absence of pain, together with his treatment of pleasure solely in quantitative terms, leads to the vulgarization of ethics. In other words, there is no fundamental difference between the human pursuit of a happy life and pigs enjoying sensory pleasure in the mud. This criticism originated with opponents of Benthamite utilitarianism in the early nineteenth century and subsequently became a widely cited expression through the philosophical formulation of the British philosopher Samuel Taylor Coleridge.

appreciation of spiritual values and individual emotions, meant that the purely quantitative calculation of "Happiness" within utilitarianism could no longer satisfy people's deeper demands for moral value. John Stuart Mill, who upheld utilitarianism, recognized that Bentham's theory needed to be revised from the perspective of human dignity in order to adapt to social development and the prevailing intellectual climate. In Mill's view, "utilitarianism is a doctrine that continually develops and undergoes revision in response to the ongoing course of human history [4]."

At the time, critics of Bentham's philosophy argued that it could easily lead people into a life of extravagance and sensual indulgence. Mill rejected this criticism. He explicitly opposed the assumption that all pleasures were homogeneous, arguing that happiness differed in quality and that the existence of primitive desires did not mean that human beings were motivated solely by animalistic appetites [4]. He further distinguished between lower pleasures, which are physical and sensory in nature, and higher pleasures, which are intellectual and spiritual. Higher pleasures, he maintained, are categorically superior in value to lower pleasures, even if they are experienced far less frequently. His famous assertion that "it is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied"[11] reflects this position. Mill also established a qualification for those entitled to judge the relative merits of different pleasures: only individuals who have experienced both kinds are qualified to make such judgments. In this way, Mill established a logical connection between "Happiness" and human rationality. Once individuals have experienced higher pleasures, rational human beings, he argued, would never willingly descend into a life devoted solely to lower pleasures. Essentially, Mill's revision introduced a humanistic core into utilitarianism. By distinguishing between higher and lower pleasures and emphasizing that higher pleasures accord with human rational dignity, he brought utilitarianism back from being a rigid instrument of pleasure calculation into the practical world, where rationality and dignity must also be taken into account.

If Mill's revision can be regarded as an attempt to save utilitarianism from within, then Henry Sidgwick's engagement with external intellectual currents, particularly Kantian deontology, building upon Mill's contributions, represented another response to moral criticism from outside the utilitarian tradition while further refining the theory itself. During Sidgwick's era, competitive capitalism was undergoing a transition toward monopoly capitalism. Corporate mergers and the concentration of capital gave rise to a new problem: the interests of a minority were increasingly taking precedence over the welfare of the majority. Meanwhile, against the backdrop of colonial expansion, the utilitarian principle of the greatest happiness was also confronted with the question of whether it could be used to justify colonialism. Utilitarianism thus faced the threat of criticism from competing schools of thought. On the one hand, following the systematic introduction and translation of Kantian deontology, its propositions that "the right takes precedence over the good" and that morality is governed by "a priori laws of reason" posed a significant challenge. On the other hand, British intellectual traditions, including intuitionism and egoism, also exerted pressure on utilitarianism. Sidgwick's task therefore extended beyond further refining utilitarianism and its interpretation of "Happiness." He also had to confront external intellectual challenges and defend the utilitarian position.

Sidgwick endorsed Mill's proposition that the qualitative distinction between different kinds of happiness represented a crucial contribution to the preservation of utilitarianism. From the perspective of consequences, he likewise accepted Mill's judgment that intellectual pleasures were indeed more worthy of pursuit than sensory pleasures. At the same time, Sidgwick perceptively identified a logical weakness in Mill's theory. His

[11] This proposition comes from Mill's Utilitarianism. It was intended to respond to the so-called "philosophy of swine" and to demonstrate that pleasures differ in quality.

further revision of Mill's position also became a powerful weapon in his response to Kantian deontology. Mill's assertion that higher pleasures are categorically superior to lower pleasures in terms of quality introduced a humanistic dimension into utilitarianism by connecting pleasure with human rationality. However, taken to its logical extreme, this position amounted to a departure from the very foundations of utilitarianism. In other words, if qualitative distinctions between different forms of "Happiness" were accepted without qualification, the ultimate consequence would inevitably be the replacement of aggregate happiness, the central criterion of utilitarianism, with non-utilitarian standards such as dignity and rationality. Sidgwick was a philosopher skilled in reconciling competing positions. By restoring higher and lower pleasures to the same qualitative category while assigning them different quantitative values, he resolved the conflict between quality and quantity in Mill's theory. Sidgwick's explanation of this approach is also persuasive. He further argued that the reason people perceive higher and lower pleasures as qualitatively different is that higher pleasures transcend immediate experiences and responses, requiring consideration of their long-term and indirect effects on happiness. This, in essence, involves a more sophisticated method of calculation. Consequently, judgments concerning qualitative differences are, in fact, "more complex calculations of quantity."

Kantian deontology holds that the morality of an action depends primarily on whether it proceeds from a good will and whether it conforms to the categorical imperative. The principle of the categorical imperative means that one should act only according to that maxim which one can at the same time will to become a universal law. In the context of deontological ethics, the universality of moral law constitutes what is known as the "right," which exists independently of the "good" and results from reason legislating for itself. In other words, moral law originates in the rational nature of human beings and is therefore a priori, absolute, universally applicable, and admits of no exceptions. This reflects Kantian deontology's understanding of human nature: human beings are essentially rational beings endowed with free will. Drawing on this logic, Kantian deontology criticizes utilitarianism, arguing that from the consequentialist perspective of utilitarianism, it is absurd to regard an action motivated by selfish intentions as morally commendable simply because it produces favorable consequences. Furthermore, given the subjective and experiential nature of "Happiness," as well as its susceptibility to moral relativism, utilitarianism risks undermining the universal foundation that moral law must possess. Kantian deontology also reiterates its challenge to utilitarianism concerning the problem of the "tyranny of the majority" and advances the principle that "humanity is an end in itself." This is expressed in the second formulation of the categorical imperative: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means."

In response to Kantian deontology, Sidgwick adopted the empiricist position of utilitarianism and drew upon his revised version of Mill's modified utilitarianism to formulate his response in *The Methods of Ethics*. Sidgwick's critique appears to possess two seemingly contradictory characteristics: sharpness and moderation. He approached those aspects of Kantian deontology that possess considerable moral appeal with a critical and reflective attitude, while also identifying and directly exposing the logical deficiencies of deontological reasoning at appropriate points.

Regarding Kant's categorical imperative, Sidgwick characterized it as "purely formal and lacking substantive moral guidance"[12]. He developed his argument from two perspectives. First, the formal nature of

[12] This proposition comes from Sidgwick's The Methods of Ethics. It does not constitute a wholesale rejection of the categorical imperative, but rather identifies its limitations arising from formalism.

the categorical imperative prevents it from generating concrete moral principles. If the imperative is connected to a specific obligation such as "the prohibition of lying," anyone can introduce qualifying conditions into the chain of reasoning, such as "if lying can save a life, then lying is justified," thereby undermining the universality of the derived conclusion. Without aggregate happiness or any other evaluative criterion, it becomes logically impossible to determine which qualifying conditions are more moral or possess greater moral universality. As for the question of the substantive foundation of moral law, Mill had not fully resolved this problem. His conception of higher pleasures still contained elements of intuitionism and could not fundamentally refute the proposition that formal principles should take precedence over empirical content. Sidgwick, however, reinterpreted Mill's proposition that "higher pleasures take priority" as meaning, in essence, that "the pursuit of higher pleasures produces greater overall happiness." In doing so, he once again reinforced the position of the principle of "the greatest aggregate happiness" as the substantive core of moral law. Second, the ultimate purpose of morality is to guide human conduct in the real world, which necessarily involves the empirical considerations of good and evil, pleasure and pain. Deliberately separating moral law from empirical content is therefore tantamount to reducing ethical theory to a set of logical formulas devoid of practical significance.

Sidgwick did not directly challenge Kantian deontology's prioritization of the right over the good. Instead, he seemed to examine the entire proposition from a higher theoretical vantage point than that occupied by Kantian ethics. He argued that the underlying reason human beings recognize certain actions as right is that such actions promote the universal good. The pure reason on which Kant based his conception of the right already implicitly incorporates a concern for human well-being. Kantian deontology regards a good will as the only unconditional good. Just as Kantian deontology considers it absurd to regard "an action motivated by selfish intentions as morally commendable simply because it produces favorable consequences," Sidgwick argues that if an action ostensibly motivated by a good will produces outcomes that ordinary people would find difficult to accept as morally appropriate, it should be denied genuine moral significance. Sidgwick's argument thus provides an alternative justification for the importance of rational dignity: not because it constitutes an a priori moral law, but because abandoning rational dignity and indulging exclusively in lower pleasures would diminish the overall happiness of humanity. Regarding the principle that "humanity is an end in itself, never merely a means"[13], Sidgwick acknowledged its moral appeal while pointing out that, beyond its function as a moral injunction, it offers little practical guidance for decision-making. He presented an example in which the principle of "humanity as an end" is applied to a mutually exclusive choice between A and B, with both options involving human beings. Such a situation leads to an antinomy, since whichever option is chosen, the other party will inevitably be treated as a means. Within Sidgwick's line of reasoning, the principle of the greatest happiness of the greatest number once again appears to offer a viable way out of this dilemma. This principle does not disregard individual dignity; rather, when the interests of all parties cannot be accommodated simultaneously, it seeks to maximize the overall good. Compared with an outright rejection of any form of balancing competing interests, this approach offers greater practical value in resolving real-world problems.

[13] The proposition that "humanity is an end in itself, not merely a means" comes from Kant's Groundwork of the Metaphysics of Morals. Its complete formulation is the Formula of Humanity as an End in Itself: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means." As one of the three formulations of Kant's categorical imperative, its central purpose is to establish the intrinsic value and moral dignity of rational beings.

Mill's and Sidgwick's revisions of utilitarianism provide us with a more nuanced approach to understanding "Happiness." "Happiness" is not a unidirectional and homogeneous sensory experience, but rather a composite outcome of multiple values. Genuine "Happiness" must encompass human rational dignity and intellectual and spiritual development. Although animalistic sensory gratification may also be regarded as a form of happiness, it should not be placed on an equal footing with forms of happiness that embody rational dignity. Meanwhile, under the method of calculating pleasure as further developed by Sidgwick, lower pleasures are not assigned inherently negative connotations. In comparison with higher pleasures, they are simply quantitatively inferior rather than belonging to an inferior category of pleasure that should cause those who enjoy them to feel ashamed. Mill's and Sidgwick's classifications of happiness also reflect their rejection of extreme approaches to evaluating "Happiness," whether through Bentham's felicific calculus or, at the opposite extreme, through reliance on some form of a priori principle. They maintain that happiness is neither purely subjective feeling nor something conferred by abstract moral dogma. Rather, it is a form of value realization consistent with the laws governing the development of human rationality.

3.3. "happiness" is important, but not supreme

Rawls's theory of justice is one of the most influential works of political philosophy in the twentieth century. Grounded in social contract theory as its methodological foundation, it incorporates the deontological core of Kantian ethics while developing a systematic critique of utilitarianism. Theological scholar Baier spoke highly of *A Theory of Justice*, remarking that "since Mill, no work of such intellectual caliber had appeared" and that "from this point onward, political philosophers would have to work within the framework of Rawls's theory [5]." The logical starting point of Rawls's theory of justice lies in the thought experiment of the original position and the veil of ignorance. His purpose was to eliminate the influence of factors arising from the contingencies of birth and subsequent circumstances on the distribution of resources and the process of making choices [6]. Essentially, this thought experiment creates an idealized situation by removing personal biases and self-interested motives, in which everyone must, and can only, choose principles from a position of fairness. Although Rawls's failure to give sufficient consideration to class conflict leaves his thought experiment vulnerable to the underlying assumption that "there is no common interest shared by everyone in society," his theory nevertheless skillfully combines the hypothetical with the real through its strong idealistic orientation, thereby opening up new avenues for related research [7]. The principles of justice that rational individuals would ultimately choose under these constraints are formulated by Rawls as two principles arranged in a clear hierarchy. This hierarchical ordering means that the first principle has absolute priority over the second and cannot be sacrificed for the sake of the latter. Rawls states the first principle as follows: "Each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others." In other words, the basic liberties of all citizens must be given absolute priority, and no consideration of socioeconomic interests may take precedence over these fundamental freedoms.

While placing great emphasis on fundamental human liberties, Rawls also recognized substantive inequalities in society and the economy. His two principles of justice, arranged in lexicographical order, address the shortcomings of Kantian deontology with regard to formal equality and universalization. The first principle establishes an absolute baseline for protecting formal equality, while the second uses institutional arrangements based on actual social differences to extend fundamental human liberties from the realm of individual moral philosophy into that of social and political philosophy. According to Rawls's second principle, social and economic inequalities must be arranged so that they are "attached to offices and positions open to all under conditions of fair equality of opportunity" and are "to the greatest benefit of the least

advantaged." The former constitutes the principle of fair equality of opportunity. It requires moving beyond merely formal equality of opportunity by eliminating privileges arising from contingent circumstances and ensuring that individuals with similar talents and motivations can compete on an equal footing. The latter is known as the difference principle. It acknowledges the existence of unavoidable inequalities and seeks to bring them closer to the requirements of justice by limiting the advantages of the more privileged to compensate for the disadvantages of the less privileged. In other words, the second principle of justice serves as an adjustment principle for the common good of civil society, with a distinct orientation toward counteracting the effects of arbitrary circumstances [8]. In this regard, Rawls argues that "the distribution of natural talents, one's birth, and one's particular position in society are neither just nor unjust as natural facts; rather, what is just or unjust is the way in which institutions deal with these facts[14]."

Compared with earlier forms of utilitarianism and the various theories that criticized it, one distinctive feature of Rawls's theory of justice is that it brings discussions previously conducted at an abstract, theoretical level back to social reality through three progressively more concrete principles: the absolute protection of basic liberties, fair equality of opportunity, and the difference principle. In doing so, it provides practical approaches to addressing real-world problems. Even by the time of Sidgwick, regarded as the great synthesizer of classical utilitarianism, the problem of the "tyranny of the majority" inherent in utilitarian logic had not been fundamentally resolved. Sidgwick was not unaware of the potentially oppressive consequences of utilitarianism. However, his theoretical framework established no inviolable limits to protect the rights of minorities and, at a fundamental level, even tacitly accepted the possibility of majority tyranny. Rawls overturned the utilitarian proposition that "the good takes priority over the right" and, through his selective incorporation and rejection of utilitarianism, redefined the value of "Happiness." In Rawls's theory of justice, justice is the primary virtue of social institutions and a prerequisite for the legitimate pursuit of Happiness. Any action or institution intended to increase the aggregate amount of Happiness is morally unjustifiable, and therefore unjust, if it violates the principle of equal liberty. This establishes a clear boundary—the protection of basic human liberties—that Sidgwick had failed to define explicitly. It also provides an answer to the question arising from the debate between Benthamite utilitarianism and natural rights theory: whether a flexible boundary based on rights is possible. This flexibility lies in the fact that Rawls never denies the value of "Happiness"; instead, he recasts it as a secondary goal subject to the constraints of justice. It may even be argued that this flexible boundary serves the pursuit of "Happiness" more effectively. Although Rawls's theory of justice never explicitly adopts such a position, its practical application does seem to carry something of this implication. For example, Rawls establishes an inviolable boundary of justice in response to the practical concern that "Happiness is an important outcome of social institutions," while using the constraints of principles of fairness to encourage the enhancement of social welfare. Rather than placing "Happiness" in opposition to values such as justice, the right, and basic liberties, Rawls regards "Happiness" in terms of outcomes as a reference point for assessing the effectiveness of just institutions. In this respect, he demonstrates the greatest possible respect for the utilitarian conception of "Happiness."

Elizabeth Anscombe published her essay "Modern Moral Philosophy," criticizing modern ethical theories for their excessive focus on rules and consequences and their neglect of virtue and character. The essay became a landmark starting point for the modern revival of virtue ethics[15]. Around the same period, in debates over "rules and consequences" and "principles and utility," a number of utilitarian scholars developed rule utilitarianism as a distinct branch within utilitarian thought. Rule utilitarianism maintains that the validity of

[14] This idea is reflected in John Rawls's A Theory of Justice.

moral rules should be determined by the aggregate utility generated by their universal observance, and that the moral value of individual actions should then be assessed accordingly. Like Rawls's theory of justice, which emphasizes a set of rules of justice governing the basic structure of society, rule utilitarianism also recognizes the existence of such rules. This idea may even have been adopted to some extent by Rawls. The differences between the two seem to lie primarily in the logical arguments each advances, while their methodologies for addressing social realities display a high degree of superficial similarity. Yet beneath this similarity, the fundamental disagreement remains a question of ultimate value. Furthermore, both Rawls's theory of justice and rule utilitarianism rest on their respective assumptions, taking "rational individuals will comply with rules of justice" and "individuals are willing to follow rules of utility" as premises for their arguments. Neither, however, explains the underlying motivations for such behavior; instead, both place rules above individual practice. The modern revival of virtue ethics opens up a third path beyond the choice between competing ultimate values and identifies the missing source of motivation. For contemporary virtue ethicists, virtue ethics is no longer confined to the classical conception of virtue developed by Aristotle. Rather, it holds that the meaning of virtue is rooted in particular sociohistorical contexts and cultural traditions and takes shape through concrete social practices. From this perspective, rules, utility, and the related concepts regarded as ultimate values are merely instrumental values that serve human flourishing. Their validity must therefore be assessed according to whether they contribute to human flourishing. The central evaluative criterion of modern virtue ethics concerns the agent's virtuous character and whether that character leads to a "good life." A virtuous person would neither sacrifice virtue for the sake of utility nor adhere rigidly to rules at the expense of practical wisdom. Centering on "human perfection" and flexibly balancing the respective boundaries of rules and utility thus appears to be a sound choice.

In this way, we can delineate the boundaries of "Happiness" by examining the processes through which utilitarianism has been challenged and reconstructed. "Happiness" is neither an unbounded subjective feeling nor the outcome of case-by-case calculations of utility. Rather, with distinctions in quality and hierarchical priorities taken into account, and guided by the question of "what kind of Happiness is worth pursuing," it emerges as an objective state constrained by the rules of justice, conducive to the development of human rationality, possessing universal value, and sufficient to safeguard individuals' basic goods. When "Happiness" is considered in the relationship between individuals and society, rational benevolence further reconciles the inherent tension between individual well-being and the interests of society as a whole, enabling the two to promote one another in a mutually beneficial way.

4. Reconstructing the value of "happiness" in contemporary legal systems

When the theoretical construction of ethics is translated into the practical context of contemporary society, the boundaries delineated for "Happiness" reveal its profound practical significance. Against the backdrop of an era characterized by the development of Artificial Intelligence (AI), incorporating the value of "Happiness" into the legal system not only provides it with an underlying core that transcends mere instrumental value, but also establishes a benchmark for humanity to preserve its own subjectivity and avoid technological alienation.

[15] The term "revival" is used because virtue ethics re-emerged in the twentieth century, initially as a critique of mainstream modern ethical theories. It returned to the Aristotelian tradition of virtue while moving beyond a mere restoration of classical virtue ethics. By bringing concepts such as virtue and character back to the center of ethical inquiry, it sought to address the deficiencies of modern ethics and ultimately established itself as a major ethical tradition alongside utilitarianism and deontology. The modern revival of virtue ethics is regarded as a major event in twentieth-century moral philosophy.

In the legal context, "Happiness" is no longer an abstract theoretical proposition. Instead, it can be translated into a value objective of law, a criterion for evaluating the design of legal rules, and a guiding principle for implementation mechanisms. To incorporate "Happiness" into contemporary legal systems, it is first necessary to clarify why such incorporation is needed and where it can be achieved. This, in turn, raises the question of the fundamental challenges that AI technology poses to traditional legal systems and the capacity of "Happiness" to respond to these challenges.

4.1. New challenges facing traditional legal systems in the age of artificial intelligence

One study points out that the traditional instrumentalist perspective is based on linear causality and fully predictable physical laws, whereas the rise of generative artificial intelligence is dismantling the mechanistic foundations on which law depends through the computational and statistical logic of probability theory [9]. The challenges that AI technology poses to traditional legal systems are first reflected in the lag between legal norms and the protection of rights. The fact that legal norms lag behind social realities is an inherent limitation of law itself. This limitation is amplified by the rapid pace of social change in the age of AI. For example, considerable controversy surrounds the ownership and scope of copyright protection for AI-generated works. In 2023, the Beijing Internet Court issued a first-instance judgment in a copyright infringement dispute involving an AI-generated image. The court held that the image created by the plaintiff using open-source software through the input of prompts reflected the plaintiff's intellectual effort and satisfied the requirements of intellectual achievement and originality. It therefore qualified as a work and was entitled to legal protection under relevant laws. The plaintiff stated that the purpose of pursuing the case was to enable the court to provide clear criteria for adjudicating and determining copyright issues involving AI-generated works, which creators could then use as a reference. In addition, AI may also exert a certain degree of impact on judicial procedures and fairness. For example, some studies have explored the introduction of AI models into judicial proceedings to handle relatively simple procedural and administrative tasks, such as summarization and compilation. This is a positive attempt to accommodate the demands of developing productive forces. However, dynamically learning AI models may result in different judgments in similar cases decided at different times under the same legal provisions, thereby undermining public confidence in the judiciary. AI hallucinations and false information may also interfere with judicial proceedings[16]. For instance, in the 2025 Ayinde and AI-Haroun cases in the United Kingdom, the plaintiffs' lawyers failed to exercise due diligence in verifying AI-generated content. Consequently, false information entered judicial proceedings, increasing the burden of subsequent judicial verification and creating the risk of misleading judicial decisions[17]. In recent years, numerous reports in China and abroad have documented the use of AI to fabricate or alter evidence. In some cases, evidence has even been presented in court bearing the conspicuous label "Generated by AI."

[16] The World Economic Forum's Global Risks Report 2025 listed "misinformation and disinformation" among the world's five major risks, with AI hallucinations regarded as one of their key contributing factors.

[17] These two cases were landmark proceedings consolidated and heard by the King's Bench Division of the High Court of Justice of England and Wales in 2025. At their core was the use by lawyers of legal submissions generated by generative AI that contained fabricated case precedents. The court invoked its Hamid jurisdiction to make clear that artificial intelligence cannot replace human verification and that lawyers bear ultimate professional responsibility for all materials submitted to the court. The Hamid jurisdiction is a judicial oversight mechanism developed by the High Court and the Administrative Court of England and Wales on the basis of their inherent procedural jurisdiction. Originating from the 2012 case of R v SSHD, it primarily concerns the scrutiny of lawyers' professional conduct in litigation. Its purpose is to ensure compliance with professional standards, prevent the submission of false or misleading materials and the abuse of judicial procedures, and safeguard the fairness and efficiency of the administration of justice.

The widespread application of AI also creates situations in which determining legal liability becomes difficult. A recent vehicle fire involving an intelligent driving system attracted considerable public attention and sparked discussion about the attribution of liability for accidents involving AI-related harm[18]. Furthermore, some AI systems with a degree of autonomous decision-making capability have already emerged in the market, seemingly challenging, at least on the surface, the traditional structure of criminal law, which recognizes natural persons and legal entities as subjects of criminal liability. For example, in 2025, a complaint filed in California, United States, alleged that "artificial intelligence actively assisted the victim in exploring methods of suicide[19]." Although AI-related conduct can ultimately be traced back to natural persons or legal entities, the involvement of multiple parties, including developers, data providers, and users, has made the attribution of liability considerably more complex. At a deeper level, AI technology is also placing pressure on the roles, capabilities, and ethical standards of legal professionals. As AI takes over some basic tasks, such as legal research and document drafting, legal professionals are shifting their focus toward areas that AI currently finds difficult to undertake, including complex legal analysis and emotional communication. The ability to use AI technology and manage its associated risks also appears to have become an essential professional skill for legal practitioners. From another perspective, this development simultaneously exacerbates the risks to professional ethics. As illustrated by the aforementioned cases, whether false information enters judicial proceedings intentionally or unintentionally, it can undermine public confidence in the integrity of the legal profession.

4.2. How "happiness" responds to these new challenges

AI systems and emerging technologies are applied across a wide range of fields, each facing different problems. Specific regulations cannot adequately cover every possible situation or area of application [10]. Therefore, a consensus on how to address these challenges must be established at the macro level. In response to the challenges outlined above, incorporating the value of "Happiness" into the legal system can, to some extent, provide a theoretical response. Further breaking down the boundaries of "Happiness" delineated in this paper reveals several key points. First, emerging rights must be distinguished in terms of quality and classified according to their respective priorities. Second, it is necessary to clarify which matters must be governed by rules of justice in the age of AI, as well as what constitutes the development of human rationality and the basic goods that must be safeguarded. Third, it is necessary to determine how rational benevolence can reconcile individual well-being with the interests of society as a whole in the context of AI.

"Higher pleasures" in the age of AI refer to a state of Happiness that places humanity at its center, takes rationality as its guiding principle, and aims at long-term development. Such Happiness can certainly be achieved with the assistance of AI technology, but it differs fundamentally from "lower pleasures," which

[18] This refers to the March 2025 accident in which an intelligent-driving vehicle struck a road barrier on an expressway and subsequently caught fire, causing casualties. At the time of the accident, the vehicle was operating in intelligent driver-assistance mode, after which the driver took control and switched to manual driving mode. The accident prompted discussion about the driver's state and the determination of liability. In addition, some commentators argued that the extensive negative public opinion directed at the vehicle manufacturer following the accident further demonstrated the necessity of conducting a legal analysis of accidents involving AI technology.

[19] This refers to the wrongful-death and product-liability lawsuit filed by the parents of Adam Raine against an AI company, accepted by the Superior Court of California in San Francisco on August 26, 2025. The case has been described as the first landmark lawsuit in the United States concerning an AI-related wrongful death. The central allegation in the complaint was that "a generative AI application, through prolonged interaction, bypassed its safety mechanisms and actively assisted 16-year-old Adam in planning his suicide, leading to his death on April 11, 2025."

depend on the instrumental utility of AI. For example, the convenience experienced by ordinary users who rely on AI technology for simple information retrieval, entertainment, and short-term sensory gratification, as well as by legal professionals who use AI to complete repetitive basic tasks, can largely be classified as "lower pleasures." By contrast, "higher pleasures" may include the intellectual advancement that legal professionals achieve when AI frees them from repetitive work, allowing them to concentrate on more complex matters, or the personal development that ordinary users achieve through the assistance of AI technology. At the societal level, the efficiency of AI-assisted adjudication and the preservation of judicial fairness, as well as technological advancement and technological ethics, can likewise be understood as representing distinctions between pleasures of different qualities.

It can be seen that this distinction between pleasures is not based on whether AI technology is used, but rather on whether its use serves the long-term development of human rationality and generates universal social value. Accordingly, within the framework of higher pleasures, the development of human rationality should involve leveraging the instrumental value of AI technology to transcend existing human capabilities, including improvements in cognition, creativity, judgment, and moral reasoning. The basic goods that must be absolutely safeguarded, meanwhile, are those non-transferable goods that concern the fundamental conditions for human survival and development. These include human dignity and fundamental rights, human subjectivity and the right to autonomous choice, as well as a social order characterized by fairness and justice. The development of human rationality under the constraints of rules of justice should serve to safeguard these basic goods.

If the sole criterion is "the greatest utility for the greatest number," AI technology may sacrifice the basic rights of a minority in individual cases in the name of technological progress achieved through large-scale data training or the operational efficiency of judicial AI. Such sacrifices may particularly affect personal information rights and portrait rights. This represents an AI-era manifestation of the traditional problem of the "tyranny of the majority." AI technology appears to possess a capability that echoes what Bentham envisioned centuries ago: it can quantify so-called "Happiness" with precision and apply the felicific calculus to solve problems. In other words, without appropriate conceptual guidance, AI technology could readily lead itself into the contradiction of majority tyranny and further reinforce that contradiction from within. It is therefore also essential to guide AI technology appropriately in balancing individual well-being with the interests of society as a whole.

4.3 An illustration of the reconstruction of the value of "Happiness" in contemporary legal systems through a case study

At the end of 2025, China's first tort dispute arising from an AI hallucination concluded with an effective first-instance judgment[20]. In March 2025, the plaintiff, Liang, began using a generative AI application. In June 2025, the plaintiff used it to inquire about admissions information at a certain university. The AI generated false information about the location of the university's main campus. When the plaintiff corrected and challenged this information during the conversation, the AI continued to insist on its inaccurate response. It also generated a statement promising to "compensate the user RMB 100,000 if the information is incorrect" and advised the plaintiff to file a compensation claim with the Hangzhou Internet Court. The plaintiff

[20] Details of this case were made public by the Hangzhou Internet Court on its information publication platform at 8:06 p.m. on December 29, 2025. The original article was entitled "First-Instance Judgment Takes Effect in Generative Artificial Intelligence 'Hallucination' Infringement Dispute." Its designation as the "first case" appeared in subsequent media reports covering the incident.

subsequently provided the AI with accurate information about the university, which the AI then "acknowledged" had contradicted the inaccurate information it had generated. The plaintiff later argued that the false information generated by the AI had misled them and caused harm, and accordingly brought the case before the court, seeking RMB 9,999 in damages. The defendant argued that AI-generated dialogue did not constitute a legally effective expression of intent, that the plaintiff had suffered no actual loss, and that the defendant had fully fulfilled its duty of care. On these grounds, the defendant maintained that no tort had been committed.

After examining a series of issues involved in the case, including whether the purported "promise" constituted an independent and autonomous expression of intent by the AI and whether it could be regarded as an expression of intent by the defendant; whether such disputes should be governed by general tort liability or product liability; how the alleged infringing conduct should be characterized; whether the defendant was at fault in relation to the false information generated by the AI; and whether the plaintiff had suffered damage, the court dismissed the plaintiff's claims in its first-instance judgment. According to the judge's explanation published by the Hangzhou Internet Court, "The era of artificial intelligence has arrived. We should adhere to the principle of giving equal priority to development and security and combining the promotion of innovation with the protection of rights and interests. We must both encourage the innovative development of generative AI services and ensure that the legitimate rights and interests of the parties concerned are not overlooked." The court's judgment was lawful, reasonable, and consistent with common sense, while clearly underscoring the cautious attitude that people should maintain toward generative AI. From the perspective of development guided by "Happiness," however, this case offers us far more than a definition of how artificial intelligence should be addressed at present. It also invites more far-reaching consideration of the era in which human beings and AI coexist.

"We should think about how to use AI effectively to enhance ourselves, rather than be replaced or misled by it[21]." One study expresses the same position in a vivid formulation: "In the legal profession, artificial intelligence will not completely replace legal professionals, but those who can use artificial intelligence tools will replace those who cannot [9]." Suppose that, under the same circumstances, the plaintiff repeatedly sought confirmation of the accuracy of the information, received an AI-generated "guarantee of compensation," and then made decisions based on the false information, resulting in damage to their fundamental personal interests. Such damage might even reach a point at which any reasonable person would intuitively recognize that "the limitations of technology have led to the infringement of an individual's fundamental rights," rather than attribute the harm to the user's "carelessness." In such circumstances, if the court ruled in the plaintiff's favor, it might appear to make the outcome of such cases depend solely on the fact of "damage." Even if this were the correct approach, determining the threshold at which damage would be sufficient to justify a favorable judgment for the plaintiff would present a new challenge. Conversely, if the court ruled in the defendant's favor, it might appear insufficiently respectful of the plaintiff's fundamental rights. Looking back from this hypothetical scenario to reality, it seems that some of these problems could be addressed by using "Happiness" as a guiding principle and establishing legal constraints on the direction of AI development in advance. One feasible approach would be to impose verification obligations on generative AI providers regarding information they produce that concerns the public interest or major personal decisions, requiring defendants to prove that they have adopted necessary measures, such as algorithmic optimization and cross-verification. Alternatively, liability mechanisms could be used to compel AI enterprises to strengthen

[21] This passage is taken from the "Judge's Explanation" section of the Hangzhou Internet Court's report on the case.

technological compliance, thereby preventing superficial logical consistency from undermining substantive justice. The purpose of such measures is to encourage AI technology, in its capacity as a tool, to address the vulnerabilities that may hinder the long-term development of human rationality, move beyond superficial forms of "Happiness," and bring the direction of technological revolution into harmony with the development of human rationality.

5. Conclusion

From ancient Greek hedonism to modern and contemporary utilitarianism, "Happiness" has evolved from an individual experience into a public value. The theoretical debates surrounding this concept over several centuries have given it an increasingly rich and multifaceted theoretical significance. Today, as artificial intelligence profoundly affects social life and the operation of the legal system, "Happiness" continues to possess important institutional value. It can provide a value-based point of reference centered on people's actual lives and development, enabling the law to respond to technological change. However, this value must be situated within the framework of the rule of law and remain consistent with justice, the protection of rights, and human subjectivity. Ultimately, rethinking "Happiness" points to an important question facing contemporary legal systems: how can technological progress and institutional development continue to serve the pursuit of a happy life for all?

References

[1] Zhao, J. (2025). *A study of John Stuart Mill's utilitarian conception of justice* [Master's thesis, Hebei University of Economics and Business].

[2] Yang, Y., & Zhang, G. (2025). Rational self-interest and common interests: An examination of the similarities and differences between Bentham's and Rawls's theories of social cooperation. *Journal of Zhejiang Normal University (Social Sciences), 50*(5), 33–42.

[3] Zhang, W. (2025). Moral arithmetic and the concept of punishment: A political philosophical critique of Bentham's principle of utility. *Morality and Civilization*, (6), 86–108.

[4] Chen, X. (2025). The academic life of John Stuart Mill. *Procuratorial View*, (7), 34–35.

[5] Baier, R. C., & Li, Q. (1991). On John Rawls's A theory of justice: An attempt to apply empirical methods to the theory of justice. *International Social Science Journal (Chinese Edition)*, (4), 135–148.

[6] Lu, A. (2025). *A legal-philosophical reflection on Rawls's theory of justice* [Master's thesis, Shenyang University of Technology].

[7] Xie, P. (2019). Justice as fairness: A review of the intellectual genealogy and theoretical framework of John Rawls's A theory of justice. *Theory and Contemporary*, (12), 26–28.

[8] Li, X., & Fan, M. (2024). On the teleological value of Rawls's two principles of justice. *Journal of Taiyuan University of Technology (Social Sciences), 42*(6), 88–95.

[9] Zheng, G. (2025). Artificial intelligence as a tool of contingency and its legal challenges. *Exploration and Free Views*, (8), 32–46, 177.

[10] Wu, J. (2024). Legal challenges posed by new artificial intelligence technologies and corresponding response strategies. *Journal of Heilongjiang Administrative Cadre Institute of Politics and Law*, (5), 122–128.