Journal of Contemporary Educational Research, 2026, Volume 10, Issue 3
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Design Principles for AI-Enhanced Labor Education Evaluation Systems in Higher Vocational Colleges

Yuntao Wang[1], **Suhan Wu**[2]*

[1]Nanjing Polytechnic Institute, Nanjing 210048, Jiangsu, China
[2]School of Economics and Management, Nanjing Polytechnic Institute, Nanjing 210048, Jiangsu, China

Corresponding author: Suhan Wu, wusuhan@link.cuhk.edu.hk

Abstract: Artificial intelligence (AI) is becoming routine infrastructure in higher vocational colleges, while labor education increasingly serves holistic talent cultivation through authentic tasks and practicum experiences. However, evaluating labor education is challenging because intended outcomes combine competence and occupational dispositions, and evidence is heterogeneous, process-oriented, and distributed across sites. This paper reframes labor education evaluation as a system design problem rather than a single-instrument measurement task. It proposes a layered conceptual model that separates constructs, evidence, analytics, decision procedures, and governance, and provides a minimal evidence taxonomy to support triangulated and longitudinal interpretation. Building on this foundation, the paper consolidates eight design principles and maps them to implementable system requirements and assurance considerations, emphasizing authentic-task anchoring, stake-sensitive human-in-the-loop decisions, traceability and explainability proportional to stakes, context-aware fairness, purpose-limited data practices, contestability, and continuous monitoring. The framework offers actionable guidance for designing AI-enhanced labor education evaluation systems and identifies directions for further research on construct operationalization, evidence integration, and governance-by-design.

Keywords: AI-enhanced evaluation; Labor education; Higher vocational education; Principle design

---

Online publication: April 7, 2026

1. Introduction

Artificial intelligence (AI) is rapidly moving from experimental classroom use to routine infrastructure in higher vocational colleges, supporting teaching, learning services, and institutional decision-making [1]. In parallel, labor education has become a salient component of holistic talent cultivation in vocational settings, where students are expected to develop work habits, practical competencies, and occupational dispositions through authentic tasks, productive activities, and practicum experiences [2]. This convergence creates a timely design challenge: labor education requires evaluation systems that can document development, provide formative guidance, and enable fair and accountable decisions, yet its evidence is typically dispersed, process-oriented, and context-dependent, making conventional score-centric evaluation insufficient.

In many higher vocational contexts, labor education evaluation faces three structural tensions. First, the target construct is often underspecified. Labor education may encompass attitudes toward work, responsibility norms, safety and quality awareness, teamwork, perseverance, and task-specific operational competence, but these dimensions are not consistently operationalized into explicit indicators and rubrics. Second, the evidence base is heterogeneous and difficult to standardize: artifacts, process logs, reflections, observations, peer feedback, and enterprise mentor comments are informative but challenging to compare and audit. Third, evaluation is embedded in multi-stakeholder governance, where teachers, counselors, enterprise mentors, and administrators may interpret standards differently, while students require transparency, feedback, and contestability. These tensions indicate that labor education evaluation is a system-level problem rather than a purely measurement issue [3].

AI-enhanced evaluation systems are frequently proposed to address these challenges by aggregating multi-source evidence, supporting portfolio structuring, generating timely feedback, detecting missing documentation, and assisting rubric-based judgment [4]. However, AI can also introduce new risks if deployed without robust design and governance. Over-automation may dilute teachers' professional judgment; opaque models can weaken contestability; biased data or proxy variables may disadvantage particular groups; and expanded data collection increases privacy and security exposure. AI, therefore, does not automatically improve evaluation quality; it amplifies the assumptions embedded in constructs, evidence pipelines, decision workflows, and institutional controls.

Despite growing work on AI in assessment and learning analytics, translating the aims of labor education into AI-supported evaluation systems that remain valid, improvement-oriented, and trustworthy requires a coherent integration of assessment logic with system design and governance responsibilities. Much of the existing discussion emphasizes isolated techniques—such as automated feedback, text analysis of reflections, dashboarding, or algorithmic scoring—while labor education, by contrast, is practice-enacted, evidence-situated, and oriented toward both competence development and value formation, which calls for explicit principles on authenticity, evidence triangulation, proportional human oversight, and procedural safeguards. Responding to this need, this paper proposes design principles for AI-enhanced labor education evaluation systems in higher vocational colleges by conceptualizing evaluation as a chain of linked design decisions spanning construct definition, evidence specification, analytics support, decision procedures, and governance controls. It contributes a layered conceptual model connecting constructs, evidence, analytics, decisions, and governance, a guideline set foregrounding validity, authenticity, transparency, fairness, privacy, and contestability, and implementation-oriented recommendations that translate these principles into system functions, workflows, and assurance mechanisms, including a synthesizing table mapping principles to functional components and safeguards.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature and identifies the research gap. Section 3 frames the evaluation problem and introduces the layered conceptual model. Section 4 presents the design principles and governance-aware implementation guidelines. Section 5 concludes with the main results, implications, limitations, and directions for future research.

2. Literature review

This literature review covers three thematic strands: labor education evaluation in vocational contexts, assessment foundations for system design, and AI-enhanced assessment together with trustworthy-AI considerations in education. For each strand, we summarize the key concepts and recurring debates that are most relevant to evaluating practice-based outcomes supported by heterogeneous evidence. We then synthesize these insights to articulate the research gap that motivates the conceptual model and design principles developed in the subsequent sections.

2.1. Labor education evaluation in higher vocational colleges

In higher vocational colleges, labor education aims to cultivate productive habits, responsibility, safety, and quality awareness, collaboration, perseverance, and competence in work processes through practice-based activities [5]. It overlaps with work-based and experiential learning traditions, yet it places stronger emphasis on the integration of practical engagement with value formation and habit development. As a result, labor education outcomes are less visible in discrete knowledge demonstrations and more often manifested in sustained behaviors and situated performance.

Evaluation is complicated by three structural features. First, labor education frequently occurs across distributed contexts such as campus tasks, training centers, and practicum sites, where learning opportunities and supervision can differ substantially [6]. Second, available evidence is heterogeneous and uneven in quality, including artifacts, process records, reflections, observation notes, peer feedback, and workplace mentor comments; each source captures only part of the construct and is difficult to standardize [7]. Third, evaluation commonly serves mixed purposes, combining formative support for development with summative signaling for certification, recognition, or accountability [8]. These conditions jointly suggest that labor education evaluation is best approached as a system-level design issue that must coordinate construct definition, evidence management, interpretation procedures, and stakeholder roles.

2.2. Assessment foundations for system design

Assessment theory provides key foundations for evaluation systems that must interpret heterogeneous, practice-based evidence [9]. Validity-oriented perspectives emphasize that evaluative judgments depend on clear construct definitions, representative evidence, and appropriate decision procedures [10]. For labor education, this requires making intended outcomes explicit and ensuring that collected evidence aligns with what the evaluation claims to measure. Formative assessment frames evaluation as continuous evidence gathering and feedback for improvement, which fits labor-related dispositions and habits that develop over time [11]. Authentic and performance-based assessment further supports using meaningful tasks aligned with vocational practice, but it also heightens the need for transparent rubrics and moderation to sustain consistency across sites and evaluators [12]. Finally, triangulation and portfolio-based approaches help address construct complexity by combining multiple evidence sources and documenting development longitudinally, implying that system design should support evidence quality, rubric-guided judgment, and auditability [13].

2.3. AI-enhanced assessment and trustworthy AI in education

Research on AI in education and learning analytics suggests that AI can support evaluation systems by aggregating multi-source evidence, organizing portfolios, summarizing process information, generating formative feedback, and flagging missing documentation [14]. These capabilities are particularly attractive in vocational contexts where teachers face heavy workloads and learning evidence is dispersed across multiple sites. At the same time, the trustworthy-AI literature highlights that AI-supported evaluation raises governance concerns beyond technical performance. Key issues include transparency and explainability, fairness and differential impact across learner groups, privacy and data protection, accountability for decisions, and mechanisms for contestation [15]. A recurring implication is that AI should function as decision support rather than a replacement for professional judgment, especially when evaluative outcomes have meaningful consequences for students [16]. Trustworthy deployment, therefore, depends on coupling AI capability with institutional procedures that ensure oversight, documentation, auditability, and avenues for appeal [17].

2.4. Research gap

Existing work offers valuable insights into vocational practice-based formation, assessment validity, and evidence design, and AI-enabled analytics with ethical principles. Yet there remains limited guidance on how to integrate these insights into a coherent, system-ready set of design principles that connects constructs, heterogeneous evidence, AI-assisted functions, and governance safeguards in higher vocational labor education evaluation. This gap motivates a design-principles approach developed in the remainder of the paper.

3. Problem framing and conceptual model

This section frames labor education evaluation in higher vocational colleges as a system design problem rather than a single-instrument measurement task. The core challenge is to convert a practice-based, context-sensitive educational aim into an evaluation process that is interpretable, usable for improvement, and defensible under multi-stakeholder scrutiny. To do so, the section clarifies evaluation targets and boundary conditions, specifies the roles of key stakeholders and decision contexts, proposes a system-ready evidence model for labor education, and integrates these elements into a layered conceptual model that will later support the formulation of design principles.

3.1. Evaluation targets and boundary conditions

A practical evaluation system begins with a stable articulation of what counts as "labor education outcomes." In higher vocational settings, these outcomes usually combine competence-oriented and disposition-oriented components. The former concerns performance in authentic work processes, including procedural correctness, quality awareness, and the ability to execute tasks with appropriate tools and standards. The latter concerns work habits and occupational dispositions such as responsibility, persistence, collaboration, and responsiveness to feedback. These outcomes should be defined at a level that is shared across majors, while remaining open to contextual operationalization through locally appropriate indicators and rubrics.

Boundary conditions are equally important because they prevent evaluation drift and reduce governance risks. Evaluation should remain purpose-limited and avoid becoming generalized behavioral surveillance. Consequential judgments should preserve accountable human review, with AI positioned as support rather than a substitute for professional judgment. Finally, requirements should scale with stakes: as consequences increase, stronger expectations for transparency, documentation, review, and contestability become necessary to sustain legitimacy.

3.2. Stakeholders and decision contexts

Labor education evaluation is inherently multi-stakeholder. Students are both evaluated individuals and primary evidence producers, which makes the visibility of standards and feedback mechanisms central to improvement. Teachers and instructors are principal evaluators who apply rubrics with contextual knowledge from teaching and supervision. Workplace mentors may contribute practicum-based input that enriches evidence but also increases variability in documentation and interpretation. Counselors or student advisors often use evaluation outputs for developmental guidance, while administrators and quality assurance roles focus on aggregated information for program improvement and accountability. Supporting roles in IT and data governance are also essential to ensure that evaluation remains secure, auditable, and procedurally fair.

Decision contexts differ by purpose and stakes and should therefore be treated distinctly. Formative use prioritizes feedback, reflection, and improvement planning, where timeliness and evidence completeness matter more than strict comparability. Summative use emphasizes milestone judgments and recognition, where consistency and moderation are more salient. High-stakes use additionally requires explainability, reviewability, and contestability, with clear documentation of evidence-informed decisions. Without such differentiation, systems risk either over-regulating learning support or under-protecting consequential decision-making.

3.3. Evidence model and data sources for labor education

Because labor education outcomes are enacted through practice, evaluation relies on heterogeneous and distributed evidence. To make this evidence system-ready, it is necessary to specify a clear taxonomy of acceptable evidence types together with their evaluative value and minimum controls. **Table 1** summarizes four commonly used evidence categories in higher vocational contexts.

Table 1. Evidence taxonomy for labor education evaluation

Evidence category	Typical examples	Interpretive purpose
Performance artifacts	Task deliverables; products; service outputs; project results	Outcome quality and task alignment; shared rubrics across sites; evidence relevance boundaries
Process evidence	Workflow records; safety/quality logs; iteration history; checklists	Execution quality and compliance; completeness and proportionality; avoid surveillance-like collection
Observational judgments	Rubric ratings; structured notes by teachers/mentors	Context-aware appraisal; rater calibration and moderation; rationale traceability for consequential use
Reflective/dialogic evidence	Reflections; self-assessment; peer feedback; response-to-feedback records	Developmental meaning-making; language sensitivity; supportive use in high-stakes contexts

Table 1 provides a minimal taxonomy for organizing the heterogeneous evidence typically used in labor education evaluation. The "Interpretive purpose" column clarifies what each evidence category is most suitable for supporting: artifacts emphasize outcome quality, process evidence illuminates execution and compliance, observational judgments capture contextualized appraisal, and reflective/dialogic evidence documents developmental change and responsiveness to feedback.

This taxonomy also highlights why defensible evaluation cannot rely on a single source. Each category has characteristic limitations, so robust interpretation depends on combining complementary evidence across categories, particularly in consequential decisions. These considerations motivate the layered conceptual model in the next subsection, which separates constructs, evidence, analytics, decision procedures, and governance to keep interpretation coherent and accountable.

3.4. Layered conceptual model: Construct–Evidence–Analytics–Decision–Governance

Based on the preceding framing, this paper adopts a layered conceptual model that separates five elements often conflated in practice. The construct layer specifies intended outcomes and performance standards, stabilizing meaning and reducing drift toward convenience-driven proxies. The evidence layer specifies what is collected and curated, and under what quality criteria it becomes interpretable and verifiable. The analytics layer provides AI-assisted functions—such as organization, summarization, feedback support, and consistency prompts—while remaining constrained by construct definitions and evidence rules. The decision layer defines who makes judgments, under what procedures, and with what documentation, including moderation and stake-sensitive workflows. The governance layer specifies the controls that make the system trustworthy in use, including access management, auditability, transparency artifacts, and contestation procedures.

The value of the model lies in clarifying interfaces among layers. Analytics outputs should be anchored in explicit constructs and defensible evidence standards; decision procedures must constrain how AI outputs are used; and governance controls should be embedded across evidence capture, analytics, and decision workflows. Without these interfaces, systems tend to over-collect data without interpretive clarity or over-automate judgments without accountability.

3.5. Design requirements derived from the conceptual model

The layered model yields design requirements that are system-ready and can later be translated into explicit design principles. First, the system must support explicit construct definition and rubric management, including stable indicator libraries, transparency of standards to students, and mechanisms for evaluator alignment across sites. Second, it must provide an evidence infrastructure that supports longitudinal portfolios and triangulation, enabling reviewers to examine development over time rather than relying on isolated snapshots. Third, AI functions should be positioned primarily as assistive capabilities that improve evidence handling and formative guidance, and any AI-assisted scoring or classification should be limited, reviewable, and coupled with human justification in consequential contexts. Fourth, decision workflows should be differentiated by stakes, with stronger requirements for documentation, review, and contestability as consequences increase. Finally, governance must be treated as a design component rather than a policy afterthought, requiring role-based access, audit trails, and mechanisms that allow students to understand and contest evaluative outcomes.

4. Design principles and implementation considerations

This section translates the conceptual model into actionable guidance for designing AI-enhanced labor education evaluation systems in higher vocational colleges. The principles below are stated at a level that is system-ready: each principle clarifies what the system should prioritize, what it should avoid, and how governance requirements shape functional choices. The aim is not to standardize labor education across majors, but to provide a defensible design logic that remains robust under heterogeneous evidence, distributed practicum contexts, and multi-stakeholder decision-making.

4.1. Design principles for AI-enhanced labor education evaluation systems

This subsection condenses the proposed guidelines into eight system-ready principles. **Table 2** maps each principle to implementable system requirements and baseline assurance considerations, so that the principles remain actionable rather than purely rhetorical.

Table 2. Principles to system requirements and assurance considerations

Design principle	System requirements	Assurance considerations
G1. Construct and rubric clarity	Outcome definitions; indicator/rubric library; rubric version control	Prevent proxy-driven drift; ensure standards are transparent and stable
G2. Authentic-task anchoring	Task templates linked to rubrics; evidence submission tied to task criteria	Avoid completion-only signals; interpret evidence relative to task conditions
G3. Triangulated and longitudinal evidence	Portfolio structure across evidence types; time-stamped records; trajectory views	Reduce single-source bias; support developmental interpretation over snapshots
G4. Stake-sensitive human-in-the-loop decisions	Differentiated workflows by stakes; human sign-off; moderation mechanisms	Preserve accountable judgment; avoid over-automation in consequential use
G5. Traceability and explainability proportional to stakes	Evidence-to-rubric links; rationale recording; decision/change logs	Enable audit and contestation; limit opaque AI influence
G6. Context-aware fairness across sites	Practicum context metadata; stratified reporting where needed	Avoid penalizing site constraints; prevent hidden contextual penalties
G7. Purpose-limited data governance and security	Evidence relevance rules; minimization; retention schedules; role-based access	Reduce surveillance risk; protect privacy; ensure secure and compliant use
G8. Contestability and continuous improvement	Review/appeal channel; evidence supplementation; monitoring dashboards; periodic updates	Protect student agency; detect drift and unintended effects; sustain legitimacy

Table 2 summarizes the proposed guidelines in a system-ready form by linking each design principle to concrete system requirements and the safeguards needed for defensible use. The three columns are intended to be read together: principles define the educational intent, requirements specify what the system should implement, and assurance considerations clarify the conditions under which those functions remain legitimate and accountable.

Two implications are central. First, construct and rubric clarity (G1) is a precondition: without stable standards, AI support is likely to drift toward convenient proxies. Second, evaluation should be anchored in authentic tasks and interpreted through triangulated, longitudinal evidence (G2–G3), while consequential judgments remain human-led and increasingly traceable as stakes rise (G4–G5). The remaining principles make explicit that fairness and governance are embedded design concerns rather than afterthoughts: context-aware interpretation (G6), purpose-limited data governance and security (G7), and contestability with continuous monitoring and improvement (G8) collectively sustain trustworthiness over time.

4.2. Human-in-the-loop evaluation workflow across decision stakes

A governance-aware evaluation system should make its workflow explicit, specifying how evidence is transformed into feedback and decisions under human accountability. In routine formative use, the workflow can be lightweight: tasks are defined with rubric-aligned expectations, evidence is submitted and checked for completeness, and AI may assist with organizing portfolios or drafting feedback for teacher review. In summative contexts, the workflow should strengthen procedural safeguards by requiring structured rubric-based judgment and, where appropriate, moderation to reduce variability across evaluators and sites. For high-stakes outcomes, the workflow must additionally guarantee reviewability and contestability through rationale recording, traceable links to evidence, and a clear pathway for re-checking or supplementing evidence. Treating these workflows as stake-sensitive variants prevents both over-governance of low-stakes learning support and under-protection of consequential decisions.

4.3. System capabilities to support evidence use and feedback

The principles in **Table 2** imply a reference system logic centered on three capabilities: rubric-mediated interpretation, portfolio-based evidence organization, and constrained AI assistance. First, the system should maintain stable outcome definitions and rubrics to anchor interpretation and enable cross-site comparability. Second, it should organize heterogeneous evidence into longitudinal portfolios that support triangulation rather than single-source judgment. Third, AI should be implemented primarily as assistive functionality—indexing, summarization, completeness prompting, and feedback drafting—operating within constraints defined by rubrics and decision procedures. This framing avoids a feature-driven build and instead positions system capabilities as instruments for realizing the principles, especially construct clarity, authentic-task anchoring, and stake-sensitive human review.

4.4. Embedded governance and operational assurance

Governance should be treated as an internal design requirement rather than an external policy layer. At minimum, the system should enforce role-based access, retention discipline, and auditability of evidence use and decision changes. It should also provide transparency artifacts that communicate evaluation standards and explain how evidence supports judgments, with traceability strengthened as decision stakes rise. For AI-assisted components, operational assurance requires documenting where AI is used, what outputs are produced, and how those outputs are bounded by human-led procedures. Finally, continuous monitoring is necessary to sustain legitimacy: institutions should periodically review evidence quality, inter-rater consistency, differential impacts across practicum contexts, and user perceptions of fairness and transparency, using these findings to update rubrics, workflows, and governance settings.

5. Conclusion

5.1. Main results

This paper addressed the design challenge of building AI-enhanced labor education evaluation systems in higher vocational colleges that remain interpretable, improvement-oriented, and trustworthy under heterogeneous evidence and multi-stakeholder governance. It reframed labor education evaluation as a system design problem, and proposed a layered conceptual model that separates constructs, evidence, analytics, decision procedures, and governance. The paper also introduced a minimal evidence taxonomy to support triangulated, longitudinal interpretation, and consolidated eight design principles (G1–G8) mapped to implementable requirements and assurance considerations. Together, these contributions position AI as constrained decision support within human-led workflows, with traceability, fairness, purpose limitation, contestability, and continuous monitoring embedded as core design commitments.

5.2. Implications for future research

Future research can advance this agenda by developing practical indicators and rubric exemplars that balance cross-major consistency with task and practicum variation. Additional work is needed on defensible evidence-integration and moderation strategies, including how triangulation and portfolio practices influence interpretive consistency and student development. Research should also examine which AI assistive functions meaningfully reduce evaluator workload without eroding professional judgment, and how explanation designs support transparency and contestability in different stakes contexts. Finally, studies on governance-by-design—covering institutional capacity, monitoring routines, and lifecycle assurance—are needed to sustain trustworthiness as AI tools and educational practices evolve.

Funding

This research was supported by the Higher Education Reform Research Key Project of Jiangsu Higher Education Association (No.2025JGZZ30), Special Research Project on "High-Quality Public Course Teaching Reform in Jiangsu Higher Education Institutions" (No.2024GZJX011), and Jiangsu Provincial Education Science Planning Project of Jiangsu Academy of Educational Sciences (No.C/2025/02/96).

Disclosure statement

The authors declare no conflict of interest.

References

[1] Zawacki-Richter O, Marín VI, Bond M, et al., 2019, Systematic Review of Research on Artificial Intelligence Applications in Higher Education–Where Are the Educators? International Journal of Educational Technology in Higher Education, 16(1): 1–27.

[2] Liu J, Zhao W, Gao G, 2022, Exploration of Practical Strategies of School Labor Education under The Information Environment. Advances in Multimedia, 2022(1): 8025036.

[3] Xin Y, Corpuz GV, 2024, The Challenge of Labor Education in Higher Vocational Colleges. Frontiers in Educational Research, 7(3).

[4] Chi Y, 2025, AI-Enhanced Blended Learning Models: Testing and Evaluating an Innovative Educational Framework, Proceedings of the 2025 International Conference on Digital Technology and Educational Psychology (DTEP 2025), 15.

[5] Billett S, 2013, Learning Through Practice: Beyond Informal and Towards a Framework for Learning Through Practice. Revisiting Global Trends in TVET: Reflections on Theory and Practice, 123.

[6] Smith R, Betts M, 2000, Learning as Partners: Realising The Potential Of Work-Based Learning. Journal of Vocational Education and Training, 52(4): 589–604.

[7] Zhang S, Li X, Zhang C, et al., 2023, Measurement of Factor Mismatch in Industrial Enterprises with Labor Skills Heterogeneity. Journal of Business Research, 158: 113643.

[8] Wu S, Duan J, Luo M, 2024, Evaluating and Analyzing Student Labor Literacy in China's Higher Vocational Education: An Assessment Model Approach. Frontiers in Education, 2024: 1361224.

[9] Messick S, 1995, Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning. American Psychologist, 50(9): 741.

[10] Kane MT, 2013, Validating the Interpretations and Uses of Test Scores. Journal of Educational Measurement, 50(1): 1–73.

[11] Schoepp K, Tezcan-Unal B, 2017, Examining the Effectiveness of a Learning Outcomes Assessment Program: A Four Frames Perspective. Innovative Higher Education, 42(4): 305–319.

[12] Vantassel-Baska J, 2021, Using Performance-Based Assessment to Document Authentic Learning, Alternative Assessments with Gifted and Talented Students, Routledge, 285–308.

[13] Kaphle P, 2025, Teachers' Perceptions and Practices of Portfolio-Based Assessment—A Narrative Inquiry, dissertation, Kathmandu University School of Education.

[14] Darvishi A, Khosravi H, Sadiq S, et al., 2022, Incorporating AI and Learning Analytics to Build Trustworthy Peer Assessment Systems. British Journal of Educational Technology, 53(4): 844–875.

[15] Matta SS, Bolli M. 2023, Trustworthy AI: Explainability & Fairness in Large-Scale Decision Systems. Review of Applied Science and Technology, 2(04): 54–93.

[16] Pasquale F, 2019, Professional Judgment in an Era of Artificial Intelligence and Machine Learning. Boundary 2: An International Journal of Literature and Culture, 46(1): 73–101.

[17] Ferdaus MM, Abdelguerfi M, Loup E, et al., 2026, Towards Trustworthy AI: A Review Of Ethical and Robust Large Language Models. ACM Computing Surveys, 58(7): 1–43.