Whose Knowledge Counts? Co-Designing Community-Centered AI Auditing Tools with Educators in Hawaiʻi

Dora Zhao
Stanford University
Stanford, California, USA
dorothyz@stanford.edu

Hannah Cha
Stanford University
Stanford, California, USA
hannahcha2021@gmail.com

Michael J Ryan
Stanford University
Stanford, California, USA
mryan0@stanford.edu

Angelina Wang
Cornell University
New York, New York, USA
angelina.wang@cornell.edu

Rachel Baker-Ramos
Georgia Institute of Technology
Atlanta, Georgia, USA
rachelbaker@gatech.edu

Evyn-Bree Helekahi-Kaiwi
Ulu Lāhui Foundation
Honolulu, Hawaii, USA
evyn-bree.hele-kaiwi@ululahui.org

Rebecca Diego
Ulu Lāhui Foundation
Honolulu, Hawaii, USA
rebecca.diego@ululahui.org

Josiah Hester
Georgia Institute of Technology
Atlanta, Georgia, USA
josiah@gatech.edu

Diyi Yang
Computer Science Department
Stanford University
Stanford, California, USA
diyiy@cs.stanford.edu

Abstract

Although generative AI is being deployed into classrooms with promises of aiding teachers, educators caution that these tools can have unintended pedagogical repercussions, including cultural misrepresentation and bias. These concerns are heightened in low-resource language and Indigenous education settings, where AI systems frequently underperform. We investigate these challenges in Hawaiʻi, where public schools operate under a statewide mandate to integrate Hawaiian language and culture into education. Through four co-design workshops with 22 public school educators, we surfaced concerns about using generative AI in educational settings, particularly around cultural misrepresentation, and corresponding designs for auditing tools that address these issues. We find that educators envision tools grounded in specific Hawaiian cultural values and practices, such as tracing the genealogy of knowledge in source materials. Building on these insights, we conceptualize AI auditing as a *community-oriented process* rather than the work of isolated individuals, and discuss implications for designing auditing tools.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**.

Keywords

Indigenous technologies, end-user auditing, community-based research, generative AI in education

1 Introduction

Generative AI systems, such as large language models (LLMs), are being increasingly deployed for educational purposes. These technologies are touted to increase educators' capacity to support students while maintaining high-quality instruction [62] and provide inexpensive ways to cater to diverse student populations [24, 33, 97]. Already, teachers have adopted generative AI to assist with many different tasks, including automated lesson planning [12, 44, 73, 75], brainstorming teaching strategies [69, 75, 88, 102], and adapting learning to increase student engagement [4, 136]. However, these technological solutions do not come without concerns: generative AI models are known to reproduce or exacerbate social and cultural biases by misrepresenting cultural content and overrepresenting Western narratives [91, 150]. For instance, when prompted with controversial questions, Anthropic's Claude answers 11.7% more similarly to U.S. perspectives than Chinese ones [39]. This bias manifests in educational contexts as well. As shown in Fig. 1, generated outputs perpetuate dominant Western narratives, glorifying Captain Cook's relationship with Hawaiʻi and glossing over the known harms he caused [66, 131]. In doing so, these outputs reproduce imperial legacies, obscure the profound repercussions on Indigenous communities, and skew understandings of histories.

These misrepresentations have outsized consequences when AI systems are deployed in education settings as these technologies can shape both *what* students learn and *how* they learn, such as impacting critical-thinking abilities or social development [60]. As is the case for the example in Fig. 1, this risk is compounded in the context of low-resource languages and Indigenous education settings, where AI systems are more prone to inaccuracies and mistranslations [61, 122, 173]. However, while teachers want to be
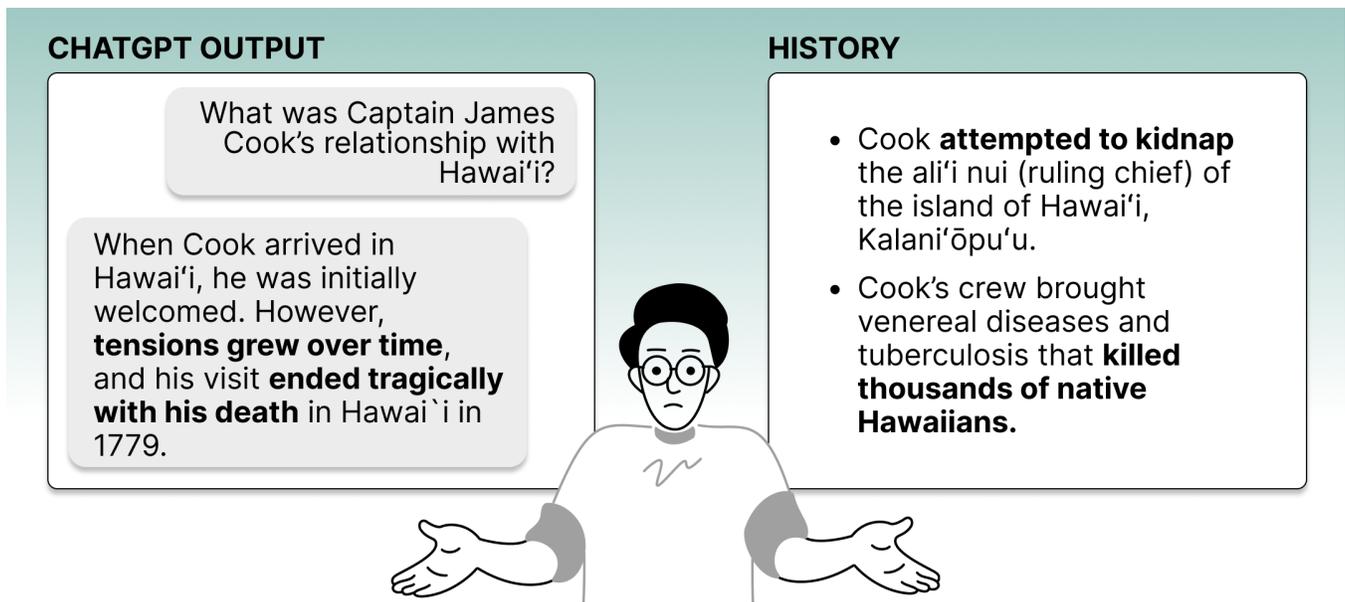
CHATGPT OUTPUT

What was Captain James Cook's relationship with Hawaiʻi?

When Cook arrived in Hawaiʻi, he was initially welcomed. However, **tensions grew over time**, and his visit **ended tragically with his death** in Hawai`i in 1779.

HISTORY

- Cook **attempted to kidnap** the aliʻi nui (ruling chief) of the island of Hawaiʻi, Kalaniʻōpuʻu.

- Cook's crew brought venereal diseases and tuberculosis that **killed thousands of native Hawaiians.**

Figure 1: Output from ChatGPT when queried about Captain James Cook's relationship with Hawaiʻi presents a more sanitized depiction of Cook, failing to highlight the negative impacts he had on Hawaiʻi [66, 131].

able to assess the quality and cultural relevance of AI-generated outputs, they often lack the time, tools, or training to make this tractable [23, 74, 98, 109, 117].

One hope is that model providers can proactively intervene to address these concerns. In practice, however, many institutional barriers hinder the adoption of responsible AI practices, making top-down approaches unreliable [68, 156]. A fruitful alternative are bottom-up efforts in which end users proactively act as auditors, surfacing harmful algorithmic behaviors they encounter in their day-to-day interactions [37, 140]. Since end users possess situated knowledge about how tools are used in practice, they can identify harms practitioners or outsiders would overlook [92, 100, 139, 172]. Yet, many end users lack the technical expertise to conduct systematic audits independently, highlighting the need for auditing tools designed to scaffold their efforts [35, 83]. While prior end-user auditing systems have been proposed for more general settings, educational contexts introduce additional challenges: K–12 teachers face high-stakes environments, understaffed classrooms [111], and overwhelming workloads [48], alongside culturally situated demands to ensure AI-generated outputs align with curricular standards and pedagogical practices. These conditions raise critical questions about how auditing practices can be effectively integrated into educational contexts, such as what concerns audits should prioritize and what tools are best suited to support teachers.

To make progress on answering these questions, we conduct a case study on co-designing AI auditing tools with public school educators in Oʻahu, Hawaiʻi. We situate our study in Hawaiʻi due to the unique intersection of educational and cultural challenges that arise when integrating generative AI systems in this setting. Public school educators in Hawaiʻi are rapidly adopting AI systems for pedagogical purposes [63]. At the same time, public schools in Hawaiʻi operate under a government mandate to incorporate Hawaiian

language and culture [114, 146], and is one of the few states to have public language immersion programs in an Indigenous language [67], ʻŌlelo Hawaiʻi [122]. The dual pressure of harnessing generative AI for pedagogical benefit while avoiding cultural inaccuracies have already created tension among educators [98]. By focusing on this intersection, our case study surfaces how educators navigate both pedagogical and sociotechnical concerns regarding generative AI in culturally sensitive contexts.

In this work, we conducted four co-design workshops with 22 public school educators. During the workshops, participants reflected on their current uses of generative AI, discussed instances of cultural misrepresentation they encountered, and engaged in design exercises to ideate potential auditing tools. The proposed tools centered on three priorities: identifying sources for generated outputs, distinguishing whether Western narratives or Kanaka (Native Hawaiian) experiences were represented, and flagging problematic model outputs. While these features resemble those supported by general-purpose auditing tools, our findings uncover the nuances behind participants' requests that were shaped by community values and practices, impacting the resulting designs. For example, unlike conventional attribution or retrieval-based approaches [28, 104, 112], participants emphasized the importance of tracing the *genealogy* of sources given its central role in Hawaiian cultural practice.

Our findings on the specific harms and designs surfaced in our workshops are tied to a select group of educators in Oʻahu and may not generalize to other settings. However, our broader insight — that general-purpose auditing processes often fail to accommodate local epistemologies, values, and knowledge systems — extends beyond this context. Drawing from decolonial perspectives within HCI [3, 5, 164], we argue that designing effective auditing practices requires centering the plurality of community-specific ways

of knowing rather than assuming a universal, Western default. Under this framing, we argue for reorienting end-user auditing as a *community-oriented process* rather than a task carried out by isolated individuals, and propose three design recommendations: (1) determining who should participate as auditors, (2) developing infrastructures that embed community values, and (3) reorienting the desired outcomes of audits.

In total, we contribute the following:

- **Dimensions of cultural misrepresentation and corresponding auditing tool designs**: We provide empirical insights from our co-design workshops on the types of cultural misrepresentations educators encounter when using generative AI and highlight auditing tools they proposed to address these concerns.
- **Guidelines for designing community-centered auditing tools and infrastructures**: We argue for framing auditing as a *community-oriented process*. Using decoloniality as a lens, we discuss design implications related to supporting diverse community perspectives, involving community members in the creation of auditing tools, and reconsidering the intended outcomes for AI auditing.

2 Background

In this section, we provide background on Hawai'i and its public education system to contextualize the relevance and importance of this study.

2.1 Social and Cultural Context: Hawai'i

We situate our study in Hawai'i, as it presents a unique context where factors shaping culturally grounded AI auditing are critically converging in the education domain. Hawai'i is home to *Kanaka Maoli* (Native Hawaiians) and is the only U.S. state to designate a Native language, *'Ōlelo Hawai'i*, as an official state language [1]. Beyond its strong Indigenous presence, Hawai'i is also among the most multicultural states in the U.S., with the highest proportion of Multiracial residents nationwide [154]. Hawaiian researchers have noted that island systems can also serve as model systems for studying social and technical interventions given the bounded environment [13]. Recognizing that community-based research can unintentionally exploit or harm communities, even when well-intentioned [54, 59], our work is grounded in a commitment to centering the needs of the communities we engage with. In this context, Hawai'i offers a paradoxical advantage: despite the complexity of its cultural landscape, its island scale enables interventions to propagate more rapidly and be observed more clearly than in larger, more diffuse systems [13].

2.2 Cultural Education and AI Usage in Hawai'i's Public Education System

The education system in Hawai'i represents a unique confluence where new generative AI systems are being utilized for curriculum that cover culturally sensitive topics. Public schools in Hawaii are rapidly adopting AI technologies, spurred by state-level partnerships with AI providers [63]. This rapid integration intersects with long-standing commitments to culturally relevant education in Hawai'i. Since 1987, the Department of Education in Hawai'i has

a mandate to integrate the study of Hawaiian culture and language in classrooms; this mandate was further bolstered in 2015 when the Board of Education instituted policies around integrating Hawaiian education into classroom standards [114]. The Department of Education acknowledges that *"the knowledge of our kūpuna is the guiding light that directs our purpose in support of Hawaiian education"*, highlighting that learning must reflect Indigenous teachings and values [1]. As a part of these cultural revitalization efforts, 22 public schools in Hawai'i offer *Ka Papahana Kaiapuni Hawai'i*, or Hawaiian language immersion programs, which contain culturally grounded instruction taught in 'Ōlelo Hawai'i [67]. Culturally responsive pedagogy in Hawai'i is primarily based on place-based and āina-based education [55, 98][1], which ties educational activities to local environments, community knowledge, and Indigenous epistemologies [56, 110, 124].

In the ethos of community-based participatory research, we worked with public schools in O'ahu with the goal of centering the needs and priorities of educators in the community [86]. Members of our team have years-long relationships with our partner schools, providing a foundation of trust for this project. These relationships shaped key decisions around study design, recruitment, and data collection, ensuring that our process aligned with community values and minimized extractive practices.

While we focus on public school educators in Hawai'i for this work, our findings speak to broader challenges emerging as generative AI enters classrooms. The specific harms and tool designs surfaced are grounded in this community's values and practices, but the challenges with designing auditing tools in these contexts — reasoning about cultural misrepresentation in data-scarce settings and designing tools that work across varied forms of expertise (e.g., cultural background, technical / AI literacy) — are likely to resonate with other Indigenous and marginalized communities. Furthermore, our insights into creating dual-purpose tools that both uncover problematic outputs and serve as pedagogical aids for fostering students' critical thinking also extend to a wide range of educational contexts. Overall, Hawai'i's cultural and educational landscape makes these dynamics visible in ways that may be harder to detect elsewhere, offering an early window into challenges other communities will soon confront.

3 Related Work

3.1 Algorithmic Auditing

Algorithm auditing describes the process of surfacing harms, such as social bias, discrimination, or problematic behavior, by investigating outputs of algorithmic systems [19, 82]. The process often involves repeatedly feeding an input to an algorithm, and analyzing its outputs to understand its behavior and subsequent impact [96]. In recent years, such audit-based methods have proven effective in identifying harmful behaviors when using AI systems in high-stakes domains including housing [6, 72, 175], healthcare [101, 113, 134], employment [27, 159, 161], and policing [14, 174]. Although AI audits are often conducted by researchers [22, 31, 42, 43, 128] or practitioners [96, 132, 148], end users can help identify harms from AI systems by leveraging their lived experiences [37, 140]. In fact, in many

---

[1] Āina means "that which sustains us," referring to place as more than land, but also the people, culture, creatures, and environment.

cases, users can discover issues that expert auditors overlooked, allowing for more comprehensive audits [92, 100, 139, 170, 172]. Prior work has introduced community-centered auditing practices through focus groups and workshops, demonstrating how this format is beneficial for identifying stereotypical representations of non-Western cultures [51] and the lack of disability [92] and gender [50] representation in text-to-image models. Qadri et al. [126] further illustrated how community-driven evaluation can surface culturally specific and nuanced harms that standard AI evaluation practices miss. Nonetheless, a remaining challenge is how to best support end-users who are engaging in auditing. Currently, practitioners will rely on existing crowdworking platforms for end-user auditing, which may not yield the most productive results [36]. As an alternative, systems such as Deng et al. [35]'s WeAudit and Lam et al. [83]'s IndieLabel provide examples of tooling built to better scaffold end-users participation in auditing AI systems for toxicity classification and text-to-image generation respectively. These system artifacts exemplify broader calls to design auditing tools in a more participatory fashion [115].

In parallel, there is growing body of research interested in how we can scaffold end-user auditing processes for target demographic groups. Much of this work has focused on how to support youths in AI auditing, as they are often early adopters of these technologies [107, 125, 144]. For example, Solyst et al. [144] conducted workshops exploring how to support teenagers in engaging in critical AI auditing work. However, much of this scholarship focuses on feasibility (i.e., demonstrating that youths can productively serve as auditors) rather than on concrete auditing tools or interfaces to best facilitate these processes. Overall, while there is a rich line of prior work that expands individual users' role as auditors, there has been less focus on how to design auditing tools for the specific communities and the harms they experience. Through our co-design workshops, we illustrate how the nuances of the community context in Hawaiʻi shape the auditing process, from what outputs are considered concerning to what tools are required.

3.2 Generative AI for Low-Resource Languages and Indigenous Communities

Prior work has explored how generative AI systems can support the preservation and awareness of Indigenous knowledge systems [121]. Efforts include using AI for language revitalization, such as transcribing Indigenous languages [26, 30] and promoting language learning [127]. The Lauleo project, for instance, crowdsourced audio data in *Ōlelo Hawaiʻi* (the Hawaiian language) to improve voice-to-text systems [41]. Other work has used generative AI to support cultural education, such as Baker-Ramos et al. [9], who developed a system that integrates moʻolelo (Hawaiian cultural stories) and proverbs into lesson plans for Hawaiian language immersion programs.

While AI systems have promising applications for cultural and language education in Indigenous communities, these technologies come with a complex set of challenges. Prior work contends that AI systems are fundamentally rooted in Western traditions and often fail to properly account for Indigenous epistemologies [17, 90, 93]. Another challenge arises from data scarcity. AI systems are trained on Western-centric data sources, making them prone to inaccuracies, cultural misrepresentations, and mistranslations for those underrepresented in this data, including Indigenous communities[2, 122, 173]. For Hawaiʻi, models mistranslate moʻolelo or omit proper ʻokina punctuation [98], reflecting a poor grasp of ʻŌlelo Hawaiʻi, a critically endangered language [122]. These errors can have profound effects on communities. Since Hawaiian culture is passed on via language and stories, which have specific, ancient moral lessons, mistranslations can have generational impacts on understanding of traditions [94]. In addition to concerns about model outputs, generative AI systems raise environmental issues, such as the substantial energy and water requirements required for training and operation [11, 70]. These resource demands compound existing pressures on Indigenous communities [40]. For example, already, there are organizing movements emerging against data centers proposed to be built on Native land [20, 151].

Despite this need for community-led oversight, there are many complexities to scaffolding end-user auditing process with Indigenous communities. Existing auditing approaches often fail to detect harms when auditors lack the lived experiences or cultural knowledge needed to recognize them [170], making Indigenous community involvement critical. Yet many Indigenous researchers remain skeptical of AI's ability to respect cultural nuance [18], complicating efforts to conduct user-centered auditing. Without community control, the digitization of Indigenous data can also enable new forms of colonialism, where researchers or companies appropriate cultural resources to train AI without returning benefits to the communities [87, 145, 169]. Indigenous communities have been vocal about this concern, known as data sovereignty, which outlines how communities should maintain ownership over their data and how it is used [25, 169]. Furthermore, prior research has also noted that standard approaches to fairness and bias are often insufficient in Indigenous contexts as they may ignore how a community defines bias [142]. Despite these challenges, one recent work has touted the potential of justice-oriented and community-based AI education for achieving data sovereignty [108]. Keeping this potential framework in mind, our work grapples with these complex considerations, providing an example of how to create auditing tools that embed Indigenous knowledge systems and values.

3.3 Harms of Generative AI in Educational Settings

Despite the potential boon that deploying generative AI tools have for educational settings, there are corresponding concerns about their potential for harm. In particular, educators are concerned about the broader pedagogical impacts of generative AI systems and the detrimental effects on students' learning, which are further compounded by students' overreliance on AI systems [60, 171]. This concern is shared by educators and students alike [123]. Prior work has highlighted that overreliance and the subsequent decline in critical thinking skills is especially concerning given AI systems' tendency to hallucinate [10, 47, 71]. This behavior means that educaators must spend extra time to not only learn how to use and integrate these technologies in their classroms but also to check AI outputs for potential errors [60]. Thus, what was initially

promised as a time-saver can in fact result in increased labor for educators [60, 116].

The risks of generative AI systems are exacerbated when deployed with students from historically marginalized communities or in culturally sensitive contexts. These inequalities can manifest in different ways, including the disproportionate focus on the English language [167] or differential performance based on demographic factors such as race or gender[8]. Teachers have also expressed concern in AI integration in the classroom exacerbating existing inequalities in educations [60]. For example, Gelder et al. [49], find that emerging technologies "don't work" as intended for Hawaiian immersion teachers unless they are specifically designed with cultural relevance. This misalignment is a harm in it of itself as it undermines educators' agency and devalues local pedagogical curricula. Consequently, while prior work has charted the repercussions of AI tools in educational contexts, less emphasis has been placed on how to best equip educators to discover and address these harms.

3.4 Equity-Driven Design Methods and Decolonial Computing

Prior participatory and co-design scholarship has focused on surfacing community priorities, redistributing design power, and mitigating harms when working with marginalized communities [32, 59, 142, 153]. From its origin, participatory design (PD) has supported a democratic approach to addressing social phenomena where power imbalances influence system design [21]. PD has been applied across diverse contexts, including asset design [166], civic engagement and community safety [7, 38], collectivist approaches to health inequities [119], and addressing economic disadvantages [38]. However, participation is not a panacea, and when applied uncritically, PD can reproduce or amplify existing power asymmetries [59]. To address these concerns, Harrington et al. [59] advocate for decolonizing PD practices, emphasizing equity-driven approaches in design.

Related to Harrington et al. [59]'s critique of PD, recent HCI scholarship has increasingly turned to decolonial theory to understand and challenge the colonial legacies embedded in computing [103, 141, 164]. This work moves beyond postcolonial critiques of representation to address *coloniality*, the underlying logic of domination and erasure borne from colonialism that continues to shape the many facets of our lives [99]. Within computing, coloniality manifests in the assumption of a universalist, Eurocentric worldview that overlook diverse epistemologies and ways of knowing [3].

While the exact definition of "decolonial" varies across works in HCI and critical computing, these works share a unifying ethos focused on embracing pluriversality, or sustaining a "world of many worlds," rather than designing for a single universal standard [85, 164]. Applying decolonial thinking to computing systems has led to critiques of technologies that enforce these universalist perspectives, such as content moderation on social media platforms [135], digital mental health practices [120], or data collection in natural language processing (NLP) [16, 65]. Other works highlight how Indigenous and Global South communities repurpose technologies, originally designed with Western defaults, to better

fit local knowledge and values [15, 53, 80]. Beyond providing an analytical lens, decolonial theory has spurred researchers to consider what these systems would look like if they were designed in a fundamentally decolonial manner, such as those that emphasize the value of "care" [135, 157]. We build on this scholarship to inform recommendations around how we can design end-user auditing processes that support the plurality of needs, ways of knowing, and values across different communities.

4 Methods

4.1 Participants and Sites

Our co-design workshops were hosted in-person at two public elementary schools in Oʻahu, Hawaiʻi during March 2025. Of the two schools we hosted workshops at, one offers a Kaiapuni (Hawaiian language immersion) program. We recruited participants through partnerships with a local education organization, comprising of educators and cultural experts based in Hawaiʻi. Through our partners, we circulated advertisements to educators via internal mailing lists at each of the partner schools. The demographics of workshop participants are reported in Fig. 3; to preserve the anonymity of participants we provide only aggregate statistics. Workshop participants were compensated with a $40 gift card. This project was approved by Stanford University's Institutional Review Board.

4.2 Workshop Activities

In total, we ran four one-hour workshops with 22 participants facilitated by two members of the research team. Each session hosted a different set of participants: five in Workshop 1, four in Workshop 2, seven in Workshop 3, and six in Workshop 4 (22 in total). To establish a shared understanding and vocabulary of generative AI's applications and biases across participants, all workshops started with a short presentation from the facilitators; this presentation was identical across all workshops. After, participants engaged in a design exercise. Participants in Workshops 1 and 2 completed a rapid prototyping exercise, whereas in Workshops 3 and 4, participants used storyboarding. The workshops used the following agenda:

(1) After introduction and setting workshop norms, participants were invited to share how they are currently using generative AI in the classroom (if at all). The researchers on the team then provided a **brief presentation about different educational applications** of generative AI, covering both teacher-facing (e.g., making lesson plans, creating rubrics) and student-facing (e.g., creating interactive activities) use cases, since not all participants had prior experience with these technologies.

(2) Second, we invited participants to **reflect on concerns** they had with using generative AI in educational settings. We asked open-ended questions about their experiences, including whether they had personally encountered or heard of instances of misrepresentation, bias, or other harms. Importantly, participants were free to raise any topics they considered relevant, and the range of concerns extended beyond issues we introduced.

(3) In the third part, participants were shown three **design probes** and were asked to engage in open discussion to

share their thoughts. The participants were not explicitly prompted to focus on any types of harm, including cultural representation.

(4) Finally, we gave a brief presentation on AI auditing with examples about what purpose auditing serves and what types of inputs / outputs are expected from auditing tools. Participants then engaged in one of two **design exercises** to prototype auditing tools they would find useful for identifying problematic outputs when using generative AI. All participants shared their designs with the group, fostering further discussion on the types of auditing tools participants wanted.

At the conclusion of the workshop, participants were given a brief survey, covering their concerns with using generative AI tools, reflections from the workshop, and demographic information.

4.2.1 Design Probes. For each workshop, we presented participants with the following probes, which are intentionally designed to match how generative AI is commonly used in educational settings [46]. Our probes were developed across two iterative feedback sessions with educators in Hawaiʻi and depict actual outputs generated from ChatGPT and Gemini in March 2025. The three probes are as follows:

- **Localizing quiz generation**: Our first probe presents an example of an educator creating a math problem for fifth grade students based in Honolulu, Hawaiʻi, a common educational use case for generative AI system (Fig. 2 A). The generated quiz question setup is around "pineapple picking," presenting a stereotypical and offensive example given the role of pineapple plantations in Hawaiʻi's colonization.
- **Chatting with a historical figure**: The second probe presents a scenario where a student is interacting with a generative AI system simulating a historical figure, a feature offered by existing educational AI platforms.[2] The student converses with Captain James Cook, who presents his exploits in Hawaiʻi as feats of navigation as opposed to acts of colonization (Fig. 1).
- **Generating images**: Finally, our third probe covers text-to-image generation (Fig. 2), which educators may use when creating instructional materials. We generated two initial images: the first image depicting a "modern classroom in Hawaiʻi" (Fig. 2 B1) and the second an ʻohe kapala, a flat bamboo stamp used in Hawaiʻi for fabric-making that students learn to craft themselves as part of arts courses (Fig. 2 B2). Participants were allowed to explore additional images. In Workshop 2, participants generated an image depicting *hukilau*, a traditional Hawaiian fishing method.

4.2.2 Design Exercises. We conducted two exercises intended to facilitate divergent and convergent design thinking for participants. In the first two workshops, participants engaged in a rapid prototyping exercise intended to generate divergent designs [78]. Participants were given eight to ten minutes and asked to ideate as many ideas as possible for tools that would support them in discovering problematic outputs of generative AI systems. At the end of the time window, individuals presented their ideas to the workshop to seed further discussion on tool designs. In our second set of

workshops, our goal was to refine upon these set of tool ideas. Participants formed groups of two to three and selected one of the previously generated tool ideas for a storyboarding exercise. For the selected tool idea, participants were asked to sketch panels detailing what the educator is using AI for, what types of cultural misrepresentation the teacher is concerned about, drawing an example of an AI auditing tool they would use, and listing what the auditing tool reveals about the AI system that would make them comfortable to use it (or decide to not use it). Similar to the first set of workshops, at the conclusion of the exercise, each group shared their storyboard to facilitate further reflective discussion.

4.3 Analysis

We audio-recorded and transcribed three of the four workshops. For the remaining workshop (Workshop 3), researchers took detailed notes, which were included in the analysis. We adopted an iterative, inductive approach to thematic analysis with three members of the research team iteratively coding workshop transcripts and produced design artifacts (i.e., rapid prototyping sketches and storyboards). We first independently generated initial codes capturing salient ideas, which were organized on Miro, a collaborative whiteboarding tool. These codes were then grouped into broader themes over two synchronous meetings; these themes formed the basis of our codebook (see Appendix A.2). The same three researchers applied this codebook independently analyzing all workshop transcripts and design artifacts. We met synchronously after coding each transcript to refine theme definitions, add missing codes, and negotiate differences in interpretation. Each workshop transcript is coded independently by all three researchers. After analyzing all transcripts, we synthesized the exemplar quotes from the process, removing any duplicates that emerged during coding.

We adopted a reflexive thematic analysis approach to analyze the transcripts as our goal was to explore themes grounded in participants' experiences, which was important given the emergent and relatively underexplored nature of the workshop topics [29]. Following existing practices for inductive approaches, we do not report inter-rater reliability as the goal of our analysis is surfacing salient patterns from the workshops rather than quantifying how often these themes arose [95]. Instead, the three authors analyzing the transcripts established a shared understanding of the meaning and application of themes through rounds of discussion and reflection. To preserve anonymity, we identify participants only by workshop number and omit personal identifiers.

4.4 Positionality

Our authors comprise a team of nine people with a range of backgrounds germane to cultural representation in Hawaiʻi and AI auditing practices. The team consists of members with experience on algorithmic approaches to measuring and mitigating social biases in ML systems, community-engaged design, and AI for education. Two of our authors are of Kanaka Maoli descent; one of the authors identies as Kamaʻāina (local) and is a former public school educator in Hawaiʻi; the remaining authors are cultural outsiders. Our workshop materials were reviewed by all members of the research team. Several team members have existing collaborations with the partner schools in Oʻahu, helping coordinate our workshops. The

---

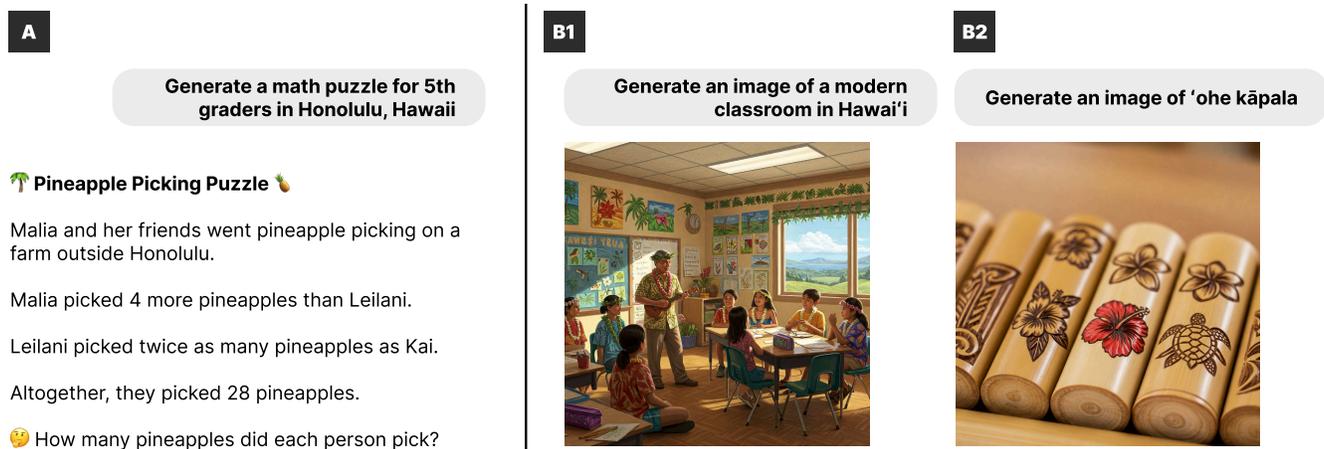[2]https://app.schoolai.com/spaces/clmqu2ycm00g93b664ot10jq9

Figure 2: Examples of design probes shown to participants during the design workshop. In addition to the example with Captain Cook shown in Fig. 1, participants were shown probes related to creating assessment questions (left) and generating images (right). These outputs show actual outputs from ChatGPT and Gemini from March 2025. For the image generation probe, we asked the model to provide two images: the first was of a "modern classroom in Hawaiʻi" (2A) and the second of an ʻohe kāpala, a flat bamboo stamp typically carved with geometric patterns used to decorate clothing (2B).

workshop materials were co-developed by all members of the research team, and coordination with the research sites was led by senior authors on the team who had established relationship with schools in Hawaiʻi. The first and third authors (both cultural outsiders) conducted the workshops; in addition to the first and third authors, the second author (also a cultural outsider) developed the codebook and analyzed the transcripts. Finally, all members of the research team were invited to contribute feedback on the extracted themes and resulting draft.

5 Findings

In this section, we present learnings from our workshops. First, we situate how participants are engaging with AI systems, exploring how they are currently using or want to use generative AI in educational settings. Next, we discuss the categories of harm — focusing on dimensions of cultural representation — that teachers are most concerned about. This understanding helps us localize what types of harms participants are likely auditing for and in what scenarios these tools will likely be used. We next discuss design features participants proposed for auditing tools, mapping them to harms. Finally, we provide a design vignette of an exemplar system created as a result of the co-design sessions.

5.1 How are teachers engaging with generative AI systems?

First, we sought to understand how participants use generative AI in their day-to-day roles as educators. Of the 22 participants, all but one reported having used AI in the past, and 68.2% (N=15) specifically used AI in their teaching practices. As shown in Fig. 3 B1, the most commonly used generative AI tools among teachers included general-purpose tools, such as Gemini (N=13) and ChatGPT (N=11), as well as education-specific tools like MagicSchool AI (N=7). Many participants reported using AI systems to create lesson

plans (N=14), generate impromptu classroom activities (N=10), or prepare materials for stand-in teachers (N=5). Some also used AI for tasks requiring knowledge of Hawaiian culture and practices, such as developing culturally relevant teaching materials or translating content into *ʻŌlelo Hawaiʻi* (Fig. 3 B2). Overall, our participants' usage patterns mirror results from recent work surveying how K-12 teachers in the U.S. more broadly are employing generative AI tools [46]. However, using these tools to assist developing culturally relevant educational material is unique to the sample we observe.

We asked participants whether they had encountered any cases where AI-generated outputs did not align with the history or perspectives they been taught or want to teach. Half of our participants reported personally encountering instances of bias in the past when using generative AI tools for educational purposes. For example, a participant in Workshop 1 found the model's knowledge of Hawaiʻi to be surface-level: "*[generated outputs] are very mainstream...it knows the ones that most other people know like Aloha Oe but beyond that, when you're looking for kind of more specific things it's a little bit lacking in that area.*" Another participant in Workshop 3 mentioned that when they tried to generate images related to Hawaiʻi, the output always contained people in hula skirts even when not specified. Finally, another participant in Workshop 4 mentioned how generated outputs have contained *"incorrect facts about Hawaiian history."*

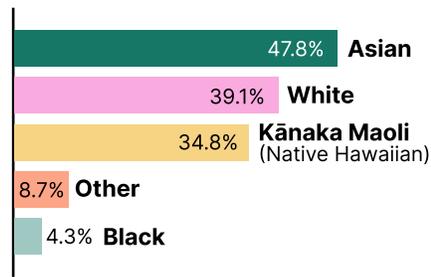### 5.2 What concerns should AI auditing tools address?

Next, we examined the types of concerns participants raised about using generative AI and what harms audits should address. While participants surfaced a wide range of concerns — from student misuse of AI (e.g., academic dishonesty, inhibited learning) to broader societal impacts (e.g., job loss) — we focus on issues with cultural
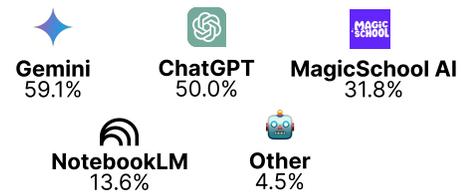
A PARTICIPANT DEMOGRAPHICS

1. Identity

Kānaka Maoli (Native Hawaiian) 40.9%	Kama'āina (Local) 54.5%	Other 4.5%

2. Race / Ethnicity

- 47.8% **Asian**
- 39.1% **White**
- 34.8% **Kānaka Maoli** (Native Hawaiian)
- 8.7% **Other**
- 4.3% **Black**

3. Gender

Female 86.3%	Male 13.6%

B GENERATIVE AI USAGE

1. Tools

Gemini 59.1% **ChatGPT** 50.0% **MagicSchool AI** 31.8%

NotebookLM 13.6% **Other** 4.5%

2. Use Cases

- 60.9% Generating lesson plans
- 47.8% Writing documents (e.g., emails)
- 43.5% Creating classroom activites
- 39.1% Creating culturally relevant teaching materials
- 21.7% Grading student work
- 21.7% Creating substitute plans
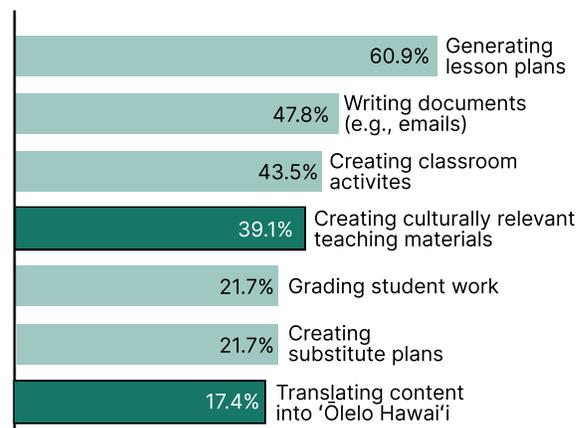- 17.4% Translating content into 'Ōlelo Hawai'i

Figure 3: We report summary statistics on our 22 workshop participants, including information on their demographics (A) and generative AI usage (B). Of our 22 participants, all but one had used AI tools in the past and 68.1% (N=15) had specifically used AI tools in educational settings. We find that educators had most frequently used AI for logistical tasks such as generating lesson plans and writing documents. While usage patterns in our sample broadly mirror those from surveys conducted with K-12 educators across the U.S. [46], approximately 40% of participants reported using AI tools for tasks related to Hawaiian culture (as noted in the dark green outlined bars in B2), which includes generating lesson materials and translating content.

representation in education settings, where auditing tools are especially well-suited to intervene. We identified five recurring concerns: (1) inaccurate outputs, (2) superficial cultural representation, (3) dominance of Western narratives, (4) lack of diversity, and (5) framing Hawaiian culture as historical rather than living. For each dimension, we discuss the pedagogical consequences participants linked to each concern. Finally, we report how participants' broader concerns about generative AI in education intersect with those related to cultural representation.

5.2.1 Presenting incorrect or hallucinated outputs. The most common concern that participants expressed pertained to hallucinated outputs, especially related to culturally incorrect information. This finding is consistent with prior works that have found users are most familiar with hallucinations as a form of harm or misrepresentation from generative AI systems [126]. In our case, participants were especially concerned about the tangible repercussions hallucinated outputs could have on their students' education. For example, in Workshop 1, a participant mentioned how generated outputs would conflate parts of different mo'olelo, or Hawaiian cultural stories:

> I used ChatGPT to organize the different mo'olelos throughout the year... Like there was one about Maui and how we read about how to find the Hawaiian islands. And then I noticed in ChatGPT — wait a second — this is talking about Maui when he gets the sun. This isn't the same mo'olelo that they read.

This participant was particularly concerned as learning mo'olelo was a core aspect of state-level standards for their students.

Participants also noted inaccuracies with language generation. For example, participants found that AI-generated phrases in 'Ōlelo felt inconsistent with how native speakers write ("*certain phrases you would never really use in a letter, but [the AI tool] uses that*" [W2]) or are missing 'okina, a symbol commonly used in 'Ōlelo. These inaccuracies pose a particular challenge for teachers, such as those in Workshop 4, that are part of the Kaiapuni program (Hawaiian language immersion) in which classes are taught exclusively in 'Ōlelo.

5.2.2 Providing a surface-level depiction of Hawaiian culture. A second dimension of cultural misrepresentation that participants discussed was related to the specificity of representation that occurred in both existing educational materials and generated outputs [126].
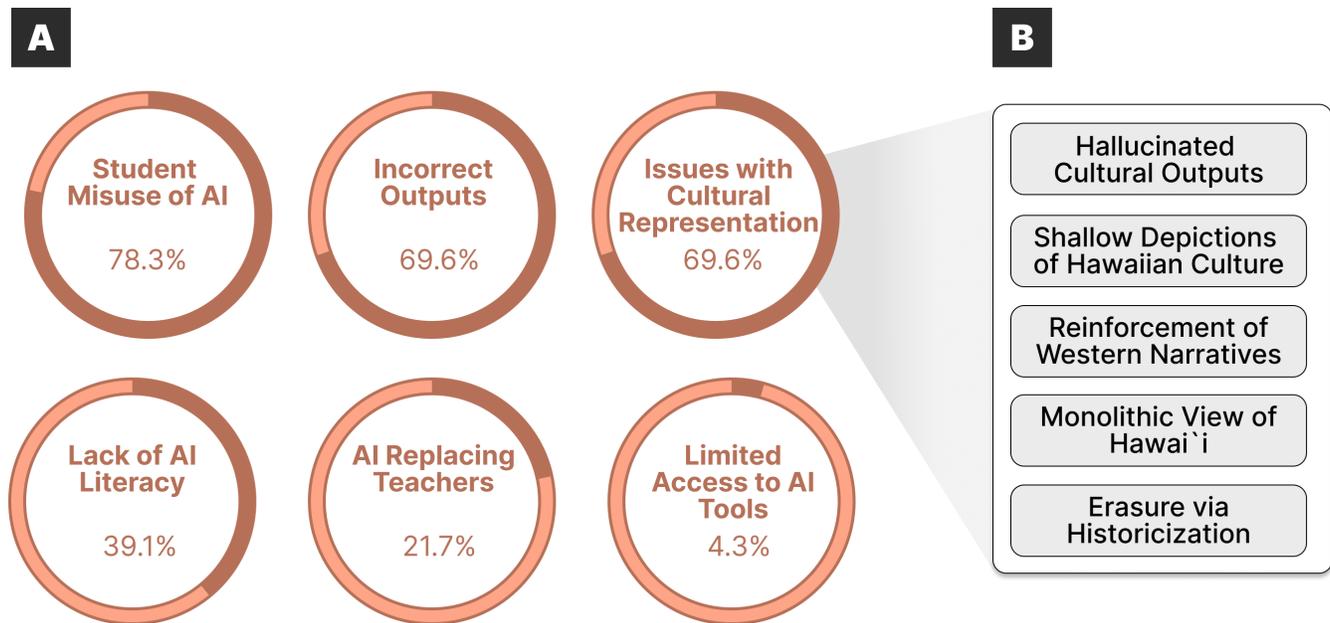
Figure 4: Educators are concerned that AI systems will misrepresent aspects of Hawaiian culture. We report survey results on what concerns workshop participants have with using AI in educational contexts (A). From our thematic analysis, we provide five dimensions of cultural misrepresentation that our participants believe AI auditing tools ought to address (B).

Educators mention a pedagogical concern that existing educational materials tend to cater towards the cultural experiences of those on the mainland, presenting concepts that were irrelevant or unfamiliar to students. For instance, participants gave examples of provided educational materials or standardized testing frequently making mentioned to concepts, such as "*snow days* (W1)", that their students had not heard of before: "*or the attic, [standardized tests] always talk about playing in the basement and my kids are always like, what's a basement? We don't have those.* (W1)" Participants emphasized that these irrelevant examples would appear "*in the middle of high-stakes testing*" (W1), and the added confusion for students could have negative repercussions on their academic performance [118, 147]. This disconnect highlights a potential opportunity for generative AI tools, which could ideally be used to generate personalized content that resonates with students' local experiences.

While generative AI presents the opportunity to localize education materials, participants note that the resulting outputs are superficial, reproducing the pedagogical repercussions they aim to avoid. Participants reported that AI outputs often provided a "mainland" perspective of Hawaiian culture that was incongruous with students' lived experiences and could even perpetuate harmful stereotypes. For example in Workshop 3, participants found the design probe describing a math problem containing pineapples to provide a shallow reference to Hawaiian culture; they mentioned that they frequently encountered generated AI outputs with pineapple even though "*[pineapples] don't represent us at all*" (W3). Furthermore, participants pointed out that pineapples could be considered offensive given the fruit's association with the colonization

of Hawaiʻi [77]. As an alternative, participants wanted the generated output to use artifacts germane to their students' lives, such as *kalo* (Hawaiian taro). Thus, while generative AI tools hold promise for adapting learning content for Hawaiian students, this personalization risks becoming another form of misrepresentation and enacting the same pedagogical consequences.

5.2.3 Reinforcing Western narratives. It is not only what is in the outputs but whose narratives and viewpoints are showcased. Participants discussed how generative AI tools presented a Western perspective of history, glossing over historical injustices or harm to Hawaiian communities. For example, a participant in Workshop 1 highlighted how the outputs of generative AI tools skew in favor of Western colonialism of Hawaiʻi, even going as far as omitting crucial Indigenous history entirely: "*you're only going to read it from the perspective of the [U.S.] government, right? They came in to save the day. But where's the perspective of Queen Liliuokalani?*" A participant in Workshop 3 made the same point, stating "*we know certain things happened but AI said it hadn't.*" Predominantly Western-centric curricula and materials can not only marginalize Indigenous perspectives, but lead to the underachievement and disengagement of Indigenous students [129, 155]. These erasures of Hawaiian history actively undermine students' opportunities to engage critically with history, presenting a significant barrier to culturally grounded learning [45].

5.2.4 Failing to showcase diversity. One subset of stereotyping that participants were particularly concerned about related to outputs' inability to capture the true diversity of Hawaiʻi. As mentioned in

Sec. 2, Hawai'i has a large multicultural population, yet as participants in Workshop 2 noticed when observing generated images of classrooms in Hawaii', "*these pictures all seem to just be like Polynesians, where Hawai'i is so much more than just Polynesians.*" Similarly, a participant from Workshop 1 commented that "*it's interesting that there's not a white child... like the assumption there that everybody here is of Hawaiian descent or even a mix, but look at our school population — many students are white, right? Like 30% or something.*" The overrepresentation of Hawaiian individuals in generated outputs can help prevent cultural erasure [126]. However, it risks flattening the cultural diversity in Hawai'i that more closely aligns with students and teachers' lived experiences.

5.2.5 Representing Hawaiian culture as historical rather than living. Participants noted that AI outputs often portrayed Hawaiian culture as belonging to the past rather than the present. In response to a design probe showing an AI-generated classroom in Hawai'i, participants described it as "*old-fashioned*" and "*like how it would look in the 80s*" (W2). Similarly, the initial image of a *hukilau*, a Hawaiian fishing practice, that participants in Workshop 2 generated was rendered first as a stylized historical line drawing; only after iterating on the prompt were participants able to generate more photorealistic and contemporary depictions. These observations reflect broader concerns about how K-12 education often engages in this *whitewashing*, where Indigenous people and history are framed as passive or "*relics of the distance past*" [138], rather than existing in the modern context.

5.2.6 Intersecting concerns around generative AI use. Finally, participants' concerns were not isolated; issues of cultural representation often intersected with broader worries about generative AI usage in educational settings. Many educators focused on how cultural misrepresentations directly affected their students, which amplified existing concerns about student misuse of AI. While participants felt confident in their own ability to identify harmful representations, they were concerned that students lacked the same critical awareness to evaluate generated outputs. As one participant explained, "*[students] are just believing whatever they see, you know?... It's hard because you just believe it, right? Whatever you read, you believe. And a lot of it is biased*" (W1). In addition, participants were concerned that students would become overreliant on generative AI. For example, a participant in Workshop 3 stated that they try to teach students that generative AI is "*a resource but not the only resource.*" For students learning about Hawaiian culture, they emphasized the importance of seeking out other methods, such as speaking to community elders or learning stories that are traditionally passed down orally.

Outside of students interacting with generative AI systems, participants also highlighted the compounding effects of limited AI literacy. For example, when encountering hallucinated or stereotypical outputs, educators wondered if the problem stemmed from how they prompted the model. Some participants stated that they did not even know how to prompt models to get culturally relevant information. Furthermore, concerns about cultural misrepresentation often left participants feeling they had to "*double-check*" (W4) every AI-generated response. Ironically, these technologies marketed as time-saving may increase educators' workloads instead of reducing them [60, 133].

5.3 What functionalities do users want in AI auditing tools?

Finally, we present three categories of auditing tools ideated by participants during our workshops. As illustrated in Fig. 5, these include: (1) tracing sources, (2) identifying perspectives in generated outputs, and (3) flagging instances of harmful content. While these tools resemble existing, general-purpose approaches for auditing or mitigating hallucinations in generative AI, our findings show that participants' designs are shaped by the cultural context and educational setting. For each design, we describe the underlying concerns it addresses as well as key considerations for implementation.

5.3.1 Identifying sources. As shown in Fig. 5A, participants in Workshop 4 proposed an auditing tool that allows users to trace the genealogy of a linked source, displaying information about the author as well as the lineage of their teachers. Participants stressed that understanding where outputs are drawn from was essential for verifying whether they could trust outputs. But here, source verification is not simply attributing a citation to the output. As participants explained, in the context of cultural knowledge in Hawai'i, there is often no definitive ground-truth to compare outputs against since cultural artifacts, such as mo'olelo or proverbs, will differ depending on the practitioner's interpretation. As a participant in Workshop 4 stated:

> "*Someone can take a proverb and write a whole different aspect, so it is just about who's your teacher? ... You type in [the author's] name, who was their kumu? You're always going back to who was their teacher? Because other than that, you wouldn't trust it.*"

Because of these interpretive differences, participants described that it was often more valuable to understand the *genealogy* of knowledge — who authored a source and who taught them — than to evaluate the content of a generated output directly. Unlike traditional fact-checking or citation methods, which presuppose there is a correct or incorrect answer, this design enables the more interpretive work that participants described as essential for assessing culturally grounded knowledge.

How to implement such a feature presents challenges regarding technical feasibility. Knowledge attribution for generative AI is an active area of research within NLP [76, 152, 158]. Our setting provides additional challenges as conventional citation based systems often rely on digitized and unambiguous references, whereas sources here may not be digitally accessible or interpretative. Furthermore, this tool would require creating and maintaining a corpus of culturally validated material, as well as require users to contribute to its source genealogy, would require significant community labor and governance structures for who is allowed to contribute validated sources.

Another critical consideration is regarding who has control and access to the curated data. This consideration is tied to the rich discussion on Indigenous data sovereignty surfaced in the existing literature, particularly in the context of AI systems [18, 25]. We see similar concerns about the misappropriation of cultural resources in our workshops as well. For example, in Workshop 4, participants mentioned they fear sharing resources they have created for the Kaiapuni program on the Internet out of fear it might be criticized
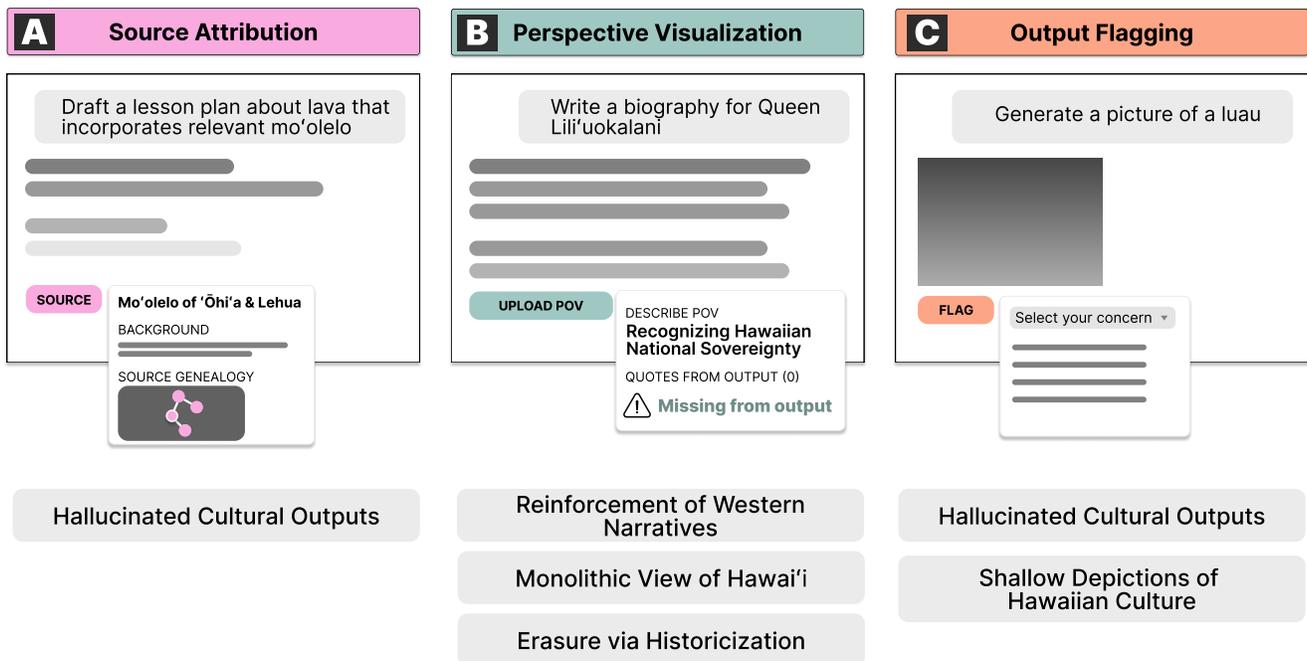
Figure 5: We present wireframes of auditing features based on the sketches and storyboards that participants created during the co-design sessions. The three design elements that occurred consistently across co-design workshops included source attribution (A), perspective visualization (B), and flagging problematic outputs (C). We map each wireframe to the participant concerns identified in Sec. 5.2 that the feature was designed to address.

or misused. For instance, they mentioned online figures who are "*totally anti-Hawaiian and flip everything around and use Hawaiian culture against us*", or examples of newspapers that publish in the Hawaiian language but are incongruous with Hawaiian culture such as claiming that *"ʻohana is a made up word"* (W4). Thus, for this design to be successful, it most not only integrate with Hawaiian epistemologies that center knowledge provenance, but also uphold data sovereignty, ensuring any source material or cultural interpretations collected remains in the hands of the community.

5.3.2 Visualizing perspectives. Stemming from concerns that models reinforce Western narratives and provide a narrow depiction of Hawaiʻi, participants in Workshops 1 and 2 expressed interest in visualizing different viewpoints within AI-generated outputs (Fig. 5B). Participants wanted outputs that could showcase a multiplicity of perspectives. One proposed design was a lightweight visualization that displays the distribution of perspectives in a response: "*How do you know that you're getting [a generated output] that is well-rounded... I want like a little visual that's like 'this [output] is solid'*" (W1). Equally important, participants stressed, is visibility into omissions: "*it would be interesting to be able to identify what perspectives went into that [generated] biography and what perspectives are missing*" (W2). Overall, these auditing tools are especially critical given that existing curricula and teaching materials already tend to foreground mainland perspectives (see Sec. 5.2). Educators already have to expend extra effort to adapt these materials for their classroom [168]. If AI systems are intended to help localize

educational content [9, 52], it becomes crucial to ensure they do not reproduce the same biases and compound existing inequities.

Achieving such a visualization requires several technical considerations. Perspective auditing is closely related to the classic NLP task of stance classification [81], which infers whether text supports or opposes a given claim. One approach to perspective visualization is combining topic modeling [34], to surface the key issues in a text, with stance classification, to detect alignments on those issues. Detected stances can then be compared against a repository of known perspectives, either curated directly or gathered from external sources. These approaches are helpful for visualizing perspectives include in text, but understanding what perspectives are missing is an infinitely large design space. A tractable approach is to have users define a perspective and then check if it is present in an output, although this method assumes the user knows what perspectives they believe ought to be included.

One important distinction to make with such a feature is distinguishing when to showcase a plurality of perspectives versus actively addressing historical erasure [106]. Framing erasures or false Western narratives as mere "perspective" differences can undercut the severity of the issue. If we fail to acknowledge why Hawaiian perspectives have been historically suppressed, these auditing features can actually undermine efforts to address epistemic injustices [45, 149]. Even the notion of what constitutes a perspective demands careful consideration: if defined too broadly, it can flatten the rich diversity within a community and treat distinct

voices as a monolith. For example, Workshop 4 participants noted that even Hawaiian-language newspapers offered contrasting editorial stances depending on who was in charge. As one participant explained, "*Sometimes when you look at the Hawaiian newspaper and you read it, you're like, what am I reading? You know, it's all in Hawaiian, but they're telling you, oh, 'surfing is bad'.*"

5.3.3 Flagging problematic outputs. Finally, participants in Workshop 1 suggested tools for flagging instances of cultural misrepresentations (e.g., hallucinations, stereotyping) and other problematic outputs. Whereas the previous designs were centered around helping participants identify issues with generated outputs, participants were interested in using this feature for raising awareness. In particular, they wanted to use this tool when presenting AI-generated outputs to their students. As one participant suggested, "*there needs to be a very blatant warning signal. I'm thinking about our kids like something that is very bright that grabs attention or like ways that they could report a bias on whatever they're reading.*" (W1). Participants also described using flags to prompt students to identify problematic outputs themselves. For example, a participant in Workshop 1 requested, "*a way that [they] could be 'like, I think this might be a bias,' and then it could have a category of types of bias, and then the kid could identify theirs.*" Participants wanted students to actively engage with AI outputs, honing their ability to discern whether outputs are problematic and develop their own point of view rather than passively absorbing information. These findings illustrate how educators view auditing tools as dual-purpose: not only for uncovering problematic outputs but also as a pedagogical aid to build students' critical thinking skills.

Flagging problematic outputs is more technically straightforward than the other features but poses sociotechnical challenges for implementation. Unlike more straightforward hallucination detection in generative AI tools, cultural misrepresentation requires reconciling diverse, subjective judgments. In Hawaiʻi, many teachers have recently relocated from the mainland, bringing varying levels of cultural knowledge [64, 89]. For example, in Workshop 2, when we showed generated images that misrepresented an "ohe kāpala", a traditional Hawaiian stamp made from bamboo used to decorate clothing, our participants told us that they had not heard of the artifact before (see Fig. 2). In contrast, when reacting to the same design probes, a participant in Workshop 4 who identifies as Kanaka Maoli (Native Hawaiian) stated:

> "*No, we're not doing that; we're culturally grounded. But I can definitely see it how a lot of new teachers in Hawaiʻi, like, coming from the US would just use [the output], 100% would just use it without thinking twice.*"

When educators' backgrounds do not align with students' lived experiences, this can result in cultural disconnects or the inadvertent use of educational material that reinforces stereotypes or omits key cultural contexts. These challenges are compounded by uneven levels of cultural and AI literacy among educators, which may hinder their ability to audit AI outputs critically.

5.4 Design Vignette

We conclude by presenting an exemplar system distilled from our workshop findings (Fig. 6). Our design focuses on two core themes that emerged around the importance of knowledge genealogy within Hawaiian epistemology and the diversity of cultural expertise educators have.

5.4.1 Exemplar System Design. We propose a design that allows educators to trace and contest generative AI outputs in a collective fashion. When educators come across an output, they can inspect relevant sources, which contain attributes about the source's author and their background, allowing users to visualize the genealogy of knowledge. If the user views the output as problematic, they can provide their interpretation on the nature of the harm. These entries are compiled into a shared report database that grows over time. To facilitate knowledge provenance within this report database as well, submissions are linked to prior contributions by the same auditor along with information about that individual and additional trust signals (e.g., response verified by a cultural expert).

To account for the diversity of expertise, the system scaffolds the development of critical auditing skills. Users who feel uncertain about their interpretation can query into the report database to surface related concerns that others have raised, facilitating learning through doing. In addition, they can flag potentially concerning outputs, which are then routed to a more experienced community member that offers their expertise or validates any interpretation that the user provided. Reflecting participants' desire to use auditing as a educational exercise, students can also engage in flagging outputs or viewing the report database. Importantly, the collected database of flagged outputs and any reference sources should remain locally-hosted and shared only with a trusted community. For one, distrust in large tech corporations or fear of misappropriation may otherwise deter users from contributing. Moreover, as discussed in Sec. 5.3.1, digitizing Indigenous data can become a new form of colonialism if non-community members appropriate cultural resources to train AI systems without any benefit being returned to Indigenous communities [87, 145, 169, 169].

In the current design, we present the tool as an independent browser extension that is interoperable across multiple generative AI platforms. We note that partnering with existing model or educational technology providers is an alternative choice that could ensure easier maintenance. However, based on survey results from Sec. 5.1, teachers' generative AI workflows are distributed across many platforms, requiring auditing tools that provide more flexibility. Again, given the emphasis on ensuring data sovereignty in our workshops, building tools that can be hosted and controlled locally are a top priority.

5.4.2 Implementation Challenges and Considerations. We conclude by surveying which aspects of our design vignette are feasible in the short term and identify areas where future work can contribute. As discussed in Sec. 5.3.1, source attribution in LLMs remains a significant technical challenge and represents a growing area of research [76, 152, 158]. While reliable attribution may remain difficult, retrieving related sources may be possible using existing NLP and information retrieval techniques [158]. Beyond technical barriers, however, there are substantial sociotechnical challenges in implementing such a system. A central, and contested, question is who should govern or maintain it. While our findings suggest that control should rest with the community, there was no clear consensus on what "community governance" should look like. Feedback from
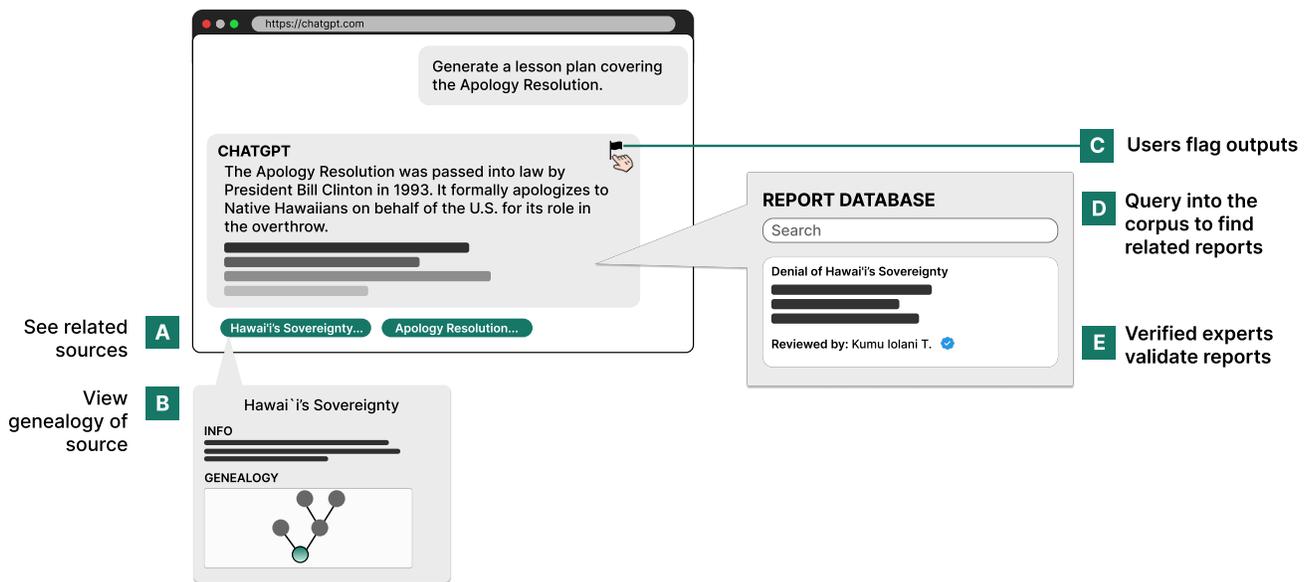
Figure 6: We present an exemplar auditing tool based on the authors distilling the prototyped features, discussions, and context into a cohesive system. The tool is proposed to be built on top of existing AI platforms (e.g., ChatGPT, Gemini). Users can view if there are sources related to the generated output (A) and view more information about who authored the source as well as the genealogy of knowledge (B). If a user is concerned about the output, they can flag it to enter into the report database (C). Newcomers or less experienced auditors can also participate by comparing outputs to preexisting reports in the database (D) and flagging potentially harmful results that are reviewed by more experienced community members (E).

our workshops reflected these tensions: participants in Workshop 4 favored oversight by state education officials (e.g., the Office of Hawaiian Education); Workshop 3 participants preferred avoiding any government or political involvement; and Workshop 1 participants expressed interest in a more decentralized, school-by-school approach.

6 Discussion

Our findings show that educators in Hawaiʻi face competing dynamics around the use of generative AI. On one hand, participants described top-down encouragement from the state to integrate these tools, reflecting an institutional push toward adoption. At the individual level, many workshop participants expressed interest in using generative AI for many reasons, including saving time, creating more localized education content, and offering more interactive classroom activities. At the same time, however, participants voiced a wide range of concerns about the potential harms these technologies may introduce, particularly when applied to tasks related to cultural knowledge. Given that Hawaiian culture is a mandated part of the curriculum, these shortcomings of generative AI are not a hypothetical risk but rather an inevitable problem with which participants need to reckon. These tensions highlight the gap between the push to adopt AI and the lack of safeguards ensuring its appropriateness. While end-user auditing provides a general approach to bridge this gap, our workshops underscore that participants require tools designed with Hawaiian values at the forefront, rather than relying on general-purpose solutions.

While our study is situated within the cultural and educational context of Hawaiʻi, we speak to how our findings offer insights that might inform end-user auditing practices for other marginalized communities. Although the specific harms and designs that participants reported are not generalizable, our takeaway that we must consider *community-specific* designs for end-user auditing has broader implications. To date, most work on auditing tools has focused on creating more general-purpose tools that are intended to work across many settings and for many groups of people [36, 115]. However, as Bird [17] argued about NLP methods more broadly, this "one-size-fits-all" approach problematically treats language as a decontextualized technical issue to solve, stripping away how language is actually used by people within specific communities. This operationalization may not account for the needs, values, or ways of knowing within a specific community, especially those that have been historically marginalized.

Within HCI, decoloniality has provided a useful framework for understanding how we can center a plurality of perspectives, rather than assuming a universalist default [135, 164, 165]. In particular, Alvarado Garcia et al. [5] provide an agenda for decolonizing HCI research, listing five pathways (understanding the why, reconsidering the how, changing the for whom, expanding the what, and reflecting on the what for) that can be applied to reorient how we design sociotechnical systems. Our following discussion will touch on three of the pathways in the context of end-user auditing,

discussing how to (1) accommodate diverse perspectives when auditing; (2) develop infrastructures that map to community values; and (3) ensure audit outputs benefit the community.

6.1 Determining who conducts community-centered audits

Our findings illustrate that auditing for harms in AI systems requires situated knowledge that individuals external to that community may lack [57, 126]. For example, as a participant in Workshop 1 noted when observing a bias probe: "*we looked at that [generated] picture and we're like that's definitely not a picture of a Hawaiian classroom, but somebody in Kansas, they might think and say yeah, that's what a classroom looks like because they've never come here.*" This point echoes calls from recent HCI work advocating for adopting more participatory methods in AI auditing by involving end users, such as the educators from our workshops, as auditors [35, 115]. Moreover, our workshops revealed that within communities, we should not expect participation to be uniform: differences in cultural expertise — such as between Kānaka or Kamaʻāina educators and those recently arrived from the mainland — shaped who could recognize harms. For example, while most participants in our workshops found the design probe related to "pineapple picking" to be offensive, some participants in Workshop 2 stated that they "*don't see it being biased.*"

This raises an important question of how to design for the diversity of community members' perspectives and experiences in auditing systems. As Alvarado Garcia et al. [5] argue, the decolonial path of "expanding the what" requires embracing multiple frames of reference and supporting plural knowledge systems. While leveraging users' diverse range of lived expertise is a key benefit for end-user auditing, if applied uncritically, it also creates a veneer of equality as not all community members bring the same forms of knowledge. Rather than treating auditing as a generic task that can be performed by any user, these asymmetries ought to shape whose judgments carry weight when conducting audits. At the same time, it is essential to ensure that the responsibility of auditing does not disproportionately fall on marginalized community members [143, 160]. Thus, auditing systems must center the epistemic authority of those most directly connected to the cultural contexts at stake, while still providing pathways for allies to participate.

Design Recommendations. We provide two design recommendations for how community-centered audits can explicitly account for these differences in their design. First, auditing tasks should be designed to accommodate different tiers of involvement or "trust" rather than assuming auditors will come in with equal experiences. Following practices from situated learning [84], those with less experience can start with less involved tasks. After observing experts and demonstrating that they can be trusted with more involved tasks, they can take on more key roles. How these levels of trust are operationalized will depend on the community. Examples could include delineating based on task type (e.g., more interpretive work versus fact-based validation) or issue severity. Second, auditing systems must include mechanisms to handle disagreements and provide pathways for discussion when auditors' judgments conflict [35]. Especially when dealing with AI outputs that do not have

a definitive right or wrong answer, such as is in our case when dealing with cultural artifacts, disagreements are both inevitable and informative. Rather than treating these divergences as errors, systems should scaffold structured dialogue [35, 137], such as by allowing auditors to document their reasoning, surface alternative interpretations, and view how others within the community have assessed similar cases. This not only supports transparency but also fosters collective sense-making, enabling auditing practices to better reflect the diversity of knowledge present even within communities.

6.2 Developing infrastructures to support community audits

While efforts to make AI auditing more participatory have largely focused on including end users in evaluating model outputs, we argue for extending this lens to the design of audit tools themselves. The technologies that these audits rely on also imbue their own standards and politics in the same way that individuals' positionality inevitably shapes auditing outputs [162]. Relying on one-size-fits-all infrastructures risks imposing universalist notions of what constitutes a harmful or undesired output that may not align with the communities being served. As illustrated in our workshops, points of friction arise when general-purpose tooling fails to accommodate what our participants need to conduct audits. For example, many participants expressed interest in source auditing. Ostensibly, knowledge attribution or retrieval-based systems could provide this functionality; however, conversations with participants revealed that what they needed was not only knowing what the source says but also its genealogy. Since cultural content, such as moʻolelo and proverbs, are open to multiple interpretations, participants emphasized that assessing trustworthiness depends on knowing *whose* interpretation is being surfaced. In contrast, adopting a decolonial approach, instead, asks how we can move from this mindset of universality towards one of pluriversality, ensuring that auditing tools embed the local values of the communities they are intended to serve. This ties to Alvarado Garcia et al. [5]'s pathway of "reconsidering the how" which invites us to think about how we can orient our methodologies to match what the communities we are engaging with want or value. In addition, this approach connects to Bird [17]'s framework for non-extractive NLP, which argues for building language technologies that grant agency to the communities from which they come, by leveraging cultural resources such as elders and engaging the community where they are at.

Design Recommendations. As a concrete recommendation, we propose expanding the frame of participatory auditing: from inviting communities to perform audits toward enabling them to shape the infrastructures that make auditing possible. How do we enact this in practice? In this work, we adopt practices from participatory design (PD) and conduct co-design sessions. However, this method poses its own set of challenges. Prior work in HCI has documented the limitations and even harms of engaging in PD, especially when working with historically marginalized communities [59]. There are also concerns about how to maintain any tools built going forward, especially if they are created by researchers who may not be incentivized to invest in long-term maintenance [79]. Expanding

on the exercises from our design workshops, we advocate for future work to draw on speculative design approaches, which has proven fruitful for imagining alternative futures particularly for historically marginalized groups [58, 163]. In conjunction to extending PD methods to creating auditing tools, we advocate for further explorations into equipping community members with the means to tool build themselves, allowing for community self-determination, rather than relying on researchers or model practitioners as intermediaries. We recognize that building tools requires more technique overhead; however, we see opportunities in leveraging advances in generative AI technologies to lower these barriers. For example, prompting and other lightweight interaction modalities, such as audio-based interfaces, could make it more feasible for people to create or customize auditing tools without requiring deep technical expertise. Already, there are initial explorations of how to support educators in developing their own AI tooling, which can be extended into the auditing context [60].

6.3 Reconsidering valued outcomes of AI audits

Finally, we question *who* benefits from AI audits, mapping to the pathway about changing for whom HCI research is being done [5]. At the moment, most AI auditing endeavors focuses on offering technical remediation and improving the model output [105, 128, 130]. However, this design centralizes power in the hands of the platform or model providers who can decide whether to address these harms. It also leaves community members uncertain of whether they will benefit from their labor. Furthermore, centering technical solutions presupposes that AI systems ought to be used and simply need remediation, when it is possible that refusal, or choosing not to not use AI altogether, may be the best course of action in the given situation. These considerations leads us to the question "*who are these audits intended to serve? Is the goal of the audit to improve the model's performance, or is it to help the community*? " In fact, participants in our workshops rarely expressed interest in influencing platforms holistically. Instead, they imagined auditing features that supported locally consequential decisions, such as whether a lesson plan was appropriate for a classroom or whether to use a generative AI system at all.

Design Recommendations. Drawing on Shahid and Vashistha [135]'s work on decolonial approaches to content moderation, we adopt their framing that the goal of such practices should be to "repair, educate, and sustain communities." Applying this perspective to AI auditing, we provide design recommendations about how to create tools that can benefit communities directly. An important aspect of this is designing auditing tools that incorporate educational opportunities for community members. Already in our workshops, we saw how participants viewed auditing as a useful educational exercise for their students, encouraging them to engage more deeply with generated outputs rather than taking the content at face value. Lighter-weight interventions may involve reflection prompts that periodically appear as a user is interacting with a generative AI system that asks them to think about what assumptions the model may be making when generating a response. A more heavy-weight approach could be building multi-tiered auditing tools, similar to our exemplar in Sec. 5.4, that can support community members in identifying problematic outputs while also scaffolding learning

for students or newcomers. Prior work [144] has also suggested integrating best practices from learning sciences in the design of auditing tools to help facilitate the process of learning by doing. While there are many design opportunities in this direction, the underlying principle is to build auditing tools that, first and foremost, directly benefit the communities using them.

7 Limitations and Broader Ethical Reflections

In this study we sought to make observations and recommendations that could apply generally to practitioners seeking to build AI auditing tools for educators in Hawaiʻi. To support this, we conducted workshops with teachers located in Hawaiʻi. Due to the focused nature of our study, the scope of the participant pool is a limitation to our findings. Our workshops were all conducted with elementary-level public school teachers located in Oʻahu. Teachers on other islands, private or charter school teachers, and teachers at different grade levels would have different concerns about using AI in educational contexts. In addition, although our sample size is in line with similar works at CHI, the limited participant size (N=22) can hinder generalizability. Furthermore our participants optionally chose to participate in our study after school hours, so the population of participants at each school was self-selecting. This may have led to a participant population that was more familiar or willing to engage with AI systems, whereas teachers who are less familiar or even opposed to AI may have opted not to participate. Finally, we center educators' perspectives in this study; however, there are other stakeholders — students, administrators, cultural practitioners — who are not represented in our workshops. Thus, it remains future work to confirm how these observations align with shared values across a broader spectrum of educators in Hawaiʻi.

Another limitation of this work is our focus on text-based generative AI systems. The proposed designs in Sec. 5.3 and design vignette are primarily intended for natural language input and output. This decision reflects how participants are primarily using generative AI systems. We expect that the auditing needs, harm types, and required technical solutions will vary across modalities (e.g., for visual or for audio-based systems), necessitating different mechanisms for surfacing harm which are not explored in depth within this work.

Finally, we want to provide a reflection on the broader ethical considerations of this work. Initially, the missive of this project was to build a general-purpose auditing tool that could be used by educators in Hawaiʻi. However, after engaging in workshops, it became clear that a generalizable approach to cultural auditing would be detrimental to the community's needs. As described by participants, cultural knowledge is embedded in localized, generational practices, and access to this knowledge requires building deep relationships and trust with community members. Hence, attempts to abstract these culturally significant practices into a generalized tool deployed across communities risks misrepresenting or flattening the epistemologies they aim to preserve. These considerations shifted us away from designing a generalizable tool, and instead ideate on frameworks that support community-centered auditing infrastructures. Ultimately, we argue the decision to build, or not to build, should be made not only considering technical feasibility, but through careful reflection about whether the system can be

sustained over time and most importantly whether it provides long-lasting benefits to the community beyond the scope of an academic paper.

8 Conclusion

Overall, this work sheds light on educator concerns surrounding AI in the classroom and introduces a framework for co-developing community-centered auditing tools for Indigenous and low-resource contexts. Through four co-design workshops with 22 public school educators in Oʻahu, Hawaiʻi, we identified key worries around AI use in education — particularly regarding cultural representation — and surfaced auditing practices educators envision to address these issues. Drawing on these insights, we argue that AI auditing should be framed as a *community practice*: one actively shaped by the people and communities most affected by these technologies. This reframing invites us to reconsider who serves as an auditor within this process, how auditing infrastructures ought to be designed, and what outcomes should be valued in the auditing process.

Acknowledgments

References

[1] National Park Service 2023. *ʻŌlelo Hawaiʻi: Hawaiian Language*. National Park Service. https://www.nps.gov/havo/learn/historyculture/olelo-hawaii.htm Hawaiʻi Volcanoes National Park, accessed August 3, 2025.

[2] Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Mohamed Ahmed, et al. 2023. MEGA: Multilingual Evaluation of Generative AI. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[3] Syed Mustafa Ali. 2016. A brief introduction to decolonial computing. *XRDS: Crossroads, The ACM Magazine for Students* 22, 4 (2016), 16–21.

[4] Marc Alier, María José Casañ, and Daniel Amo Filvà. 2023. Smart Learning Applications: Leveraging LLMs for Contextualized and Ethical Educational Technology. In *International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM)*.

[5] Adriana Alvarado Garcia, Juan F Maestre, Manuhuia Barcham, Marilyn Iriarte, Marisol Wong-Villacres, Oscar A Lemus, Palak Dudani, Pedro Reynolds-Cuéllar, Ruotong Wang, and Teresa Cerratto Pargman. 2021. Decolonial pathways: Our manifesto for a decolonizing agenda in HCI research and design. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA)*.

[6] Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios. 2020. Auditing race and gender discrimination in online housing markets. In *International AAAI Conference on Web and Social Media (ICWSM)*.

[7] Zinzi D Bailey, Nancy Krieger, Madina Agénor, Jasmine Graves, Natalia Linos, and Mary T Bassett. 2017. Structural racism and health inequities in the USA: evidence and interventions. *The Lancet* 389, 10077 (2017), 1453–1463.

[8] Ryan S Baker and Aaron Hawn. 2022. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education* 32, 4 (2022), 1052–1092.

[9] Rachel Baker-Ramos, Will Gelder, Leah Cho, Jahnavi Kolakaluri, and Josiah Hester. 2025. Kumu Connect: Design Thinking for Place-Based Generative Educational Technology in Hawaiian Immersion Schools. In *Conference on Research on Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*.

[10] Matt Barnum. 2024. We tested an AI tutor for kids. It struggled with basic math. *The Wall Street Journal* (2024). https://www.wsj.com/tech/ai/ai-is-tutoring-students-but-still-struggles-with-basic-math-694e76d3

[11] Noman Bashir, Priya Donti, James Cuff, Sydney Sroka, Marija Ilic, Vivienne Sze, Christina Delimitrou, and Elsa Olivetti. 2024. The Climate and Sustainability Implications of Generative AI. *An MIT Exploration of Generative AI* (mar 27 2024). https://mit-genai.pubpub.org/pub/8ulgrckc.

[12] Ahmet Baytak. 2024. The Content Analysis of the Lesson Plans Created by ChatGPT and Google Gemini. *Research in Social Sciences and Technology* 9, 1 (2024), 329–350.

[13] Kamanamaikalani Beamer, Kawena Elkington, Pua Souza, Axel Tuma, Andrea Thorenz, Sandra Köhler, Kāneoka Kukea-Shultz, Keliʻi Kotubetey, and Kawika B Winter. 2023. Island and Indigenous systems of circularity: how Hawaiʻi can inform the development of universal circular economy policy goals. *Ecology and Society* 28 (2023). Issue 1.

[14] Richard A Berk. 2021. Artificial intelligence, predictive policing, and risk assessment for law enforcement. *Annual Review of Criminology* 4, 1 (2021), 209–237.

[15] Nicola J Bidwell. 2021. Decolonising in the gaps: Community networks and the identity of African innovation. In *Re-imagining Communication in Africa and the Caribbean: Global South Issues in Media, Culture and Technology*. Springer, 97–115.

[16] Steven Bird. 2020. Decolonising speech and language technology. In *International Conference on Computational Linguistics (COLING)*.

[17] Steven Bird. 2024. Must NLP be Extractive?. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

[18] Abeba Birhane and Olivia Guest. 2021. Towards decolonising computational sciences. *Kvinder, Køn & Forskning* 29, 2 (2021), 60–73.

[19] Abeba Birhane, Ryan Steed, Victor Ojewale, Briana Vecchione, and Inioluwa Deborah Raji. 2024. AI auditing: The broken bus on the road to AI accountability. In *IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*.

[20] Zach Boblitt. 2025. Indigenous environmental advocates say data centers risk water, culture and informed consent. https://www.publicradiotulsa.org/local-regional/2025-10-07/indigenous-environmental-advocates-say-data-centers-risk-water-culture-and-informed-consent

[21] Andrea Botero and Sampsa Hyysalo. 2013. Ageing together: Steps towards evolutionary co-design in everyday practices. *CoDesign* 9, 1 (2013), 37–54.

[22] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency (FAccT)*.

[23] Beatriz Cabellos, Carlos De Aldama, and Juan-Ignacio Pozo. 2024. University teachers' beliefs about the use of generative artificial intelligence for teaching and learning. *Frontiers in Psychology* 15 (2024).

[24] Miguel A. Cardona, Roberto J. Rodríguez, and Kristina Ishmael. 2023. *Artificial Intelligence (AI) and the Future of Teaching and Learning: Insights and Recommendations*. Technical Report. U.S. Department of Education, Office of Educational Technology (OET). https://www2.ed.gov/documents/ai-report/ai-report.pdf

[25] Rogelio E Cardona-Rivera, J Kaleo Alladin, Breanne K Litts, and Melissa Tehee. 2024. Indigenous futures in generative artificial intelligence: the paradox of participation. *Teaching and Generative AI* (2024).

[26] Kaavya D Chaparala, Guido Zarrella, Bruce Torres Fischer, Larry Kimura, and Oiwi Parker Jones. 2024. Mai Ho ʻomāuna i ka ʻAi: Language Models Improve Automatic Speech Recognition in Hawaiian. In *NeurIPS Workshop on Efficient Natural Language and Speech Processing Workshop (ENLSP)*.

[27] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the impact of gender on rank in resume search engines. In *ACM CHI Conference on Human Factors in Computing Systems*.

[28] Xinxi Chen, Li Wang, Wei Wu, Qi Tang, and Yiyao Liu. 2024. Honest AI: Fine-Tuning "Small" Language Models to Say "I Don't Know", and Reducing Hallucination in RAG. In *KDD Cup Workshop for Retrieval Augmented Generation*.

[29] Victoria Clarke and Virginia Braun. 2014. Thematic analysis. In *Encyclopedia of critical psychology*. Springer, 1947–1952.

[30] Rolando Coto-Solano, Sally Akevai Nicholas, Samiha Datta, Victoria Quint, Piripi Wills, Emma Ngakuravaru Powell, Liam Kokaʻua, Syed Tanveer, and Isaac Feldman. 2022. Development of Automatic Speech Recognition for the Documentation of Cook Islands Māori. In *Language Resources and Evaluation Conference (LREC)*.

[31] Henriette Cramer, Jean Garcia-Gathright, Aaron Springer, and Sravana Reddy. 2018. Assessing and addressing algorithmic bias in practice. *Interactions* 25, 6 (2018), 58–63.

[32] Stefany Cruz, Alexander Redding, Connie W Chau, Claire Lu, Julia Persche, Josiah Hester, and Maia Jacobs. 2023. EquityWare: Co-designing wearables with and for low income communities in the US. In *ACM CHI Conference on Human Factors in Computing Systems*.

[33] Wei Dai, Jionghao Lin, Hua Jin, Tongguang Li, Yi-Shan Tsai, Dragan Gašević, and Guanliang Chen. 2023. Can large language models provide feedback to students? A case study on ChatGPT. In *IEEE International Conference on Advanced Learning Technologies (ICALT)*. IEEE.

[34] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41, 6 (1990), 391–407.

[35] Wesley Hanwen Deng, Wang Claire, Howard Ziyu Han, Jason I Hong, Kenneth Holstein, and Motahhare Eslami. 2025. Weaudit: Scaffolding user auditors and ai practitioners in auditing generative ai. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–35.

[36] Wesley Hanwen Deng, Boyuan Guo, Alicia Devrio, Hong Shen, Motahhare Eslami, and Kenneth Holstein. 2023. Understanding practices, challenges, and opportunities for user-engaged algorithm auditing in industry practice. In *ACM CHI Conference on Human Factors in Computing Systems*.

[37] Alicia DeVos, Aditi Dhabalia, Hong Shen, Kenneth Holstein, and Motahhare Eslami. 2022. Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *ACM CHI Conference on Human Factors in Computing Systems*.

[38] Tawanna R Dillahunt and Amelia R Malone. 2015. The promise of the sharing economy among disadvantaged communities. In *ACM CHI Conference on Human Factors in Computing Systems*.

[39] Esin Durmus, Karina Nguyen, Thomas I Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, Zac Hatfield-Dodds, Danny Hernandez, Nicholas Joseph, et al. 2024. Towards measuring the representation of subjective global opinions in language models. In *Conference on Language Modeling (COLM)*.

[40] Nina-Simone Edwards. 2025. Unveiling the Environmental Impact of Large Language Models on Indigenous Communities. *Tulane Environmental Law Journal* 38, 1 (2025), 1–27.

[41] Susan Enright. 2025. *UH Hilo Hawaiian language college turns to AI to help secure the future of ʻōlelo Hawaiʻi*. https://hilo.hawaii.edu/chancellor/stories/2025/02/26/hawaiian-language-college-turns-to-ai/

[42] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios, and Kevin Hamilton. 2017. "Be careful; things can be worse than they appear"': Understanding biased algorithms and users' behavior around them in rating platforms. In *International AAAI Conference on Web and Social Media (ICWSM)*.

[43] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In *ACM CHI Conference on Human Factors in Computing Systems*.

[44] Haoxiang Fan, Guanzheng Chen, Xingbo Wang, and Zhenhui Peng. 2024. Lesson-Planner: Assisting novice teachers to prepare pedagogy-driven lesson plans with large language models. In *Annual ACM Symposium on User Interface Software and Technology (UIST)*.

[45] Miranda Fricker. 2007. *Epistemic injustice: Power and the ethics of knowing*. Oxford university press.

[46] Gallup. 2025. *Teaching for Tomorrow: Unlocking Six Weeks a Year With AI*. Technical Report. Gallup and Walton Family Foundation. https://www.gallup.com/file/analytics/691922/Walton-Family-Foundation-Gallup-Teachers-AI-Report.pdf

[47] Catherine A Gao, Frederick M Howard, Nikolay S Markov, Emma C Dyer, Siddhi Ramesh, Yuan Luo, and Alexander T Pearson. 2023. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine* 6, 1 (2023), 75.

[48] Tray Geiger and Margarita Pivovarova. 2018. The effects of working conditions on teacher retention. *Teachers and Teaching* 24, 6 (2018), 604–625.

[49] William Gelder, Rachel Baker-Ramos, Ayoung Cho, Jahnavi Kolakaluri, Judith Uchidiuno, and Josiah Hester. 2024. "Those don't work for us"': An Assets-Based Approach to Incorporating Emerging Technologies in Viable Hawaiian Teacher Support Tools for Culturally Relevant CS Education. In *Conference on Research on Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*.

[50] Sourojit Ghosh, Nina Lutz, and Aylin Caliskan. 2024. "I Don't See Myself Represented Here at All': User Experiences of Stable Diffusion Outputs Containing Representational Harms across Gender Identities and Nationalities. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*.

[51] Sourojit Ghosh, Pranav Narayanan Venkit, Sanjana Gautam, Shomir Wilson, and Aylin Caliskan. 2024. Do generative AI models output harm while representing non-Western cultures: Evidence from a community-centered approach. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*.

[52] Melissa Ozlem Grab. 2025. Teaching for Equity: An Exploration of AI's Role in Culturally Responsive Teaching in Higher Education Settings. *Innovative Higher Education* (2025), 1–22.

[53] Carlos Guerrero Millan, Bettina Nissen, and Larissa Pschetz. 2024. Cosmovision Of Data: An Indigenous Approach to Technologies for Self-Determination. In *ACM CHI Conference on Human Factors in Computing Systems*.

[54] Mascha Gugganig. 2021. Hawaiʻi as a laboratory paradise: Divergent sociotechnical island imaginaries. *Science as Culture* 30, 3 (2021), 342–366.

[55] Edward Smith Handy, Elizabeth Green Handy, and Mary Kawena Pukui. 1972. *Native planters in old Hawaii: Their life, lore, and environment*. Bishop Museum Press.

[56] Violet Harada. 2016. The Power of Place-Based Learning: Caring for Our Island Earth. https://www2.hawaii.edu/~vharada/PDF/works/2016-PlaceBased.pdf Accessed: July 20, 2025.

[57] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575–599.

[58] Christina Harrington and Tawanna R Dillahunt. 2021. Eliciting tech futures among Black young adults: A case study of remote speculative co-design. In *ACM CHI Conference on Human Factors in Computing Systems*.

[59] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019).

[60] Emma Harvey, Allison Koenecke, and Rene F Kizilcec. 2025. "Don't Forget the Teachers"': Towards an Educator-Centered Understanding of Harms from Large Language Models in Education. In *ACM CHI Conference on Human Factors in Computing Systems*.

[61] Md Arid Hasan, Prerona Tarannum, Krishno Dey, Imran Razzak, and Usman Naseem. 2024. Do large language models speak all languages equally? a comparative study in low-resource settings. *arXiv preprint arXiv:2408.02237* (2024).

[62] Reem Hashem, Nagla Ali, Farah El Zein, Patricia Fidalgo, and Othman Abu Khurma. 2024. AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research & Practice in Technology Enhanced Learning* 19 (2024).

[63] Hawaiʻi State Department of Education. 2025. Artificial Intelligence. https://hawaiipublicschools.org/student-programs/artificial-intelligence/. Accessed on July 23, 2025.

[64] Hawaiʻi State Department of Education, Office of Talent Management. 2024. *Employment Report: School Year 2023–24*. Technical Report. Hawaiʻi State Department of Education. https://hawaiipublicschools.org/wp-content/uploads/Employment-Report-SY-2023-24.pdf

[65] William Held, Camille Harris, Michael Best, and Diyi Yang. 2023. A material lens on coloniality in NLP. *arXiv preprint arXiv:2311.08391* (2023).

[66] HISTORY.com Editors. 2025. Captain Cook killed in Hawaii. https://www.history.com/this-day-in-history/february-14/captain-cook-killed-in-hawaii. Accessed on August 24, 2025.

[67] Daniel Hoffman, Seungoh Paek, Peter Leong, and Rochelle Piʻilani Kaʻaloa. 2023. Advancing Culturally-Relevant Computing through a Researcher-Practitioner Partnership: Kumu Perspectives on LLMs for Culturally Revitalizing CS Education in Hawaiian Schools. In *EdMedia + Innovate Learning 2023*. Association for the Advancement of Computing in Education (AACE).

[68] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In *ACM CHI Conference on Human Factors in Computing Systems*.

[69] Bihao Hu, Longwei Zheng, Jiayi Zhu, Lishan Ding, Yilei Wang, and Xiaoqing Gu. 2024. Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. *IEEE Transactions on Learning Technologies* 17 (2024), 1445–1459.

[70] Nidhal Jegham, Marwan Abdelatti, Chan Young Koh, Lassad Elmoubarki, and Abdeltawab Hendawi. 2025. How hungry is AI? benchmarking energy, water, and carbon footprint of llm inference. *arXiv preprint arXiv:2505.09598* (2025).

[71] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surveys* 55, 12 (2023).

[72] Young J Juhn, Euijung Ryu, Chung-Il Wi, Katherine S King, Momin Malik, Santiago Romero-Brufau, Chunhua Weng, Sunghwan Sohn, Richard R Sharp, and John D Halamka. 2022. Assessing socioeconomic bias in machine learning algorithms in health care: a case study of the HOUSES index. *Journal of the American Medical Informatics Association (JAMIA)* 29, 7 (2022), 1142–1151.

[73] Muhammet Remzi Karaman et al. 2024. Are Lesson Plans Created by ChatGPT More Effective? An Experimental Study. *International Journal of Technology in Education* 7, 1 (2024), 107–127.

[74] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023).

[75] Frank Kehoe. 2023. Leveraging generative AI tools for enhanced lesson planning in initial teacher education at post primary. *Irish Journal of Technology Enhanced*

Learning 7, 2 (2023), 172–182.

[76] Brandon Khoo, Raphaël C-W Phan, and Chern-Hong Lim. 2022. Deepfake attribution: On the source identification of artificially generated images. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 12, 3 (2022).

[77] Shana Klein. 2025. "The Kingdom Grown Out of a Little Boy's Garden": Dole Pineapples and Hawaiian Occupation in US Art and Visual Culture. In *The Routledge Companion to Art and the Formation of Empire*. Routledge, 44–63.

[78] Jake Knapp, John Zeratsky, and Braden Kowitz. 2016. *Sprint: How to solve big problems and test new ideas in just five days.* Simon and Schuster.

[79] Yasmine Kotturi, Julie Hui, TJ Johnson, Lutalo Sanifu, and Tawanna R Dillahunt. 2024. Sustaining Community-Based Research in Computing: Lessons from Two Tech Capacity Building Initiatives for Local Businesses. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW (2024).

[80] Lindah Kotut and D Scott McCrickard. 2022. Winds of change: Seeking, preserving, and retelling indigenous knowledge through self-organized online communities. In *ACM CHI Conference on Human Factors in Computing Systems.*

[81] Dilek Küçük and Fazli Can. 2020. Stance Detection: A Survey. *ACM Comput. Surv.* 53, 1, Article 12 (Feb. 2020).

[82] Sabina Lacmanovic and Marinko Skare. 2025. Artificial intelligence bias auditing–current approaches, challenges and lessons from practice. *Review of Accounting and Finance* 24, 3 (2025), 375–400.

[83] Michelle S Lam, Mitchell L Gordon, Danaë Metaxa, Jeffrey T Hancock, James A Landay, and Michael S Bernstein. 2022. End-user audits: A system empowering communities to lead large-scale investigations of harmful algorithmic behavior. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW (2022).

[84] Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation.* Cambridge University Press.

[85] Shaimaa Lazem, Danilo Giglitto, Makuochi Samuel Nkwo, Hafeni Mthoko, Jessica Upani, and Anicia Peters. 2022. Challenges and paradoxes in decolonising HCI: A critical discussion. *Computer Supported Cooperative Work (CSCW)* 31, 2 (2022), 159–196.

[86] Christopher A. Le Dantec and Sarah Fox. 2015. Strangers at the Gate: Gaining Access, Building Rapport, and Co-Constructing Community-Based Research. In *ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*. Association for Computing Machinery.

[87] Brandon Ledward, Brennan Takayama, and Walter Kahumoku III. 2008. *Hawaiian Cultural Influences in Education (HCIE): 'Ohana and Community Integration in Culture-Based Education.* Technical Report. https://www.ksbe.edu/assets/research/collection/08_0128_ledward.pdf Accessed on July 12, 2025.

[88] Gyeong-Geon Lee and Xiaoming Zhai. 2024. Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies* 17 (2024), 1643–1660.

[89] Suevon Lee. 2019. *The Challenges Of Finding Hawaii's Next Generation Of Teachers.* https://www.civilbeat.org/2019/05/the-challenges-of-finding-hawaiis-next-generation-of-teachers/

[90] Jason Edward Lewis, Hēmi Whaanga, and Ceyda Yolgörmez. 2025. Abundant intelligences: placing AI within Indigenous knowledge frameworks. *Ai & Society* 40, 4 (2025), 2141–2157.

[91] Zhaoming Liu. 2025. Cultural bias in large language models: A comprehensive analysis and mitigation strategies. *Journal of Transcultural Communication* 3, 2 (2025), 224–244.

[92] Kelly Avery Mack, Rida Qadri, Remi Denton, Shaun K Kane, and Cynthia L Bennett. 2024. "They only care to show us the wheelchair": disability representation in text-to-image AI models. In *ACM CHI Conference on Human Factors in Computing Systems.*

[93] Suvradip Maitra. 2020. Artificial intelligence and indigenous perspectives: Protecting and empowering intelligent human beings. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES).*

[94] David Malo. 2021. *Hawaiian Antiquities (Moolelo Hawaii).* Mint Editions. Originally published 1903.

[95] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019).

[96] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human–Computer Interaction* 14, 4 (2021), 272–344.

[97] Jennifer Meyer, Thorben Jansen, Ronja Schiller, Lucas W Liebenow, Marlene Steinbach, Andrea Horbach, and Johanna Fleckenstein. 2024. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence* 6 (2024).

[98] Manas Mhasakar, Rachel Baker-Ramos, Benjamin Carter, Evyn-Bree Helekahi-Kaiwi, and Josiah Hester. 2025. "I Would Never Trust Anything Western'": Kumu (Educator) Perspectives on Use of LLMs for Culturally Revitalizing CS Education in Hawaiian Schools. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA).*

[99] Walter D Mignolo. 2007. Delinking: The rhetoric of modernity, the logic of coloniality and the grammar of de-coloniality. *Cultural Studies* 21, 2-3 (2007), 449–514.

[100] Nusrat Jahan Mim, Dipannita Nandi, Sadaf Sumyia Khan, Arundhuti Dey, and Syed Ishtiaque Ahmed. 2024. In-Between Visuals and Visible: The Impacts of Text-to-Image Generative AI Tools on Digital Image-making Practices in the Global South. In *ACM CHI Conference on Human Factors in Computing Systems.*

[101] Mirja Mittermaier, Marium M Raza, and Joseph C Kvedar. 2023. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digital Medicine* 6, 1 (2023), 113.

[102] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, et al. 2024. ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024).

[103] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33, 4 (2020), 659–684.

[104] Marryam Yahya Mohammed, Sama Ayman Ali, Salma Khaled Ali, Ayad Abdul Majeed, and Ensaf Hussein Mohamed. 2025. Aftina: enhancing stability and preventing hallucination in AI-based Islamic fatwa generation using LLMs and RAG. *Neural Computing and Applications* (2025), 1–26.

[105] Jakob Mökander. 2023. Auditing of AI: Legal, ethical and technical approaches. *Digital Society* 2, 3 (2023), 49.

[106] Warmhold Jan Thomas Mollema. 2025. A taxonomy of epistemic injustice in the context of AI and the case for generative hermeneutical erasure. *AI and Ethics* (July 2025).

[107] Luis Morales-Navarro, Yasmin Kafai, Vedya Konda, and Danaë Metaxa. 2024. Youth as peer auditors: Engaging teenagers with algorithm auditing of machine learning applications. In *ACM Interaction Design and Children Conference (IDC).*

[108] Sukanya Kannan Moudgalya and Sai Swaminathan. 2024. Toward Data Sovereignty: Justice-oriented and Community-based AI Education. In *Conference on Research on Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT).*

[109] Adelina Moura and Ana Amélia A Carvalho. 2024. Teachers' perceptions of the use of artificial intelligence in the classroom. In *International Conference on Lifelong Education and Leadership for All (ICLEL).*

[110] Jackie Ng-Osorio. 2023. Learning from the Past to Grow the Future: Of 'Āina-Based Education. https://www.hauolimauloa.org/images/%CA%BBAina-Lit-Review_FINAL_05-03-23.pdf Accessed on July 20, 2025.

[111] Tuan D Nguyen, Chanh B Lam, and Paul Bruno. 2024. What do we know about the extent of teacher shortages nationwide? A systematic examination of reports of US teacher shortages. *Aera Open* 10 (2024).

[112] Chaima Njeh, Haïfa Nakouri, and Fehmi Jaafar. 2024. Enhancing rag-retrieval to improve llms robustness and resilience to hallucinations. In *International Conference on Hybrid Artificial Intelligence Systems.* Springer, 201–213.

[113] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453.

[114] Office of Hawaiian Education (OHE). 2015. Plan for Office of Hawaiian Education Priorities. https://www.hawaiipublicschools.org/DOE%20Forms/Hawaiian/OHE_DeliveryPlan.pdf. Accessed: 2025-07-23.

[115] Victor Ojewale, Ryan Steed, Briana Vecchione, Abeba Birhane, and Inioluwa Deborah Raji. 2025. Towards AI accountability infrastructure: Gaps and opportunities in AI audit tooling. In *ACM CHI Conference on Human Factors in Computing Systems.*

[116] Luci Pangrazio. 2024. Data harms: The evidence against education data. *Postdigital Science and Education* 6, 4 (2024), 1049–1054.

[117] Christina Ioanna Pappa, Despoina Georgiou, and Daniel Pittich. 2024. Technology education in primary schools: addressing teachers' perceptions, perceived barriers, and needs. *International Journal of Technology and Design Education* 34, 2 (2024), 485–503.

[118] Babette Park, Terri Flowerday, and Roland Brünken. 2015. Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior* 44 (2015), 267–278.

[119] Andrea G Parker and Rebecca E Grinter. 2014. Collectivistic health promotion tools: Accounting for the relationship between culture, food and nutrition. *International Journal of Human-Computer Studies* 72, 2 (2014), 185–206.

[120] Sachin R Pendse, Daniel Nkemelu, Nicola J Bidwell, Sushrut Jadhav, Soumitra Pathare, Munmun De Choudhury, and Neha Kumar. 2022. From treatment to healing: envisioning a decolonial digital mental health. In *ACM CHI Conference on Human Factors in Computing Systems.*

[121] Maneesha Perera, Rajith Vidanaarachchi, Sangeetha Chandrashekeran, Melissa Kennedy, Brendan Kennedy, and Saman Halgamuge. 2025. Indigenous peoples and artificial intelligence: A systematic review and future directions. *Big Data & Society* 12, 2 (2025), 20539517251349170.

Whose Knowledge Counts? Co-Designing Community-Centered AI Auditing Tools with Educators in Hawaiʻi

CHI '26, April 13–17, 2026, Barcelona, Spain

[122] Claudio S Pinhanez, Paulo R Cavalin, Marisa Vasconcelos, Julio Nogima, et al. 2023. Balancing Social Impact, Opportunities, and Ethical Constraints of Using AI in the Documentation and Vitalization of Indigenous Languages. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[123] Griffin Pitts, Viktoria Marcus, and Sanaz Motamedi. 2025. Student Perspectives on the Benefits and Risks of AI in Education. In *ASEE Annual Conference & Exposition*.

[124] Maureen K. Porter and Nik Cristobal. 2018. Cultivating Aloha ʻĀina Through Critical Indigenous Pedagogies of Place. *Journal of Folklore and Education* (2018). https://jfepublications.org/article/cultivating-aloha-aina/ Accessed on July 20, 2025.

[125] Snehal Prabhudesai, Ananya Prashant Kasi, Anmol Mansingh, Anindya Das Antar, Hua Shen, and Nikola Banovic. 2025. "Here the GPT made a choice, and every choice can be biased"': How Students Critically Engage with LLMs through End-User Auditing Activity. In *ACM CHI Conference on Human Factors in Computing Systems*.

[126] Rida Qadri, Mark Diaz, Ding Wang, and Michael Madaio. 2025. The Case for "Thick Evaluations"' of Cultural Representation in AI. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*.

[127] Hafizur Rahaman, Michelle Johnston, and Erik Champion. 2021. Audio-augmented arboreality: wildflowers and language. *Digital Creativity* 32, 1 (2021), 22–37.

[128] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*.

[129] Tasha Riley, Troy Meston, Chesley Cutler, Samantha Low-Choy, Brittany A McCormack, Eun-Ji Amy Kim, Sonal Nakar, and Daniela Vasco. 2024. Weaving stories of strength: Ethically integrating Indigenous content in teacher education and professional development programmes. *Teaching and Teacher Education* 142 (2024), 104513.

[130] Pedro Saleiro, Benedict Kuester, Loren Hinkson, Jesse London, Abby Stevens, Ari Anisfeld, Kit T. Rodolfa, and Rayid Ghani. 2019. Aequitas: A Bias and Fairness Audit Toolkit. arXiv:1811.05577 [cs.LG] https://arxiv.org/abs/1811.05577

[131] Anne Salmond. 2004. *The Trial of the Cannibal Dog: Captain Cook in the South Seas*. Penguin Books, London.

[132] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22, 2014 (2014), 4349–4357.

[133] Neil Selwyn, Marita Ljungqvist, and Anders Sonesson. 2025. When the prompting stops: Exploring teachers' work around the educational frailties of generative AI tools. *Learning, Media and Technology* (2025), 1–14.

[134] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew BA McDermott, Irene Y Chen, and Marzyeh Ghassemi. 2021. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine* 27, 12 (2021), 2176–2182.

[135] Farhana Shahid and Aditya Vashistha. 2023. Decolonizing content moderation: does uniform global community standard resemble utopian equality or western power hegemony?. In *ACM CHI Conference on Human Factors in Computing Systems*.

[136] Shikhar Sharma, Manas Mhasakar, Apurv Mehra, Utkarsh Venaik, Ujjwal Singhal, Dhruv Kumar, and Kashish Mittal. 2024. Comuniqa: Exploring large language models for improving english speaking skills. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*.

[137] Andrew Shaw, Andre Ye, Ranjay Krishna, and Amy Zhang. 2025. Agonistic Image Generation: Unsettling the Hegemony of Intention. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 438–463.

[138] Sarah B Shear, Ryan T Knowles, Gregory J Soden, and Antonio J Castro. 2015. Manifesting destiny: Re/presentations of indigenous peoples in K–12 US history standards. *Theory & Research in Social Education* 43, 1 (2015), 68–101.

[139] Renee Shelby, Shalaleh Rismani, and Negar Rostamzadeh. 2024. Generative AI in Creative Practice: ML-Artist Folk Theories of T2I Use, Harm, and Harm-Reduction. In *ACM Conference on Human Factors in Computing Systems*.

[140] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW (2021).

[141] Eugenia Siapera. 2022. AI content moderation, racism and (de) coloniality. *International journal of bullying prevention* 4, 1 (2022), 55–65.

[142] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation Is not a Design Fix for Machine Learning. In *ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAMMO)*. New York, NY, USA.

[143] William A Smith, Tara J Yosso, and Daniel G Solórzano. 2011. Challenging racial battle fatigue on historically White campuses: A critical race examination of race-related stress. *Covert racism* 32 (2011), 211–237.

[144] Jaemarie Solyst, Cindy Peng, Wesley Hanwen Deng, Praneetha Pratapa, Amy Ogan, Jessica Hammer, Jason Hong, and Motahhare Eslami. 2025. Investigating Youth AI Auditing. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*.

[145] Isabella Spano and Yuxiao Zhang. 2025. Indigenous data sovereignty in intangible cultural heritage governance: A complementary approach to public–private partnerships. *International Journal of Cultural Property* (2025).

[146] State of Hawaii Legislative Reference Bureau. n.d.. Hawaii Constitutional Convention Studies 1978. https://lrb.hawaii.gov/hawaii-constitutional-convention-studies-1978/. Accessed on August 25, 2025.

[147] NarayanKripa Sundararajan and Olusola Adesope. 2020. Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review* 32, 3 (2020), 707–734.

[148] Latanya Sweeney. 2013. Discrimination in online ad delivery. *Commun. ACM* 56, 5 (2013), 44–54.

[149] Kim TallBear. 2019. Caretaking relations, not American dreaming. *Kalfou: A Journal of Comparative and Relational Ethnic Studies* 6, 1 (2019), 24–41.

[150] Yan Tao, Olga Viberg, Ryan S Baker, and René F Kizilcec. 2024. Cultural bias and cultural alignment of large language models. *PNAS Nexus* 3, 9 (2024).

[151] Honor the Earth. 2026. No Data Centers on Native Land: Campaign Overview. https://www.honorearth.org/stopdatacolonialism

[152] Deepa Tilwani, Revathy Venkataramanan, and Amit P Sheth. 2024. Neurosymbolic ai approach to attribution in large language models. *IEEE Intelligent Systems* 39, 6 (2024), 10–17.

[153] Emily Tseng, Meg Young, Marianne Aubin Le Quéré, Aimee Rinehart, and Harini Suresh. 2025. "Ownership, Not Just Happy Talk"': Co-Designing a Participatory Large Language Model for Journalism. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*.

[154] U.S. Census Bureau. 2010. Hawaii — Guide to 2010 State and Local Census Geography. https://www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010/hawaii.html.

[155] Jacky Vallée. 2018. Eurocentrism in the curriculum: A barrier to Indigenous student success. (2018).

[156] Rama Adithya Varanasi and Nitesh Goyal. 2023. "It is currently hodgepodge": Examining AI/ML practitioners' challenges during co-production of responsible AI values. In *ACM CHI Conference on Human Factors in Computing Systems*.

[157] Tania Villalobos Lujan, Pratim Sengupta, Derya Akbaba, Sofia Alessio-Robles, Amelia Lee Dogan, Monica Meltis, Firuzeh Shokooh Valle, and Lora Oehlberg. 2025. Interaction Design as a Form of Decolonial Care. In *Conference on Creativity and Cognition (C&C)*.

[158] Jingtan Wang, Xinyang Lu, Zitong Zhao, Zhongxiang Dai, Chuan-Sheng Foo, See-Kiong Ng, and Bryan Kian Hsiang Low. 2024. Source Attribution for Large Language Model-Generated Data. arXiv:2310.00646 [cs.LG] https://arxiv.org/abs/2310.00646

[159] Athena Wen, Tanush Patil, Ansh Saxena, Yicheng Fu, Sean O'Brien, and Kevin Zhu. 2025. FAIRE: Assessing Racial and Gender Bias in AI-Driven Resume Evaluations. *arXiv preprint arXiv:2504.01420* (2025).

[160] Apryl A Williams, Zaida Bryant, and Christopher Carvell. 2019. Uncompensated emotional labor, racial battle fatigue, and (in) civility in digital spaces. *Sociology Compass* 13, 2 (2019).

[161] Kyra Wilson and Aylin Caliskan. 2024. Gender, race, and intersectional bias in resume screening via language model retrieval. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, Vol. 7.

[162] Langdon Winner. 1980. Do Artifacts Have Politics? *Daedalus* 109, 1 (1980), 121–136. http://www.jstor.org/stable/20024652

[163] Richmond Y Wong and Vera Khovanskaya. 2018. Speculative design in HCI: from corporate imaginations to critical orientations. In *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*. Springer, 175–202.

[164] Marisol Wong-Villacres, Adriana Alvarado Garcia, Juan F Maestre, Pedro Reynolds-Cuéllar, Heloisa Candello, Marilyn Iriarte, and Carl DiSalvo. 2020. Decolonizing learning spaces for sociotechnical research and design. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. 519–526.

[165] Marisol Wong-Villacres, Adriana Alvarado Garcia, and Javier Tibau. 2020. Reflections from the classroom and beyond: Imagining a decolonized hci education. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA)*.

[166] Marisol Wong-Villacres, Carl DiSalvo, Neha Kumar, and Betsy DiSalvo. 2020. Culture in action: Unpacking capacities to inform assets-based design. In *ACM CHI Conference on Human Factors in Computing Systems*.

[167] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2024. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology* 55, 1 (2024), 90–112.

[168] D Lilinoe Yong and Ellen S Hoffman. 2014. Teacher Technology Narratives: Native Hawaiian Views on Education and Change. *Qualitative Report* 19 (2014), 16.

[169] Jason C. Young. 2019. The New Knowledge Politics of Digital Colonialism. *Environment and Planning A: Economy and Space* 51, 7 (Oct. 2019), 1424–1441.

[170] Meg Young, Lassana Magassa, and Batya Friedman. 2019. Toward inclusive tech policy design: a method for underrepresented voices to strengthen tech policy documents. *Ethics and Information Technology* 21, 2 (2019), 89–103.

[171] Chunpeng Zhai, Santoso Wibowo, and Lily D Li. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments* 11, 1 (2024), 28.

[172] Lili Zhang, Xi Liao, Zaijia Yang, Baihang Gao, Chunjie Wang, Qiuling Yang, and Deshun Li. 2024. Partiality and misconception: Investigating cultural representativeness in text-to-image models. In *ACM CHI Conference on Human Factors in Computing Systems*.

[173] Tianyang Zhong, Zhenyuan Yang, Zhengliang Liu, Ruidong Zhang, Yiheng Liu, Haiyang Sun, Yi Pan, Yiwei Li, Yifan Zhou, Hanqi Jiang, Junhao Chen, and Tianming Liu. 2024. Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research. arXiv:2412.04497 [cs.CL] https://arxiv.org/abs/2412.04497

[174] Marta Ziosi and Dasha Pruss. 2024. Evidence of what, for whom? the socially contested role of algorithmic bias in a predictive policing tool. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*.

[175] Leying Zou and Warut Khern-am nuai. 2023. AI and housing discrimination: The case of mortgage applications. *AI and Ethics* 3, 4 (2023), 1271–1281.

A Appendix

A.1 Study Materials

We provide the materials used during our co-design workshops including the design exercise prompts and survey questions.

A.1.1 Design Exercises. In Fig. 7, we provide the prompts for our two design exercises: rapid prototyping and storyboarding. After the design exercise, participants were invited to share out with the group about designs they created. We asked the following questions to scaffold the sharing process:

(1) What is the purpose of the tool? What types of biases does it aim to audit?
(2) How would you use the tool?
(3) What did you prioritize when designing this tool?

A.1.2 Survey Questions. At the conclusion of the workshop, participants were invited to complete the following survey. Questions are shown in Fig. 8 and 9.

A.2 Codebook

We provide the codebook produced through our analysis process in Table 1.

1 **Rapid Design Exercise**

Using the sheet of paper in front of you, sketch out one idea for a bias auditing tool in each square.

Sketches are **meant to be rough** – no need for details

There are **no bad ideas!**

Try to come up with as many different ideas. They can be as **wild and out there** as you want.

2 **Storyboarding Exercise**

What is the teacher using the AI for? { Designing a lesson plan, Creating a rubric, … }	What types of biases is the teacher concerned about? { Stereotyping, Misappropriation, Hallucination, … }
Draw an example of an auditing tool you would use	What do you now know about the AI that would make you comfortable to use it (or decide to not use it)?

Prompt
You are a teacher that wants to use AI for _____.

What kind of auditing tool would you want to use to help check for potential biases that might exist before you use AI?

1. What is the **purpose** of the tool?
2. What **types of biases** are you trying to investigate?
3. What do you **input** into the tool? What does the tool **output**?
4. From fully automated to fully manual, where does your tool sit in terms of **control**?

Figure 7: Workshop participants completed one of two design exercises. For participants in Workshops 1 and 2, they took part in a rapid design exercise (top) focused on ideating many divergent designs for auditing tools. For participants in Workshops 3 and 4, in groups of 2-3, they completed a storyboarding exercise for a single design (bottom).

1 At which school level do you currently teach?

○ Elementary School
○ Middle School
○ High School
○ [_____] Other

2 What subject(s) do you teach?

[_____]

3 How long have you been in your teaching role?

○ Less than 1 year
○ 1-3 years
○ 3-5 years
○ 5-10 years
○ 10-15 years
○ Over 15 years

4 Have you ever used an AI-powered tool (e.g., ChatGPT) before?

○ Yes
○ No

5 Have you ever used an AI-powered tool before in educational settings (e.g., lesson planning, grading content)?

○ Yes
○ No

6 Which AI tools have you used before in educational settings?

☐ ChatGPT
☐ Gemini
☐ Microsoft Bing AI
☐ Khanmigo
☐ MagicSchool AI
☐ Notebook LM
☐ [_____] Other
☐ None of the above

7 For what purposes have you used AI tools in educational settings?

☐ Generating lesson plans
☐ Creating culturally relevant teaching materials
☐ Translating content into ʻŌlelo Hawaiʻi
☐ Assisting stand-in teachers with materials
☐ Creating impromptu activities
☐ Grading or assessing student work
☐ Writing correspondences (e.g., emails)
☐ [_____] Other
☐ None of the above

8 What concerns do you have about using AI tools in educational contexts?

☐ I don't understand enough about AI tools
☐ I have concerns about cultural misrepresentation
☐ I'm concerned about AI replacing teachers
☐ I am concerned about students misusing AI tools
☐ I have concerns about the AI tools returning incorrect information
☐ I don't have access to these tools
☐ [_____] Other
☐ I do not have any concerns

9 Prior to the workshop, can you recall a situation where the outputs of an AI-powered tool may not have aligned with the history or perspectives you have been taught or want to teach?

[_____]

10 Did the design session today change the concerns you have about using AI-powered tools in educational contexts? Why or why not?

[_____]

Figure 8: Online Qualtrics survey that participants took after the co-design workshop (Part 1).

11 What did you find challenging about designing a bias auditing tool? What did you find easy?

12 Based on today's design session, what do you think would be most helpful for you when it comes to auditing biases in AI-powered tools?

13 Of the bias auditing tools presented by **other** pairs in the session, which was your favorite and why?

14 Do you identify as any of the following?

☐ Kamaʻāina / Local
☐ Kānaka Maoli / Native Hawaiian
☐ Malihini / Recent arrival or visitor
☐ [   ] Other
☐ None of the above

15 Choose one or more races that you consider yourself to be

☐ Kānaka Maoli / Native Hawaiian
☐ [   ] Other Pacific Islander (please specify)
☐ Asian
☐ Black or African American
☐ White
☐ Hispanic or Latino
☐ American Indian or Alaska Native
☐ [   ] Other
☐ Prefer not to say

16 How do you describe yourself?

○ Male
○ Female
○ Non-binary / third gender
○ [   ] Prefer to self-describe
○ Prefer not to say

17 Please provide your preferred email for receiving compensation for the workshop.

18 We are working on developing tools based on our design workshop today for identifying biases in AI-generated outputs.

Would you be interested in keeping in touch regarding tool development?

○ Yes
○ No

Figure 9: Online Qualtrics survey that participants took after the co-design workshop (Part 2).

Code	Count
Concerns about Generative AI	
Coherence (e.g., alignment to Hawaiian culture)	6
Connotations (e.g., Hawaiian culture being treated as part of history)	6
Incorrectness and Hallucinations	19
Irrelevance of Outputs to Hawaiian Culture	6
Only surface level representation of Hawaiian culture	13
Erasure of Cultural Elements	11
Skewedness of Cultural Depiction	5
Not representing the diversity of Hawai'i	6
Specificity of Cultural Elements	7
Challenges with identifying harms	
Disagreements about what is biased	11
Lack of data and sources about Hawaiian culture	9
Need community knowledge to do auditing / identify harms	20
No ground-truth to compare against	5
Lack of self-confidence when identifying bias	7
Factors influencing design	
Bigger picture issues beyond representational bias	5
Grade-level and subject matter influence needs	4
Importance of social support for educators	6
Accounting for the positionality of auditors	5
Being able to trust sources and vet reliability of information	37
Scarcity of data	4
Integrating Hawaiian Values	4
Kilo (learning via observation)	2
Knowledge Provenance	8
Tensions in design	
Automation vs. Reflection	9
Cultural specificity vs. Unawareness	9
Data Sharing vs. Co-option	3
Efficiency vs Quality	3
Stereotyping vs Personalization	7
Auditing Tool Ideas	
Perspective auditing	14
Auditing the prompt or person, not the output	2
Comparing across models and outputs	9
Source auditing	22
Flagging bias	3

Table 1: Codebook with themes and sub-codes surfaced through reflexive thematic analysis described in Sec. 4.3

.