Reflections on Judgment in the Age of AI

Problem Formation, Aesthetic Judgment, and the Education of Judgment

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Artificial intelligence is altering the production of candidate problems, candidate solutions, and scholarly outputs at the same time. Search, synthesis, hypothesis generation, proof search, coding, analysis, and writing can increasingly be scaled through machine assistance, while their material, experimental, and epistemic costs remain uneven across domains. This essay considers what this redistribution of scarcity may mean for the evolution of scholarship. It approaches judgment as a plural practice that enters inquiry before, during, and after solution. Drawing on Dewey, Peirce, and Kuhn, the discussion treats problem formation as part of inquiry itself: problems acquire form through attention, inherited expectations, encounters, and acts of selection. Kant's accounts of reflective and aesthetic judgment then provide a point of departure for examining how a possible problem can acquire aesthetic salience before its utility or broader significance is known. Weber, Berlin, and Arendt support a further move toward plural forms of judgment and the mediation of aesthetic, epistemic, normative, practical, economic, and historical considerations. The educational discussion brings this framework to classics, contingency, and AI-assisted learning. Classics and disciplinary traditions condense prior judgments and make anomalies intelligible; encounters with an unfinished world can revise inherited problem-spaces. AI enters this circulation as a generator, solver, evaluator, and increasingly capable participant in inquiry. The essay proposes that judgment education should cultivate participation in the formation, comparison, mediation, and revision of problem-spaces as scholarship becomes increasingly abundant in both questions and answers.

Keywords: artificial intelligence; judgment; problem formation; aesthetic judgment; education

Essay Note

This text is a philosophical essay and era commentary with a deliberately limited scope. It develops provisional reflections on how AI may be changing the internal organization of scholarly work and on the place of judgment within that change. Historical, philosophical, and contemporary examples clarify the argument and support a situated interpretation of the present transition. The essay marks its extensions of established philosophical concepts as proposals of the present text and identifies the cited authors only with positions supported by the referenced sources. The claims remain deliberately scoped and revisable.

Statement on the Use of Language Models

Exploratory discussion and preparation of this essay involved OpenAI's ChatGPT. ChatGPT supported conceptual dialogue, source discovery followed by online verification, argumentative criticism, and drafting assistance in LaTeX. The author selected the questions, directed the philosophical commitments, evaluated the proposed arguments, and bears responsibility for the manuscript, its interpretations, claims, conclusions, and errors.

1 The Changing Economy of Scholarship

Scholarship has always been shaped by scarcity. Attention is finite; so are years of training, access to instruments, opportunities to test an idea, and the number of alternatives a researcher can seriously compare. The *economy of scholarship* in this essay refers broadly to the distribution of these scarce intellectual and material capacities across inquiry. Money and markets form only one part of that economy. Artificial intelligence matters because it is beginning to alter several of those distributions at once. Systems now participate, with uneven reliability and substantial domain dependence, in literature search, hypothesis generation, formal proof, algorithm design, data analysis, experimental planning, and the composition of research outputs. AlphaProof substantially advanced automated formal mathematical reasoning and, as part of a combined system, contributed to a silver-medal-level result on the 2024 International Mathematical Olympiad problems.[1] Co-Scientist has been developed to generate, rank, critique, and refine scientific hypotheses and research proposals, with experimental validation reported in several biomedical settings.[2] Robin links literature-based hypothesis generation to data analysis and iterative experimental interpretation in a semi-autonomous scientific workflow.[3] Taken together, these systems matter less as isolated benchmarks than as signs that candidate questions and candidate answers can both be produced at greater scale.

The word *cheaper* also requires care. Computation is not free. Advanced AI systems depend on costly infrastructures of chips, electricity, data, engineering, model development, and institutional organization. Empirical validation, laboratory work, field access, expert oversight, and the consequences of error remain expensive. The narrower change concerns marginal human effort: once suitable infrastructure exists, the time and coordination required to produce another hypothesis, proof attempt, code implementation, synthesis, or written formulation can fall sharply in some domains. A researcher can request dozens of alternative formulations where only one or two might previously have been attempted; a formal system can explore large families of variants; an agentic workflow can continue searching after a person has left the desk. AI therefore redistributes scarcity within inquiry, and the remaining scarcities acquire new relative weight.

Seen historically, scholarship has repeatedly changed when the location of scarcity changed. Writing externalized memory. Printing enlarged the reproducibility and circulation of texts. Journals and libraries reorganized access to dispersed findings. Statistical and computational tools expanded the scale of calculation, while digital databases reduced the time required to locate literature. Each transformation altered what could reasonably be attempted and which activities could become routine. AI belongs to this longer history, but its present significance lies in the fact that it increasingly enters activities often treated as intellectually generative: proposing directions, searching solution spaces, interpreting intermediate results, and producing revised proposals.

Generation and solution are also becoming less separable. A hypothesis can be scored and revised automatically; a failed proof attempt can produce new variants; experimental output can be returned to a system that proposes follow-up tests. Co-Scientist explicitly combines generation, debate, ranking, and refinement, while AlphaProof searches formal proof spaces and Robin links proposed interventions, experimental results, analysis, and revised hypotheses.[2, 1, 3] The scholarly cycle can therefore contain more machine-mediated iteration before a human researcher intervenes again.

For the evolution of scholarship, the important question is what becomes relatively scarce when such operations become easier to reproduce. Experimental access may become more constraining. Verification may become more constraining. So may attention, responsibility, the formulation of objectives, and the ability to decide which among many technically possible directions deserves continued time. The observation supports a modest structural claim. AI systems already rank alternatives, critique hypotheses, compare candidate solutions, and apply learned or supplied evaluation criteria, so judgment remains open to machine participation. As possibilities and solution attempts become abundant, acts of selection, framing, continuation, interruption, and redirection become more visible as parts of scholarship in their own right.

This visibility unsettles a familiar picture of expertise. A mature discipline presents itself through stabilized concepts, exemplary methods, canonical results, and recognized open questions. Such organization is indispensable. Without it, each generation would spend much of its time rediscovering what others already know. Yet the clarity of a mature problem-space can hide the process by which its questions became visible. Once a question has been named, formalized, reproduced in textbooks, and surrounded by a literature, it can appear as though the world had always contained it in that form.

AI can intensify this effect because it produces maps with remarkable speed. A system can summarize central debates, list research gaps, propose hypotheses, compare methods, and generate plausible continuations of an existing literature. These maps can be extraordinarily useful. Their fluency also makes it easier to forget that every map records selections. Concepts, instruments, institutions, and earlier judgments determine what enters the representation. A polished account of a field is still an account of a field that has already learned to see certain things.

The resulting philosophical question concerns the status of the problem itself. If candidate problems and candidate solutions can both be generated in increasing abundance, what makes one direction worth sustained attention? Technical difficulty cannot answer by itself. A problem may be difficult because it discloses a deep structure, because it connects domains, because it exposes a contradiction, or simply because a large search space has been constructed around it. Solvability is equally insufficient. The fact that a problem can be solved does not tell us why it deserved years of attention, public resources, or a place in a discipline.

Judgment enters here before the familiar judgment of whether an answer is correct. A researcher may find a problem elegant, scientifically revealing, socially urgent, ethically troubling, economically imprudent, historically timely, or institutionally premature. These assessments belong to different evaluative languages, and later sections will consider their relations. For now, the narrower point is enough: scholarship cannot be understood as the possession of a ready-made problem followed by the production of a solution. The problem itself has a history, and so does the decision to keep attending to it.

The next movement therefore goes behind the apparently given question. Before a problem is solved, it must acquire a form in which it can be pursued. Something is noticed as indeterminate, surprising, anomalous, compelling, or otherwise worthy of inquiry. Philosophies of inquiry and scientific discovery help recover this earlier movement, which textbooks, benchmarks, and AI-generated research agendas can easily hide.

2 The Formation of Problems

A mature field often presents its problems with deceptive clarity. The textbook states the exercise, the article states the research question, the grant proposal states the gap, and the benchmark states the task. By the time inquiry reaches these forms, much of the earlier uncertainty has disappeared. The problem has been separated from its surroundings, named, bounded, and made discussable. This finished appearance encourages a simple sequence: first there is a problem, then there is an attempt to solve it. Philosophies of inquiry and scientific development suggest a less linear picture. A problem can itself be an achievement of inquiry.

2.1 Indeterminate Situations and Problem Formulation

John Dewey's *Logic: The Theory of Inquiry* offers a useful starting point because it places problem formation inside inquiry. Dewey describes inquiry as the transformation of an indeterminate situation into one whose relevant distinctions and relations have become sufficiently determinate for coordinated thought and action.[4] The initial condition is an unsettled situation: habitual expectations no longer organize experience adequately, a relation remains obscure, or action does not proceed as anticipated. Inquiry begins by giving form to that indeterminacy.

Problem formulation is decisive in this account. Dewey emphasizes that the way a problem is conceived governs which suggestions appear relevant, which data are selected, and which hypotheses are entertained.[4] A poorly framed problem can organize rigorous reasoning around the wrong distinctions, while reformulation can change the field of possible answers before any answer has been established. Problem formation is therefore part of the intellectual work through which inquiry acquires direction.

This perspective also complicates a clean division between question generation and question solving. The two can be distinguished, but the formation of a problem may contain anticipations of possible solutions, while attempts at solution can reveal that the problem was badly formed. A contradictory result may motivate a revised question; a failed experiment may expose an unnoticed condition; a proof attempt may show that one formulation conflated two structures. Inquiry can return to its own beginning.

2.2 Surprise, Abduction, and the Opening of Inquiry

Charles Sanders Peirce approaches discovery through a different but complementary route. In his mature discussions of abduction, inquiry can begin when an observed fact conflicts with what had been expected. Abduction proposes a hypothesis under which the surprising fact would cease to be surprising; the hypothesis is introduced provisionally and becomes a candidate for subsequent deductive development and inductive testing.[5] The force of the episode lies less in a recipe for invention than in the movement from disturbance to a new way of organizing what has been encountered.

Surprise alone does not constitute discovery. Most unexpected events do not become research programmes. Some are measurement errors, accidents, irrelevant irregularities, or phenomena already explained elsewhere. An unexpected occurrence becomes intellectually generative only against a background of expectations and available conceptions. The investigator must recognize that an explanation is called for, imagine a hypothesis that would render the occurrence intelligible, and decide that the hypothesis deserves provisional examination. Peirce's own account keeps this move answerable to later testing.[5]

For the present essay, the important point is relational. A phenomenon becomes surprising only relative to some prior sense of what would ordinarily occur. Theory, practical experience, instruments, disciplinary habits, and tacit familiarity can all contribute to that background. The unexpected is therefore encountered through an organized field of anticipation. A novice and an expert can observe the same trace, institutional practice, or experimental deviation and inhabit very different problem-spaces because only one of them knows enough to recognize what does not fit.

This gives contingency a more precise role. Problems can arise from unplanned encounters without turning randomness into a method. A failed experiment, an unfamiliar social practice, or an unexpected regularity becomes a possible problem when someone can register its difference from expectation and begin to ask what would make sense of it. The ability to notice surprise already depends on formation.

2.3 Paradigms, Puzzles, and Anomalies

Thomas Kuhn adds a historical and disciplinary dimension. In *The Structure of Scientific Revolutions*, normal science is organized around shared achievements that provide exemplary ways of seeing and solving problems. Scientists trained within a paradigm do not encounter an unlimited field of equally available questions. Their training teaches them which puzzles are legitimate, what kinds of solutions count as successful, which instruments and techniques are appropriate, and what forms of discrepancy deserve sustained attention.[6] Normal science is productive precisely because many fundamentals are provisionally settled; researchers can work on specialized puzzles without reopening the conceptual basis of the field each morning.

Kuhn's account also shows why an anomaly is more than an unusual datum. A discrepancy becomes anomalous against expectations stabilized by a research tradition. Many discrepancies can be accommodated, postponed, attributed to experimental difficulty, or treated as local failures. Some become persistent and consequential enough to disturb confidence in an existing pattern of problem solving.[6] Their significance depends on where they occur within an established practice and on how a community comes to interpret them.

By the time a student enters a field, many prior judgments have therefore been sedimented into examples, canonical methods, standard exercises, review articles, benchmark datasets, and recognized open questions. Exemplary problem solutions teach more than propositional knowledge; they train researchers to perceive resemblance, relevance, and admissible moves within a disciplinary practice.[6] Such inheritance enables cumulative work. It also means that the apparent givenness of a problem-space is historically produced.

Kuhn's account also preserves the substantial problem-solving power of normal science, and his later clarifications resist the idea that paradigms make rational comparison impossible. The narrower implication is enough: recognition is mediated by learned expectations. A field makes some questions highly visible and leaves others difficult even to formulate.

2.4 Problem Judgment under Scalable Generation

Dewey, Peirce, and Kuhn differ in method, vocabulary, and philosophical commitment. Their differences matter. For the limited purpose of this essay, however, they converge on one useful correction to the ready-made problem: inquiry involves the formation, recognition, and stabilization of what later appears as the problem itself.

This correction becomes more consequential when generation is scalable. An AI system can produce hundreds of technically coherent research questions, hypotheses, or task variants. Such abundance weakens the old tendency to treat the mere existence of a difficult, well-formed question as evidence of significance. A problem can be difficult because it exposes a consequential structure; it can also be difficult because a large combinatorial obstacle has been constructed around it. The amount of work required for a solution and the worth of pursuing the problem are different judgments.

The distinction is simple but important: difficulty does not determine significance, and solvability does not determine whether a problem is worth pursuing. Difficulty can reveal the limits of a method; solvability can make an otherwise speculative programme tractable. Neither property, by itself, decides whether a problem deserves sustained attention. Once candidate problems can be produced in abundance, this insufficiency becomes harder to ignore. Here *salience* will name the more immediate fact that something stands out strongly enough to attract further attention, while *significance* names a broader appraisal of what a problem may reveal, connect, or change. The two can overlap without being identical.

Problem judgment therefore begins before the assessment of whether an answer is correct. It includes the judgment that an occurrence is worth noticing, that an indeterminate situation deserves articulation, that an anomaly is more than noise, that one formulation discloses more than another, and that a line of inquiry merits continuation. These judgments are fallible. Fashion, prestige, institutional incentives, disciplinary power, and inherited blind spots can shape them. Their fallibility makes their formation and criticism part of scholarship.

AI can participate in this process. Systems can detect outliers, compare observations with expected patterns, generate hypotheses from discrepancies, and rank candidate directions. Nothing in the argument requires problem judgment to remain an exclusively human property. What matters is the organization of the relation among expectation, event, formulation, testing, and revision. A system that produces many explanations may enlarge discovery; a system that merely proliferates plausible sentences may enlarge noise.

Selection, however, remains too general a word for the next step. Researchers do not continue with problems only because they expect high informational yield or practical return. Some questions become salient partly because they appear elegant, strange, deep, disproportionate, or unexpectedly simple. A mathematician may stay with a structure long before its usefulness is known; a physicist may be drawn to a symmetry; a political thinker may find in an image of order a form capable of organizing reflection beyond its original setting. Such responses do not by themselves establish a problem's broader significance or practical priority. They introduce another evaluative language into problem formation. The next section turns to the aesthetic judgment of problems.

3 Aesthetic Judgment of Problems

The preceding discussion explains how a possible problem can emerge from indeterminacy, surprise, and anomaly. It leaves another question open: why does one possible problem become salient while another recedes? Researchers often answer with a vocabulary that adds aesthetic terms to evidential and practical ones. A proof is elegant. A conjecture is beautiful. A symmetry is compelling. A question has depth. One formulation feels forced, while another seems to disclose the structure of the matter. Philosophical work on the aesthetics of mathematics treats such evaluations as elements of mathematical practice that can operate during reasoning itself.[9]

The proposal developed here is deliberately limited. Aesthetic judgment can participate in problem formation before the utility, tractability, or eventual significance of a problem is known. Beauty can contribute to salience while broader scholarly significance and overall worth remain open to other forms of judgment. Its philosophical interest lies earlier: aesthetic attraction can make an object of inquiry salient before the reasons for staying with it have been translated into another evaluative language. AI-generated abundance makes this older feature of inquiry easier to notice.

3.1 Reflective Judgment and the Absence of a Given Rule

Kant's account of reflective judgment offers a useful philosophical starting point because it distinguishes two situations in which judgment operates differently. In determining judgment, a universal is already available and a particular is subsumed under it. In reflecting judgment, by contrast, the particular is given while the relevant universal must be sought. Kant introduces this distinction in the *Critique of the Power of Judgment* and gives reflective judgment roles that extend beyond aesthetics, including the organization of empirical inquiry and the search for systematic order in nature.[7] The distinction is valuable here because problem formation often occurs under conditions closer to the second case than the first. A researcher can confront a phenomenon before possessing a settled concept of what kind of problem it is, which rule should govern it, or even which disciplinary vocabulary will eventually prove adequate.

The analogy has limits. A contemporary research problem differs from a Kantian aesthetic object, and the present essay draws from Kant only a structural resource for thinking about orientation when a rule is incompletely supplied in advance. Scalable AI generation can intensify precisely this situation. If a system produces a hundred plausible conjectures, hypotheses, model variants, or experimental directions, pre-existing criteria can rank some dimensions while leaving others open. Consistency, novelty relative to a corpus, expected cost, estimated information gain, and compatibility with available methods may all be formalized. The appearance of a problem as especially elegant, compelling, or revealing can still outrun the criteria that were available before the problem appeared.

Kant's aesthetics adds another distinction that matters for the present argument. His account places a judgment of beauty in a different relation to purpose from a judgment of the good. Beauty is associated with purposiveness while its judgment is left free of a specified external end that the object must serve.[7] The present essay uses this distinction only as a limited philosophical resource. It helps articulate a familiar research experience: a question can seem worth pursuing before the researcher can explain its eventual use. Its attractiveness can precede forecasts of applications, citations, economic return, and even a clear prediction of the theorem or theory that may emerge from it.

This gives the language of aesthetics a distinct place alongside the language of utility. Aesthetic judgment can have useful consequences and may guide attention toward structures that later prove theoretically fruitful. Its role in discovery often begins earlier, when researchers judge before they know what the judgment will yield. The possibility of being drawn to a problem without a determinate external purpose is therefore philosophically relevant to the formation of inquiry.

3.2 The Beauty of Problems

Mathematics provides unusually explicit testimony to this phenomenon. Henri Poincaré, reflecting on mathematical discovery, treated the feeling of mathematical beauty, harmony, and elegance as an aesthetic sensibility involved in the selection of combinations that rise to conscious attention.[8] His formulation is historically important because beauty appears inside the process of discovery and can precede the retrospective appreciation of a finished proof. Aesthetic sensibility can therefore participate in what is noticed, retained, and pursued.

Poincaré also moved from this selective role toward a stronger alignment between beauty and mathematical usefulness. The present essay need not follow him that far. In fact, the AI setting gives a reason to keep the concepts apart. A candidate problem may be difficult without being deep, useful without being beautiful, beautiful without an evident use, or practically valuable while aesthetically unremarkable. Contemporary philosophy of mathematics likewise treats mathematical beauty as a real subject of dispute precisely because it is not obviously reducible to a single epistemic property such as simplicity, explanatory power, or truth.[9] The persistence of the dispute is itself instructive: mathematical practice contains aesthetic evaluations, while the relation between those evaluations and other epistemic virtues remains contested.

This becomes more visible when difficulty can be manufactured or searched at scale. Once a system can vary assumptions, construct combinatorial instances, generate conjectures, or explore large formal spaces, difficulty loses some of its old rhetorical force. A hard problem can still matter enormously. The point is that hardness alone cannot tell us why one difficult problem deserves years of attention while another does not. The same is true of solvability. A problem may become easy after the right representation is found, yet the representation may reveal something profound. Another problem may resist attack for a long time because its formulation creates friction without exposing a structure that researchers judge illuminating.

Aesthetic language often appears at this threshold. Researchers describe an argument as clean, a connection as unexpected, a definition as natural, a construction as ugly, or a theorem as beautiful. These words do more than report pleasure. They can register a relation between parts and whole, an economy of means, a surprising connection, or a sense that a formulation reveals structure through the way it is expressed. These characteristics do not yield a complete theory of beauty, and different mathematical communities can disagree about them. Their plurality is compatible with the narrower point: judgment about a problem's aesthetic character can shape whether inquiry continues.

Curiosity belongs near this discussion. It is tempting to define curiosity informationally, as a desire to reduce uncertainty. That captures something important, yet scholarly curiosity often has a more selective texture. Researchers are surrounded by uncertainties they do not care to resolve. What distinguishes the uncertainty that becomes absorbing? Sometimes the answer is practical urgency, sometimes theoretical centrality, sometimes moral concern. At other times a structure simply appears compelling. One wants to know what is there, even before one can say what obtaining the answer would accomplish. In such cases curiosity carries an aesthetic dimension: the problem attracts attention through the way it appears, not merely through a calculation of expected return.

The AI era does not abolish this phenomenon. It may complicate it. A system can propose questions that no human researcher would have generated unaided. It may rank them by learned proxies for novelty, elegance, or promise. Human researchers may in turn acquire new tastes through repeated exposure to machine-generated structures. The relevant contrast is therefore not human aesthetic judgment against mechanical generation. A more interesting possibility is co-evolution: AI changes the space of available problems, and the changed space of problems alters what human communities learn to find elegant, natural, surprising, or worth pursuing.

3.3 Beauty before Full Symbolization

A further difficulty appears when aesthetic judgment is reduced to the application of the predicate *beautiful*. The word is already a symbol, and a highly compressed one. An encounter with a proof, a physical pattern, a political image, or a mathematical structure may contain attraction, surprise, balance, tension, strangeness, or a felt coherence before the person can give an adequate account of why attention has stopped there. The eventual sentence, “this is beautiful,” organizes the experience. The experience may have begun earlier, before that symbolic formulation became available.

The proposal can be stated cautiously: beauty may occur before the word “beauty” is spoken. Here *occur* does not mean that beauty has been established as a mind-independent property waiting inside an object. Nor does it mean that language plays no role in aesthetic experience. Languages, traditions, mathematical notations, artistic forms, and disciplinary educations deeply shape what can be noticed. The claim concerns sequence and remainder. Aesthetic manifestation may precede its adequate symbolization, and the subsequent symbol may fail to exhaust what made the encounter arresting.

This matters for problem formation because a problem can likewise begin before it is well named. A researcher may notice a mismatch, linger over an unexpected regularity, or sense that two domains fit together before possessing a clear research question. The sequence can therefore run from encounter to salience, from salience to a provisional judgment, and from judgment to a more explicit formulation. Symbolization stabilizes the problem and makes it shareable. It does not necessarily create the whole experiential salience of the encounter from nothing.

The distinction also prevents a simple identification of aesthetic competence with verbal competence. An agent that produces sophisticated descriptions of elegance may be predicting a community’s use of aesthetic vocabulary; an agent that never uses the word *beauty* may nevertheless organize attention through internal representations that function aesthetically in some meaningful sense. Current systems do not settle this philosophical question. They make it harder to assume that the answer can be read directly from linguistic fluency. If aesthetic judgment includes sensitivity to relations that are only later compressed into symbols, then the relevant question for AI becomes how such sensitivity is formed, coupled to experience, and revised.

This is also where the limits of a closed symbolic account become visible. Aesthetic judgment may have internal standards while particular aesthetic manifestations continue to exceed a finite vocabulary. The two kinds of closure should therefore be kept separate. A practice can be aesthetically reflexive: its judgments and modes of attention can themselves be criticized, refined, and defended in aesthetic terms. Symbolic completeness is a separate and stronger condition. The possibility of an aesthetic remainder makes a modest claim about the limits of available vocabulary and the field of experience that vocabulary organizes.

3.4 Aesthetic Reflexivity and Its Limits

A different evaluative question arises when aesthetic judgment is asked to justify what it is *for*, or whether scholarship could continue without it. Such questions are legitimate, yet they move the discussion into another grammar of evaluation. *Useful*, *necessary*, *productive*, and *socially important* belong to evaluative grammars distinct from *beautiful*. They can become decisive when institutions allocate resources or when inquiry affects others. Their force in those settings leaves aesthetic standing intelligible on its own terms.

The reflexive claim can be stated within the aesthetic language itself: the aesthetic judgment of problems can itself be aesthetically beautiful. A researcher may find beauty in an object of inquiry and also in the practice of remaining receptive to the beauty of possible problems. This is a local aesthetic claim with no automatic authority over research budgets, curricula, or public priorities. Aesthetic attention can therefore be intelligible within its own evaluative language without borrowing utility, productivity, civilizational survival, or generativity as its final warrant.

Such reflexivity also falls short of absolute self-sufficiency. A practice can possess internal norms while its actual formation depends on history, language, bodies, institutions, technologies, teachers, and encounters with the world. Self-norming does not entail self-causing. Mathematical taste is educated; scientific taste changes; new instruments disclose structures that earlier researchers could not have encountered; AI may introduce new objects and relations into the field of appreciation. Aesthetic autonomy is therefore compatible with historical and ontological embeddedness.

A deeper metaphysical possibility opens at this point, although the essay will leave it open. If beauty can manifest before it is fully symbolized, an aesthetic event can be pictured relationally. Observer, object, prior learning, cultural form, material setting, and historical moment can all participate in the conditions under which something appears as beautiful. One could extend this toward a relational account in which aesthetic manifestation is locally realized within a wider spacetime whole. That stronger proposal requires an ontology beyond the present argument. The weaker point is sufficient here: aesthetic judgment happens through relations that can change, and its occurrence exceeds the image of a free-floating symbol or a sealed private preference.

The same embeddedness marks the limit of aesthetic autonomy. Beauty does not decide whether public resources should be devoted to a project, whether a method is just, whether its risks are acceptable, or whether another need deserves priority. These questions call for other forms of judgment. Aesthetic judgment can remain irreducible to utility while accepting that it shares the world with epistemic, normative, economic, practical, and historical considerations.

Scholarship can therefore be attracted by beauty and still confront obligations that beauty alone cannot settle. A scientific question may be elegant while its institutional realization is prohibitively expensive. A theoretically illuminating experiment may impose unacceptable risks. A profound project may compete with urgent social needs. Once inquiry enters common life, the problem is no longer whether aesthetic judgment is aesthetically intelligible on its own terms. It is how several evaluative languages can meet without one silently becoming the measure of all the others. The next section turns to that mediation.

4 Plural Forms of Judgment and Their Mediation

Aesthetic judgment can help explain how a problem acquires aesthetic salience before its utility is known. It cannot by itself settle what should happen once inquiry enters a shared world of finite resources, institutional obligations, competing needs, and unequal exposure to risk. A beautiful scientific question may be expensive to investigate. A theoretically revealing intervention may be difficult to justify ethically. A programme with substantial epistemic promise may compete with other forms of research or with urgent public needs. At this point the language of beauty remains meaningful, but it no longer exhausts the judgment required.

Once inquiry enters common life, plurality appears at the level of the questions being asked. Aesthetic judgment concerns whether and how a problem, form, or relation appears beautiful, elegant, compelling, or otherwise aesthetically salient. Epistemic judgment concerns what can reasonably be learned, explained, or supported. Normative judgment asks what ought to be done or permitted. Economic and practical judgment considers costs, capacities, opportunity costs, timing, and implementation. Historical-contextual judgment situates a choice within inherited institutions and trajectories. These categories overlap, and they are not offered as a complete taxonomy. Their provisional separation serves a narrower purpose: to keep distinct evaluative questions visible before any attempt to translate them into a common metric.

4.1 Value Spheres and Plural Values

Max Weber's account of modernity provides one historical-philosophical route into this problem. In *Science as a Vocation*, and in related methodological writings, Weber describes a world in which science can clarify facts, means, consequences, and internal relations without supplying a single authoritative answer to ultimate questions of value. Modern disenchantment does not culminate in a universal scientific value-order. It leaves human beings confronting differentiated and sometimes antagonistic value spheres, including scientific, ethical, religious, and aesthetic commitments.[10] The significance of this diagnosis for the present essay lies less in accepting Weber's larger tragic view of modernity than in recognizing a limit to disciplinary translation. A scientific description can inform a practical choice without converting the ends of that choice into scientific facts.

This limit matters whenever scholarship moves toward public action. Research can estimate the probability that an intervention will achieve a stated objective, identify side effects, and compare alternative means. Such work can transform a decision. Yet the objective itself may remain contested, and several objectives may be worth pursuing at once. More accurate knowledge does not necessarily collapse these disagreements. It can sharpen them by making sacrifices visible.

Isaiah Berlin's value pluralism makes the point more explicit. Berlin rejected the monistic expectation that genuine values must ultimately fit together within one harmonious order. His pluralism allows genuine values to conflict, and his discussions of incommensurability resist the assumption that there is always a common currency in which their relative importance can be established in the abstract.[11] This view has generated substantial philosophical controversy, and the present essay adopts only a moderate consequence. Deliberation can proceed in the absence of one master measure, while value conflict deserves treatment as more than a hidden optimization problem waiting for the correct coefficients.

The phrase *languages of judgment* is a heuristic introduced by this essay. Weber and Berlin provide philosophical resources for the plurality at issue, while the phrase itself belongs to the present argument. Different evaluative practices can organize relevance differently. When a scientist says that a problem is beautiful, an economist asks about opportunity cost, and a public official asks about distributive consequences, they may be addressing the same project through different questions. Disagreement can occur within each language, and empirical mistakes remain possible. Plurality still permits criticism. A judgment can be internally intelligible while leaving other evaluative questions raised by the same action unanswered.

This distinction also guards against a reverse reduction. The argument of the previous section was that aesthetic judgment need not justify itself by proving its usefulness. The present section adds the complementary claim that aesthetic standing does not automatically confer practical priority. Local autonomy is compatible with cross-domain limitation. A judgment can possess reasons of its own and still require mediation when it enters a decision whose consequences extend beyond that domain.

4.2 Judgment of Particulars and a Shared World

Hannah Arendt's reflections on judgment provide a second bridge. Her theory of judgment remained unfinished, and it would be misleading to present it as a complete method for reconciling plural values. Nevertheless, her lectures on Kant's political philosophy develop a suggestive account of judging particulars in a shared world. Arendt drew especially on Kant's account of aesthetic judgment and on the idea of an enlarged mentality: judging requires an ability to consider a matter from standpoints other than one's own and to test one's judgment against a public horizon of possible communication.[12]

The relevance to the present argument is structural. Public judgment cannot be reduced to the private intensity with which a value is experienced. A researcher may be profoundly convinced of a project's intellectual beauty and epistemic promise. That conviction remains one perspective among others when the project requires common resources, changes institutional priorities, or distributes costs and benefits across people who do not share the same disciplinary commitments. Enlarging judgment does not mean replacing conviction with opinion polling. It means allowing the presence of others to alter what counts as an adequate consideration of the particular case.

This move also changes how one thinks about expertise. Expertise can deepen judgment by revealing consequences invisible to non-specialists, by distinguishing serious from superficial objections, and by clarifying the internal structure of a scientific opportunity. Yet expert knowledge is not automatically identical with public judgment. The latter must also ask how specialized reasons relate to other legitimate concerns. Conversely, public deliberation that ignores technical structure can misdescribe the choice it claims to judge. The problem is therefore reciprocal: expert judgment requires exposure to a shared world, while public judgment requires enough understanding to avoid treating scientific possibilities as interchangeable symbols.

Arendt's account is especially useful after the discussion of aesthetic judgment because it shows a path from taste to plurality without turning aesthetic judgment into political command. Aesthetic judgment, in her reading of Kant, becomes philosophically interesting partly because it reaches beyond merely private preference toward communicability and the imagined standpoints of others.[12] The present essay extends this insight cautiously. Different forms of judgment need not become identical in order to encounter one another. Mediation can begin from the ability to recognize that a particular decision appears differently when viewed from another evaluative position.

4.3 Scientific Beauty and Public Choice

A public controversy over a proposed next-generation particle collider in China illustrates the distinction. In 2016, Chen-Ning Yang published an essay arguing that China should not begin building an ultra-large collider at that time. His argument did not reject the continuation of high-energy physics. It raised questions about construction costs, the opportunity costs for other basic sciences, China's contemporary development needs, the expected near- and medium-term benefits to ordinary life, and the uncertainty of the relevant particle-physics expectations.[13] Yang also identified alternative directions in high-energy physics that he regarded as scientifically worthwhile.

Yifang Wang, then director of the Institute of High Energy Physics and a leading advocate of the Circular Electron Positron Collider project, publicly disagreed. His response contested several empirical and financial premises, defended the scientific case for the collider, argued that the proposed costs were manageable, and presented the project as an opportunity for China to advance at the frontier of high-energy physics and related technologies.[14] The subsequent CEPC programme continued to develop its conceptual design, with a formal Conceptual Design Report released in 2018.[15]

The philosophical use of the example does not depend on deciding which side was correct. The dispute included disagreements about facts, forecasts, costs, scientific prospects, and institutional capacity. It also displayed different judgments about how such considerations should bear on a public decision. A project can be judged scientifically beautiful and epistemically promising while still raising questions about timing, opportunity cost, distributive priority, and institutional readiness. Social need, in turn, does not automatically defeat the case for a large scientific project. The lesson is that no single judgment named in advance settles the whole case.

The collider controversy belongs here because it makes visible a transition from an internal scientific question to a public decision. Within the field, researchers ask what structures are worth investigating, what an experiment could reveal, and which theoretical possibilities are compelling. At the level of public action, a society must allocate resources among heterogeneous claims. The internal scientific judgment remains present and enters a wider field of judgment alongside other considerations.

The same structure appears at smaller scales. A laboratory chooses between an elegant high-risk experiment and a less glamorous validation study. A university chooses between supporting frontier research and maintaining teaching infrastructure. A government decides whether to fund a mission whose scientific returns are uncertain but potentially transformative. In each case, asking only which project is most beautiful, most publishable, cheapest, or most immediately useful can obscure the plural structure of the decision. The relevant judgments do not become identical merely because an institution must eventually act.

4.4 Mediation without Reduction

The language of *mediation* is intended to preserve this plurality. Mediation is different from merely adding considerations to a common score. A weighted model can be useful when variables are commensurable enough for the purpose at hand, and many practical decisions appropriately use such models. The philosophical mistake would be to infer from the usefulness of a decision model that the underlying values have therefore been fully translated into one substance.

Consider the difference between saying that a research programme has a certain financial cost and saying that its beauty, justice, epistemic promise, and social urgency have each been assigned a monetary equivalent. The first can be indispensable. The second is a much stronger claim. Similar problems arise when a single notion of impact is asked to absorb scientific novelty, educational value, institutional prestige, public welfare, environmental effects, and distributive justice. Measurement can support judgment while simultaneously hiding the loss incurred by translation.

Mediation therefore requires at least three acts. The first is differentiation: participants must identify which kind of judgment is being made and what it can legitimately establish. The second is translation: reasons formed in one evaluative language must become sufficiently intelligible to participants situated in others. The third is practical settlement: because action cannot wait for perfect philosophical reconciliation, a particular decision must eventually establish priorities, compromises, limits, or postponements. None of these acts guarantees harmony. Some choices leave genuine remainder and regret because pursuing one value sacrifices another that was also real.

This account gives a more demanding meaning to judgment in the age of AI. An optimization system can return a highly precise recommendation while leaving implicit the evaluative language in which the recommendation is optimal. A model can rank projects by expected publication impact, predicted social benefit, cost-effectiveness, probability of technical success, or some composite objective. Each ranking can be technically sophisticated. A prior question remains: *optimal in which language of judgment?*

That question asks optimization to make its evaluative and epistemic framing visible. Once the framing is visible, AI can participate in mediation through more than the pursuit of a single metric. It can expose trade-offs, generate alternative formulations, model consequences under different objectives, and reveal cases in which conclusions are sensitive to value assumptions. It may even help human communities discover evaluative patterns that they had not articulated clearly. Multiple objectives still leave open whether any final meta-objective can absorb them without remainder.

Judgment, on this account, includes the capacity to recognize heterogeneous forms of judgment, to understand what each reveals and leaves out, and to mediate among them in the particulars of action. A universal value function is one possible formal device for some decisions, while the broader practice of judgment may exceed it. Such judgment can be human, institutional, AI-assisted, or eventually partly machine-generated. Its quality depends on preserving distinctions that keep one successful language from quietly becoming the measure of all the others.

This conclusion returns the essay to AI from a different direction. The first sections considered AI as a generator and solver of scholarly problems. The present section adds a further layer: AI increasingly enters the processes by which problems, projects, and priorities are ranked. The next section therefore asks how AI should be understood within an ecology of judgment in which problem generation, solution, evaluation, and mediation may all be distributed across human and machine agents.

5 AI within the Ecology of Judgment

The preceding sections treated judgment as part of scholarship before asking who, or what, performs it. That order matters. Once judgment is understood as selection among candidate directions, recognition of salience, evaluation through several languages, and mediation among conflicting considerations, it becomes difficult to reserve the category in advance for human beings alone. Contemporary AI systems already perform operations that resemble parts of this structure. They generate alternatives, compare them, rank them, criticize them, revise them, and use experimental or documentary feedback to alter subsequent proposals. Whether these operations amount to judgment in every philosophical sense remains open. Their practical significance is already enough to complicate a picture in which machines produce possibilities while humans alone decide what they mean.

A more revealing question is therefore structural. What kind of judgment is being performed, through which criteria, with what access to the world, and with what capacity for revision? An AI system that ranks hypotheses according to scientific plausibility participates in a different evaluative practice from one that ranks projects according to expected social benefit. A system that can revise a hypothesis after laboratory evidence occupies a different place in inquiry from a model that only scores text. The phrase *ecology of judgment* is used here for this distribution of evaluative activity across people, institutions, models, instruments, inherited standards, and encounters with the world. It does not presuppose equality among these participants. It draws attention to their relations.

5.1 AI as a Participant in Judgment

Recent scientific AI systems make the transition from generation to selection visible. Co-Scientist does more than produce a list of hypotheses. Its architecture includes agents that critique proposals, conduct pairwise comparisons, organize an Elo-based tournament, rank hypotheses, and iteratively evolve top candidates. Expert scientists can intervene with reviews and feedback that alter subsequent ranking and refinement.[2] Robin similarly generates multiple therapeutic candidates and then evaluates and ranks them before experimental testing. In the reported system, candidate drugs were compared through an LLM-judged tournament using criteria that included scientific rationale, pharmacological profile, and the methodology of the supporting literature; human scientists then reviewed the ranked list before laboratory testing.[3]

These examples are modest compared with the philosophical breadth of judgment discussed earlier. They operate within specified research goals and structured criteria. Yet precisely for that reason they are instructive. They show that the boundary between "generation" and "judgment" is operationally porous. Once a system must decide which hypothesis deserves more computation, which candidate deserves an experiment, or which interpretation should guide the next cycle, evaluation becomes part of the generative process itself. A scientific agent cannot explore indefinitely. Scarcity re-enters inside the machine as compute budgets, context limits, experimental cost, ranking thresholds, and selection rules.

This point also weakens an easy human exceptionalism. If judgment were defined only as ranking alternatives under explicit or learned criteria, current systems already perform limited forms of it. If judgment requires reflective awareness, aesthetic manifestation, responsibility, embodiment, or participation in a shared world, the question becomes more demanding and may remain unsettled. A final metaphysical definition is unnecessary for the present argument. AI is already moving from assistance at the edges of inquiry toward participation in the evaluative loops through which inquiry allocates attention.

The development also clarifies why the language of "human in the loop" can be too coarse. In Robin, for example, humans execute laboratory experiments and review selected candidates, while AI systems search literature, propose hypotheses, judge pairs of candidates, analyse data, synthesize interpretations, and generate new hypotheses.[3] The meaningful issue is therefore not whether a human appears somewhere in the loop. It is which judgments are delegated, which remain shared, which are checked only after the fact, and which criteria organize the handoffs among agents.

5.2 Criteria and the Shape of Judgment

Every ranking system makes a field of relevance more explicit. The Robin tournament did not rank candidate therapies according to an abstract notion of "goodness." It used a particular scientific rationale, pharmacological considerations, and evidence from the literature.[3] Co-Scientist's generation and ranking mechanisms likewise operate relative to a research goal supplied by scientists and to published criteria including alignment with that goal, plausibility, and novelty; its reflection agent also critiques correctness, quality, novelty, and explanatory potential.[2] These systems therefore illustrate, in concrete engineering form, a point already visible in the philosophical discussion: an evaluative result cannot be separated from the language in which relevance has been organized.

This does not make machine judgment uniquely suspect. Human evaluators also carry criteria, habits, institutional incentives, disciplinary tastes, and cognitive biases. Research on LLM-as-a-judge systems is useful here because it prevents a romantic contrast between neutral human judgment and biased machine judgment. Chen and colleagues found that both human and LLM judges were susceptible, to differing degrees, to perturbations associated with several forms of bias in their evaluation setting.[16] The result does not establish equivalence between human and machine judgment. It supports a narrower conclusion: evaluative reliability depends on the structure of judging practices, not merely on the biological or artificial identity of the judge.

For scholarship, a more revealing question than "Can AI judge?" is "What has been made judgeable?" A system can rank hypotheses only after some combination of prompts, training, tools, criteria, comparators, reward signals, and institutional goals has made certain features legible to it. The same is true, less formally, of a trained researcher. Disciplinary education teaches what counts as elegant, rigorous, significant, feasible, publishable, or urgent. AI makes this architecture easier to inspect because the criteria are often partially encoded in prompts, scoring procedures, tournaments, reward functions, and data pipelines.

This visibility can be an advantage. A hidden evaluative convention can become discussable once it is operationalized. An institution that asks an AI system to rank research proposals may be forced to specify whether novelty, feasibility, cost, social benefit, scientific depth, or strategic priority should matter. The act of specification can expose conflicts that were previously concealed inside committee habits or disciplinary prestige. Yet formalization can also create a false sense of completeness. A criterion becomes easy to optimize once it has been named, while qualities that resist current representation may disappear from the decision simply because they are harder to encode.

The earlier discussion of aesthetic judgment becomes relevant here. If the appearance of a problem as beautiful can precede a full account of why it is worth attention, then a ranking system built only from already articulated criteria may miss precisely those encounters in which a new evaluative language is beginning to form. The issue is not that AI must remain incapable of such encounters. Future systems may develop richer perception, longer histories, embodied interaction, or internal representations that support unfamiliar forms of salience. The issue is that criteria which are adequate for yesterday's judgments can become filters against tomorrow's problems.

5.3 Revisability and World-Coupling

A promising way to compare forms of judgment is to ask how they can be disturbed. A judgment system that only reorders a fixed database according to a fixed score can be highly effective within its domain and still remain closed to events that do not fit the score. By contrast, scientific inquiry often advances when an experiment resists expectation, when a field observation does not fit a category, or when a practical intervention produces consequences absent from the model. For inquiry that remains open to revision, judgment therefore requires more than repeated application of a criterion. There must be some route by which the criterion, the problem formulation, or the field of relevant objects can itself become revisable.

Robin offers a limited but concrete example of such coupling. Experimental results are returned to the system, analysed through multiple trajectories, synthesized, and used to generate revised hypotheses and proposed follow-up assays.[3] Co-Scientist likewise supports interaction with expert feedback and external tools during iterative hypothesis refinement.[2] These systems are carefully engineered workflows with bounded roles. Their design still shows why the distinction between a model and its environment is philosophically consequential. An AI system connected to experiments, instruments, human criticism, and institutional consequences inhabits a different ecology of judgment from one confined to static text prediction.

World-coupling need not mean possession of a human-like body. The relevant capacity is broader: a judging practice that is to remain open must have some route by which it can encounter resistance from something not exhausted by its prior representations. Laboratory data can provide such resistance; so can a field site, another person, a new archive, a sensor, a legal decision, a market shock, or a political event. The form varies by domain. What matters is that an unexpected event can alter an answer and, when necessary, the problem or the criteria by which answers are judged.

This suggests a stronger formulation of revisability. In some contexts, judgment includes changing a ranking after receiving additional evidence and also reconsidering the objective, the language of evaluation, or the boundary of the problem. Human beings achieve such revisability unevenly. Institutions can become rigid; disciplines can protect canonical puzzles; funding systems can reward what they already know how to count. AI can reproduce these closures, amplify them, or sometimes help expose them. The more informative comparison therefore lies among judgment ecologies with different degrees of susceptibility to relevant disturbance and different capacities to revise inherited criteria.

5.4 The Pre-Organized Problem World

AI can make a field extraordinarily easy to navigate. A learner or researcher can ask for the central debates, canonical papers, unresolved questions, methodological options, promising gaps, likely objections, and candidate research designs within minutes. Such maps are valuable precisely because scholarship has become too large for any individual to survey unaided. They lower the cost of orientation and can connect literatures that disciplinary training keeps apart.

The philosophical risk lies in the success of the map. A fluent and comprehensive representation of a field can make its boundaries appear natural. Research gaps generated from existing literature are gaps inside a literature. Hypotheses generated from indexed evidence begin from what has already become retrievable. Ranking mechanisms prioritize what can be compared under available criteria. Even systems capable of unusual combinations still operate through some representational and evaluative history. None of this renders their proposals derivative in a simple sense. It means that a highly capable model can organize an inherited problem-space so well that the organization itself becomes difficult to see.

The same danger existed long before AI. Textbooks, journals, departments, grant panels, and professional societies have always organized visibility. AI changes the scale and intimacy of that organization. A student can now encounter a field first through an AI-generated synthesis, sometimes before years of dispersed reading have built a personal map of the literature. A researcher can ask an agent to propose the "most promising" directions before personally encountering the archive, laboratory, community, or landscape from which a different problem might emerge. The convenience is real. So is the possibility that the inherited map arrives before the world has had much chance to interrupt it.

This does not justify a romantic return to unmediated experience. There is no view from nowhere, and the previous discussion of Dewey, Peirce, Kuhn, and the classics already showed that prior concepts are conditions for noticing many anomalies at all. The educational problem is therefore one of circulation. AI can help organize inherited knowledge; inherited knowledge can make contingencies intelligible; encounters with the world can reveal where both the inheritance and the model have become too narrow; renewed judgment can then revise the problem-space.

Seen in this way, the evolution of scholarship under AI is a reconfiguration of the ecology in which judgments are generated, compared, delegated, challenged, and revised. Human and machine judgment increasingly coexist within that ecology. AI can enlarge the number of candidate problems and candidate solutions, and it can also enlarge the number of candidate evaluations. The next educational question is therefore unavoidable. If learners increasingly enter scholarship through systems that can already generate, solve, rank, and explain, what kind of formation allows them to participate in judgment and critically examine the outputs they inherit?

6 The Education of Judgment

The educational question changes once problems are treated as historically formed objects of inquiry. A curriculum can teach students to solve problems that a discipline has already stabilized. It can also teach them to evaluate evidence, identify fallacies, compare explanations, and use increasingly capable AI systems. These remain important achievements. Yet a learner who becomes excellent only after the problem has been specified may still lack experience in the earlier movement of inquiry. That movement includes noticing that something has become strange, deciding that an interruption deserves attention, determining which language of judgment is relevant, and allowing an inherited problem-space to be revised.

The educational problem is therefore a problem of circulation. Books, disciplines, and classics are among the conditions that make many experiences intelligible in the first place. Yet inherited knowledge can become so coherent that its selected questions begin to look like the natural inventory of reality. Judgment grows when inheritance and encounter remain able to disturb one another. AI can widen access to texts, comparisons, explanations, and candidate interpretations; it can also place a pre-organized map of a field before the learner so early and so fluently that little time remains for what the map leaves unarticulated.

6.1 Books as Sedimented Judgment

A textbook presents problems in a deceptively finished form. Definitions have been stabilized, examples selected, notation regularized, and exercises placed after the concepts required to solve them. A classic presents a different kind of compression, but it too reaches the reader only after some questions, distinctions, arguments, and forms of attention have been preserved while innumerable alternatives have disappeared from view. Here *classics* refers broadly to enduring works and exemplary texts through which a field transmits concentrated histories of inquiry; it does not imply a fixed or closed canon. The page therefore contains more than information. It contains the residue of earlier judgments about what deserved formulation, continuation, transmission, and dispute.

This makes books indispensable and potentially misleading in the same movement. They are indispensable because no learner can personally reenact the full history through which a field acquired its distinctions. A student of physics does not need to rediscover conservation laws from unaided observation; a student of political thought need not encounter every historical crisis before learning the conceptual possibilities opened by earlier thinkers. Inherited knowledge concentrates judgments made across lives that the learner cannot repeat.

Yet compression conceals its own history. Once a problem has been named, formalized, and placed inside a discipline, it can appear to have been waiting there all along. The learner sees the surviving problem while the uncertain encounter from which it emerged recedes from view. The result can be an epistemic optical illusion: the world begins to resemble the table of contents. Scholarship then appears to consist primarily in moving through an already articulated list of questions.

The discussion of Dewey and Kuhn earlier in this essay gives reasons to resist that illusion. Dewey's pattern of inquiry begins from an indeterminate situation in which a problem still has to be formulated, and that formulation can itself be revised as inquiry proceeds.[4] Kuhn's account of anomaly similarly shows that a phenomenon becomes scientifically anomalous only against a sufficiently articulated field of expectation; novelty is recognizable because researchers have learned what ordinarily ought to happen.[6] The inherited map supplies distinctions through which some departures become visible, while its partial character leaves room for discovery beyond what it already organizes.

An education of judgment should make this double character of books explicit. Students need to learn the answers a tradition preserves and also ask how its questions became visible, what forms of judgment selected them, and what kinds of experience might still lie outside their present vocabulary. The aim is a form of inheritance that remains conscious of its own incompleteness.

6.2 Tradition, Horizon, and Encounter

Gadamer's philosophical hermeneutics offers a useful vocabulary for this relation between inheritance and openness. In *Truth and Method*, understanding is historically situated through prior involvements and tradition. Gadamer's account of historically effected consciousness and horizon emphasizes that what can be understood is conditioned by inherited concepts and histories that the interpreter never fully stands above.[17] The horizon remains transformable. Encounter with what is unfamiliar can place inherited prejudgments in question and alter the horizon within which understanding proceeds.[17]

This structure matters for education because it avoids two symmetrical fantasies. The first imagines that genuine discovery requires escaping inherited knowledge so that the learner can meet the world without mediation. Such a learner would often be unable to recognize the significance of what is encountered. The second imagines that enough inherited knowledge can eventually make the world fully legible in advance. This turns tradition from a condition of understanding into a substitute for encounter.

Judgment develops between these positions. Prior knowledge allows a learner to notice a deviation, connect an event with a longer history, distinguish a familiar pattern from a meaningful exception, and compare several interpretations. Encounter supplies the possibility that those very distinctions may become inadequate. The learner therefore does not alternate between knowledge and ignorance. The more interesting circulation is between an inherited horizon and events capable of changing it.

This is also why the language of "contingency" in this essay should not be confused with randomness alone. A contingency may be accidental with respect to a plan and still become meaningful through judgment. A failed experiment, an unexpected institutional practice, an unfamiliar landscape, a conversation across cultures, or an observation that does not fit a textbook distinction can become intellectually generative because the learner possesses enough background to feel the mismatch. The same event may remain noise to someone with no relevant background and remain invisible to someone whose background has hardened into expectation.

6.3 Classics and Contingency

The educational implication is a productive reciprocity between classical learning and encounters that cannot be fully scheduled. Classics provide concepts, historical depth, examples of judgment, and access to questions refined across generations. They also disclose how earlier thinkers transformed particular experiences into objects of inquiry. Their value exceeds the answers they contain: they educate perception by showing what earlier minds learned to notice.

Contingent encounters perform a different role. They remind the learner that inherited visibility is selective. A field site can expose categories that fit poorly; an experiment can resist an elegant theory; a legal or political institution can behave differently from its formal description; a conversation can reveal that a supposedly universal problem was framed from a narrower social experience. Such encounters need not overthrow a tradition. Often they deepen it. A classic reread after an unfamiliar experience may become a different text because the reader now brings a different question to it.

The relation can be expressed without turning it into a formula: classics teach us how to see; contingency reminds us that the world contains more than what the classics have already seen. The first half guards against romanticizing immediacy. The second guards against mistaking accumulated judgment for a completed world.

This reciprocity also helps clarify what judgment education adds to many familiar exercises in critical thinking. Such exercises often begin with a claim, argument, source, or proposed solution already in view. Judgment in the present discussion includes those capacities but also reaches earlier: it includes noticing that a situation deserves inquiry, recognizing which forms of judgment are being used, asking which ones have been omitted, and determining when an inherited problem formulation itself requires revision. Reasoning can be rigorous within a poorly chosen problem. Education therefore has reason to cultivate sensitivity to the formation and transformation of the problem-space alongside rigor within it.

6.4 AI in a Circulation of Learning

AI changes this educational circulation because it can now occupy several positions within it at once. It can retrieve and summarize inherited knowledge, compare traditions, generate counterarguments, propose candidate problems, simulate alternative explanations, assist with coding and proof, and organize large bodies of scholarship. These capacities can make the encounter between a learner and a tradition far more expansive than was previously practical. A student can cross linguistic and disciplinary boundaries quickly enough to notice relations that conventional curricula keep far apart.

The same capacities create a temptation to let organization precede encounter. An AI system can tell a student what the "main questions" of a field are before the student has read much of the field. It can summarize a classic before the difficulty of the text has had time to form a question. It can also propose research gaps before the learner has encountered the archive, laboratory, institution, or community in which a different gap might become salient. The problem is not that these services are inaccurate by definition. Many will be excellent. Their very excellence can make the inherited organization of knowledge feel complete.

A more promising educational use of AI is therefore circulatory. Inherited knowledge can be explored with AI, while encounters with the world can produce anomalies or questions. AI can then help search whether similar questions have histories, compare interpretations, formalize possibilities, or expose counterexamples. Judgment can test which directions deserve continuation and which evaluative languages must enter; renewed study can return the learner to both the literature and the world. No element in this circulation needs to be treated as sovereign.

Such an education would also change what it means to teach AI literacy. Knowing how to prompt, verify, cite, or detect model errors remains necessary. Judgment education adds a different question: what has the system already decided to make visible? A generated answer carries a framing of the problem, a selection of relevant evidence, and often an implicit ordering of values. Learners should be able to inspect that framing, compare it with other languages of judgment, and recognize when an answer is excellent relative to a problem that itself deserves reconsideration.

This educational task does not preserve a human monopoly over judgment. AI systems may become increasingly capable of recognizing anomalies, revising criteria, developing unfamiliar aesthetic preferences, and participating in the mediation described in the preceding sections. The educational reason to cultivate human judgment is therefore more modest and more practical. If people are to remain participants in the evolution of scholarship, they need capacities for recognizing, contesting, and reshaping the problem-spaces in which human and artificial agents increasingly work together.

The result is an education oriented toward participation in a changing ecology of inquiry. Inherited knowledge can orient attention without closing it; AI can expand inquiry without defining its horizon in advance; encounters with the world can still revise what counts as a problem. This brings the essay back to its larger concern: how scholarship can remain open to a world that exceeds every current map of it.

7 Scholarship Open to the World

The evolution of scholarship under AI is often described as a story of substitution: tasks once performed by researchers are delegated to machines, and the frontier between human and artificial labour moves. A deeper reorganization is also underway. When candidate questions, candidate solutions, comparisons, proofs, syntheses, and evaluations can all be produced in greater abundance, the structures that orient inquiry become more visible. Scholarship is shaped by what can be generated or solved and also by what receives attention, which formulation survives, which criteria count, and what is allowed to interrupt them.

Judgment has therefore appeared throughout this essay without being reserved as a protected human remainder. AI can already rank, criticize, revise, and recommend; future systems may participate in richer aesthetic, epistemic, practical, or public judgments. The more durable issue is openness. An inquiry can become extraordinarily efficient inside a problem-space while becoming less sensitive to events that do not fit it. It can also remain revisable, allowing inherited concepts, model-generated maps, institutional priorities, and its own criteria of relevance to be disturbed by what they had not organized in advance.

Problem formation makes this openness concrete. Mature disciplines rightly stabilize questions, methods, and examples so that each generation need not begin again. Yet the resulting clarity can hide its own history. A problem-space is an achievement of attention and inheritance, not an inventory of everything that can become visible. AI can render that inheritance with unprecedented fluency. Its maps may be excellent and still remain maps.

Aesthetic judgment offered one account of how a possible problem can acquire salience before its utility is known. Public action then revealed why one evaluative language is insufficient: beauty, evidence, responsibility, cost, urgency, justice, timing, and institutional capacity can all bear on the same project without becoming interchangeable. Judgment therefore requires recognition of their claims, their limits, and the points at which action requires mediation among them. Such mediation preserves differences while still allowing practical settlement.

Education matters because openness is learned as well as declared. Classics and disciplinary traditions provide distinctions through which an anomaly can be recognized as more than noise. Encounters with experiments, institutions, landscapes, archives, other cultures, failures, and unforeseen events remind the learner that inherited visibility is selective. AI can enrich this circulation by widening access, comparing distant literatures, generating counterexamples, and exposing assumptions. It can also make an inherited problem-space feel complete before the world has had much chance to resist it.

The educational aim proposed here is continued participation in the formation of inquiry. Difficulty has value when it serves learning, discovery, or discipline, and human judgment remains one participant in an increasingly mixed ecology of inquiry. A learner should be able to inherit questions while keeping their boundaries revisable, use AI while distinguishing its organization of knowledge from the world it represents, and recognize when a polished answer belongs to a problem whose framing still deserves judgment.

The age of AI may be judged, in part, by what kinds of scholarship remain possible after generation and solution become easier to scale. Speed can narrow a horizon, or it can release time for wider comparison and deeper encounter. Abundance can harden inherited criteria, or it can expose them to more alternatives than a researcher could previously entertain. No technical property of AI decides between these possibilities on its own. Their direction depends on how systems of inquiry remain connected to histories they inherit, judgments they contest, and events they did not foresee.

The world does not arrive already divided into the problems that appear in books, benchmarks, grant calls, or model outputs. Those problems deserve serious inheritance precisely because they carry concentrated histories of attention and judgment. They also deserve exposure to what can still change them. An education of judgment may begin there: to inherit questions without mistaking them for the limits of the world, and to remain available to the possibility that something not yet named can still become a problem.

References

[1] Thomas Hubert, Rishabh Mehta, Laurent Sartran, et al. Olympiad-level formal mathematical reasoning with reinforcement learning. *Nature*, 651:607–613, 2026. doi:10.1038/s41586-025-09833-y.

[2] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, et al. Accelerating scientific discovery with Co-Scientist. *Nature*, 655:487–496, 2026. doi:10.1038/s41586-026-10644-y.

[3] Ahmed E. Ghareeb, Benjamin Chang, Liam Mitchener, et al. A multi-agent system for automating scientific discovery. *Nature*, 655:497–505, 2026. doi:10.1038/s41586-026-10652-y.

[4] John Dewey. *Logic: The Theory of Inquiry*. Henry Holt and Company, New York, 1938.

[5] Charles Sanders Peirce. *Collected Papers of Charles Sanders Peirce*. Charles Hartshorne and Paul Weiss, eds., vols. 1–6; Arthur W. Burks, ed., vols. 7–8. Harvard University Press, Cambridge, MA, 1931–1958.

[6] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. 4th ed., 50th Anniversary Edition, with an introductory essay by Ian Hacking. University of Chicago Press, Chicago, 2012.

[7] Immanuel Kant. *Critique of the Power of Judgment*. Paul Guyer, ed. and trans.; Eric Matthews, trans. Cambridge University Press, Cambridge, 2000. doi:10.1017/CBO9780511804656.

[8] Henri Poincaré. *Science and Method*. Francis Maitland, trans., with a preface by Bertrand Russell. T. Nelson and Sons, London, 1914.

[9] Angela Breitenbach and Davide Rizza. Introduction to Special Issue: Aesthetics in Mathematics. *Philosophia Mathematica*, 26(2):153–160, 2018. doi:10.1093/philmat/nkx019.

[10] Max Weber. *The Vocation Lectures: Science as a Vocation; Politics as a Vocation*. David Owen and Tracy B. Strong, eds.; Rodney Livingstone, trans. Hackett Publishing Company, Indianapolis, 2004.

[11] Isaiah Berlin. *The Crooked Timber of Humanity: Chapters in the History of Ideas*. Henry Hardy, ed. John Murray, London, 1990.

[12] Hannah Arendt. *Lectures on Kant's Political Philosophy*. Ronald Beiner, ed. University of Chicago Press, Chicago, 1982. doi:10.7208/chicago/9780226231785.001.0001.

[13] Chen-Ning Yang. China Should Not Build an Ultra-Large Collider Today [中国今天不宜建造超大对撞机]. *Caixin*, September 5, 2016; originally published in *The Intellectual* [知识分子]. `https://china.caixin.com/2016-09-05/100985407.html`.

[14] Yifang Wang. Now Is the Time for China to Build a Large Collider [中国建造大型对撞机，今天正是时机]. September 4, 2016; reproduced by the Institute of High Energy Physics, Chinese Academy of Sciences. `https://www.ihep.cas.cn/dkxzz/cepc/cmsm/201610/t20161020_4682605.html`.

[15] Institute of High Energy Physics, Chinese Academy of Sciences. CEPC Design Report Released. November 2018. `https://english.ihep.cas.cn/nw/han/y18/201811/t20181114_298193.html`.

[16] Guiming Hardy Chen, Shunian Chen, Ziche Liu, Feng Jiang, and Benyou Wang. Humans or LLMs as the Judge? A Study on Judgement Bias. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 8301–8327, 2024. doi:10.18653/v1/2024.emnlp-main.474.

[17] Hans-Georg Gadamer. *Truth and Method*. 2nd revised ed., Joel Weinsheimer and Donald G. Marshall, trans. Continuum, London, 2004.