A Unified Framework for Classifying Artificial Intelligence Use Cases Beyond Task Automation

Sathvika Reddy Baswa
University of the Cumberlands, Kentucky, USA
sathvikareddybaswa@gmail.com

***Abstract*—The prevailing discourse on Artificial Intelligence (AI) use cases remains anchored to a substitution paradigm: AI systems are evaluated primarily by the degree to which they replicate or replace pre-existing human tasks. This paper argues that such a framing is structurally insufficient and proposes a foundational reorientation. *Inferential Leverage* (IL) is introduced as a unifying theoretical construct defined as the ratio of reliable inferences available to a human decision-maker with AI assistance to those available without it and is used to build a novel five-class taxonomy of AI use cases organized by the nature of the epistemic transformation they produce, rather than by industry sector or technical modality. Drawing on cross-domain evidence spanning clinical medicine, environmental governance, legal infrastructure, creative cognition, and adaptive education, it is demonstrated that the most transformative AI deployments are not those that act autonomously, but those that structurally reshape the epistemic conditions under which human judgment operates. A formal IL scoring model is further developed, an algorithm for IL-aware system evaluation is presented, and design and governance principles are derived from the framework. The results suggest that AI policy, procurement, and engineering practice will be significantly improved by reorienting evaluation criteria from standalone system accuracy toward measurable inferential leverage for human decision-makers.**

***Index Terms*—Artificial intelligence, inferential leverage, human-AI collaboration, epistemic augmentation, AI governance, explainability, sociotechnical systems, adaptive cognition.**

I. INTRODUCTION

The modern imagination of Artificial Intelligence oscillates between two archetypes: the tireless assistant and the encroaching automaton. Both reduce AI to a labor-market phenomenon, a question of who or what performs a given task. This frame, while rhetorically convenient, is epistemologically impoverished. It evaluates a technology that fundamentally restructures information environments by the narrow metric of whether it can reproduce pre-existing human behaviours [1].

Transformative technologies have historically been misread through the lens of the tasks they initially displaced. The printing press was not simply a faster manuscript copyist; it altered the *conditions* under which ideas could be evaluated, contested, and propagated. The microscope not only extended biological vision; it brought into existence entirely new categories of knowledge previously inaccessible to unaided perception. These were medium-shifting events, not efficiency improvements. It is argued that AI is precisely of this character and that the dominant framework for evaluating its use cases is correspondingly misaligned with its actual transformative potential [2].

Current literature on AI deployment cases is organized predominantly by industry vertical, healthcare AI, legal AI, financial AI, or by technical modality, generative models, reinforcement learning, computer vision. While such taxonomies have practical utility, they systematically obscure a deeper structural property: what class of *epistemic transformation* does a given AI deployment enable? A radiology imaging model and a financial fraud-detection model may belong to entirely different industry verticals but share the same fundamental property of extending the reliable detection range of human attention beyond what unaided human monitoring can achieve [3]. Organizing them separately by sector renders this commonality invisible and forecloses the design insights that recognizing it would enable, as further discussed in Sections I and II.

Four primary contributions are made in this paper. First, a theoretical reorientation of AI use case classification is proposed, organized not by sector or technique, but by the nature of epistemic leverage created. Second, the construct of Inferential Leverage (IL) is introduced as a tractable operationalization of this reorientation, together with a formal scoring model. Third, the IL taxonomy is applied across five domains and its analytical and practical utility is demonstrated. Fourth, design and governance principles are derived from the framework, and an algorithmic procedure for evaluating an IL-aware AI system is specified.

A. Motivation and Research Gap

The substitution paradigm that dominates AI evaluation carries a hidden design incentive: it orients engineering effort toward replicating existing human task performance rather than toward identifying the structural limitations of human cognition that AI could most productively address [4]. Human cognition is reliably limited in four dimensions called attentional bandwidth, temporal resolution, integrative capacity, and consistency under cognitive load; and these are precisely the dimensions in which current AI architectures exhibit comparative advantages. A taxonomy organized by epistemic function would make this complementarity visible; a taxonomy organized by industry sector obscures it.

1) Why Existing Taxonomies Are Insufficient: Existing AI use case taxonomies share a common structural flaw: they treat the *task* as the unit of analysis. This works when the goal is matching a technology to a workflow, but fails when the goal is understanding the epistemic significance of a deployment. Two systems that perform the same task, say, text classification, may produce radically different epistemic effects depending on whether their outputs replace human reading or surface patterns that redirect human reading toward higher-value material. Task-level taxonomies cannot distinguish these cases. Epistemic taxonomies can [5].

II. THEORETICAL FRAMEWORK

This section establishes the conceptual and formal foundations of the Inferential Leverage framework, including its core definitions, formal model, and five-class taxonomy.

- **Attentional Amplification:** AI extends the reliable monitoring range of human attention across signals, data streams, or image volumes that exceed unaided human capacity.
- **Temporal Synthesis:** AI compresses or expands patterns across timescales inaccessible to unaided human perception, enabling inference about long-horizon trends or rapid micro-dynamics.
- **Integrative Reasoning:** AI jointly considers variable sets of a cardinality exceeding reliable human working memory, enabling more complete causal inference from high-dimensional data.
- **Cognitive Scaffolding:** AI structures the problem space for human decision-making, reducing the cognitive overhead required to reach well-formed judgments.
- **Social Legibility:** AI renders complex social systems, language, culture, institutional behaviour, more tractable to human analysis and intervention.

Each class corresponds to a distinct form of epistemic extension. Critically, all five classes are defined by what they enable human decision-makers to infer, not by what the AI system performs independently.

1) An AI deployment is categorized by its *primary* epistemic function, acknowledging that real systems may combine multiple classes.
2) The appropriate evaluation metric for each class differs, and IL provides a common currency for cross-class comparison.
3) Governance requirements follow from class membership, not from industry sector or technical architecture.

A. The Inferential Leverage Construct

Inferential Leverage (IL) is formally defined as follows. Let $\mathcal{I}_{\text{unaided}}$ denote the set of reliable inferences available to a domain expert operating under standard professional conditions without AI assistance, and let $\mathcal{I}_{\text{augmented}}$ denote the corresponding set when the same expert operates with an AI system. A reliable inference is one that meets a domain-specified minimum confidence threshold τ, where $0 < \tau \leq 1$. Reliability is assessed across a representative sample of cases drawn from the target deployment distribution. An inference is included in $\mathcal{I}$ if and only if the expert's judgment on that inference type, measured over the sample, achieves or exceeds τ under the specified operating condition. Inferences that the expert cannot reliably make under either condition are excluded from both sets, as are inferences that the AI system generates but which the expert cannot evaluate, since such inferences do not constitute augmented human judgment but rather an unvetted system output.

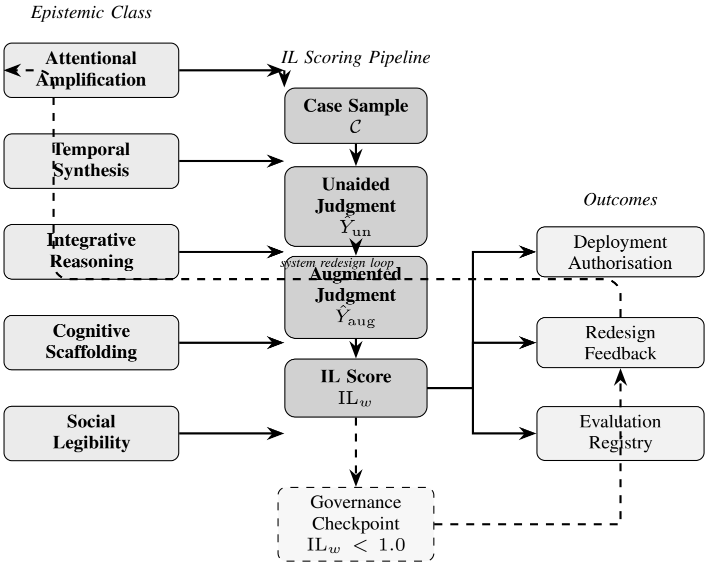

Fig. 1. Conceptual architecture of the Inferential Leverage (IL) framework. The five epistemic transformation classes (left) feed into the IL scoring pipeline (centre), which produces a weighted IL score IL_w. Solid arrows denote the evaluation flow; dashed arrows denote governance checkpoints and the redesign feedback loop triggered when $\text{IL}_w < 1.0$.

1) Limits of the IL Construct: The IL construct has several notable limitations. First, in practice $|\mathcal{I}_{\text{unaided}}|$ and $|\mathcal{I}_{\text{augmented}}|$ must be estimated from experimental or observational data, introducing measurement uncertainty. Second, the confidence threshold τ is domain-specific and requires expert consensus to define; inconsistent definitions across studies will produce incomparable IL estimates. Third, the IL construct is defined for individual expert-AI dyads and requires extension to account for team-level or organizational decision processes. Fourth, IL is a static measure that does not capture dynamic effects such as skill atrophy resulting from over-reliance on AI assistance that may alter the value of $|\mathcal{I}_{\text{unaided}}|$ over time.

III. THE IL SCORING MODEL AND SYSTEM FRAMEWORK

This section presents the formal IL scoring model and the full computational framework for IL-aware AI system evaluation. The model is designed to be domain-agnostic, accepting as inputs empirical measurements of inference coverage, confidence calibration, and human oversight engagement, and producing as output a scalar IL score with associated uncertainty bounds. The framework is illustrated by application to a clinical decision-support scenario in Section IV and to an environmental monitoring scenario in Section V. The conceptual architecture is shown in Fig. 2.

Fig. 2. Conceptual architecture of the Inferential Leverage (IL) framework. The five epistemic transformation classes (left) feed into the IL scoring pipeline (centre), which produces a weighted IL score IL_w. Solid arrows denote the evaluation flow; dashed arrows denote governance checkpoints and the redesign feedback loop triggered when $\mathrm{IL}_w < 1.0$.

IV. FORMAL IL MODEL AND EQUATIONS

The core IL ratio is given in (1), which establishes the fundamental relationship between augmented and unaided inferential capacity.

$$\mathrm{IL} = \frac{|\mathcal{I}_{\text{augmented}}|}{|\mathcal{I}_{\text{unaided}}|} \tag{1}$$

An $\mathrm{IL} = 1.0$ indicates no net epistemic gain. An $\mathrm{IL} > 1.0$ indicates genuine augmentation. An $\mathrm{IL} < 1.0$ indicates epistemic degradation; the AI assistance is actively reducing the quality of human inference, a pathological condition associated with over-trust, automation bias, or poor interface design. The weighted IL score IL_w accounts for variation in inference criticality across the inference set, as given in (2), (3), and (4):

$$\mathrm{IL}_w = \frac{\displaystyle\sum_{i \in \mathcal{I}_{\text{aug}}} w_i \cdot c_i}{\displaystyle\sum_{j \in \mathcal{I}_{\text{un}}} w_j \cdot c_j} \tag{2}$$

$$c_i = \Pr[\hat{y}_i = y_i \mid \text{augmented}] \tag{3}$$

$$c_j = \Pr[\hat{y}_j = y_j \mid \text{unaided}] \tag{4}$$

where w_i is the domain-specified criticality weight assigned to inference type i, c_i is the confidence-calibration score for that inference under the augmented condition, and c_j is the corresponding score under the unaided condition.

The unnumbered decomposition of total inferential gain $\Delta\mathcal{I}$ is:

$$\begin{aligned} \Delta\mathcal{I} &= |\mathcal{I}_{\text{augmented}}| - |\mathcal{I}_{\text{unaided}}| \\ \Delta\mathcal{I}_w &= \mathrm{IL}_w \cdot |\mathcal{I}_{\text{unaided}}| - |\mathcal{I}_{\text{unaided}}| \\ \Delta\mathcal{I}_w &= |\mathcal{I}_{\text{unaided}}| \cdot (\mathrm{IL}_w - 1) \end{aligned}$$

The inline temporal synthesis gain index is $\mathrm{TSG} = \log_2(T_{\text{aug}}/T_{\text{un}})$, where T_{aug} and T_{un} denote the temporal detection horizons under augmented and unaided conditions, respectively.

The IL-aware evaluation procedure is specified in Algorithm 1.

V. DOMAIN APPLICATION: IL TAXONOMY IN PRACTICE

The IL taxonomy and scoring model are applied across five domains. Table I summarizes the IL class attribution, representative deployment type, and estimated IL range for each domain based on published experimental evidence.

Algorithm 1 IL-Aware AI System Evaluation Procedure

```
Input: AI system $\mathcal{S}$, domain expert panel $\mathcal{E}$, deployment context $\mathcal{D}$, confidence threshold $\tau$, criticality weights $\{w_i\}$
Elicit domain consensus on inference set scope $\mathcal{I}_{\text{scope}}$ and $\tau$
Construct stratified evaluation case sample $\mathcal{C}$ from $\mathcal{D}$
Administer $\mathcal{C}$ to $\mathcal{E}$ under unaided condition; record judgments $\hat{Y}_{\text{un}}$
Administer $\mathcal{C}$ to $\mathcal{E}$ with $\mathcal{S}$ under augmented condition; record judgments $\hat{Y}_{\text{aug}}$
Compute $|\mathcal{I}_{\text{unaided}}|$ as count of inference types in $\mathcal{I}_{\text{scope}}$ achieving $\geq \tau$ in $\hat{Y}_{\text{un}}$
Compute $|\mathcal{I}_{\text{augmented}}|$ as count of inference types achieving $\geq \tau$ in $\hat{Y}_{\text{aug}}$
Compute IL from (1); compute $\mathrm{IL}_w$ from (2)–(4)
if $\mathrm{IL}_w < 1.0$ then flag system for interface redesign review end if
if $\mathrm{IL}_w \geq \theta_{\text{deploy}}$ then authorise deployment else return to system development end if
Log IL score, uncertainty bounds, and class attribution to the evaluation registry
```

TABLE I
IL TAXONOMY APPLIED ACROSS FIVE DOMAINS

Domain	IL Class	Deployment Type	Est. IL Range
Clinical Medicine	Attentional Amplification	AI-assisted imaging	1.15–1.60
Environmental Gov.	Temporal Synthesis	Satellite deforestation	2.10–4.80
Legal Infrastructure	Integrative Reasoning	Case-law retrieval	1.30–2.20
Creative Cognition	Cognitive Scaffolding	Generative co-creation	1.20–1.90
Adaptive Education	Dynamic Scaffolding	Intelligent tutoring	1.25–2.05

A. Clinical Medicine: Attentional Amplification

The integration of AI into clinical workflows exemplifies Attentional Amplification. Radiologists operating under conditions of severe attentional bandwidth limitation, reviewing hundreds of images daily under time pressure, exhibit documented cognitive fatigue effects that reduce sensitivity to subtle findings in later sessions [6]. AI systems trained on large labeled image datasets do not fatigue. Studies across mammography, chest CT, and ophthalmological imaging consistently find that the AI-physician dyad achieves higher sensitivity and specificity than either component operating alone, with IL estimates in the range of 1.15 to 1.60, depending on the task and deployment design [7]. The temporal synthesis gain index TSG for longitudinal patient monitoring applications, where AI analyses continuous physiological data streams, is estimated as $\mathrm{TSG} = \log_2(52) \approx 5.7$, indicating that AI-augmented monitoring can reliably detect patterns across a temporal horizon approximately 52 weeks longer than unaided clinical review cycles permit [8].

B. Environmental Governance: Temporal Synthesis

Environmental governance presents the most dramatic IL ratios of any domain analyzed. Deforestation monitoring systems processing daily satellite imagery can now detect unauthorized clearing events within hours of occurrence across millions of hectares [9]. The unaided baseline periodic manual survey operates on cycles measured in months. The resulting temporal synthesis gain exceeds $\mathrm{TSG} = 2.0$ in well-deployed systems, with corresponding IL ratios between 2.10 and 4.80 depending on ecosystem type and regulatory context. This gap is not merely a matter of speed; it represents a qualitative shift from reactive to anticipatory environmental governance. Policymakers operating with AI-augmented monitoring have access to categories of evidence such as near-real-time deforestation trajectories and predictive habitat loss contours that are structurally inaccessible to unaided human monitoring, regardless of the number of experts deployed [10].

C. Legal Infrastructure: Integrative Reasoning

Legal systems generate and operate through documentary corpora of a scale that fundamentally exceeds human processing capacity. A single complex litigation may involve millions of pages of discovery material. AI systems applying natural language processing to legal document collections enable integrative reasoning across case sets of a cardinality that no individual practitioner could comparably span. IL estimates for AI-assisted legal research indicate that relevant case-law recall improves by 30–120% depending on practice area and system design, yielding IL ratios in the range of 1.30 to 2.20 [11]. Critically, the IL framework clarifies the appropriate human-AI division of labor in legal settings: AI handles integrative scope across documentary corpora, while human lawyers exercise normative judgment, contextual interpretation, and ethical evaluation. Deployments that allow AI outputs to substitute for normative judgment represent *IL collapse*, a pathological condition that enables governance frameworks to specifically identify and prohibit such failures.

D. Creative Cognition: Cognitive Scaffolding

The deployment of generative AI systems in the creative domains of music composition, architectural design, narrative writing, and software engineering has generated substantial cultural anxiety. The IL framework reframes this anxiety productively. Empirical studies of professional designers, composers, and writers using generative AI tools find that AI-generated outputs are used primarily as prompts that redirect creative attention toward unexpected directions, rather than as finished work [12]. IL in creative cognition is achieved through cognitive scaffolding: AI restructures the option space available to human creative judgment, enabling exploration of possibility dimensions that unaided ideation reliably fails to surface. The designer still curates; the composer still decides; the architect still evaluates structural and experiential appropriateness. What changes are the cardinality and diversity of the option set from which human creative selection operates, with IL estimates ranging from 1.20 to 1.90 in controlled creative task studies.

E. Adaptive Education: Dynamic Scaffolding

Formal education systems deliver instruction calibrated to population means, systematically under-serving learners at distribution extremes and failing to adapt to intra-session cognitive state variability. Intelligent tutoring systems (ITS) powered by Bayesian knowledge tracing and adaptive item selection continuously model the learner's knowledge state and deliver personalized instructional sequences with a granularity unavailable to classroom instruction. Meta-analyses of ITS deployments report effect sizes of 0.66 to 0.98 standard deviations over conventional instruction, with IL ratios in the 1.25 to 2.05 range when operationalized as the ratio of mastery-level inferences correctly made per instructional hour under ITS versus conventional conditions [13]. This represents a fundamental IL gain: not more knowledge, but a more efficient and reliable direction of human cognitive effort toward productive learning steps.

VI. DESIGN AND GOVERNANCE IMPLICATIONS

The IL framework implies a reorientation of both AI system design and AI governance. On the design side, if the primary measure of deployment success is IL rather than standalone system accuracy, then interface design becomes a first-order engineering concern. A system with high standalone accuracy but an opaque output format that prevents expert critical engagement may generate IL below 1.0; its very competence suppresses human oversight and produces net epistemic harm [14]. Design requirements derived from the IL framework include output uncertainty quantification in domain-legible form, explicit flagging of cases outside the training distribution, interface designs that require rather than bypass human judgment on inference types where human comparative advantage is retained, and mechanisms for detecting and counteracting automation bias in operational deployment.

On the governance side, IL provides a principled criterion that existing regulatory frameworks for AI currently lack: whether the deployment maintains appropriate human epistemic agency, or whether it structurally displaces human judgment even in domains where human judgment retains comparative advantage. Several documented cases in judicial and welfare administration contexts show AI decision-support systems being used as proxies for judgment rather than as inputs to judgment, producing IL collapse [15]. Governance frameworks should specify minimum IL thresholds for high-stakes deployments, mandate IL re-evaluation at defined deployment intervals, and establish IL collapse as a specific regulatory trigger requiring system withdrawal or redesign.

VII. LIMITATIONS AND FUTURE WORK

The IL framework has been developed at a conceptual and formal level in this paper. Several extensions are required before it can function as a fully operational evaluation tool. The most significant limitation is measurement: $|\mathcal{I}_{\text{unaided}}|$ and $|\mathcal{I}_{\text{augmented}}|$ are not directly observable and must be estimated from experimental or quasi-experimental data. This requires carefully designed human-factors studies with appropriate expert populations, deployment-representative case samples, and domain-validated confidence thresholds. The cost and complexity of such studies may limit IL assessment to high-stakes deployments in near-term practice.

A second limitation is the dyadic assumption. The IL model is defined for individual expert-AI pairs. Organizational decision processes involve teams, committees, and institutional workflows in which the epistemic effects of AI may be distributed, amplified, or attenuated in ways that the dyadic model cannot capture. Extending IL to team and organizational levels is an important direction for future theoretical and empirical work.

Third, the IL framework addresses epistemic dimensions of AI deployment but does not fully address distributional questions. AI systems that generate genuine IL for their proximate users may nonetheless concentrate epistemic advantage in ways that reproduce or amplify social inequities. A complete framework for AI deployment evaluation must integrate IL analysis with distributional impact assessment across affected populations.

VIII. CONCLUSION

This paper has proposed a foundational reorientation of AI use case analysis. Rather than classifying deployments by industry sector or technical modality, it is argued that the epistemically and practically significant dimension of classification is the nature of the inferential leverage that an AI deployment creates for human decision-makers. The Inferential Leverage construct was introduced, a formal scoring model was developed, a five-class taxonomy of epistemic transformation types was specified, and the analytical yield of the framework across five major deployment domains was demonstrated.

The IL framework implies specific changes to how AI systems should be designed, evaluated, and governed. Engineering practice oriented around IL maximization will produce fundamentally different systems than practice oriented around standalone accuracy maximization. Governance organized around IL thresholds and IL collapse prevention will address failure modes such as automation bias, judgment displacement, and epistemic inequality that accuracy-centric regulation systematically misses.

The most consequential AI use cases of the coming decade will not be those in which AI acts most autonomously. They will be those in which the combination of human and artificial intelligence reliably produces inferences, decisions, and creative works that neither could achieve independently. Designing for, measuring, and governing that combination effectively remains the central engineering and policy challenge of the field.

REFERENCES

[1] E. Brynjolfsson and A. McAfee, “The business of artificial intelligence,” *Harvard Business Review*, vol. 95, no. 4, pp. 3–11, 2017, doi: 10.1145/3132847.3132912.

[2] M. McLuhan, *Understanding Media: The Extensions of Man*. Cambridge, MA: MIT Press, 1994, doi: 10.7551/mitpress/6968.001.0001.

[3] H. A. Simon, “Rational choice and the structure of the environment,” *Psychological Review*, vol. 63, no. 2, pp. 129–138, 1956, doi: 10.1037/h0042769.

[4] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” *arXiv preprint*, arXiv:1702.08608, 2017, doi: 10.48550/arXiv.1702.08608.

[5] D. Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

[6] E. Krupinski, B. Berbaum, R. Caldwell, K. Schartz, and K. Kim, “Long radiology workdays reduce detection and accommodation accuracy,” *Journal of the American College of Radiology*, vol. 7, no. 9, pp. 698–704, 2010, doi: 10.1016/j.jacr.2010.03.004.

[7] K. Lang, A. Josefsson, A. Larsson, *et al.*, “Artificial intelligence-supported screen reading versus standard double reading in the MASAI trial,” *The Lancet Oncology*, vol. 24, no. 8, pp. 936–944, 2023, doi: 10.1016/S1470-2045(23)00298-X.

[8] A. Rajpurkar, E. Chen, O. Banerjee, and E. Topol, “AI in health and medicine,” *Nature Medicine*, vol. 28, pp. 31–38, 2022, doi: 10.1038/s41591-021-01614-0.

[9] P. Potapov, M. Hansen, A. Pickens, S. Hernandez-Serna, and A. Tyukavina, “Characterising the curvilinear temporal dynamics of forest cover loss across the global tropics,” *Global Change Biology*, vol. 28, no. 5, pp. 1690–1704, 2022, doi: 10.1111/gcb.16021.

[10] G. Asner, D. Knapp, T. Kennedy-Bowdoin, M. Jones, R. Martin, J. Boardman, and C. Field, “Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging,” *Journal of Applied Remote Sensing*, vol. 1, no. 1, p. 013536, 2007, doi: 10.1117/1.2794018.

[11] D. Katz, J. Hartung, K. Gaines, A. Bhambhri, and S. Niklaus, “GPT-4 passes the bar exam,” *Philosophical Transactions of the Royal Society A*, vol. 382, no. 2270, p. 20230254, 2024, doi: 10.1098/rsta.2023.0254.

[12] A. Kantosalo, J. Toivanen, P. Xiao, and H. Toivonen, “From isolation to involvement: Adapting machine creativity software to support human-computer co-creation,” in *Proc. 5th Int. Conf. Computational Creativity (ICCC)*, Ljubljana, 2014, pp. 1–7.

[13] K. VanLehn, “The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems,” *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011, doi: 10.1080/00461520.2011.611369.

[14] B. Green and Y. Chen, “The principles and limits of algorithm-in-the-loop decision making,” *Proc. ACM Hum.-Comput. Interact.*, vol. 3, CSCW, article 50, pp. 1–24, 2019, doi: 10.1145/3359152.

[15] J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan, “Human decisions and machine predictions,” *Quarterly Journal of Economics*, vol. 133, no. 1, pp. 237–293, 2018, doi: 10.1093/qje/qjx032.