Received: 27 July 2024 | Accepted: 6 January 2025

DOI: 10.1111/bjet.13568

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Value-sensitive design of chatbots in environmental education: Supporting identity, connectedness, well-being and sustainability

Ha Nguyen[1] | **Victoria Nguyen**[2] | **Sara Ludovise**[3] | **Rossella Santagata**[2]

[1]School of Education, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2]University of California-Irvine, Irvine, California, USA

[3]Orange County Department of Education, Costa Mesa, California, USA

Correspondence
Ha Nguyen, School of Education, University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599, USA.
Email: ha.nguyen@unc.edu

Funding information
National Science Foundation, Grant/Award Number: 2241596

While offering the potential to support learning interactions, emerging AI applications like Large Language Models (LLMs) come with ethical concerns. Grounding technology design in human values can address AI ethics and ensure adoption. To this end, we apply Value-Sensitive Design—involving empirical, conceptual and technical investigations—to centre human values in the development and evaluation of LLM-based chatbots within a high school environmental science curriculum. Representing multiple perspectives and expertise, the chatbots help students refine their causal models of climate change's impact on local marine ecosystems, communities and individuals. We first perform an empirical investigation leveraging participatory design to explore the values that motivate students and educators to engage with the chatbots. Then, we conceptualize the values that emerge from the empirical investigation by grounding them in research in ethical AI design, human values, human-AI interactions and environmental education. Findings illuminate considerations for the chatbots to support students' identity development, well-being, human–chatbot relationships and environmental sustainability. We further map the values onto design principles and illustrate how these principles can guide the development and evaluation of the chatbots. Our research demonstrates how to conduct contextual, value-sensitive inquiries of emergent AI technologies in educational settings.

KEYWORDS
artificial intelligence, chatbot, environmental education, value-sensitive design

Practitioner notes

What is already known about this topic

- Generative artificial intelligence (GenAI) technologies like Large Language Models (LLMs) can not only support learning, but also raise ethical concerns such as transparency, trust and accountability.
- Value-sensitive design (VSD) presents a systematic approach to centring human values in technology design.

What this paper adds

- We apply VSD to design LLM-based chatbots in environmental education and identify values central to supporting students' learning.
- We map the values emerging from the VSD investigations to several stages of GenAI technology development: conceptualization, development and evaluation.

Implications for practice and/or policy

- Identity development, well-being, human–AI relationships and environmental sustainability are key values for designing LLM-based chatbots in environmental education.
- Using educational stakeholders' values to generate design principles and evaluation metrics for learning technologies can promote technology adoption and engagement.

INTRODUCTION

Large language models (LLMs) can simulate human perspectives in contexts ranging from political science to public speaking and teacher training (Jansen et al., 2023; Markel et al., 2023; Park & Choi, 2023). We leverage LLMs (eg, OpenAI's GPT-4) to develop chatbots within a high school environmental science curriculum. The chatbots represent different perspectives on how climate change affects local ecosystems. Students engage with the chatbots in scaffolded conversations, to refine their scientific models of climate change's root causes, impacts and solutions.

Although promising, leveraging LLMs comes with ethical considerations (Navigli et al., 2023; Stahl & Eke, 2024), including transparency (Graf & Bernardi, 2023; Wu et al., 2022), accuracy (Byrd, 2023) and stereotype (Cheng et al., 2023; Nguyen et al., 2024). To address these concerns, researchers have called for grounding the design of AI systems (including LLMs applications) in human values (Veale & Binns, 2017; Vernim et al., 2022; Wambsganss et al., 2021). Value-Sensitive Design (VSD) provides a principled approach to exploring values, defined as 'what is important to people in their lives, with a focus on ethics and morality' (Friedman & Hendry, 2019, p. 24). We apply VSD to investigate the values that guide the design and evaluation of AI chatbots in environmental education.

We follow VSD's tripartite approach—involving empirical, conceptual and technical investigations—for value exploration. We perform an empirical investigation to explore the values that motivate students and educators to engage with the chatbots. We elicit these values through participatory design (DiSalvo et al., 2017), involving high school students, teachers, informal educators and marine scientists. We conceptualize the values that emerge from the empirical investigation by grounding them in research in ethical AI design, human values, human–AI interactions and environmental education. Finally, we map the values onto design principles and illustrate how these principles can guide the development and evaluation of the chatbots. The current paper focuses on the design process and does not address learning from chatbots' interactions. We will report on learning outcomes in future work.

Our research demonstrates how to conduct contextual, value-sensitive inquiries of AI technologies in education contexts, with three main contributions. First, applying participatory design approaches—with a focus on values—allows us to align the chatbot-embodied values with those of education stakeholders. Second, grounding inquiries within a specific instructional context reveals important and novel insights to design for. Beyond values commonly stated in AI ethics frameworks like transparency and trust, findings illuminate consideration for the chatbots to support students' identity development, well-being, human–chatbot relationships and environmental sustainability. Finally, prior research that leverages VSD does not always use conceptual and empirical findings to substantively inform the technical investigation (Gerdes & Frandsen, 2023; Winkler & Spiekermann, 2021). To the best of our knowledge, this is the first example in designing value-sensitive pedagogical chatbots that maps from value elicitation to design principles, prototyping and evaluation.

BACKGROUND

To ground the value exploration, we review the literature on AI ethics. We survey environmental education research to better align the chatbot-embodied values with our instructional context. We turn to VSD for a systematic approach to incorporating stakeholders' perspectives into technology design.

Values in AI design

Advances in AI, specifically LLMs, have enabled new forms of human–AI learning interactions, with promise for natural language understanding, personalization and all-time availability to support learners (Kasneci et al., 2023; Moore et al., 2023). LLM-enabled tools can improve learning (Meyer et al., 2024) and student engagement (Kazemitabaar et al., 2024). However, LLMs might showcase *bias* in their training data and output content that exacerbates stereotypes or discriminations against marginalized voices (Gadiraju et al., 2023; Stahl & Eke, 2024). The models' output might include inaccuracy, raising concerns about users' *trust* to rely on the systems (Shah et al., 2024). Additionally, the models' training data and decision-making are not always *transparent* (Wu et al., 2022). There are *privacy* risks, as LLMs can memorize details from users' conversations and reveal the information when responding to another user (Carlini et al., 2021).

These multiple concerns have led to efforts to explicitly ground the design of AI systems in ethical values (Gallegos et al., 2024; Hagendorff, 2020; Tomašev et al., 2020). Wambsganss et al. (2021) conducted a systemic literature review and user interviews to explore values and design principles for AI conversational agents. Mapping the findings to the Organization for Economic Co-operation and Development's (OECD) AI principles, these authors found several values under human-centred design and fairness (eg, accessibility, bias prevention,

human rights), transparency (eg, trust, explainability, communication), robustness (eg, privacy, reliability, security) and accountability (eg, auditability, reporting, responsibility). These values can lead to design principles that guide AI systems. For example, to enhance transparency, designers can implement system feedback that explains AI's output mechanisms (Wambsganss et al., 2021).

Values in environmental education

Environmental protection (ie, protection of environments to preserve natural habitats) and *sustainability* (ie, utilizing environmental, social and economic resources in the present, while maintaining them for future generations) are inherent values in environmental education (Lewis et al., 2008; Tilbury, 1995; Tilbury & Wortman, 2008). Attitudes and actions towards environmental protection and sustainability are linked to different values (Amérigo et al., 2007; De Groot & Steg, 2009). People with *anthropocentric* values argue that nature should be preserved because it offers utility to humans (Turner et al., 2003). Those with *biospheric* values show concerns for environmental well-being (De Groot & Steg, 2009). *Egoistic* values consider one's own well-being, while *altruistic* values attend to others' well-being (Schultz, 2001).

These multiple values are important to consider in our chatbot design, as different chatbot profiles can embrace unique and intersecting values in their exchanges with students (De Dominicis et al., 2017). For example, a chatbot with biospheric values may converse about concerns for life forms, while another chatbot with altruistic values may discuss social responsibility (Kim & Stepchenkova, 2020; Schultz & Zelezny, 2003).

While our initial literature survey surfaces several values to consider in chatbot design, value perceptions are highly contextual and driven by the stakeholders directly or indirectly impacted by the technology (Le Dantec et al., 2009). We use value-sensitive design to ground the technology design in human values.

Value-sensitive design (VSD)

VSD presents a theoretically grounded and systematic approach to centring human values throughout the technology design process (Friedman, 1996; Friedman et al., 2017). VSD employs a tripartite approach, where designers iterate between conceptual, empirical and technical investigations (Friedman & Hendry, 2019). The conceptual phase elicits values through theoretical and philosophical investigations, with consideration for the stakeholders that are directly and indirectly influenced by the design. The empirical investigation concerns the collection of empirical data with stakeholders' involvement. Designers can leverage multiple methods in this phase, such as inviting stakeholders to sketch scenarios (Woelfer et al., 2011), co-create prototypes (Yoo et al., 2013) and discuss value priorities and tensions (Friedman et al., 2006). Designers can also employ toolkits, such as the Envisioning Cards (Friedman & Hendry, 2012) or Metaphor Cards (Logler et al., 2018), to introduce topics conducive to value generation. Drawing from insights from the conceptual and empirical explorations, the technical investigation involves creating new, value-centred designs (Strikwerda et al., 2022) or redesigning existing technologies (Vernim et al., 2022; Wynsberghe, 2017). Designers can start with any phase (conceptual, empirical, technical) and iterate as the design space evolves.

We select VSD to guide our chatbot design for two reasons. First, emergent research has demonstrated the potential of VSD in designing AI systems (Dexe et al., 2020; Gerdes & Frandsen, 2023; Vernim et al., 2022; Wambsganss et al., 2021) and learning analytics (Chen

& Zhu, 2019; Prieto et al., 2023; Viberg et al., 2023). Viberg et al. (2023) argued for culture-sensitive learning analytics design that focused on the values and needs of the target users, to enhance users' engagement with learning analytics tools. As an empirical example, Prieto et al. (2023) applied VSD to discover values such as *self-direction* and *sense of progress* from surveys, interviews and diary data with graduate students. The authors then used these values to generate design insights for educational technologies, to promote students' persistence and well-being.

Second, VSD aligns with and complements our focus on participatory design (PD) in educational contexts (DiSalvo et al., 2017). Both approaches employ a grounded, bottom-up approach to integrating stakeholders' knowledge and expertise throughout the design process. Compared to PD, VSD considers a wider range of stakeholders. While PD has a substantial commitment to values like participation and democracy (Bødker et al., 2022), VSD attends to a broader range of values emerging from the tripartite investigations (Borning & Muller, 2012). This emphasis on values serves as an anchor for our chatbots' conceptualization, development and evaluation.

MATERIALS AND METHODS

Our investigation is guided by the following research questions (RQs):

RQ1: What values guide the design of AI chatbots in environmental education?
RQ2: How do these values inform the design principles and evaluation of the chatbots?

Study setting

This paper presents the development stage of a multi-year project to create an AI-guided, high school curriculum that promotes science communication around climate issues in (Western State), United States. We designed several chatbots (enabled by OpenAI's GPT-4) to embody different perspectives about climate change and introduce students to various data sources (eg, scientific evidence, personal anecdotes). The chatbots support students to construct a scientific model depicting climate change's root causes and impacts on biotic and abiotic components, economics, culture, and health and well-being. Through chatbot interactions, students gain insights specific to their local context, (PLACE), to refine the components and causal relationships between components in their models.

The research team designed the chatbots and curriculum with a design team of three high school students (two in 10th grade; aged 16, one in 12th grade; aged 18), four undergraduates, two high school teachers and five environmental educators (with expertise in outdoor education, sportfishing and marine biology). They represented diverse backgrounds (two identified as Hispanic, six of Asian and Asian American descents and six White). To recruit the high school students, we disseminated fliers to local science teachers in our network and conducted interviews to select students with interest in environmental science and AI in education. The team included both direct (eg, students, teachers) and indirect stakeholders of the technology (eg, scientists to be represented by the chatbots). We engaged the design team in idea conceptualization, prototyping and evaluation over eight sessions (17 hours; spread over 6 months) to surface which perspectives the chatbots might represent and what values guided the design of the chatbot dialogues and interface. The research procedures were approved by the Institutional Review Board (#2801). All names reported are pseudonyms.

Procedures

We examined RQ1 about values guiding chatbot design through empirical and conceptual investigations. Following an initial literature review (Background section), we moved to an empirical investigation and a second round of conceptual investigation. This process allowed us to decentre the expertise of the researchers (Le Dantec et al., 2009), and instead, ground value exploration in emerging stakeholders' insights. Figure 1 outlines our procedures.

Empirical investigation

From the earlier sessions (sessions 1–3), the design team decided on five chatbot profiles: college student, social media influencer, fisherman, civil engineer and kelp researcher. The profiles covered different aspects of climate change's root causes and impacts, including marine life (fisherman; kelp researcher), infrastructure (civil engineer), livelihood (fisherman), and culture and community (influencer; student). They represented diverse demographics (eg, age, education backgrounds) and experiences to discuss climate change over time and space.

For the empirical investigation, we focused on the design sessions that leveraged different VSD methods to envision how students might interact with each chatbot. The activities involved in-person, small-group discussions (4–5 participants per group; three groups per session). We examined how participants engaged with design fiction (sessions 4 and 5; 75 minutes/group), metaphor cards (session 6; 35 minutes/group) and discussion centred around AI ethics (session 7; 20 minutes). While value-oriented interviews are common in VSD studies (Winkler & Spiekermann, 2021), the selected activities promoted small and whole-group interactions that aligned with the collaboration focus of the co-design sessions. They explicitly invited the design team to collaboratively ideate and articulate underlying values, concerns and possible interactions with the chatbots in open-ended structure (Baumer et al., 2020; Peters et al., 2021). This open-endedness supported preliminary value exploration, compared to other VSD methods that employed researcher-created scenarios and design solutions (eg, value scenarios, value dams and flows, prototypes, Friedman & Hendry, 2019). Data included transcripts from audio recordings (total 6.5 hours) and design artefacts (eg, metaphor cards, digital notes).

In design fiction (Muller & Liao, 2017), participants write or sketch stories about fictional technologies, to reveal the underlying values attached to the AI applications. In our case, participants wrote character cards that outlined the chatbots' backgrounds, expertise and desired interactions without emphasizing what was technologically feasible. We

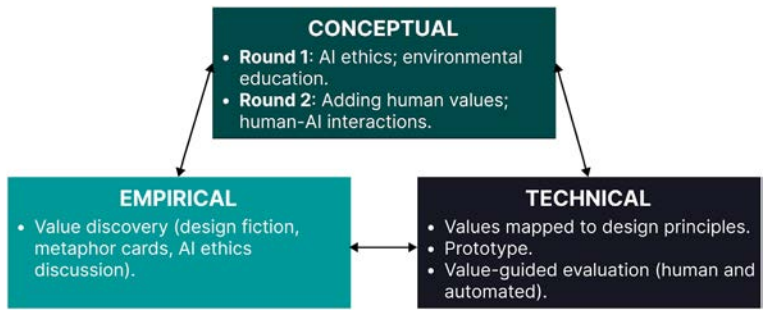

FIGURE 1 Research procedures.

included several questions to support the discussion, for example, 'Who is this person?', 'What values do they hold?', 'What expertise do they have regarding climate change in [PLACE]?' and 'What questions will a high school student ask this person, and how will they respond?'

Meanwhile, metaphor cards (Lockton et al., 2019; Logler et al., 2018) invited participants to find visual examples for the chatbot interactions. We followed Lockton et al.'s (2019) approach to pair an image with an abstract concept (eg, 'How could *a burning candle* be a metaphor for *climate change*?') to spark inspiration for new design ideas. In the first round of this activity, we asked participants to come up with their own images to prompts such as: '___ could be a metaphor for interactions with the *fisherman* chatbot' (Appendix A). In the second round, they used images that the researchers had curated. These images presented more distant connections to the chatbot interactions (eg, mazes, black holes), to further elicit design ideas.

Finally, we facilitated AI ethics discussion (Payne, 2019). The discussion questions invited participants to experiment with different LLM applications, reflect on which stakeholders might be negatively or positively impacted by the technologies and outline how the technologies can do the most harm and good in the short (eg, a few months) and long term (eg, 50 years).

The first two authors applied inductive coding (Thomas, 2006) of the data across the three VSD activities to explore the values that stakeholders embraced. This process started with scanning data (107 transcript pages from design sessions that included the three design activities) to identify text segments related to the RQs. In two iterations, we separately coded all data line-by-line to generate initial codes about values associated with chatbot design and memos of researchers' making sense of the data. The first iteration involved coding the metaphor cards and a subset of the design fiction, and the second iteration involved the rest of the design fiction and AI ethics discussion. In each iteration, we compared the researchers' individual code lists and organized them (eg, grouping 'curiosity', 'seeking knowledge' and 'problem-solving' to denote self-driven actions). In a third discussion, we refined, named and defined the categories drawing from prior literature (see next section on Conceptual Investigation). We recoded the data with the developed categories and resolved four cases of discrepancy through discussion. Appendix B outlines the code iterations and Appendix C presents the coded excerpts ($n = 250$).

Conceptual investigation

We conceptualized the values emerging from the empirical investigation by situating them within prior literature. We focused on several sources: AI ethics (eg, Hagendorff, 2020; OECD, 2024; Tomašev et al., 2020; Umbrello & Poel, 2021; Wambsganss et al., 2021), environmental education (Lewis et al., 2008; Schultz & Zelezny, 2003; Tilbury, 1995), human values (Friedman & Hendry, 2019; Schwartz, 2012) and human–AI relationships (Bae Brandtzæg et al., 2021; Skjuve et al., 2021; Zimmerman et al., 2023). Table 1 lists the values.

The empirical investigation informed the broadened conceptual investigation. Our investigation extended to human–AI relationships based on insights from design participants. For example, *connectedness* emerged as a value, as students might perceive the chatbots as social, anthropomorphize them (ie, attribute beliefs and emotions to chatbots), and share their thoughts and feelings in the interactions (Christoforakos et al., 2021; Grové, 2021). Because LLMs can memorize conversational history and maintain coherent exchanges, student–chatbot relationships can evolve over extended interactions (Clark et al., 2019). We consider these research areas to conceptualize the values in the Findings.

TABLE 1 Values guiding the chatbot design.

Values	*n*	Definition (representative literature)
Identity	75	One's perception of who they are, in relation to past, present and possible future selves (Briggs & Thomas, 2015; Oyserman et al., 2006). Identity includes inherited attributes (eg, background), experiences, family ties and how individuals represent themselves
Connectedness	56	A social bond formed to make someone feel heard and valued (Christoforakos et al., 2021) through storytelling, mentorship and relatability
Self-direction	44	One's independent thought and action (eg, curiosity, problem-solving, self-initiated inquiries; Prieto et al., 2023; Schwartz, 2012)
Trust	23	One's perception of reliance on another entity (human/technology) to show vulnerability or extend goodwill (Friedman & Hendry, 2019; Wambsganss et al., 2021)
Human well-being	28	The physical, material and psychological state of oneself (personal well-being), as well as others (others' well-being; Friedman & Hendry, 2019; OECD, 2024)
Environmental sustainability	24	Protection of nature to meet the present needs, while not negatively impacting future generations (Tilbury, 1995); can focus on anthropocentric, biocentric, egoistic and altruistic values

Note: *n* = frequencies of coded segments under each value.

Technical investigation

The technical investigation incorporated findings from the empirical and conceptual investigations to answer RQ2: How do emerging values inform chatbots' design principles and evaluation? For this, we mapped the values to design principles for the chatbots. This analysis illustrates how to translate ethical AI values that are often abstract into concrete design decisions. Further, we demonstrated how the values and design principles informed our ongoing evaluation.

To generate chatbot responses, researchers can carefully craft prompts to specify the interaction contexts and the tasks for the LLMs (ie, prompt engineering). These prompts can be refined to tweak the instruction, add limitations (eg, what not to output) and provide example responses. We followed Jurenka et al.'s (2024) approach to conducting LLMs' evaluations based on clearly defined pedagogical principles. We used the values emerging from empirical and conceptual investigations to create rubric criteria for human evaluation. The evaluators (five trained researchers, all majoring in environmental science and education) answered Yes/No to whether the responses met the value criteria and rewrote the responses to reflect the values if needed. We established inter-rater agreement (50 responses; 13% of the dataset; Krippendorff's α range 0.67–0.95) and conducted one round of human evaluation (375 chatbot responses; 75 responses per profile). Following this first evaluation round, we iterated upon the prompt instruction for the chatbots and conducted a second, automated round of evaluation.

We used similar value-guided criteria for our automated evaluation. We constructed prompts for different API calls (GPT-4o; temperature = 1; Appendix D). Each prompt included a sample student–chatbot conversation (one turn per student/chatbot) and instruction for the LLMs to gauge whether the chatbot's response met a value criterion. For example, the instruction might state: 'Answer with Yes or No if the chatbot's response promotes *self-direction* from students. Response can ask students if they want to learn more or what questions they have about the topic'. While the human evaluation provides rich examples to improve the chatbots' responses, the automated evaluation offers a scalable approach to assessing prompt iterations.

RESULTS

Values emerging from the empirical and conceptual investigations (RQ1)

When we started this research, we expected to find values common in AI ethics frameworks like privacy, transparency and bias (Kasneci et al., 2023; Navigli et al., 2023; Stahl & Eke, 2024). Our findings revealed additional, unexpected values, such as how the chatbots can support *identity development*—how students view themselves through identifying with the different chatbots. Results demonstrated nuances in *human–AI interactions*, to promote connectedness, self-direction and trust. Further, they showed how technology might support *human well-being* and *environmental sustainability* (Table 1; see Appendix C for coded excerpts).

Identity

The empirical investigation revealed a focus on *identity*, or how people perceive their sense of self over time. The design participants described the chatbots' identity in terms of inherited backgrounds (eg, age, race/ethnicity), experiences (eg, locations, careers, hobbies), expertise and family connections. For example, the metaphor cards that participants co-created highlighted the personas' experiences and expertise. They compared the *kelp researcher* chatbot to 'an iceberg'—possessing a 'deeper understanding of how climate change works', while linking the *fisherman* to a 'naturalist or a guide' and the *influencer* to a 'mountain climber' with vast knowledge of nature. Discussions about identity were present throughout the design fiction. The following exchanges between Marvin and Elena (environmental educators) highlight the experiences, family ties and knowledge that the *college student* chatbot may bring:

Marvin: Maybe theygrew up collecting seashells and visiting tidepools with their siblings and notice those changes? […]

Elena: Someone who's first generation who may not have a lot of conversations about these topics, seashells were their way to notice these changes.

These multiple identity facets allow students to identify with the chatbots. Participants made intentional choices for place association (whether the chatbots lived by the coast or within in-land communities), so they could be relatable to students with different lived experiences. For instance, George (teacher) and Ricky (undergraduate) discussed how the *civil engineer* chatbot might live on the coast but grew up in a landlocked area, to bring multifaceted perspectives about coastal and in-land infrastructure.

Important discussions emerged about how the chatbot conversations presented opportunities for students to construct their own identity. George emphasized that the *kelp researcher* chatbot should convey to students that 'science can be anywhere'. Meanwhile, Nina, Ana, Sophie (high school students) and Erin (environmental educator) excitedly brainstormed how the *influencer* chatbot could hold multiple occupations (eg, content creation, spearfishing, and scientific diving with deep understanding of local ecosystems). These discussions can be linked to the construct of *possible selves*, defined as one's images of who they might be in the future (Oyserman et al., 2006). Students with science possible selves are more likely to show interest in and form friendships around scientific activities (Robnett & Leaper, 2013). The chatbots' multiple identities may invite students to imagine different possible selves.

Human–AI interactions: Connectedness, self-direction and trust

Connectedness

We defined *connectedness* as a bond formed to make someone feel heard and valued through storytelling, mentorship and relatability (Christoforakos et al., 2021). Storytelling emerged as a dominant interaction strategy. The metaphor cards portrayed the *influencer* chatbot as a 'film maker' and a 'campfire' (connecting people, showing links to 'community and the environment'). Another card visualized the *fisherman* as a bookshelf, a 'wise old man/grandpa that has stories'. Participants recounted stories to enrich the chatbots' knowledge in the design fiction. For instance, Erin (environmental educator) proposed that the *fisherman* chatbot showed 'nostalgic reflection' about changes to fish population, 'in a short timeframe, 30 years, when they were a kid remembering how it was to fish' and now realizing 'climate change's impact on livelihood'.

Additionally, the chatbots were described as mentoring figures with relatability to students' experiences. For example, a metaphor card described the *college student* as an 'older sibling or friend who can be a role model'. After reviewing the metaphor cards, Nina (high school) noted how the chatbots could pose questions to build relatability: 'values are really, really important. The chatbot could start a conversation with what do you find important about the environment? … to help the student feel more emotional about what it's asking'. Sophie (high school) observed:

> I feel like an influencer would kind of have like similar experiences with being outdoors as a high schooler would. A high schooler probably looks up to the things that they do outside so probably have questions like—when you go diving, when you go hiking on the coast, like what are the things you see?

Self-direction

Across data sources, we found examples of multi-turn exchanges between students and the chatbots to support *self-direction*, which can be associated with curiosity, seeking and constructing knowledge, and problem-solving (Prieto et al., 2023; Schwartz, 2012). Metaphor cards such as 'law school classroom', 'curious learner', 'puzzle pieces' and 'maze' described how different chatbots (*kelp researcher*, *fisherman*, *civil engineer*) might not give answers, but 'question the students' and 'solve climate change problems in their community'. This translated into how the environmental educators and teachers authored the student–chatbot conversations in the design fiction. For instance, the *kelp researcher* chatbot might draw out students' observations:

> As you were walking to class, and now that you're all wet, you were thinking about: Why is it raining like this here at this time of year? And how does that work? And, you know, what is the climate change connections to it?
>
> (George, teacher)

Daniel (environmental educator) provided another example for this chatbot:

> Let's take it back to the atmospheric river. Why do you think this is happening? And student writes something, it may not be fully correct, but it could be like, that's a very interesting idea, why don't we investigate that? Can you explain your reasoning?

Relatedly, in the AI ethics discussion, Ricky (undergraduate) raised concerns that students might overly rely on AI and only use the technology to generate answers. Dana (teacher)

acknowledged this concern but highlighted LLMs' capacity for natural language understanding to generate long-term impact and 'make learning more conversational, not transactional'.

Trust

The design team noted the importance of the chatbots to convey truthful, transparent information for students to *trust* the technology (Wambsganss et al., 2021). They came up with metaphors to describe this value, including a 'reporter' (*fisherman*) and a flashlight 'shining light on the information' (researcher). In the AI ethics discussion, George and Dana (teachers) highlighted the need to 'authenticate the chatbots' responses', 'whether you're human or AI, you got to prove it'. The student–chatbot interactions that Sam, Daniel and Erin (environmental educators), Sophie (high school), and Maya and Lay (researchers) created in the design fiction reflected a similar focus. These participants discussed how the *kelp researcher* and *college student* chatbots might cite external, verifiable resources and the *influencer* chatbot might insert multimedia elements (eg, videos, images) of their work.

Human well-being

Findings echo prior research's emphasis on human *well-being* (supporting one's physical, material and psychological state) as an important construct for technology design (Friedman & Hendry, 2019). The chatbots might embrace human well-being as a value, with metaphors for 'activism' and 'time capsule' to relay the impact of climate change on communities. In the design fiction, the *civil engineer* chatbot took on the responsibility of 'building a better world':

> But they that brings a lot of, it's a lot of responsibility. Because not only are like, in our case, you want to design for the environment, you're talking about protecting people's lives […] Stories about the greater purpose of who a civil engineer is.
>
> (George, teacher)

Meanwhile, Daniel and Elena (environmental educators) brainstormed how the *kelp researcher* chatbot might provide in-depth information about climate change's impact, including 'how frequent storm events can cause water quality issues' affecting coastal and inland communities. Further, the chatbots can encourage students to reflect on their experiences, to connect climate change to *personal well-being* and *others' well-being*. Consider the following excerpt involving three high school students and an environmental educator designing the *influencer* chatbot:

> I'm feeling like someone who's an outdoor influencer in some way, seeing the changes, whether it's where they're traveling or what they do recreationally […]
>
> (Erin, educator)

> Not only do certain people get affected, but everybody gets affected.
>
> (Nina, student)

> … an influencer would be a good way to hit the mental health aspect of how climate change can also affect people […] It's just like a concern of when I'm older and I have kids, like what is the world that they're gonna have to live in like?
>
> (Sophie, student)

The students stressed that the chatbots should convey a sense of hope. They constructed dialogues for the *college student* chatbot to discuss climate change solutions and the future world that they imagined living in.

Environmental sustainability

There was a strong emphasis on *environmental sustainability*. The metaphor cards for this value included a 'scale' ('balancing between community needs and resources') and 'phoenix' ('our earth can bounce back from certain things, but what happens when it can't anymore?'). Different stances linked to attitudes towards sustainability—anthropocentric, biospheric, egoistic and altruistic—were present in the design fiction. The *influencer* chatbot embraced anthropocentric values and came from a culture that relied on seafood as the main source of protein. The group built on a suggestion from Nina and Ana (high school) to leverage their heritage as Filipino Americans and emphasize sustainability from individuals' and communities' perspectives:

> There are a lot of seafood and cooking influencers, that will care about like sustainable seafood or waste […]
>
> (Maya, researcher)

> It'd have more of an impact on the person communicating if they were able to talk about a cultural impact it has had on them and not just them but the culture.
>
> (Sophie, high school)

Meanwhile, the *fisherman* chatbot might showcase anthropocentric and biospheric values. They might communicate about how 'the livelihood and sustainable fishery are impacted by a changing climate' while expressing a strong commitment to protecting nature (Erin, educator). Maya (researcher) and Pam (undergraduate) discussed altruistic values, such as how the *college student* chatbot might focus on sustainability for future generations. George (teacher) and Ricky (undergraduate) brought up altruistic and biospheric values in connecting the *civil engineer* chatbot to sustainability issues like green energy and sustainable design.

Technical investigation: How values inform design principles and evaluation (RQ2)

Design principles

The values uncovered in RQ1 informed our design principles (Figure 2). The researchers drafted the principles and revised them in conversation with the design team. The principles guided our pipeline to develop the chatbots' knowledge base. To establish *connectedness* (chatbots can cite personal and local examples; DP3) and *trust* (chatbots can provide evidence for their responses; DP7), we linked the LLMs to external data sources that the models might cite in their responses. For this, we leveraged Retrieval-Augmented Generation (RAG; Lewis et al., 2020) with LangChain, so the chatbots could reference external vector databases and use this additional context for response generation (Figure 3). The databases included local news and scientific papers, videos to illustrate learning concepts and interviews with (PLACE)'s residents and environmental professionals.

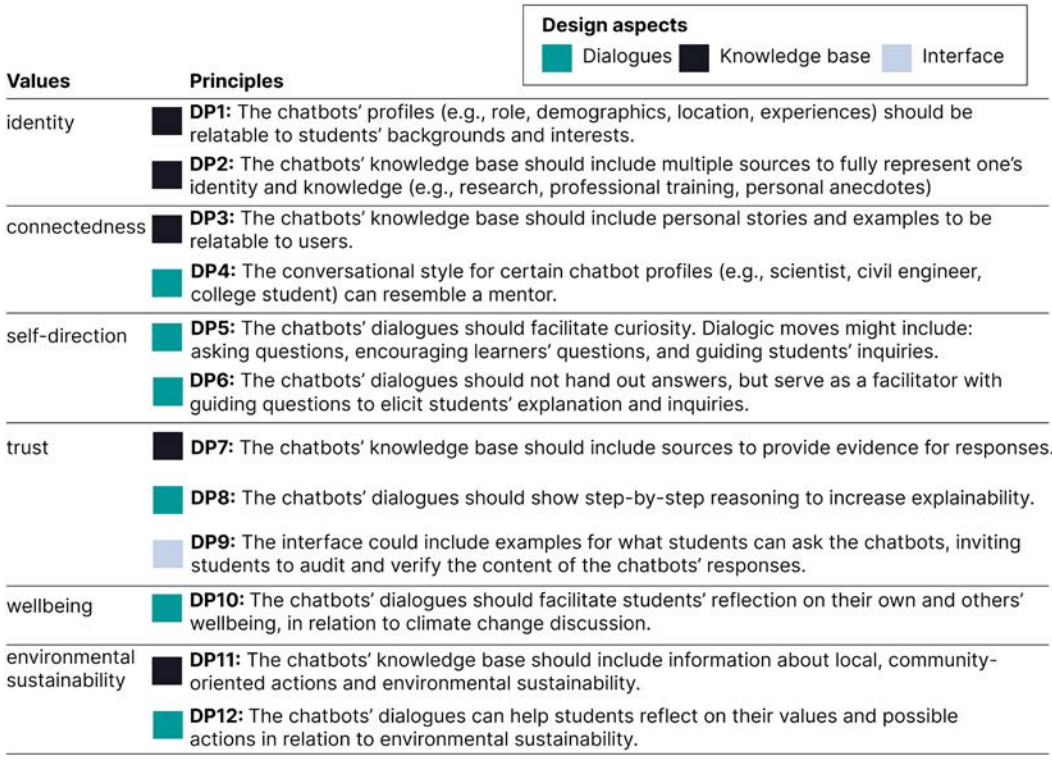

FIGURE 2 Design principles.

The principles also influenced how we engineered the chatbots' dialogues. For example, to promote *self-direction*, the chatbots can encourage students to ask questions and lead the interactions (DP5). The chatbots invite students to reflect on their own values and actions in connection with *environmental sustainability* (DP12). As another example of establishing *trust*, interface design can present examples (eg, sentence starters) that invite students to verify the responses' accuracy (DP9).

To illustrate how the values and design principles inform chatbot development, consider interactions between a student and the *civil engineer* chatbot (Figure 4). Note how the chatbot stays consistent with its defined role as a civil engineer (eg, 'I am a civil engineer; I focus on water quality'; DP1), invites students' questions (eg, 'What would you like to learn more about?'; DP5), and cites external sources and personal experiences (DP3, DP7). The right side of the image shows example questions within the interface for students to critique the chatbot (DP9).

Value-guided evaluation of chatbot dialogues

The values also informed the criteria for our rubric to evaluate the chatbots' responses, specifically whether responses (1) reveal an *identity* aspect (and which aspect), (2) build rapport with students (*connectedness*), (3) support *self-direction*, (4) include accurate information (*trust*), (5) include evidence (such as external links) to support claims (*trust*), (6) invite students to reflect on *well-being* and (7) promote *environmental sustainability*. We used the rubric for human and automated evaluations. Below, we provide two examples (one human; one automated evaluation) of *identity* discussion (rubric criterion #1) in chatbot responses, to show how VSD can be applied to ongoing technology evaluation.

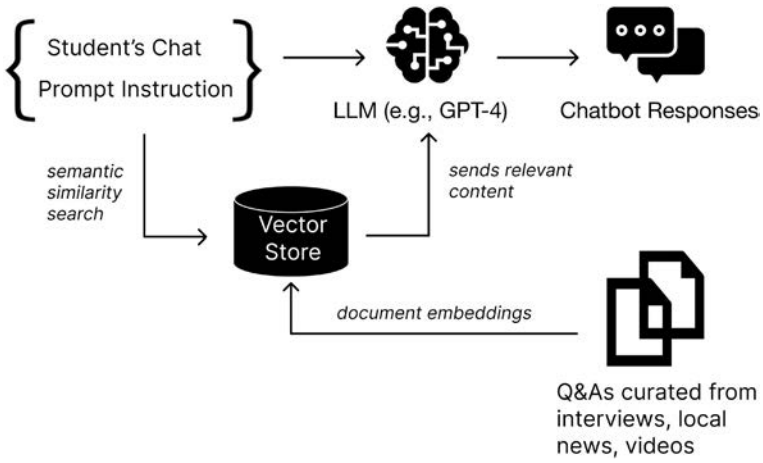

FIGURE 3 Pipeline for chatbots' knowledge base.

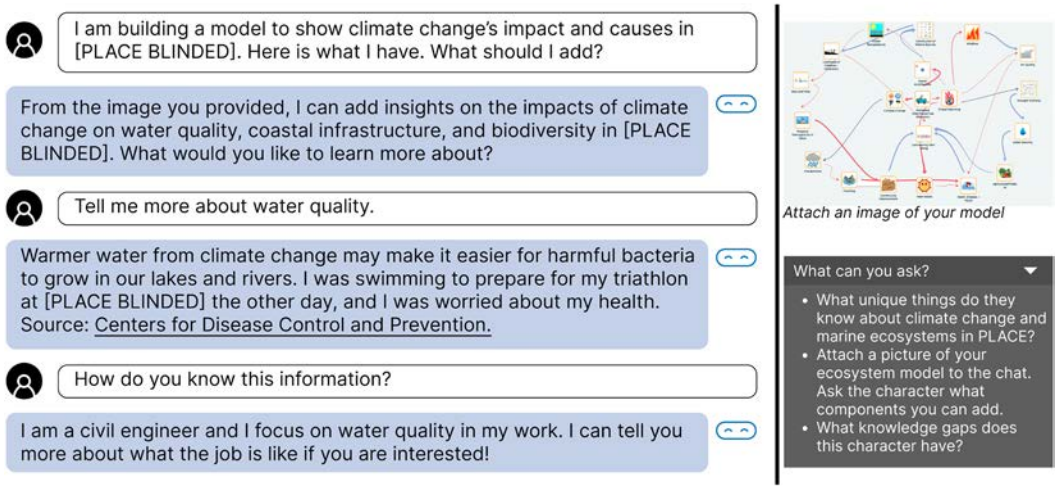

FIGURE 4 Example student–chatbots' interactions. We add a RAG pipeline for interpreting the images that students upload (using GPT-4Vision), so the chatbots can refer to the image description in their responses.

Human evaluation provides quality examples of the desired responses to incorporate into future chatbots' knowledge base and prompt instruction. Figure 5 shows an example human evaluation when interacting with the *college student* chatbot. The evaluators re-wrote the responses to share personal experiences regarding climate change's impact and thicken the identity connections. They added external sources and corrected the inaccuracy in the responses. The example illustrated a finding from human evaluation that many responses mentioned identity aspects superficially (44 of 375 coded responses; 11.73%). This finding informed iteration two, when we connected the chatbots to external databases including interviews from community stakeholders. For example, to enrich the *influencer* chatbot's database, we interviewed three content creators affiliated with local ocean education organizations about how their personal backgrounds affected their social media communication.

Meanwhile, the automated evaluation allowed us to assess improvement between prompt iterations using value-guided criteria. We prompted GPT-4o to detect identity aspects in

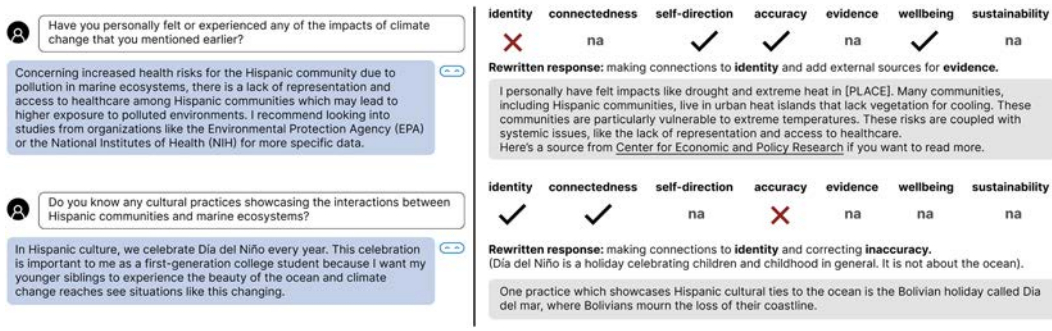

FIGURE 5 Examples of value-guided human evaluation (adapted from Jurenka et al., 2024).

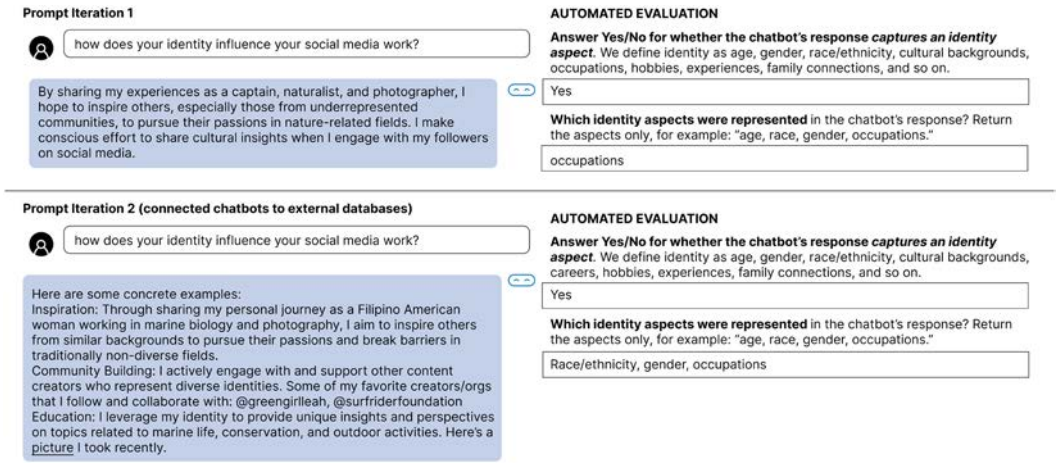

FIGURE 6 Examples of value-guided automated evaluation.

both iterations (Figure 6). The model showed substantial agreement with human coders using iteration one's data (Cohen's $\kappa = 0.63–0.88$). The automated evaluation revealed similar frequencies of *identity* discussion in both iterations (42.13% of responses). It further showed that iteration two's responses more frequently highlighted occupations as an identity facet (32%; iteration one: 24%) with deeper elaboration. Responses cited resources (rubric #5; 25.87%; iteration one: 8.80%) and posed questions to students (rubric #3) in relation to the chatbots' occupations (23.04%; iteration one: 9.87%). The automated evaluation also informed subsequent design activities. While identity connections in iteration two were richer, responses rarely promoted reflection on well-being (rubric #6; 5.87%). In our current testing, we invite the design team to review responses' connections to well-being and rewrite follow-up questions to deepen such connections.

DISCUSSION

The motivation for our work was to introduce students to varying perspectives about climate change and its interactions with physical and social-ecological systems. One way to achieve this vision is to include representation from local scientists (Ardoin et al., 2020). However, bringing professionals to students is logistically challenging. We explore the

design principles for LLM-based chatbots to address this challenge. The current paper adds to efforts to develop AI chatbots in environmental education, to portray perspectives on carbon emissions (Menkhoff & Gan, 2023) or answer questions about climate reports (Vaghefi et al., 2023). While related work has focused on the content of chatbots' responses (Muccione et al., 2024; Vaghefi et al., 2023), our findings illuminate additional considerations for interaction design. Discussions about environmental sustainability can be coupled with storytelling, resources and questions to students, to build *connectedness*, *trust* and reflection on *well-being*. Such interactions are critical to align emerging AI technologies with values within environmental education.

Our work has key methodological contributions to design and educational technology research. First, we illustrate how to translate AI ethics frameworks into concrete principles for designing educational technology. We leverage VSD and PD methods spanning multiple months to carry out empirical inquiries of design stakeholders' values. We conduct further conceptual investigation to define the values in context and uncover new insights about identity, human–AI interactions and environmental education. Revisiting the conceptual investigation after the empirical investigation reflects our commitment to value discovery and centring stakeholders' voices (Le Dantec et al., 2009). Scholars have highlighted the lack of guidelines for AI ethics in education that cover both technical and pedagogical aspects (Holmes et al., 2022). We provide an example of leveraging VSD and PD to make explicit the learning goals, pedagogical choices and learners' agency guiding chatbot design. For instance, drawing from stakeholders' insights about *self*-direction, we prompt the chatbots' dialogues to promote knowledge construction.

Second, we show how the uncovered values and design principles can guide the development cycle of education technologies, from conceptualization and prototyping to evaluation. Most research leveraging VSD has focused on empirical and conceptual investigations, and most technical investigations have only involved prototyping (Gerdes & Frandsen, 2023). Our illustration of applying the values to evaluating the chatbots' dialogues responds to increasing calls to align LLMs' evaluation metrics with pedagogical principles and perspectives from learners and educators (Demszky et al., 2023; Jurenka et al., 2024). Design practitioners can incorporate VSD and PD in workshops and user interviews within shorter design cycles (He et al., 2024; Leiser et al., 2023). They can leverage existing benchmark datasets (Hendrycks et al., 2021; Jurenka et al., 2024) to evaluate the technology for both learning performance and value alignment with education stakeholders.

Limitations and future research

This work has two main limitations. First, the empirical investigation was grounded in the perspectives of a small group of stakeholders and might not capture a full range of experiences and values. Second, human values are in flux and influenced by current events. Chatbots with a set knowledge cut-off date in their training data may not always be aware of these events and values. Building pipelines for iterative value sampling, evaluations, knowledge updates and response refinement might address this limitation.

Future research can include a broader range of stakeholders and instructional contexts to illustrate the application of VSD and PD to the conceptualization, development, evaluation and improvement of educational technologies. Future work can report on design iterations stemming from human and automated evaluation in more depth. Studies can examine how students interact with and learn from the chatbots, including how they perceive and respond to the designed values, given varied knowledge and disposition to environmental science and AI.

CONCLUSION

Concerns about LLMs' bias, transparency and accuracy become even more problematic when the models are simulating human perspectives, as learners interacting with the technologies may trust AI's outputs uncritically. A value-sensitive approach to technology development helps generate design principles that address ethical concerns. This approach surfaces important pedagogical insights about how AI can support identity development, well-being, human–AI relationships and sustainability. Our work provides an example of translating values into the conceptualization, development and evaluation of learning technologies, to increase technology adoption and engagement.

FUNDING INFORMATION
This work was supported by the National Science Foundation through grant #2241596.

CONFLICT OF INTEREST STATEMENT
The authors do not have any conflict of interest to disclose.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

ETHICS STATEMENT
The research procedures have been approved by the Institutional Review Board at University of California-Irvine (#2801).

ORCID
Ha Nguyen https://orcid.org/0000-0001-7138-1427

REFERENCES
Amérigo, M., Aragonés, J. I., de Frutos, B., Sevillano, V., & Cortés, B. (2007). Underlying dimensions of ecocentric and anthropocentric environmental beliefs. *The Spanish Journal of Psychology*, *10*(1), 97–103. https://doi.org/10.1017/S1138741600006351

Ardoin, N. M., Bowers, A. W., & Gaillard, E. (2020). Environmental education outcomes for conservation: A systematic review. *Biological Conservation*, *241*, 108224. https://doi.org/10.1016/j.biocon.2019.108224

Bae Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K., & Følstad, A. (2021). When the social becomes non-human: Young people's perception of social support in chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).

Baumer, E. P., Blythe, M., & Tanenbaum, T. J. (2020). Evaluating design fiction: The right tool for the job. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 1901–1913).

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2022). What is participatory design? In S. Bødker, C. Dindler, O. S. Iversen, & R. C. Smith (Eds.), *Participatory design* (pp. 5–13). Springer International Publishing. https://doi.org/10.1007/978-3-031-02235-7_2

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1125–1134).

Briggs, P., & Thomas, L. (2015). An inclusive, value-sensitive design perspective on future identity technologies. *ACM Transactions on Computer-Human Interaction*, *22*(5), 1–28.

Byrd, A. (2023). Truth-telling: Critical inquiries on LLMs and the corpus texts that train them. *Composition Studies*, *51*(1), 135–142.

Carlini, N., Tramèr, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, Ú., Oprea, A., & Raffel, C. (2021). Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)* (pp. 2633–2650). https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

Chen, B., & Zhu, H. (2019). Towards value-sensitive learning analytics design. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. https://doi.org/10.1145/3303772.3303798

Cheng, M., Piccardi, T., & Yang, D. (2023). CoMPosT: Characterizing and evaluating caricature in LLM simulations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Christoforakos, L., Gallucci, A., Surmava-Große, T., Ullrich, D., & Diefenbach, S. (2021). Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Frontiers in Robotics and AI*, *8*, 640444. https://doi.org/10.3389/frobt.2021.640444

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., & Munteanu, C. (2019). What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

De Dominicis, S., Schultz, P. W., & Bonaiuto, M. (2017). Protecting the environment for self-interested reasons: Altruism is not the only pathway to sustainability. *Frontiers in Psychology*, *8*, 1065. https://doi.org/10.3389/fpsyg.2017.01065

De Groot, J. I. M., & Steg, L. (2009). Morality and prosocial behavior: The role of awareness, responsibility, and norms in the norm activation model. *The Journal of Social Psychology*, *149*(4), 425–449. https://doi.org/10.3200/SOCP.149.4.425-449

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, *2*(11), 688–701.

Dexe, J., Franke, U., Nöu, A. A., & Rad, A. (2020). Towards increased transparency with value sensitive design. In *International Conference on Human-Computer Interaction* (pp. 3–15). Springer International Publishing.

DiSalvo, B., Yip, J., Bonsignore, E., & Carl, D. (2017). Participatory design for learning. In B. DiSalvo, J. Yip, E. Bonsignore, & D. Carl (Eds.), *Participatory design for learning*. Routledge.

Friedman, B. (1996). Value-sensitive design. *Interactions*, *3*(6), 16–23.

Friedman, B., & Hendry, D. (2012). The envisioning cards: A toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1145–1148). https://doi.org/10.1145/2207676.2208562

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. MIT Press.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends in Human–Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Friedman, B., Kahn, P. H., Jr., Hagman, J., Severson, R. L., & Gill, B. (2006). The watcher and the watched: Social judgments about privacy in a public place. *Human Computer Interaction*, *21*(2), 235–272. https://doi.org/10.1207/s15327051hci2102_3

Gadiraju, V., Kane, S., Dev, S., Taylor, A., Wang, D., Denton, E., & Brewer, R. (2023). "I wouldn't say offensive but…": Disability-centered perspectives on large language models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 205–216).

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, *50*(3), 1097–1179. https://doi.org/10.1162/coli_a_00524

Gerdes, A., & Frandsen, T. F. (2023). A systematic review of almost three decades of value sensitive design (VSD): What happened to the technical investigations? *Ethics and Information Technology*, *25*(2), 26.

Graf, A., & Bernardi, R. E. (2023). ChatGPT in research: Balancing ethics, transparency and advancement. *Neuroscience*, *515*, 71–73. https://doi.org/10.1016/j.neuroscience.2023.02.008

Grové, C. (2021). Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in Psychiatry*, *11*, 606041. https://doi.org/10.3389/fpsyt.2020.606041

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

He, J., Houde, S., Gonzalez, G. E., Silva Moran, D. A., Ross, S. I., Muller, M., & Weisz, J. D. (2024). AI and the future of collaborative work: Group ideation with an LLM in a virtual canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (pp. 1–14).

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., & Steinhardt, J. (2021). Measuring mathematical problem solving with the math dataset. *arXiv Preprint arXiv:2103.03874*.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., & Bittencourt, I. I. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *1*, 23–32.

Jansen, B. J., Jung, S., & Salminen, J. (2023). Employing large language models in survey research. *Natural Language Processing Journal*, *4*, 100020.

Jurenka, I., Kunesch, M., McKee, K. R., Gillick, D., Zhu, S., Wiltberger, S., Phal, S. M., Hermann, K., Kasenberg, D., & Bhoopchand, A. (2024). Towards responsible development of generative AI for education: An evaluation-driven approach. *arXiv Preprint arXiv:2407.12687*.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274.

Kazemitabaar, M., Ye, R., Wang, X., Henley, A. Z., Denny, P., Craig, M., & Grossman, T. (2024). CodeAid: Evaluating a classroom deployment of an LLM-based programming assistant that balances student and educator needs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–20). https://doi.org/10.1145/3613904.3642773

Kim, M.-S., & Stepchenkova, S. (2020). Altruistic values and environmental knowledge as triggers of pro-environmental behavior among tourists. *Current Issues in Tourism*, *23*(13), 1575–1580. https://doi.org/10.1080/13683500.2019.1628188

Le Dantec, C. A., Poole, E. S., & Wyche, S. P. (2009). Values as lived experience: Evolving value sensitive design in support of value discovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1141–1150).

Leiser, F., Eckhardt, S., Knaeble, M., Maedche, A., Schwabe, G., & Sunyaev, A. (2023). From ChatGPT to FactGPT: A participatory design study to mitigate the effects of large language model hallucinations on users. In *Proceedings of Mensch und Computer 2023* (pp. 81–90).

Lewis, E., Mansfield, C., & Baudains, C. (2008). Getting down and dirty: Values in education for sustainability. *Issues in Educational Research*, *18*(2), 138–155.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474.

Lockton, D., Singh, D., Sabnis, S., Chou, M., Foley, S., & Pantoja, A. (2019). New metaphors: A workshop method for generating ideas and reframing problems in design and beyond. In *Proceedings of the Conference on Creativity and Cognition* (pp. 319–332). https://doi.org/10.1145/3325480.3326570

Logler, N., Yoo, D., & Friedman, B. (2018). Metaphor cards: A how-to-guide for making and using a generative metaphorical design toolkit. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 1373–1386).

Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023). GPTeach: Interactive TA training with GPT-based students. In *Proceedings of the Tenth ACM Conference on Learning @ Scale* (pp. 226–236). https://doi.org/10.1145/3573051.3593393

Menkhoff, T., & Gan, B. (2023). Engaging students through conversational chatbots and digital content: A climate action perspective. In *Proceedings of the 9th International Conference on Human Interaction and Emerging Technologies (IHIET-AI 2023)*, Lausanne, Switzerland, April 13–15, *70* (pp. 334–347). https://doi.org/10.54941/ahfe1002960

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (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*, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Moore, S., Tong, R., Singh, A., Liu, Z., Hu, X., Lu, Y., Liang, J., Cao, C., Khosravi, H., Denny, P., Brooks, C., & Stamper, J. (2023). Empowering education with LLMs—The next-gen interface and content generation. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (pp. 32–37). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_4

Muccione, V., Vaghefi, S. A., Bingler, J., Allen, S. K., Kraus, M., Gostlow, G., Wekhof, T., Colesanti-Senni, C., Stammbach, D., Ni, J., Schimanski, T., Yu, T., Wang, Q., Huggel, C., Luterbacher, J., Biesbroek, R., & Leippold, M. (2024). Integrating artificial intelligence with expert knowledge in global environmental assessments: Opportunities, challenges and the way ahead. *Regional Environmental Change*, *24*(3), 121. https://doi.org/10.1007/s10113-024-02283-8

Muller, M., & Liao, Q. V. (2017). *Exploring AI ethics and values through participatory design fictions*. Human Computer Interaction Consortium.

Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory and discussion. *ACM Journal of Data and Information Quality*, *15*(1), 21.

Nguyen, H., Nguyen, V., López-Fierro, S., Ludovise, S., & Santagata, R. (2024). Simulating climate change discussion with large language models: Considerations for science communication at scale. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 28–38). https://doi.org/10.1145/3657604.3662033

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, *91*(1), 188–204.

OECD. (2024). OECD AI principles overview. https://oecd.ai/en/ai-principles. Retrieved July 1, 2024.
Park, J., & Choi, D. (2023). AudiLens: Configurable LLM-generated audiences for public speech practice. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–3). https://doi.org/10.1145/3586182.3625114
Payne, B. H. (2019). *An ethics of artificial intelligence curriculum for middle school students*. MIT Media Lab Personal Robots Group.
Peters, D., Loke, L., & Ahmadpour, N. (2021). Toolkits, cards and games–a review of analogue tools for collaborative ideation. *CoDesign*, *17*(4), 410–434.
Prieto, L. P., Rodríguez-Triana, M. J., Dimitriadis, Y., Pishtari, G., & Odriozola-González, P. (2023). Designing technology for doctoral persistence and well-being: Findings from a two-country value-sensitive inquiry into student progress. In O. Viberg, I. Jivet, P. J. Muñoz-Merino, M. Perifanou, & T. Papathoma (Eds.), *Responsive and sustainable educational futures* (pp. 356–370). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-42682-7_24
Robnett, R. D., & Leaper, C. (2013). Friendship groups, personal motivation, and gender in relation to high school students' STEM career interest. *Journal of Research on Adolescence*, *23*(4), 652–664.
Schultz, P. W. (2001). The structure of environmental concern: Concern for self, other people, and the biosphere. *Journal of Environmental Psychology*, *21*(4), 327–339. https://doi.org/10.1006/jevp.2001.0227
Schultz, P. W., & Zelezny, L. (2003). Reframing environmental messages to be congruent with American values. *Human Ecology Review*, *10*(2), 126–136.
Schwartz, S. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, *2*(1), 11. https://doi.org/10.9707/2307-0919.1116
Shah, S. B., Thapa, S., Acharya, A., Rauniyar, K., Poudel, S., Jain, S., Masood, A., & Naseem, U. (2024). Navigating the web of disinformation and misinformation: Large language models as double-edged swords. *IEEE Access*, *1*, 1–21. https://doi.org/10.1109/ACCESS.2024.3406644
Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion—A study of human-chatbot relationships. *International Journal of Human-Computer Studies*, *149*, 102601.
Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT—Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, *74*, 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700
Strikwerda, L., van Steenbergen, M., van Gorp, A., Timmers, C., & van Grondelle, J. (2022). The value sensitive design of a preventive health check app. *Ethics and Information Technology*, *24*(3), 38. https://doi.org/10.1007/s10676-022-09662-x
Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246.
Tilbury, D. (1995). Environmental education for sustainability: Defining the new focus of environmental education in the 1990s. *Environmental Education Research*, *1*(2), 195–212. https://doi.org/10.1080/1350462950010206
Tilbury, D., & Wortman, D. (2008). How is community education contributing to sustainability in practice? *Applied Environmental Education & Communication*, *7*(3), 83–93. https://doi.org/10.1080/15330150802502171
Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., van der Haert, F. C., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., de Wever, W., & Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, *11*(1), 2468. https://doi.org/10.1038/s41467-020-15871-z
Turner, R. K., Paavola, J., Cooper, P., Farber, S., Jessamy, V., & Georgiou, S. (2003). Valuing nature: Lessons learned and future research directions. *Ecological Economics*, *46*(3), 493–510. https://doi.org/10.1016/S0921-8009(03)00189-7
Umbrello, S., & Van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, *1*(3), 283–296.
Vaghefi, S. A., Stammbach, D., Muccione, V., Bingler, J., Ni, J., Kraus, M., Allen, S., Colesanti-Senni, C., Wekhof, T., & Schimanski, T. (2023). Chatclimate: Grounding conversational AI in climate science. *Communications Earth & Environment*, *4*(1), 480.
van Wynsberghe, A. (2017). Designing robots for care: Care centered value-sensitive design. In *Machine ethics and robot ethics*. Routledge.
Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2), 2053951717743530. https://doi.org/10.1177/2053951717743530
Vernim, S., Bauer, H., Rauch, E., Ziegler, M. T., & Umbrello, S. (2022). A value sensitive design approach for designing AI-based worker assistance systems in manufacturing. *Procedia Computer Science*, *200*, 505–516. https://doi.org/10.1016/j.procs.2022.01.248
Viberg, O., Jivet, I., & Scheffel, M. (2023). Designing culturally aware learning analytics: A value sensitive perspective. In O. Viberg & Å. Grönlund (Eds.), *Practicable learning analytics* (pp. 177–192). Springer International Publishing. https://doi.org/10.1007/978-3-031-27646-0_10
Wambsganss, T., Höch, A., Zierau, N., & Söllner, M. (2021). Ethical design of conversational agents: Towards principles for a value-sensitive design. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation through*

information systems (pp. 539–557). Springer International Publishing. https://doi.org/10.1007/978-3-030-86790-4_37

Winkler, T., & Spiekermann, S. (2021). Twenty years of value sensitive design: A review of methodological practices in VSD projects. *Ethics and Information Technology*, *23*, 17–21.

Woelfer, J. P., Iverson, A., Hendry, D. G., Friedman, B., & Gill, B. T. (2011). Improving the safety of homeless young people with mobile phones: Values, form and function. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1707–1716). https://doi.org/10.1145/1978942.1979191

Wu, T., Terry, M., & Cai, C. J. (2022). AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–22). https://doi.org/10.1145/3491102.3517582

Yoo, D., Huldtgren, A., Woelfer, J. P., Hendry, D. G., & Friedman, B. (2013). A value sensitive action-reflection model: Evolving a co-design space with stakeholder and designer prompts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 419–428).

Zimmerman, A., Janhonen, J., & Beer, E. (2023). Human/AI relationships: Challenges, downsides, and impacts on human/human relationships. *AI and Ethics*, *4*(4), 1555–1567. https://doi.org/10.1007/s43681-023-00348-8

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nguyen, H., Nguyen, V., Ludovise, S., & Santagata, R. (2025). Value-sensitive design of chatbots in environmental education: Supporting identity, connectedness, well-being and sustainability. *British Journal of Educational Technology*, *56*, 1370–1390. https://doi.org/10.1111/bjet.13568