

Towards Inclusive Intelligence: Defining the Research Challenges at the Intersection of Artificial Intelligence, Universal Design for Learning, and Workplace Training

Aina G. Irbe
0009-0006-4040-8109
Accessible Minds Baltic SIA
Email: aina@accessiblemindstech.com

Eduardo Meza Etienne
0009-0001-6894-6391
Disability: IN DC Metro
Email: eduardo@disabilityindcmetro.org

Abstract—This position paper examines the gap at the intersection of AI, Universal Design for Learning (UDL), and workplace training. While AI capabilities, including content generation, learning material adaptation, and multimodal delivery seem to align with UDL principles, this alignment does not guarantee inclusive learning outcomes. AI-generated outputs may still overlook barriers related to disability, language, cognition, and skills transfer in workplace contexts. The paper identifies three core challenges: (1) misalignment between AI optimization targets and inclusive workplace learning outcomes; (2) accessibility barriers introduced or amplified by AI systems; and (3) the absence of a practical, validated framework for evaluating AI-supported training design. The contribution is a research agenda and a preliminary UDL-informed evaluation framework that defines measurable dimensions for assessing whether workplace training content and processes created with AI are accessible, equitable, and pedagogically sound. This work is intended to support future empirical validation and give organizations a more consistent way to review AI-created training content for accessibility, inclusion, and evidence-based learning design.

Index Terms—Artificial intelligence, Universal Design for Learning, workplace training, accessibility, learning and development.

I. INTRODUCTION

ORGANIZATIONAL use of artificial intelligence (AI) is expanding rapidly. The Stanford AI Index reports that AI adoption rose from 55% in 2023 to 78% in 2024, while generative AI use in at least one business function more than doubled, from 33% to 71% over the same period [1]. In parallel, scholarly interest in AI-supported workplace education is growing: a 2025 systematic review identified 39 studies examining AI applications in employee training, with reported impacts across cognitive, behavioral, affective, and technical-organizational outcomes [2].

Despite this growth, a central question remains unresolved: how can organizations determine whether AI-supported workplace training is accessible, inclusive, and effective for diverse learners? This gap is particularly important because workplace learning is evaluated not only by knowledge acquisition but by the ability to apply skills in real work contexts.

This paper examines that problem at the intersection of AI, Universal Design for Learning (UDL), and workplace training. Prior work suggests that AI capabilities, including content generation, personalization and customization, and multimodal delivery, could support more flexible learning environments. However, alignment between technical capability and inclusive learning outcomes should not be assumed. AI systems optimized for efficiency or surface performance metrics (e.g., completion rates or quiz scores) may still fail to address barriers related to disability, language, cognition, or workplace skill transfer.

Accessibility presents an immediate concern. AI tools used in L&D workflows can introduce or amplify barriers through inaccessible generated content, unreliable captioning, unstructured documents, or limited support for diverse interaction modes. As Treviranus argues, AI systems can reproduce exclusion when their underlying assumptions and data fail to represent human diversity [3]. Without systematic evaluation, these limitations may scale across organizational learning systems.

Currently, organizations lack a practical, integrated framework for assessing whether AI-supported training design meets both accessibility requirements and UDL-informed instructional goals. Existing approaches address parts of the problem—such as technical accessibility standards, AI governance frameworks, or training evaluation models—but do not provide a coherent method for evaluating inclusive learning outcomes in workplace contexts.

Literature was identified through searches of EBSCO Academic Search Complete, ERIC, Google Scholar, the ACM Digital Library, and IEEE Xplore, supplemented by targeted searches of practitioner outlets and policy repositories for adoption and accountability evidence not yet well represented in peer-reviewed journals. This review is a targeted narrative synthesis consistent with the challenge-paper genre, not a systematic review or meta-analysis. Peer-reviewed sources provide the evidentiary basis for the three challenges; practitioner and policy sources are cited only for contextual, adoption, or accountability claims. As discussed in Section II, the workplace UDL evidence base is limited; therefore, this review

draws on training-transfer, AI evaluation, and algorithmic hiring research to frame the need for domain-sensitive, subgroup-aware evaluation [4]–[7].

The issue is not that AI lacks the ability to create content. It is that capability does not guarantee alignment with UDL principles, freedom from new accessibility barriers, or the evaluation infrastructure needed to verify inclusive outcomes in workplace contexts. This paper addresses that gap by identifying three core challenges: (1) misalignment between AI optimization targets and inclusive workplace learning outcomes; (2) accessibility barriers introduced or amplified by AI tools; and (3) the absence of a validated evaluation framework. Section II reviews relevant background on UDL, accessibility, and AI in workplace learning. Section III develops these challenges and presents the preliminary evaluation framework. Section IV examines their broader implications, and Section V proposes the research agenda and validation directions. The paper contributes a workplace-specific framing of the problem, three research questions, and an initial framework that can be tested in future studies.

II. BACKGROUND: INTERSECTING CONTEXTS

Universal Design for Learning in the Workplace

Universal Design for Learning was developed by CAST and informed by Rose and Meyer’s learning-science and neuroscience research [8], [9]. It is organized around three principles: multiple means of representation, multiple means of action and expression, and multiple means of engagement. UDL Guidelines 3.0 extend this foundation by looking beyond individual learning differences to the structural biases and exclusions embedded in learning systems [8]. In this paper, workplace training refers to organizational L&D materials, tools, and learning experiences used for onboarding, compliance, technical skills, professional development, and job-performance support. UDL research in workplace contexts remains limited. Existing studies indicate that most UDL research has been conducted in K–12 and higher education settings, with comparatively little evidence in corporate training environments [10], [11], [12].

Bartlett and Ehrlich [13] and McNichols [14] discuss UDL in workplace training and development, but the literature remains limited compared with formal education. Al-Azawei et al. similarly found that most peer-reviewed UDL research has been conducted in traditional classroom contexts [12].

This gap is noteworthy because workplace learning differs from formal education in its emphasis on performance transfer, organizational accountability, and heterogeneous learner populations. For example, Irbe’s study of UDL in corporate technical training found domain-specific outcomes: while cognitive gains were not significantly different, psychomotor outcomes improved using UDL as a design framework [4]. Although limited in scope, such findings suggest that UDL application in workplace settings may depend on the type of learning outcome being measured.

More recent work has begun to extend UDL into organizational training. Selseleh, Fragomeni, and Gulseren position

UDL as a shared organizational responsibility rather than a concern assigned only to an instructional designer or accessibility specialist [15]. Their model connects training design to organizational outcomes over time and frames UDL as a system issue. How training design that uses AI should operationalize and measure those outcomes remains an open question.

This distinction matters because corporate L&D depends not only on what learners know but on what they can apply and transfer to job performance. The evidence base is therefore too limited to support broad claims that training design that uses AI can produce UDL-aligned workplace learning at scale without additional validation.

Accessibility and UDL are complementary but distinct. Accessibility focuses on removing physical and digital barriers—typically through compliance with technical requirements—such as the Web Content Accessibility Guidelines (WCAG) 2.2 and Section 508. UDL goes further by design: rather than retrofitting accommodations, it constructs learning experiences from the outset to offer multiple ways to support flexible engagement, representation, and expression for all learners. In workplace training, this distinction is critical: content can meet accessibility standards while still failing to provide meaningful opportunities for engagement, practice, or skills demonstration.

Accessibility Standards in Workplace Technology

Workplace learning technologies are guided by major accessibility standards and laws, including WCAG 2.2, Section 508 in the United States, and the European Accessibility Act (EAA) and EN 301 549 in the European Union. These standards define requirements for perceivable, operable, and understandable digital content but do not address instructional design decisions or learning outcomes.

Similarly, the EU AI Act introduces requirements related to risk management, transparency, and human oversight [16]. While highly relevant for governance, it does not specify how AI-generated training should support inclusive learning design or equitable outcomes. As a result, organizations must navigate a gap between compliance-oriented accessibility requirements and the broader goal of inclusive learning experiences.

AI in Workplace Training: Capabilities, Promise, and Disparity

AI is increasingly embedded in workplace L&D through tools for content generation, adaptive learning, and speech-based interaction. These systems can produce training materials, personalize learning paths, and support multimodal access through captioning, transcription, and conversational interfaces [17], [18].

Although these capabilities can be mapped conceptually to UDL principles, such alignment is not evidence of effective or inclusive learning design. The ability to generate multiple content formats or adapt difficulty does not ensure that training addresses learner variability, supports meaningful practice, or enables skills transfer. Prior research notes that generative AI tools may produce outputs that appear pedagogically

sound while requiring substantial expert review for contextual and instructional adequacy [19]–[21].

Sakowicz and Hamidi identify this convergence as promising for learners with disabilities [22], and Angeloni similarly argues that AI embedded within UDL can widen participation [23]. However, both sources condition that promise on human oversight and empirical validation [22], [23]—neither of which has been established in workplace L&D contexts.

At the same time, AI-generated outputs may introduce new forms of inequity. Performance disparities in speech recognition, limitations in generated content structure, and variability in personalization or customization can affect whether learners can access and benefit from training. These risks highlight the need to evaluate training developed using AI not only in terms of functionality, but in terms of accessibility, instructional quality, and outcome equity in workplace contexts.

III. CORE RESEARCH CHALLENGES

Misalignment Between AI Optimization and UDL Evidence

AI may help scale UDL-aligned design practices. However, when instructional designers use AI tools to develop workplace learning content, the prompts they write and the outputs they accept may be shaped by narrow, easily measured criteria such as quiz scores, completion rates, time-on-task, or surface alignment with learning objectives. These indicators are administratively convenient, but they capture only part of what inclusive learning requires. Without explicit UDL criteria, AI-generated course outlines, knowledge checks, scenarios, or slide decks may reflect the tool’s default pedagogical assumptions rather than an intentional response to learner differences.

Recent research suggests that AI-generated instructional materials are not pedagogically neutral. Chen et al. found that commercial AI lesson-plan generators tended to reproduce teacher-centered models with limited learner agency [19], while Parsons and Curry found that ChatGPT-generated instructional design work could show surface competence but still require expert review for contextual and pedagogical adequacy [20]. Related work also identifies gaps in feedback, differentiation, and adaptation for varied learner needs [21]. Together, these findings reinforce the need for human instructional judgment and UDL-informed review before AI outputs are adopted in workplace L&D.

As discussed in Section II, domain sensitivity matters because workplace learning outcomes may vary across knowledge, skill, and transfer domains. This concern is compounded by two gaps in the current evidence base. First, AI-based personalized learning research remains underrepresented in informal education and workplace contexts [18]. Second, research on AI for psychomotor skills training has focused primarily on healthcare, particularly surgical skill acquisition, with limited attention to general corporate technical training [24]. As a result, AI is increasingly used to scale content, course outlines, and knowledge checks, while the skill transfer domain most relevant to workplace performance remains comparatively underdeveloped.

The problem also extends to learner diversity. AI-supported training design may adapt content to performance metrics, but those metrics do not necessarily capture accommodation needs, disability status, language proficiency, neurodivergence, or prior experience with technology. For example, an employee with dyslexia struggling with text-heavy content may generate performance data that resembles disengagement or lack of preparation. Without accessibility-aware review and feedback loops, AI-supported design processes may fail to distinguish low performance from performance limited by inaccessible design.

The central research challenge is therefore to determine how AI-supported training design can be developed and evaluated for inclusive, transferable learning rather than narrow cognitive performance metrics alone. This challenge is captured in the first consolidated research question in Section V. Addressing it requires mixed-methods empirical research that combines experimental or quasi-experimental outcome studies with design-based research, learning analytics, accessibility evaluation, and qualitative analysis of learner experience.

New Accessibility Barriers Introduced by AI

AI deployed without inclusive design principles may create systemic, invisible barriers embedded in infrastructure that organizations have already adopted—barriers that can appear compliant on the surface while still excluding employees in practice. Automatic Speech Recognition (ASR) provides the clearest documented example. ASR underpins AI-powered captioning for training videos, voice interfaces in learning platforms, and transcription of coaching or feedback sessions. Yet ASR systems do not perform equally across speaker groups. Feng et al. found disparities associated with gender, age, regional accent, and non-native accent [25]. Koenecke et al. found that commercial ASR systems misrecognized speakers of African American Vernacular English at nearly twice the rate of white speakers [26]. Studies of speakers with disordered speech similarly show that ASR performance declines as disorder severity increases [27].

These disparities are measurable only when systems are audited appropriately. Aggregate word error rate (WER) can mask subgroup-level failures, and WER alone may not capture semantic errors, hallucinated transcriptions, or other failures that affect whether captions and transcripts are usable for learning. Robust ASR evaluation therefore requires subgroup-disaggregated results, representative samples reflecting the linguistic diversity of the intended learner population, and metrics that go beyond word-level accuracy. In corporate L&D, however, such audit conditions are rarely standardized in procurement or deployment. Organizations may therefore adopt AI captioning, transcription, or voice-interface tools without knowing whether they impose disproportionate barriers on multilingual employees, speakers from minority-language backgrounds, or employees with speech differences. This creates a validity problem: apparent learner performance may reflect inaccessible system behavior rather than actual knowledge, skill, or transfer.

Generative AI introduces additional barrier types. LLMs used to generate training content may produce inaccurate or missing image descriptions, documents without heading structure, color combinations with insufficient contrast, or dense text-only formats that disadvantage learners with learning disabilities such as dyslexia, ADHD, or cognitive processing differences. These risks do not emerge randomly — they follow directly from using AI to generate content without accessibility constraints at the authoring stage. Mallary, Moore, and McClain note that while generative AI can support UDL implementation, algorithmic bias and data privacy remain significant practical concerns [28]. Sakowicz and Hamidi similarly found that generative AI tools can assist in creating personalized content for learners with disabilities, but their effectiveness depends on prompting, constraints, and human decision-making that are not yet standardized in corporate L&D practice [22].

Addressing these risks requires layered validation across the training design workflow when AI is used, including prompt-level accessibility constraints, automated accessibility checks, assistive technology compatibility testing, human review of alt-text accuracy and cognitive load, caption quality thresholds disaggregated by speaker group, LMS publication gates, and periodic re-auditing when AI tools change. These practices are technically feasible, but remain largely absent from mainstream corporate L&D workflows. This challenge is captured in the second consolidated research question in Section V.

Lack of an Evaluation Framework

A third challenge is the lack of a validated framework for evaluating whether workplace training content and design created with AI meet accessibility standards and UDL-informed inclusion criteria. Without one, organizations cannot assess vendor claims systematically, L&D professionals cannot consistently check AI-generated content for accessibility and pedagogical soundness, and researchers cannot build a comparable evidence base across studies.

Existing frameworks address pieces of the problem but not the full workplace-training context in which AI is used to generate, adapt, or deliver content. WCAG 2.2 addresses digital content and interface accessibility; the EU AI Act addresses risk categories, transparency, human oversight, and fundamental rights [16]; and the Kirkpatrick Four-Level Training Evaluation Model, widely used in workplace L&D, evaluates reaction, learning, behavior, and results [29]. However, none of these frameworks evaluates whether AI-generated or AI-supported training is accessible, inclusive, pedagogically sound, and equitable across learner populations. UDL Guidelines 3.0 provide the relevant instructional design foundation [8], but they do not address the AI production layer, vendor claims, corporate performance-transfer demands, or domain-specific outcome differences documented in workplace training research.

This gap has a specific AI dimension. As generative AI is used to create training materials, organizations may begin treating “UDL-aligned” prompts as evidence that the final content is inclusive. Prior work argues that AI-UDL integration has transformative potential, but does not provide workplace-based validation or an evaluation framework for determining whether AI outputs actually embody UDL principles rather than merely invoking them [28], [30]. For example, an AI-generated module prompted to be “UDL-aligned” may provide text, audio, and a visual summary while still lacking scaffolding, meaningful practice, flexible expression, feedback, or authentic skills application. Inclusivity may appear to meet all the requirements on a formal checklist but still fail to give learners meaningful ways to engage, practice, respond, or demonstrate skill.

Any viable framework must therefore extend beyond interface conformance and end-of-course metrics. It must be usable by the full set of stakeholders who shape workplace learning, including instructional designers, accessibility specialists, managers, subject-matter experts, procurement teams, and vendors. It must also be domain-sensitive: a framework that treats UDL as simply present or absent cannot capture differences among knowledge retention, skills application, engagement, and transfer. Training content generated or adapted with AI that provides accessible knowledge content is not necessarily equivalent to training that supports practice, feedback, skill demonstration, and workplace performance transfer.

The research challenge is whether a UDL-informed, domain-sensitive evaluation framework for workplace training content and design processes created with AI can be developed, empirically validated, and operationalized for use in L&D practice and vendor procurement. This challenge is captured in the third consolidated research question in Section V.

Existing frameworks such as WCAG, the Kirkpatrick Four-Level Training Evaluation Model, UDL Guidelines, and AI governance models, each address part of the problem but none brings them together. Table 1 proposes a framework that does, integrating accessibility, instructional design, AI system behavior, and equitable outcomes across learner populations into a single evaluation tool for workplace training content and design processes created with AI. It maps the core risks identified in Sections II.A and II.B, including accessibility barriers, narrow performance metrics, and governance limitations, to measurable evaluation dimensions.

Each dimension links a UDL principle to a specific AI component and proposes indicators that can be assessed through automated checks, human review, assistive technology testing, or subgroup-disaggregated outcome analysis. The framework is intended as a structured research artifact rather than a validated instrument, providing a basis for empirical testing, comparison across studies, and practical use in L&D evaluation and procurement contexts.

TABLE I.
 DIMENSIONS OF A UDL-INFORMED EVALUATION FRAMEWORK FOR WORKPLACE TRAINING OUTPUTS CREATED WITH AI

Framework Dimension	UDL Guidelines 3.0 Alignment	AI Component Evaluated	Measurable Indicator	Validation Method
Content Accessibility — AI-generated learning materials meet technical accessibility standards before reaching learners	Multiple Means of Representation — Guideline 1: Design Options for Perception	Generative content tools (LLMs, document authoring pipelines)	% of AI-generated items conforming to WCAG 2.2 Criteria 1.1.1 (non-text alternatives), 1.3.1 (info and relationships), and 1.4.3 (contrast) prior to LMS publication	Automated audit + human spot-check sample
Speech and Interaction Equity — AI-mediated voice interfaces and captioning perform equitably across the learner population	Multiple Means of Action and Expression — Guideline 4: Design Options for Interaction; Guideline 5: Design Options for Expression and Communication	ASR, voice-based interfaces, AI captioning systems	Word error rate (WER) disaggregated by speaker group (accent, speech difference, language background), with a maximum acceptable disparity threshold set against a representative corporate workforce sample	Subgroup WER audit against workforce sample
Adaptation Logic — Personalization responds to access needs, not only to performance signals	Multiple Means of Engagement — Guideline 8: Design Options for Sustaining Effort and Persistence	Personalization engine and adaptive learning platform	Proportion of adaptive interventions triggered by format, modality, or pacing signals vs. cognitive performance proxies (quiz score, completion rate, time-on-task); baseline established by logging intervention triggers across a randomly sampled learner cohort	System log analysis of intervention triggers
Domain-Sensitive Outcome Measurement — The system's objective function encompasses psychomotor and affective outcomes, not only knowledge acquisition	Multiple Means of Action and Expression — Guideline 4: Design Options for Interaction; Guideline 5: Design Options for Expression and Communication; especially Consideration 5.3, Build Fluencies with Graduated Support for Practice and Performance; with affective outcomes also aligned to Multiple Means of Engagement	Learning outcome model and AI objective function	Proportion of outcome reports disaggregating results by cognitive, psychomotor, and affective domain across a randomly sampled learner cohort	Statistical analysis of outcome reporting layer
Outcome Equity — Learning outcomes are equitable across learner populations	UDL Guidelines 3.0 systemic and structural inclusion emphasis: reducing barriers rooted in bias and systems of exclusion	Analytics and reporting layer	Statistical significance of outcome differences across groups defined by disability status, language background, neurodivergence, and content domain; requires disaggregated data collection as a system design requirement, not a post-hoc audit	Disaggregated statistical analysis; requires subgroup data collection by design
Human Oversight and Audit Readiness — Organizational accountability mechanisms exist to catch and correct AI-introduced barriers	UDL Guidelines 3.0 systemic and structural inclusion emphasis, combined with EU AI Act transparency and human oversight requirements	Content generation and delivery pipeline; LMS governance layer	% of AI-generated content items with a documented human review checkpoint before publication; presence of a revertible audit trail covering the full LMS content inventory; assessable by L&D teams without specialist accessibility expertise	Process audit; assessable by L&D teams without specialist expertise

Some indicators, particularly those requiring disaggregated learner outcome data, depend on data collection and governance practices that are not yet standard in most organizations.

The indicators in Table 1 are proposed for empirical validation rather than presented as validated measures. Future pilot studies should test the framework across mid-to-large corporate organizations and multiple content domains, including technical skills training, compliance training, and onboarding. Sample sizes should be determined through an a priori power analysis based on outcome measures, expected effect sizes, learner subgroups, and planned statistical models. Validation should include inter-rater reliability testing for human review dimensions, comparison with independent accessibility audits, analysis of whether the framework distinguishes constrained from unconstrained AI-generated outputs, and practitioner usability testing with L&D teams. The organizational framing proposed by Selseleh et al. [15] and the collaborative barrier-anticipation process described by Boucher et al. [31] further suggest that the framework should be usable by cross-

functional teams spanning instructional design, accessibility, HR, procurement, and vendor management.

IV. SOCIETAL STAKES

These challenges are especially pressing in workplace contexts because affected learners are often invisible to the system. Disability non-disclosure, enterprise procurement decisions, and the embedding of AI tools across standard learning infrastructure all increase the risk that inaccessible design will scale without detection.

The population at risk is larger than organizations typically estimate. The World Health Organization (WHO) estimates that approximately 16% of the global population lives with a significant disability [32], and in corporate contexts accessibility needs are often underreported due to stigma, non-disclosure, or lack of awareness of available accommodations. Aging demographics, increasingly multilingual workforces, and growing neurodiversity awareness further expand the

population for whom inclusive L&D design is essential. Evaluation frameworks must therefore be validated against demographically diverse workforce samples, not aggregate populations alone.

The regulatory landscape is accelerating. The EU AI Act introduces phased obligations related to risk management, transparency, and human oversight [16], while the European Accessibility Act began applying accessibility requirements to covered products and services in June 2025 [33]. Recent UDL implementation research in formal education suggests that stronger implementation may improve learner performance and engagement [34], but those findings cannot be transferred to corporate L&D contexts using AI without targeted workplace research. Without validated frameworks that L&D teams can apply during procurement, design, content review, and evaluation, accessibility commitments cannot be operationalized [35].

V. RESEARCH DIRECTIONS AND CALL TO ACTION

Because this agenda sits at the intersection of workplace training, enterprise technology, accessibility governance, and learning analytics, this section translates the preceding problems into three consolidated research questions and four actionable study designs for the intelligence systems community. Jovanovic's FedCSIS work on learning analytics and AI is relevant here because it connects analytics, AI, and evidence-based learning decisions [36].

RQ1. How can AI-supported workplace training design be evaluated for inclusive learning outcomes beyond narrow cognitive metrics such as quiz scores, completion, or time-on-task?

RQ2. What accessibility barriers are introduced or amplified when AI tools are used to generate, adapt, caption, transcribe, or personalize workplace training content?

RQ3. What framework, validation methods, and governance processes are needed to determine whether workplace training content and processes created with AI are accessible, UDL-informed, equitable, and usable in practice?

The studies below map directly to these three questions.

Study 1: Domain-sensitive inclusive outcome metrics (RQ1). Research question: How does AI-supported training design affect cognitive, psychomotor, affective, and transfer outcomes in corporate L&D? Method and data: multi-site experimental or quasi-experimental comparisons of AI-supported, human-designed, and hybrid training across technical and non-technical topics, using training artifacts, prompts, learner demographics where ethically collected, pre/post assessments, performance tasks, engagement data, transfer indicators, and interviews. Metrics and contribution: cognitive retention, rubric-scored skills tasks, engagement/self-efficacy, delayed transfer evidence, and subgroup-disaggregated comparisons; the contribution is domain-sensitive objective functions for workplace training design and evaluation when AI is used.

Study 2: Inclusive AI content generation standards (RQ2 and RQ3). Research question: What constraints and validation mechanisms ensure that AI-generated workplace training content is accessible and UDL-informed? Method and data: develop an AI authoring pipeline that encodes accessibility and UDL requirements into system prompts, authoring templates, metadata, automated validators, and human approval workflows, then compare constrained and unconstrained outputs. Metrics and contribution: WCAG conformance, document structure, alt-text quality, contrast, readability, assessment diversity, UDL alignment, and assistive technology compatibility; the contribution is reusable design patterns for AI authoring in enterprise L&D.

Study 3: Bias audit protocols for workplace training tools using AI (RQ2 and RQ3). Research question: How can ASR systems, generative content tools, personalization engines, and conversational agents be audited for differential performance across learner populations and contexts? Method and data: define population strata, collect representative speech, multilingual learner inputs, generated outputs, recommendations, chatbot interactions, performance signals, and contextual metadata. Metrics and contribution: disaggregated word error rate, semantic error, accessibility errors, hallucination rates, UDL-alignment failures, misclassification of support needs, and remediation thresholds; the contribution is an algorithmic audit protocol for enterprise training tools and outputs created with AI.

Study 4: Validation roadmap for a UDL-informed AI evaluation framework (RQ3). Research question: Can the proposed framework be made reliable, usable, and predictive of inclusive learning outcomes? Method and data: staged validation through expert review or Delphi study, practitioner usability testing, pilot audits of AI-generated artifacts, inter-rater reliability testing, and outcome validation using audit scores and subgroup-disaggregated learner outcomes. Metrics and contribution: content validity, inter-rater reliability, usability, convergence with independent accessibility audits, discrimination between constrained and unconstrained outputs, and predictive validity; the contribution is a framework that links AI evaluation with procurement, governance, and quality assurance in workplace L&D.

We invite the FedCSIS community, with expertise spanning computer science, information systems, intelligence systems, Human Computer Interaction (HCI), algorithmic auditing, and trustworthy AI governance, to engage with this agenda. Many of the technical capabilities needed to address these challenges already exist; what is missing is a shared empirical agenda for ensuring that workplace training content, tools, and processes created with AI are scalable, accessible, inclusive, and valid across the diverse populations they are intended to serve.

VI. CONCLUSION

This paper demonstrates that AI use in workplace training introduces a distinct inclusion challenge: training content, tools, and design processes created with AI cannot be assumed to be

accessible, UDL-informed, or equitable simply because they are efficient, adaptive, or prompted with inclusive language.

The proposed framework provides an initial operational structure for evaluating these risks, linking AI system behavior to accessibility, instructional design, and equity of outcome. While not yet evaluated, it establishes a foundation for systematic assessment and comparison across workplace training contexts. Current evidence does not support the assumption that AI-enabled training can deliver inclusive, transferable learning outcomes at scale. Instead, it underscores the need for domain-sensitive evaluation, accessibility by design, and empirically grounded validation. Advancing this agenda requires interdisciplinary research and practical tools that enable organizations to assess, govern, and improve AI-supported training systems. Without such frameworks, AI risks scaling exclusion alongside efficiency in workplace learning.

ACKNOWLEDGMENT

No funding was received for this work.

Generative AI Disclosure Statement

The authors used a large language model tool to support drafting, editing, language refinement, and reference-checking during preparation of this manuscript. All AI-assisted text was reviewed, revised, and cross-checked by the authors before submission. The authors reviewed and approved the final text and remain responsible for the accuracy, originality, and integrity of the work.

REFERENCES

- [1] N. Maslej et al., "The AI Index 2025 Annual Report," AI Index Steering Committee, Stanford Institute for Human-Centered Artificial Intelligence, Stanford Univ., Stanford, CA, USA, 2025, doi: 10.48550/arXiv.2504.07139.
- [2] E. Kazancı Yabanova, "Artificial intelligence-supported workplace education: A systematic review of learning outcomes, opportunities, and challenges," *Educational Process: International Journal*, vol. 19, Art. no. e2025602, 2025, doi: 10.22521/edupij.2025.19.602.
- [3] J. Treviranus, "Are we teaching our machines our biases, presumptions and stereotypes?" *Global Journal of Intellectual & Developmental Disabilities*, vol. 1, no. 2, pp. 28–29, 2017, doi: 10.19080/GJIDD.2017.01.555560.
- [4] A. G. Irbe, "Application of universal design for learning in corporate technical training design: A quantitative study," Ph.D. dissertation, Capella Univ., Minneapolis, MN, USA, 2016.
- [5] B. D. Blume, J. K. Ford, T. T. Baldwin, and J. L. Huang, "Transfer of training: A meta-analytic review," *Journal of Management*, vol. 36, no. 4, pp. 1065–1105, 2010, doi: 10.1177/0149206309352880.
- [6] S. Barocas, A. Guo, E. Kamar, J. Krones, M. R. Morris, J. W. Vaughan, D. Wadsworth, and H. Wallach, "Designing disaggregated evaluations of AI systems: Choices, considerations, and tradeoffs," in *Proc. AAAI/ACM Conf. AI, Ethics, and Society (AIES)*, 2021, pp. 368–378, doi: 10.1145/3461702.3462610.
- [7] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proc. 2020 Conf. Fairness, Accountability, and Transparency (FAccT '20)*, Barcelona, Spain, 2020, pp. 469–481, doi: 10.1145/3351095.3372828.
- [8] CAST, "UDL guidelines 3.0," Center for Applied Special Technology, 2024. [Online]. Available: <https://udlguidelines.cast.org/>. [Accessed: May 10, 2026].
- [9] D. H. Rose and A. Meyer, *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA, USA: ASCD, 2002.
- [10] Q. I. Almeqdad, A. M. Alodat, M. F. Alquraan, M. A. Mohaidat, and A. K. Al-Makhzoomy, "The effectiveness of universal design for learning: A systematic review of the literature and meta-analysis," *Cogent Education*, vol. 10, no. 1, Art. no. 2218191, 2023, doi: 10.1080/2331186X.2023.2218191.
- [11] L. Zhang, R. A. Carter, J. A. Greene, and M. L. Bernacki, "Unraveling challenges with the implementation of universal design for learning: A systematic literature review," *Educational Psychology Review*, 2024, doi: 10.1007/s10648-024-09860-7.
- [12] A. Al-Azawei, F. Serenelli, and K. Lundqvist, "Universal design for learning (UDL): A content analysis of peer reviewed journals from 2012 to 2015," *Journal of the Scholarship of Teaching and Learning*, vol. 16, no. 3, pp. 39–56, 2016, doi: 10.14434/josotl.v16i3.19295.
- [13] M. Bartlett and S. Ehrlich, "Adopting a universal design for learning mindset in the workplace," *Training Industry*, Sep. 2020. [Online]. Available: <https://trainingindustry.com/articles/content-development/adopting-a-universal-design-for-learning-mindset-in-the-workplace/>. [Accessed: May 10, 2026].
- [14] D. McNichols, "Universal design for learning: Out of the classroom and into the corporate world," *Training Industry Magazine*, Spring 2016. [Online]. Available: <https://trainingindustry.com/magazine/issue/universal-design-for-learning-out-of-the-classroom-and-into-the-corporate-world/>. [Accessed: May 10, 2026].
- [15] M. Selseleh, P. Fragomeni, and D. B. Gulseren, "Universal design for learning: An integrative literature review and integrated model for organizational training and development," *Human Resource Development International*, vol. 28, no. 3, pp. 393–422, 2025, doi: 10.1080/13678868.2024.2397603.
- [16] European Parliament and Council of the European Union, "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence," *Official Journal of the European Union*, OJ L, 2024/1689, Jul. 12, 2024. [Online]. Available: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>. [Accessed: May 10, 2026].
- [17] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, "Intelligent tutoring systems and learning outcomes: A meta-analysis," *Journal of Educational Psychology*, vol. 106, no. 4, pp. 901–918, 2014, doi: 10.1037/a0037123.
- [18] K. Bayly-Castaneda, M.-S. Ramirez-Montoya, and A. Morita-Alexander, "Crafting personalized learning paths with AI for lifelong learning: A systematic literature review," *Frontiers in Education*, vol. 9, Art. no. 1424386, 2024, doi: 10.3389/educ.2024.1424386.
- [19] B. Chen, J. Cheng, C. Wang, and V. Leung, "Pedagogical biases in AI-powered educational tools: The case of lesson plan generators," *Social Innovations Journal*, vol. 30, no. 2, Apr. 2025. [Online]. Available: <https://socialinnovationsjournal.com/index.php/sij/article/view/10004>. [Accessed: May 10, 2026].
- [20] B. Parsons and J. H. Curry, "Can ChatGPT pass graduate-level instructional design assignments? Potential implications of artificial intelligence in education and a call to action," *TechTrends*, vol. 68, no. 1, pp. 67–78, 2024, doi: 10.1007/s11528-023-00912-3.
- [21] N. Kahraman and G. Kırııcı, "Evaluating the efficacy of AI-generated inquiry-based lesson plans in science," *Sakarya University Journal of Education*, vol. 15, no. 1, pp. 40–53, 2025, doi: 10.19126/suje.1463067.
- [22] M. Sakowicz and F. Hamidi, "Exploring the role of generative AI in supporting students with disabilities, through the lens of universal design for learning," *American Journal of STEM Education*, vol. 15, pp. 1–22, 2025, doi: 10.32674/yeyaaq64.
- [23] G. Angeloni, "AI and universal design for learning: Perspectives for accessible and multimodal teaching," in *Artificial Intelligence with and for Learning Sciences: Past, Present, and Future Horizons*, A. Di-pace, C. Limongelli, M. Marras, and S. M. Pagliara, Eds. Cham, Switzerland: Springer, 2026, vol. 16438, pp. 29–42, doi: 10.1007/978-3-032-17604-2_3.
- [24] M. Albaladejo-González, J. A. Ruipérez-Valiente, and F. Gómez Már-mol, "Artificial intelligence to support the training and assessment of professionals: A systematic literature review," *ACM Computing Surveys*, vol. 57, pp. 1–29, 2024, doi: 10.1145/3699712.
- [25] S. Feng et al., "Towards inclusive automatic speech recognition," *Computer Speech & Language*, vol. 83, Art. no. 101567, 2024, doi: 10.1016/j.csl.2023.101567.
- [26] A. Koenecke et al., "Racial disparities in automated speech recognition," *Proceedings of the National Academy of Sciences*, vol. 117, no. 14, pp. 7684–7689, 2020, doi: 10.1073/pnas.1915768117.
- [27] J. Tobin et al., "Automatic speech recognition of conversational speech in individuals with disordered speech," *Journal of Speech, Lan-*

- guage, and Hearing Research, vol. 67, no. 11, pp. 4176–4185, 2024, doi: 10.1044/2024_JSLHR-24-00045.
- [28] K. J. Mallary, E. J. Moore, and A. L. McClain, “Artificial intelligence and universal design for learning: Transforming teaching and learning in adult and continuing higher education,” *New Directions for Adult and Continuing Education*, no. 188, pp. 39–47, 2025, doi: 10.1002/ace.70016.
- [29] J. D. Kirkpatrick and W. K. Kirkpatrick, *Kirkpatrick’s Four Levels of Training Evaluation*. Alexandria, VA, USA: Association for Talent Development, 2016.
- [30] S. Saborío-Taylor and F. Rojas-Ramírez, “Universal design for learning and artificial intelligence in the digital era: Fostering inclusion and autonomous learning,” *International Journal of Professional Development, Learners and Learning*, vol. 6, no. 2, Art. no. ep2408, 2024, doi: 10.30935/ijpdl/14694.
- [31] A. R. Boucher, G. Rappolt-Schlichtmann, and E. Hensler, “Universal design for learning: A shared language to create a culture of collaboration and leverage interprofessional practice,” *Language, Speech, and Hearing Services in Schools*, vol. 56, pp. 581–597, 2025, doi: 10.1044/2025_LSHSS-24-00142.
- [32] World Health Organization, “Disability,” WHO Fact Sheet, Mar. 7, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/disability-and-health>. [Accessed: May 10, 2026].
- [33] European Parliament and Council of the European Union, “Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services,” *Official Journal of the European Union*, L 151, pp. 70–115, Jun. 7, 2019. [Online]. Available: <https://eur-lex.europa.eu/eli/dir/2019/882/oj/eng>. [Accessed: May 10, 2026].
- [34] P. Guo and Z. Wang, “Implementation fidelity of universal design for learning and effects on student achievement, engagement and belonging,” *Scientific Reports*, vol. 15, Art. no. 45430, 2025, doi: 10.1038/s41598-025-30079-1.
- [35] Disability:IN, “2024 Disability Equality Index Report,” 2024. [Online]. Available: <https://disabilityin.org/2024-di-report/>. [Accessed: May 10, 2026].
- [36] J. Jovanovic, “The interplay of learning analytics and artificial intelligence,” in *Proc. 19th Conf. Computer Science and Intelligence Systems (FedCSIS)*, Belgrade, Serbia, 2024, *Annals of Computer Science and Information Systems*, vol. 39, pp. 35–44, doi: 10.15439/2024F5859.