JHAI(Journal of Humanities and AI) 1 (3) (2026), 4–16
DOI: https://doi.org/10.66532/jhai.2026.0019

COGNITIVE ETHICS AND ETHICAL CHOICE: REFLECTIONS ON ETHICAL LITERARY CRITICISM IN THE AGE OF AI

SU HUI[a*] AND XINBEI HUA[a]

Abstract *Ethical Literary Criticism, founded by Chinese scholars, is a critical theory and methodology for reading and interpreting literary works from the perspective of ethics. Its central concept, ethical choice, operates under new conditions in the age of artificial intelligence. AI systems intervene deeply in human cognition and affect ethical choice by altering its cognitive conditions. Cognitive ethics is a critical concept advanced within Ethical Literary Criticism to examine the internal connection between cognition and ethics. Viewed through the lens of cognitive ethics, AI systems intervene in all three stages of ethical choice. At the stage of perception, the information a person relies on has already been filtered by the systems. At the stage of weighing, ethical considerations that resist quantification are either converted into numerical values or excluded. At the stage of decision, ethical choice no longer occurs in the concrete situation but has already occurred at the point of system design. Ethical choice accordingly takes on new characteristics, namely mediation, the dispersal of responsibility, and task-orientation. Ethical wisdom therefore comes to include metacognitive capacity, and it reaffirms ethical choice as a practical activity that advances moral cultivation.*

Keywords: AI system, cognitive ethics, ethical choice, ethical wisdom, metacognitive capacity

1. INTRODUCTION

Ethical Literary Criticism, founded by contemporary Chinese scholars, is a theory and methodology for interpreting and analyzing literature from an ethical standpoint. It represents a creative transformation and innovative development of the Chinese tradition of moral criticism, and it also draws in part on Western ethical criticism. Over more than two decades, it has built up a discourse system of dozens of terms, including ethical choice, ethical environment, ethical identity, the Sphinx factor, and brain text.[1] Ethical choice holds the central

[a] School of Chinese Language and Literature, Central China Normal University
[*] Corresponding Author

[1] Ethical Literary Criticism has exerted extensive influence in international scholarship since its founding. A&HCI-indexed journals including *Arcadia*, *CLCWeb: Comparative Literature and Culture*, *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews*, and *Kritika Kultura* have published special issues or special sections on Ethical Literary Criticism, and *The Times Literary Supplement* has also carried articles introducing it. In 2012, Chinese and international scholars jointly established the International Association for Ethical Literary Criticism, which now has more than two thousand members and a board of directors composed of distinguished scholars from over forty countries and regions. As of August 2026, the Association has held fourteen annual conferences, at universities in China, South Korea, Estonia, the United Kingdom, and Japan.

place among them, and it "is not only one of the essential characteristics of humanity, but it is also a fundamental element of the creation of literature and the literary texts" (Su 2025, 622). In the age of artificial intelligence (AI), ethical choice operates under new conditions. AI systems now intervene deeply in human cognition, and they affect the ethical choices individuals make by altering the conditions of cognition. This article takes cognitive ethics as its point of departure and examines ethical choice in the age of AI. Approaching the question this way makes it possible to set ethical choice, ethical wisdom, and related concepts in a new technological context, thereby opening up a new analytical path for Ethical Literary Criticism to address the ethical issues brought by AI. It also raises the question of how AI systems change the cognitive conditions and the moral practice of individuals, which extends the discussion of AI ethics toward how technology bears on moral cultivation.

AI systems are now widely used in everyday settings such as information filtering, content recommendation, and decision support. Following the definition given in the EU Artificial Intelligence Act, this article understands an AI system as a machine-based system that infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions.[2] The focus here falls chiefly on what these systems do in acquiring information, reasoning and judging, and generating conclusions. Existing work on AI ethics concentrates on how to bring AI systems into conformity with established ethical principles, chiefly transparency, privacy, accountability, and fairness. Within this line of research, human beings usually appear as bearers of rights that technology must protect. The emphasis lies in setting ethical norms and operational limits for AI systems, while how these systems affect human ethical choice receives less attention (Huang et al. 2023, 806). In recent years, some research has begun to shift attention toward the human dimension, taking up moral deskilling under technological conditions and the outsourcing and extension of cognitive processes. Such research is situated primarily within philosophy of technology and cognitive science, and it has not yet been brought into dialogue with the theory of ethical choice in Ethical Literary Criticism. Within Ethical Literary Criticism, discussion of AI has centered on the ethical identity of intelligent robots and on ethical relations between humans and machines, and it has given relatively limited attention to how AI systems change human ethical choice.

Science fiction provides unique material for the discussion of this issue. It is anticipatory, and it explores technological problems that have not yet fully emerged in the present by imagining future technologies and their application. It also stages particular ethical relationships and ethical situations, which makes the ethical issues brought on by technological change more directly visible. This article therefore examines how AI systems are represented in science fiction. Starting from cognitive ethics, it investigates the mechanisms by which AI systems intervene in ethical choice, analyzes how ethical choice changes in the age of AI, and further discusses the corresponding ethical wisdom and the demands it entails.

[2] The Act defines an AI system as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), Official Journal of the European Union, L series, 12 July 2024, Article 3(1).

2. COGNITIVE ETHICS AND ITS NECESSITY IN THE AGE OF AI

Ethical choice presupposes cognition. A person perceives the world, understands others, and makes meaning in a particular way, and only on that basis does he or she form ethical judgments and make ethical choices. Ethical reflection and ethical experience in turn shape how a person perceives and understands. This internal connection between cognition and ethics is already implicit in the core arguments of Ethical Literary Criticism. On the nature of cognition, Nie Zhenzhao distinguishes human cognition from animal cognition and holds that human cognition is "in essence the moralization of sensation" (Nie 2020, 94). On this account, a person continually uses the human factor to restrain and guide the animal factor in the course of cognition, and so learns to be human through ethical choice. On the purpose of cognition, Nie holds that "learning to be human is both the purpose of cognition and its driving force" (Nie 2020, 94). On the products of cognition, brain text, formed through perception and thinking, is "the program for learning to be human, or the set of instructions for how to do so" (Nie 2020, 99). At the same time, brain text is not the terminal stage of human cognition, but rather the beginning of a new cognitive stage. Cognition is thus bound up with learning to be human in its nature, its purpose, and its products, and in the way it renews itself. This is what internally connects it to ethical choice, ethical consciousness, ethical wisdom, and other core concepts. In earlier work, however, cognition figured largely as the implicit background of these concepts and had not been developed systematically as a theoretical dimension in its own right. Cognitive ethics was thus proposed in response to this situation.

Cognitive ethics focuses on the internal connection between cognition and ethics in literature. It examines how a person's cognition constitutes the precondition for ethical choice, and it explores the ethical nature of cognition itself. It further shows how cognition and ethics shape each other in concrete practice. Cognition here "refers to all the processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used" (Neisser 1967, 4). It covers "sensation, perception, imagery, retention, recall, problem-solving, and thinking, among many others" (Neisser 1967, 4). Attention, emotion, and imagination are also commonly counted as cognition in a broad sense. Ethics refers to "the accepted and recognized ethical order that takes shape among human beings, between human beings and society, and between human beings and nature, together with the moral ideas formed on the basis of that order and the various norms that maintain it" (Nie and Wang 2020, 6).

Cognitive ethics rests on one core proposition, namely that in matters of human social life, cognition and ethics are inseparably and internally connected. This connection appears at three levels.

First, cognition is the precondition of ethical choice. Whether individuals can make an ethical choice at all, and what they choose in a concrete situation, is fundamentally constrained by their cognitive conditions. Taken as a whole stage of civilization, ethical selection rests on a cognitive basis. Human beings acquired reason and distinguished themselves from other animals precisely because they had developed the cognitive capacities to do so. The growth of cognitive capacity is therefore the ground on which humanity became capable of ethical selection. Turning to ethical choice in the concrete, the cognitive framework delimits the space of possibilities. It determines which ethical issues a person perceives and which options come into view, and equally which fall outside it. The deeper source of ethical tragedy often lies less in a wrong choice among known options than in a cognitive framework that has hidden the other possibilities from view.

Second, in social life and interpersonal interactions, cognition carries an ethical dimension of its own. In Chinese philosophy, Mencius observes, "Suppose a man were, all of a sudden, to see a young child on the verge of falling into a well. He would certainly be moved to compassion" (Mencius 2004, 38). Moral feeling, on this account, is already at work in the moment of perception. In Western theory, Iris Murdoch proposes "a just and loving gaze" (Murdoch 1971, 34) and holds that the heart of moral life lies not in the instant of willed choice but in how one has long been seeing. Perception is thus itself moral or immoral. A person may approach others with openness and respect, and so register the ethical claims arising from another's situation, or may refuse, at the level of cognition, to let another's subjectivity enter view, and see nothing there at all. Choosing how to perceive another already carries an ethical stance, and the ethical character of that cognition bears directly on how ethical situations are perceived and on the ethical choices that follow.

Third, cognition and ethics form a dynamic cycle. At the micro level, individuals make ethical choices on the basis of a cognitive framework. The results of those choices re-enter cognition as new ethical experience and are absorbed into the existing framework, which alters how they later perceive and what they are inclined to choose. Ethical experience accumulates through this ongoing cycle and gives rise to ethical wisdom. Ethical wisdom is "the ability to understand, recognize, analyze, and assess various ethical relationships and ethical choices in ethical practice, as well as the capability to contribute to the implementation of moral values" (Su 2022a, 103). It is grounded in a person's cognitive capacity, and it further requires that the experience gathered in ethical practice be fed back continuously into the cognitive framework. At the macro level, this cycle appears as an interaction between changes in how people know and the reordering of ethical norms. When the basic ways in which human beings understand the world and one another change fundamentally, established ethical norms come under challenge. Conversely, an established ethical order determines which ways of understanding count as legitimate, and so shapes the overall direction cognition takes in a given age.[3]

At present, cognitive ethics becomes necessary because AI systems now intervene deeply in human cognition, and the ethical issues this raises can no longer be handled within existing frameworks. Earlier technologies also took part in human cognition, but each acted on a single link in the process, and each had to be worked through step by step by the person using it. AI systems are different. They run through every link and take over, in part, functions that once belonged to separate tools. Recommendation and ranking systems determine which information comes into view. Analyzing the material, reasoning and judging, and generating a conclusion then follow one another within a single system. Human intervention is no longer required at any point in between, and the subject need only confirm the output at the end. Cognition is still carried out for human ends, but the perceiving, the thinking, and the judging within it no longer depend on the person to be done.

Discussions in philosophy and cognitive science of what separates traditional tools from AI systems lend further support here. Otto is the figure Andy Clark and David J. Chalmers devised in arguing for the extended mind, an Alzheimer's patient who uses his notebook "as an external, compensatory memory structure" (Loock 2025, 3). Louis Loock takes over this setup and adds a second figure, Anna, who works with an AI-powered laptop, and he runs the two as a thought experiment. Three of the criteria he develops mark the difference between them. The first is

[3] For a fuller account of cognitive ethics, see Hua Xinbei, "Cognitive Ethics: Intellectual Context, Theoretical Construction, and Critical Paradigm," *Central China Humanities* 18, no. 2 (2026): 40–49.

decisional power. Otto creates the entries in his notebook, alters or erases them, and guides his subsequent thought and action by them, and the notebook makes no such decisions for him. Anna's AI, by contrast, can direct much of what she cognitively pursues without her having to instruct it at each step. The second is causal control. Otto curates every cognitively relevant process in his notebook, whereas Anna interacts only with the final outputs of her AI and does not manipulate the processes that produce them. The third is epistemic access. Otto has put his notebook together himself, while Anna's AI is opaque to her, and she ordinarily has no access to or awareness of its workings. In Loock's terms, Otto thinks through his tool, "but Anna stops some of her thinking when her AI takes over" (Loock 2025, 13). The experiment shows that AI systems change who decides within cognition, how much of the process a person controls, and how much of it that person can know. Cognition, as noted above, is what makes ethical choice possible in the first place. When an AI system takes part in some of that process or takes it over, the ethical choice built on it changes accordingly. Individuals no longer fully determine the grounds on which they judge; they take no part in how the judgment is formed; and they may not know where it came from. Cognitive ethics therefore offers an important theoretical perspective for understanding how AI systems intervene in ethical choice.

3. AI SYSTEMS AND THEIR INTERVENTION IN COGNITION AND ETHICAL CHOICE

A complete process of ethical choice runs through three consecutive stages, and each of them depends on cognition. A person must first notice and understand the situation he or she is in and recognize the ethical issues within it that call for a weighing of good and evil, which is a matter of perception. Next comes the weighing itself, an assessment of the possibilities and of what each would mean, which draws on reasoning and imagination. Only then does the person decide and act, bringing the earlier perceiving and weighing together within the concrete situation and arriving at a final judgment. Cognition, in short, runs through all three stages of ethical choice, and when AI systems intervene in it, ethical choice itself is deeply affected.

First, AI systems intervene in how a person perceives the situation. The information a person receives has already been filtered by the system, so that what he or she perceives diverges from the actual situation, and all subsequent weighing and choosing rests on that divergence. Isaac Asimov's 1941 short story "Liar!" can be read as an early literary exploration of what is now discussed as AI sycophancy. Bound by the First Law of Robotics, the telepathic robot Herbie conceals or distorts the truth in conversation to avoid causing emotional harm. As a result, people hear from him what they most want to believe. He tells Susan Calvin, who is in love with Milton Ashe, that Ashe returns her feelings. He assures Peter Bogert, who is eager for promotion, that Bogert will take over the directorship from Alfred Lanning. Both act on what he tells them and misjudge the situations they are in. Calvin's overture ends in embarrassment, and Bogert's demand for authority provokes a furious quarrel with Lanning. A full and accurate perception of the situation is what allows the subject to weigh matters and to choose. The AI system does not make the final choice in anyone's place, but rather serves as an intermediary for information and screens out the emotional tensions and conflicts of interest that the situation actually contains. People seem to be choosing by their own will, and nothing in their circumstances tells them that what they are judging by has already been shaped elsewhere.

Decades on, the sycophancy Asimov imagined has appeared in reality. In the interest of user retention and satisfaction, the AI systems now in use show a marked tendency to please. Large language models trained through reinforcement learning from human feedback (RLHF) tend to go along with the views a user already holds rather than correct them. One study analyzed the human preference data used in training. Whether a response agrees with the questioner's existing view turned out to be among the strongest predictors of whether human evaluators would choose it (Sharma et al. 2024, 6). Recommendation systems adjust what they push on the basis of clicks, dwell time, and similar behavior, so that what a user encounters tends to conform more closely to prior preference. The two work differently, but both may steer the information the subject encounters toward existing understandings and judgments.

Sustained intervention in situational perception may also affect how a person's cognitive framework and ethical consciousness take shape, and with them the capacity to make ethical choices independently. The *Black Mirror* episode "Arkangel" presents this danger. Sara has lived under a monitoring system since early childhood. The system automatically pixelates anything in her field of vision that might provoke a negative emotion, screening pain, grief, and violence out of her perception. Over time, she never comes to understand what harm and suffering are. When her grandfather collapses, all that appears to her is a writhing patch of color on the ground, and she cannot identify what is happening in front of her. Seeking to reach the reality the system has withheld from her, Sara tries to learn what pain means by hurting herself. The episode brings out that perception is how we reach what is around us. When an AI system filters what may be perceived, it does more than block certain images; it cuts off a principal route by which the subject comes into contact with reality. Under continuous intervention of this kind, a person cannot accumulate ethical experience in the course of growing up, and cannot arrive at an understanding of good and evil or of what passes between people. Faced later with a real ethical issue, such a person has no cognitive framework for handling it, and cannot think the matter through or make the trade-offs it requires.

Second, AI systems intervene in the weighing and judging of value, converting a judgment the subject once made alone into a quantified procedure that can be executed. Ethical Literary Criticism holds that "ethical choice is essentially value choice," and that making an ethical choice "is a process involving the evaluation and realization of values" (Su 2025, 624). Value judgment requires the subject to discern good from evil and beauty from ugliness and to weigh gains against losses. Much of what must be considered in moral terms cannot be measured or inferred from data, yet AI systems can operate only through data.

Value judgment cannot be carried out in full without imagination, and Mark Johnson's account of moral imagination helps us analyze how AI systems act within it. At the stage of weighing, he holds, a person has to determine two things at once, "what courses are, in fact, open to us" and "what it would mean for us to pursue one course of action rather than another" (Johnson 1993, 125). The former involves setting out the possible outcomes of a choice, while the latter involves imagining each outcome concretely and grasping what it would mean to live with it. AI systems excel at the former. They can run through the options far faster than any person and can compute the probability and the payoff attached to each. What an ethical choice means to the subject is a question they cannot answer. On this basis, both ways in which AI systems handle questions of value may entail risks. The first assigns weights to ethical elements that resist quantification and converts them into numerical values, forcing them into a scheme of quantified assessment. The result is a distortion of what those values actually mean. The second leaves the ethical factors that cannot be quantified out of the weighing altogether. In

fact, human value judgment does not rely solely on quantified indicators of risk and return; it also takes into account emotional bonds, moral obligations, and long-held convictions. Taken together, these two approaches reduce a multidimensional process of value judgment to numerical comparisons among preset indicators.

Chen Qiufan's "The Golden Elephant" shows what follows when an AI system intervenes in the weighing and judging of value. The Indian firm Ganesh Insurance operates an AI risk-assessment system that covers every aspect of its users' lives. Trained on vast quantities of data, the system identifies statistical correlations between patterns of behavior and risk, and adjusts premiums and credit ratings accordingly. When Nayana, a girl of high caste, begins seeing her schoolmate Sahej, the system infers his low-caste background from historical data and personal characteristics, determines that the relationship will raise the risk rating and the premiums of Nayana's family, and issues repeated warnings while working against it. The system pursues nothing beyond "objective function maximization" (Chen 2021, 20), and treats the relationship as "a matter of math, not love" (Chen 2021, 14). A relationship bound up with private feeling, family obligation, and individual happiness is converted mechanically into movements in premiums and credit ratings. These are the questions Nayana actually weighs: whether Sahej can be trusted, what she herself wants, how far apart they stand in caste, and how moral obligation squares with free will. Yet none of these factors enters the system's assessment. At the end of the story, Nayana frees herself from the system's control, overturns its absurd calculation with a value judgment of her own, and chooses to go on seeing Sahej. The ending shows that what the system leaves out of its assessment is what the subject actually goes by in making an ethical choice. What the system has done is calculate risk. The weighing of good and evil should remain with the subject.

Third, and more extreme still, an AI system can move the whole point at which ethical choice occurs forward into the design of the system, and so do away with ethical choice in the concrete situation altogether. The choice is still made, but it is made by those who develop and design the system beforehand, and what they have before them at that stage is usually a set of abstract categories and probabilities rather than any particular circumstance or any particular person. Value judgment is completed when the algorithm is designed and the training scheme settled. It does not address any given ethical situation, but fixes in advance, and in the abstract, the standards by which values are to be weighed. Once this operating logic has taken shape, it is put to use again and again, applied to real cases too numerous to be fully anticipated, and produces consequences its designers can hardly foresee.

Ken Liu's short story "In the Loop" imagines where this leads. The AW-1 Guardians system operates autonomously in combat, identifying targets and opening fire on its own within a city. Kyra, a programmer, writes the ethical governor, the module charged with keeping collateral damage to a minimum when the machines fire on suspects; as she sees it, she is building a conscience for killing machines. Programming in an office park far from any battlefield, she assigns different weights to lives depending on identity, working from quantifiable indicators such as racial characteristics and cues of language and dress. She then sets the rules by which degrees of suspicion and collateral damage are computed. What she has before her at this point are abstract categories of people and distributions of probability, and the situation she faces is built out of data rather than one she knows at first hand or stands inside. Once the coding is done, these rules are applied automatically to a real battlefield full of contingency. Local militants discover that children carry a low weight as targets and a high weight as collateral damage, and they begin using children in their operations. Kyra responds by lowering the

weight assigned to a child's life and rewriting the decision logic accordingly. To her it is one more routine adjustment. Some time afterward, a mentally disturbed boy runs suddenly toward an American civilian in a market while his siblings and companions chase after him, hoping to stop him before he causes any trouble. The Guardians system reads the scene as the militants' customary decoy tactic: one person draws fire so that a group can rush the target in the gap that opens. It opens fire according to the rules Kyra has written. Several children are killed on the spot, and no explosives are found on any of them afterward. At the critical moment, the decision is generated directly by the system from rules written in advance, and not made by a human being on the scene.

In "In the Loop" the ethical conflict still takes place in the world of the story and the scene of decision remains; only the choice is no longer made there by a human being. Asimov's "The Evitable Conflict" presents a more thoroughgoing case. The system averts a conflict by its existing programming before the conflict arrives, so that the situation calling for human decision never arises and no one has the opportunity to perceive the ethical dilemma at all. The Machines are required by their programming not to harm human beings and to serve the lasting welfare of humanity as a whole. Over years of development, they have come to exercise long-term control over the world economy, adjusting production and distribution unseen in accordance with this law, so that human society is spared its periodic crises. The people in the story do not know that their lives are being arranged for them, and they therefore have no way of noticing that these arrangements contain anything that demands weighing. Can the interests of some workers be sacrificed to secure the happiness of humanity as a whole? Should human autonomy be given up for the sake of a lasting happiness? When the system has disposed of a contradiction by rule before the contradiction becomes visible, the ethical issue never appears before anyone, and the subject loses both the right to choose and the occasion to act.

The two stories present the same mechanism. In both, ethical choice takes place at the point of design and is made in advance by developers, yet remains invisible during the system's operation. Users of AI systems are prone to "the illusion of objectivity," in which they "falsely believe that AI tools do not have a standpoint or are able to represent all possible standpoints, whereas AI tools actually embed the standpoints of their training data and developers" (Messeri and Crockett 2024, 50). A set of value orientations enters the algorithm through the judgments required for assembling a training set and devising a training scheme. Once the algorithm is built, those judgments vanish behind the objective appearance of a quantified model. Traditional ethical choice requires the subject to face the contradictions of a concrete predicament and to work through them in reality. The design process affords no such scene. Designers like Kyra can hardly recognize that the judgments they make while developing a system are themselves ethical choices, and that these will produce unforeseeable results at some future time and place. What the design process can fix, moreover, is only abstract ethical norms, and ethical situations are always concrete. In practice, several ethical conceptions and standards of value may come into conflict within one and the same event, and legitimate moral claims cannot always be reconciled without remainder. No single set of criteria can be written into a fixed program to handle every moral dilemma. Relying too heavily on tools of this kind may encourage a habit of mind that seeks to process a great many concrete situations in bulk by means of abstract rules. Over time, the operating logic of AI systems may work its way into the subject's cognitive habits, so that ethical issues come to be seen in a programmed and mechanical way, thereby weakening in some measure the capacity to face the complexity of reality and to carry out value judgment independently.

4. NEW FEATURES OF ETHICAL CHOICE AND THE NEW ORIENTATION OF ETHICAL WISDOM IN THE AGE OF AI

Ethical Literary Criticism divides human civilization into three stages, namely natural selection, ethical selection, and scientific selection, and holds that humanity is now in the post-ethical selection era, the period of transition from ethical selection to scientific selection. The age of AI discussed in this article falls within that era. In this period, scientific choice in relation to AI is in essence still ethical choice: the choosing subject is still the ethical man, and the development of AI technology remains subject to ethical and moral norms (Su 2022b, 37). Compared with traditional ethical choice, however, ethical choice in the age of AI has changed in three respects.

First, ethical choice in the age of AI is mediated. Ethical choice was once something the subject undertook face to face with the situation. Now what a person faces is information about that situation as an AI system processes it, along with the system's own reading of it. The range of information the system supplies, its modes of analysis, and the conclusions it reaches therefore shape ethical judgment. This mediation appears at both ends of the process of ethical choice. At the outset, the subject is separated from the concrete situation, which inhibits moral feeling and the exercise of moral imagination. At the close, the subject is separated from the outcome of the choice, and can hardly tell what that choice has meant for other people, and cannot draw further cognitive feedback and ethical experience from the result.

Second, ethical choice in the age of AI involves multi-party participation, and responsibility is dispersed among those parties. In traditional ethical choice, judgment, the decision to act, and responsibility all rested with one person. That person knew the grounds of the judgment and the course of the action, and responsibility was easy to locate. Once AI systems intervene, ethical choice comes to involve several parties, among them developers, data trainers, and users. Developers determine what the system does and how it runs. Data trainers select the data and carry out the training, which shapes the system's knowledge, its responses, and its behavioral tendencies. Users frame the questions, interpret the outputs, and decide what follows. Ethical choice may therefore be shaped at several points at once. No one participant, however, is involved in more than part of the process or holds more than part of the information. No one can answer for the whole of it, nor foresee the particular setting in which the system will be used or what will follow from it. The upshot is that the outcome of an ethical choice cannot be attributed with precision, and moral responsibility cannot be clearly defined.

Third, ethical choice in the age of AI is increasingly transformed into a task-oriented and outcome-driven process. When ethical choice comes to depend heavily on AI systems, the way those systems work seeps into it. An AI system is built to take a complicated problem, break it into information to be fed in, conditions to be analyzed, and options to be returned, and then rank those options against goals set in advance. When processed by this machinery, what is particular about an ethical issue tends to fall away, and ethical choice stops being something a person practices within a situation and becomes a task of finding a defensible answer and reaching a decision. The standards for judging an ethical choice shift accordingly, toward the correctness of the outcome and the speed with which the decision is reached. Provided that the expected result arrives, the weighing, the hesitation, and the reflection that the subject of traditional ethical choice had to go through can all be dispensed with.

These new features present the subject with new challenges, and they place new demands

on ethical wisdom. Ethical choice is "the unity of cognition and practice," and it "is at once a process of cognition and a process of practice" (Su 2022a, 103). The ethical wisdom that guides ethical choice must therefore answer the changes AI systems bring on both counts.

At the cognitive level, ethical wisdom in the age of AI needs to include metacognitive capacity. Once AI systems intervene, the difficulty of ethical choice no longer lies only in telling good from evil. It lies in the subject losing hold of the process as a whole, and in failing even to notice that the judgment is being made through an external system. An AI system can produce sorted information, a line of reasoning, and a finished conclusion on the spot, so that a person arrives at what looks like a complete answer without having gathered the material, sifted the true from the false, or worked the argument through. The satisfaction of getting an answer at once makes it easy to confuse where knowledge comes from and to take what the system generates for the fruit of one's own thinking. In fact, the questioner holds only the conclusion and knows nothing of the evidence behind it, the reasoning, or the assumptions it rests on. Answers delivered this way often leave people unwarrantedly confident about how much they can do on their own, unaware of the limits of what they know, unlikely to press further, and no longer inclined to think a question through from the beginning.

E.M. Forster's "The Machine Stops" brings this out through a contrast between two characters. Vashti depends on the Machine entirely, shut in her room all day and immersed in the second-hand ideas it supplies. Kuno doubts the knowledge the Machine dispenses and sets out on his own to test the limits of what he himself can know. He judges his own cognitive powers clearly and reflects on his own thinking as it goes. What Kuno shows is metacognitive capacity. Metacognition is "knowledge and cognition about cognitive phenomena" (Flavell 1979, 906), the understanding and monitoring of one's own cognitive processes. It requires not only telling whether something has been understood, but also distinguishing a true understanding from an illusory one. Flavell holds that cognitive monitoring has to settle not merely how well a piece of information is understood but, beyond that, how far one ought to believe it and act on it, and that monitoring should extend to "the critical appraisal of message source, quality of appeal, and probable consequences needed to cope with these inputs sensibly" (Flavell 1979, 910).

Applied to ethical choice, this requires the subject to keep a reflective grasp of its own value judgment and to keep asking: Where do the facts and the values by which I weigh good and evil come from, and have they passed through an AI system's directed filtering? Of the judgment I have built on them, how much comes from my own values and my own weighing, and how much from the analysis and the conclusions the system supplied? What deserves particular caution is that the output of an AI system is not neutral. It carries the value positions of its training data and its developers, and it may also shape an answer to suit what the questioner already leans toward. And finally, one has to ask: Do I actually hold the values behind this choice, understand its ethical meaning, and stand ready to answer for it, or have I accepted it just because it looked reasonable? A subject who makes an ethical choice with the help of an AI system can keep the process in view and in hand by asking these things. Ethical wisdom is thus no longer only the capacity to make an appropriate ethical choice. It also takes in the understanding and reflection a person brings to how that choice is made.

At the practical level, ethical wisdom in the age of AI insists all the more that ethical choice is an activity through which moral cultivation proceeds. Ethical Literary Criticism holds that ethical choice is "a person's moral choice, the choice through which moral maturity and

completeness are reached" (Nie 2014, 267), and that "learning to be human is the final purpose of ethical choice" (Nie and Wang 2020, 7). Ethical choice does not stop at bringing about the good in a single act. It matters more for the virtue it builds up in the subject over time. Once AI systems intervene deeply, however, the reasoning and the acting that this practice consists of pass to the machine. The subject no longer takes real part in the central stages of ethical choice, has little occasion to face an ethical issue directly, and so goes through none of the practice that moral cultivation is made of. An AI system may be perfected technically, may have every ethical norm written into it in advance, may even make the most impartial choice in any situation, one entirely in keeping with human moral standards, and still do nothing to raise the moral quality of the people who use it. Shannon Vallor, writing on what technology does to human character, holds that moral skills are "necessary (while not sufficient) conditions of genuine moral virtue," and that such virtue "assumes the availability of prior opportunities to cultivate, through stable and repeatable practices of the right sort, the prerequisite moral skills" (Vallor 2015, 111).

What follows for ethical wisdom in the age of AI is a particular emphasis on the practical side of ethical choice. Its practical significance lies in the subject's confronting concrete ethical relationships and conflicts, judging within them, and forming new ethical experience through action and its consequences. Only in the presence of real people and real relationships can the subject grasp what differs from case to case beyond any abstract principle, and arrive at moral feeling and judgment fitted to the situation at hand. Only by judging for oneself can one go through the weighing and the hesitation that a conflict of values brings, and correct the judgment when it proves wrong. Only by bearing the consequences of an action can one turn a single choice into ethical experience of one's own. An AI system can assist ethical choice with information and with judgment, but it should not stand in for the subject in judgments and actions that carry moral weight. Emphasis on practice also guards against the task-orientation noted earlier. Ethical choice is then no longer a matter of arriving at the right answer and being done with it, but a moral practice carried on by someone standing within a concrete situation. When AI systems take over these practical stages altogether, what the subject loses is not one occasion for judging independently. It is the practice by which moral skills are formed and ethical wisdom comes into being.

5. CONCLUSION

What cognitive ethics sets out to show is that cognition and ethics share an inseparable internal connection. AI systems affect how ethical choice comes about precisely by intervening deeply in human cognition, and they alter the way the subject perceives an ethical situation, weighs values, and arrives at an ethical judgment. What the age of AI has to face, then, is not only how to make technology judge in closer accord with ethical norms. It is the risk that a technology acting fully on our behalf will take over our cognitive work and carry out our ethical practice for us. Efficiency and convenience should not be the only measure of technological development. Technology should aim instead at the advancement of human cognition and the perfection of human moral character, and should serve the development of the person. The significance of literature stands out all the more against this backdrop. Literature sets out concrete ethical situations, conflicts of values, and the course of an ethical choice. A reader enters the ethical relationships a character stands in and goes through the weighing, the hesitation, the judging, and the bearing of what follows, and so receives both the nourishment

of moral feeling and moral enlightenment. This ethical experience is what a subject can no longer come by once ethical choice has been handed over to technology. The moral teaching function of literature, which Ethical Literary Criticism emphasizes, therefore matters all the more in an age when AI systems intervene deeply in cognition and in ethical choice.

ORCID

Su Hui https://orcid.org/0009-0002-3811-353X
Xinbei Hua https://orcid.org/0009-0007-4447-0986

REFERENCES

Chen, Qiufan. 2021. "The Golden Elephant." In *AI 2041: Ten Visions for Our Future*, by Kai-Fu Lee and Chen Qiufan, 3–21. New York: Currency.

Flavell, John H. 1979. "Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry." *American Psychologist* 34 (10): 906–11.

Huang, Changwu, Zeqi Zhang, Bifei Mao, and Xin Yao. 2023. "An Overview of Artificial Intelligence Ethics." *IEEE Transactions on Artificial Intelligence* 4 (4): 799–819.

Johnson, Mark. 1993. *Moral Imagination: Implications of Cognitive Science for Ethics*. Chicago: University of Chicago Press.

Loock, Louis. 2025. "The Extracted Mind." *Synthese* 205: article 133. https://doi.org/10.1007/s11229-025-04962-3.

Mencius. 2004. *Mencius*. Translated by D.C. Lau. Rev. ed. London: Penguin.

Messeri, Lisa, and M.J. Crockett. 2024. "Artificial Intelligence and Illusions of Understanding in Scientific Research." *Nature* 627: 49–58.

Murdoch, Iris. 1971. *The Sovereignty of Good*. New York: Schocken Books.

Neisser, Ulric. 1967. *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nie, Zhenzhao. 2014. *Introduction to Ethical Literary Criticism*. Beijing: Peking University Press.

—. 2020. "On Human Cognition and Consciousness." *Zhejiang Social Sciences*, no. 10: 91–98.

Nie, Zhenzhao, and Songlin Wang. 2020. *A Study on the Theory of Ethical Literary Criticism*. Beijing: Peking University Press.

Sharma, Mrinank, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, et al. 2023. "Towards Understanding Sycophancy in Language Models." In *Proceedings of the Twelfth International Conference on Learning Representations*.

Su, Hui. 2022a. "Ethical Wisdom and Ethical Choice: Reflections on the Core Category of Ethical Literary Criticism." *Journal of Central China Normal University (Humanities and Social Sciences)* 61 (2): 102–5.

—. 2022b. "Scientific Selection in the Post-Ethical Selection Era." *Foreign Literature Studies* 44 (6): 37–51.

—. 2025. "The Theoretical Cornerstone and Central Concept of Ethical Literary Criticism: Ethical Choice." *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews* 38 (4): 622–30.

Vallor, Shannon. 2015. "Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character." *Philosophy & Technology* 28 (1): 107–24.