A Chatbot Won't Judge Me: An Exploratory Study of Self-disclosing Chatbots in Introductory Computer Science Classes

Quinn Goddard
Linguistics and Community Health Sciences
University of Calgary
Calgary, Canada
quinn.goddard@ucalgary.ca

Nathan Moton
Computer Science
University of Calgary
Calgary, Canada
nathan.moton@ucalgary.ca

Jonathan Hudson
Computer Science
University of Calgary
Calgary, Canada
jwhudson@ucalgary.ca

Helen Ai He
Computer Science
University of Calgary
Calgary, Canada
helen.he1@ucalgary.ca

ABSTRACT

Students in introductory Computer Science (CS) courses sometimes struggle with learning course content, but feel these struggles are uniquely theirs. To foster a more inclusive CS culture and normalize challenges in the learning process, we designed a conversational agent ("chatbot") that self-discloses information about the chatbot's own imaginary struggles with learning course material. Inspired by previous work in the mental health domain where humans reciprocated disclosure when a chatbot disclosed sensitive information, our goal was to promote student self-disclosure of learning challenges and to help students feel less alone. To inform design, we first conducted three focus groups with CS students on themes of identity and belonging. Based on these findings, we designed a self-disclosing chatbot ("Mibi") and deployed it in a pilot summer course (40 students) and a larger course (460 students) in the fall semester of 2023. Our work is the first real-world deployment of a chatbot in higher education for promoting student wellbeing, rather than assisting with practical course content. We highlight findings from this exploratory study, sharing how students engaged with Mibi, where it succeeded, where it has room to grow, and how that can inform future iterations of this promising new classroom companion for student mental health.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**.

KEYWORDS

Chatbot, Self-Disclosure, Mental well-being, Computer Science, Qualitative, CS1/CS2

WCCCE '24, May 02–03, 2024, Kelowna, BC, Canada

ACM ISBN 979-8-4007-0997-5/24/05
https://doi.org/10.1145/3660650.3660662

ACM Reference Format:
Quinn Goddard, Nathan Moton, Jonathan Hudson, and Helen Ai He. 2024. A Chatbot Won't Judge Me: An Exploratory Study of Self-disclosing Chatbots in Introductory Computer Science Classes. In *The 26th Western Canadian Conference on Computing Education (WCCCE '24), May 02–03, 2024, Kelowna, BC, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3660650.3660662

1 INTRODUCTION

Students in introductory Computer Science (CS) courses sometimes struggle with learning course content, but feel these struggles are uniquely theirs. In online course delivery, students may lack access to social and contextual cues of how others are doing in the class [23]. In online platforms, students may curate content to portray themselves in overly flattering ways [11]. In typical classrooms, a small subset of vocal, confident individuals who have previous CS or programming experience may publicly "show off" how much they know; instructors may inadvertently teach to these advanced students without an awareness of the true needs the class [4]. Together, these factors may alienate those who already feel like "outsiders" to CS, increasing feelings of isolation, and the notion that experiencing challenges, making mistakes, or failing in the learning process is abnormal or even shameful. Over time, this can lower student confidence and self-perceived ability to succeed in the discipline, and may even affect their desire to stay in the field [9].

To build more inclusive CS cultures and to normalize the encountering of challenges in the learning progress, we designed and deployed a self-disclosing conversational agent (chatbot) for use in higher education introductory CS courses. Our chatbot is programmed to self-disclose information about the chatbot's own imaginary struggles with learning CS content, where the goal is to facilitate student self-disclosure of learning challenges and to help the student recognize a sense of *common humanity* - the notion that everyone makes mistakes, struggles and fails in the learning process [16, 21]. Our chatbot was deployed in two in-person introductory CS courses (40 students, 460 students) in 2023. This paper presents findings from this exploratory study, and discusses design implications and opportunities for chatbots in higher education for fostering CS inclusivity and student wellbeing.

2 RELATED WORK: CHATBOTS IN MENTAL HEALTH AND EDUCATION

We situate our work among two areas of related literature: 1) chatbots in mental health domains, and 2) chatbots in education. In the mental health domain, Fitzpatrick et al. designed a cognitive behavioral therapy chatbot ("Woebot") for college students struggling with depression and anxiety [6]. Lee et al. designed a self-compassion chatbot ("Vincent") who was designed to use harsh and self-judgemental language to oneself to prompt care and compassion by human participants for Vincent [13]. Interestingly, the literature on chatbots has found that people apply the same social rules, norms, and expectations (as they would in human relationships) to interactions with chatbots - known as the *Computer as Social Actors (CASA)* Paradigm [7]. Previous research has even found a reciprocal self-disclosure effect when humans engage in dialogue with chatbots about sensitive or personal issues. For example, Kim et al. found that chatbots used in web surveys could prompt higher quality self-disclosure from humans, compared to web surveys without the use of chatbots [20]. Lee et al. compared none, low, and high self-disclosure experimental conditions, and found that chatbots with high self-disclosure prompted a reciprocal disclosure effect from humans, compared to chatbots with none or low self-disclosure levels [15]. A follow-up study found that a self-disclosing chatbot successfully acted as an intermediary between a human participant and mental health professional, whereby participants who developed trust with the self-disclosing chatbot then transferred that trust to a real mental health professional [14].

Outside the mental health domain, chatbots have been increasingly used in academic settings. In a 2023 scoping review of chatbots in higher education by Pereira et al., the authors identify the use of chatbots in supporting enrolment, collaborative learning, academic tasks, writing support, administrative support, or obtaining information on the degree of students' satisfaction in a course [19]. Pereira et al. categorize these chatbot applications in terms of interpersonal skills training, logistic support, educational support, or other miscellaneous purposes. Another systematic literature review on chatbots in education by Wollny et al. (2021) identified the use of chatbots to support instructors and teaching assistants in answering practical student questions on course content (such as "When is the assignment due?"), or acting in a mentoring role to support students in planning or reflecting on their learning progress [22].

To date, though, the literature is still sparse on the use and deployment of chatbots to support student wellbeing in higher education. One exception is a one-page abstract by Kumar et al. which contributes a short pilot study of students interacting with three different chatbot designs for 5 minutes in a laboratory setting to support student mental health [12]. We build off of this initial work, to offer the first design and deployment of a self-disclosing chatbot for student wellbeing in introductory CS courses at a large university in Canada.

We explore the overarching research questions: ***How receptive are students to a self-disclosing chatbot in the classroom, and what kind of interactions do students engage in with the chatbot (if any)? Specifically, are students willing to disclose emotionally vulnerable information to the chatbot?***

3 THE PRESENT STUDY: CHATBOT DESIGN, IMPLEMENTATION, AND DEPLOYMENT

To inform the design of a chatbot to prompt student self-disclosure of learning challenges, we conducted this research in three phases: 1) qualitative focus groups, 2) software design and implementation, and 3) chatbot deployment in two introductory CS courses at the University of Calgary. This research was approved by the University of Calgary's Ethics Board (REB21-1051).

3.1 Phase 1: Focus Groups

[Data Collection and Analysis.] To inform the design of our chatbot, we conducted three exploratory focus groups in June 2023. Students currently enrolled in graduate and undergraduate CS courses were invited to participate through class email. To reduce perceptions of hierarchy, recruitment emails were sent by a student (the first author who was a graduate research assistant at the time of the study), rather than a professor. All three focus groups were conducted in-person in meeting rooms on the university campus. 14 participants (5: Group 1, 5: Group 2, 4: Group 3) received a $20 gift card after consenting and before the focus group began.

Sessions were audio-recorded and subsequently fully transcribed (21,146 words). To encourage participants to feel safer opening up about their experiences at university, the interviewer was a student rather than a professor. To facilitate discussion that might otherwise be hindered by power-related social dynamics, the interviewer organized participants into either an all-female focus group (since women often feel marginalized in the CS classroom [8]) or an all-graduate-student focus group (so that undergraduate students would not feel intimidated, or potentially had one of the graduate students as a previous TA). The remaining focus group was mixed gender. Focus groups followed a semi-structured interview protocol, covering two major themes: (1) if and how students have ever felt isolated or like they "didn't belong" in a CS classroom, and (2) how receptive they would be to having a hypothetical chatbot to talk to about their struggles in a CS introductory course. This second theme took on a co-design approach, where students were given the chance to both evaluate and suggest potential chatbot functionalities.

To analyze the transcripts, three members of the research team used an inductive, open-coding method [5]. First, each member individually coded and analyzed all interviews using the process of affinity diagramming [10]. Each coder generated ad-hoc tags for thematic categories, and then hierarchically refined, re-organized, and merged the tags to identify commonalities between themes. Next, using a shared virtual whiteboard, the three coders met three times, to discuss and refine the independently-generated themes. In the findings below, we present the final themes that emerged from these collaborative analysis sessions.

[Findings.] Three themes emerged from our data analysis around belonging in CS: feelings of isolation, fear of judgement, and the willingness to disclose struggles and weaknesses to someone whom they felt comfortable with. Students often felt feelings of isolation in the CS classroom due to comparisons with others, perceiving that others are *"faster"* than them, or that the content was somehow *"easier"* for these more talented students. For example, one participant felt that the teaching style in their courses catered to

"the fastest person". Another student expressed a similar sentiment, saying that the course content "[was] going way too fast for me, I don't know if I'm just slow..."). A key contributor to this sense of alienation was a lack of prior coding knowledge, with many feeling that lectures often catered to the learning speed of the more experienced students (who are also more vocal about their knowledge in class). This led to a sense that they do not belong in the classroom (that maybe CS *"is not for me"*), a feeling often compounded by being a demographic minority within the classroom.

We found that students were largely very receptive to the idea of a classroom chatbot, identifying the potential for it to feel like an uniquely comforting conversational agent if one feels isolated in class. Specifically, students perceived a potential chatbot as a conversational agent that was not judgemental, allowing them to be free of any negative social perceptions of struggling or not understanding CS course content. For instance, one student said that *"I don't feel shy because I understand that it's just an AI, no one is seeing my question, so who is judging me?"* , while another noted *"I think I would definitely open up to the chatbot, it doesn't judge!"* This sentiment was explicitly echoed by three other students, and was met with non-verbal agreement in the focus groups.

Similarly, students perceived the anonymity offered by the chatbot as appealing, again offering a safer space to be vulnerable about their struggles, and protecting against the social fear of being singled out. For instance, some participants noted that in courses they *"try to be at the back so I'm not really singled out"*, and prefer online platforms, since as one student noted: *"it's easier sometimes to ask things on Discord [since] nobody's kind of singling you out and stuff"*. This fear of being "singled out" is nicely remedied by a chatbot, something that was noted by participants (for example, *"you didn't want to look like you had a lot of questions, but if you are anonymous or there is a chatbot, you feel just comfortable to ask as many questions [as you want]"*.

Interestingly, participants distinctly did not want the chatbot to seem overly human. For instance, they disliked the idea of it having a "life" outside of the study that it would discuss (such as family members, travels, etc.). Instead, since *"obviously it will not have emotions or those human-like attributes"*, students felt that this programmed humanity would make them more reluctant to open up to the chatbot. Other participants noted that having a fake backstory would make the chatbot *"not feel genuine"*), and that they would not want the chatbot to *"pretend that it has had similar experiences or relate to me on this level"*). Furthermore, students perceived a very human-like chatbot as carrying all the baggage around social pressure that this study aimed to eliminate in the first place. For instance, one student stated, *"I think the fact that the chatbot is not judgemental really helps. So if it's a human, I don't want to talk to it."* This highlights the unique role that chatbots can have - if they are human enough to be conversational and supportive, but not too human to make people fear the social judgement inherent in interacting with other humans.

The primary functionality that students suggested for the chatbot was the ability to answer questions about course content, allowing it to serve as a non-judgemental tutor. While the anonymity inherent in conversing with a chatbot would perhaps make students more willing to ask course content questions, a chatbot designed for learning instructional material is unrelated to the goal of this present study. For this reason, we did not expand the chatbot's functionality to include this suggestion, and instead leave it for future research to determine how best to integrate the instructional and self-disclosing roles that a chatbot can have.

3.2 Phase 2: Software Design and Implementation

The chatbot name "Mibi" was selected to ensure that (i) it wasn't a pre-existing name that a student might have preconceptions about (for instance, if the chatbot was named "Franklin" and someone had been severely bullied by a "Franklin" at some point, that might hinder their engagement with the bot), and (ii) it was reasonably gender neutral, both to avoid any of the potential intimidation that women can feel talking to men in the CS classroom, and also to shake the predominance of servile female-voiced AI assistants. Below, we discuss details of our software implementation and human-chatbot dialogue design.

[Technologies.] The research team was informed by the course instructor that Discord would be used as the primary online community discussion platform for students registered in introductory CS courses. This made Discord the ideal platform for reaching the most students with a chatbot. The chatbot could exist parallel to the Discord server operational for the course, and use Discord's participant list and direct messaging infrastructure to initiate contact with candidate participants for the study. Thus, "Mibi" was built using the Discord API, Google Dialogflow, Google Natural Language, OpenAI GPT-4, and custom applications running on Amazon Web Services (AWS) that were written in Python, Java, and JavaScript/TypeScript.

A custom Python application was built to interface with Discord's public API so that the chatbot could initiate scheduled chat sessions, send and receive messages, and keep track of conversations with students on Discord. The application initiated contact with users on the server through direct messaging informing them of the study, the study protocol, and the research team they could inquire more information from. Following our ethics protocol, "Mibi" would then refuse to continue conversation until users had consented to further communications. Students could also withdraw this consent at any time through the direct message chat interface. To allow researchers to monitor the chatbot and schedule chat sessions with users, we built a custom web application that interfaced with the Python application using Java and JavaScript/TypeScript. Figure 1 presents the infrastructure for our self-disclosing chatbot.

Chatbot conversations were pre-defined scripted question and response sequences with a single source, some alternate paths, and a final sink. To enhance the natural flow of conversations between the chatbot and students, we integrated Dialogflow, Natural Language, and GPT-4 with our chatbot. These conversational paths were implemented in Dialogflow with a goal of eliciting on-topic responses after the bot self-disclosed information in its own messaging choices.

Whenever a user responded to a conversational decision point message from the chatbot (for example, *"What are your thoughts on the course so far?"*), it would first pass the user's response to GPT-4. Using Natural Language Processing (NLP), GPT-4 determined if the response was relevant to the question and also performed

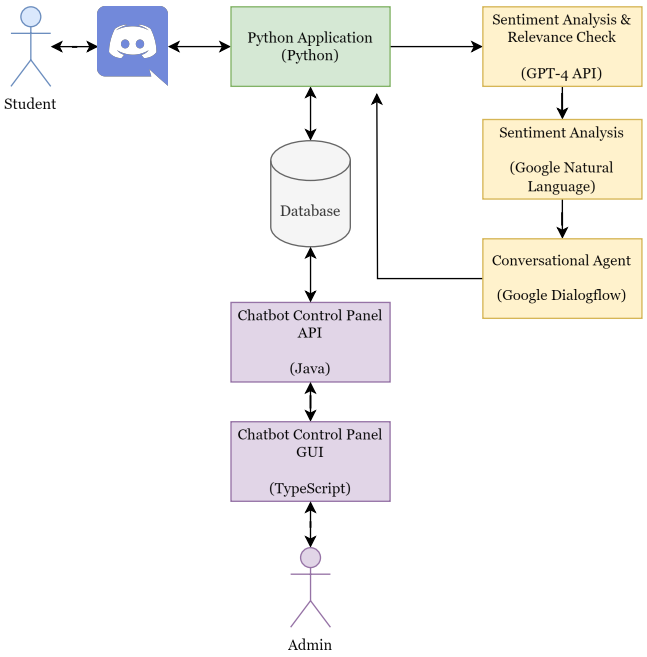

Figure 1: Infrastructure for the self-disclosing chatbot "Mibi"

an emotional sentiment analysis on the user's response. It was rated between -1 to 1, with -1 being very negative and 1 being very positive.

After confirming the response's relevance and analyzing it with GPT-4, the chatbot would then pass the same response to Natural Language, where an additional emotional sentiment analysis was performed using the same rating system. The average of these two ratings, along with the response, were then passed to Dialogflow, where Dialogflow would ultimately choose an appropriate response based on the evaluated sentiment of the student response.

If a response was off-topic, the chatbot would try to steer the conversation back by rephrasing the original question and asking the user to respond again until an on-topic response was provided (for example, *"Sorry, I didn't get that. What are your thoughts on the course again?"*). If a response was on-topic, the chatbot would respond appropriately by choosing a response based on the sentiment rating. For example, if a user responded positively (such as *"I like the course"*), the chatbot would respond with *"That's great! Keep up the good work! We've got this!"*. If a user responded negatively (such as *"This course is hard"*), the chatbot would respond with *"I'm sorry to hear that. Just keep at it. We've got this"*. After this conclusion of the decision point in the dialog flow conversational path graph, "Mibi" would then continue the conversation by asking more predefined questions within the scripted chat session.

Outside of any scripted chat sessions, if a user were to initiate a conversation with the chatbot (for example, *"How are you?"*), the chatbot would also use Dialogflow to respond with the best predefined message. If no suitable messages were found, the chatbot would simply tell the user it did not know how to answer that question and recommend asking the research team.

Some additional features were added to make chatting with the chatbot seem as realistic as possible. First, a custom message queue was built to support responses sent across multiple messages. The chatbot also waited up to a minute before responding if it detected the user it was chatting with was still typing after receiving a message. Second, the chatbot was designed to operate continuously, 24 hours a day, meaning users could respond to the chatbot at any time. Finally, if there was any downtime, the chatbot would scan for any missed messages after returning to online operation and respond appropriately.

[Dialogue Design.] To maintain control over what the chatbot says, we scripted all possible chatbot responses. Based on our focus group findings, in order to promote support and feelings of common humanity for students, the chatbot needed to have a specific character, one that was always supportive, humble, and self-deprecating. Since the chatbot is put into a role where students are potentially being vulnerable, giving a Large Language Model (LLM) free range to respond to the student could potentially be damaging to the student's mental health, either reinforcing negative self-perceptions or otherwise making the student feel insecure. As a result, the chatbot dialogue was designed with two primary guiding ideas: (i) the chatbot has no backstory or "life" outside of the course, so as to adhere to the suggestions coming out of the focus groups that it should not seem too "human", and (ii) the chatbot always struggles with course content to foster self-disclosure (and so to never make a student feel as if they are inferior to the chatbot, since that is the exact opposite goal of the study).

The chatbot was written to speak informally, sporadically using texting slang like 'lol' or 'tbh' and omitting periods at the end of sentences to make the chatbot seem more like a peer. Dialogues were structured as branching conversations, with positive and negative branches based on the sentiment of the student's response to the chatbot. An example of the chatbot's side of a conversation is given below:

> *Hey! Looks like marks from the first exam got released a while ago Were you happy with how you did?*
> [STUDENT RESPONSE]
>
> IF STUDENT'S RESPONSE HAD **POSITIVE** SENTIMENT:
>
> > *That's good, I really wish I'd done better*
>
> *I felt pretty confident going in but I guess I was wrong lol*
> *What did you find the most difficult?*
>
> IF STUDENT'S RESPONSE HAD **NEGATIVE** SENTIMENT:
>
> *Sorry to hear that :(*
> *I feel the same though, I'm not very happy with my mark*
> *We got this though, I've heard that there's a big learning curve with comp sci*
> *What did you find the most difficult?*

3.3 Phase 3: Chatbot Deployment in Higher Education CS Classrooms

3.3.1 *Summer Semester - Pilot Course.*

[Data collection and analysis.] We first tested the chatbot in an intensive summer CS course in July 2023 for high school students,

taught at the University of Calgary. The chatbot initially contacted students over Discord to ask if they wished to consent to the study. If so, students were contacted three more times during the course, to check in with the student. These contact events would occur after major assignment or exam milestones, and would ask the student about assessments they just received, as well as their thoughts on upcoming assessments and the course in general. Compensation was not provided during this pilot study. 40 high school students enrolled in the course. Out of those, 11 students engaged with the chatbot in conversation. The instructor of this course was a member of the research team; however, in accordance with our ethics protocol, he did not have access to the Discord server, the study data, and did not discuss or analyze the anonymized data with our team until final grades were released and the period for grade appeals had passed.

[Pilot Findings.] We found a divide in how students interacted with the chatbot: some tried to "mess with" the bot", attempting to see how they could break it or figure out how it worked, while others genuinely talked to the chatbot, sometimes confiding their own struggles in the chatbot and making itself vulnerable. For instance, when asked by the chatbot *"What do you think the course will be like? Any thoughts on what is going to be the most difficult?"*, one student answered: *"i feel so pressed doing the assignments on time i work bad under pressure so definitely the midterm and final are going to stress me out like crazy, thats probably the hardest"*

We also observed some students employing conversational elements that one might typically expect to be reserved for human-to-human conversations, such as emojis to convey emotion, or saying goodbye to the chatbot. However, these were only sporadically observed. For example, when the chatbot asked about thoughts on the course and its assessments, one student wrote *"glad its over with* 😭", while another wrote *"I got a 53, its a pass* 😤☝. *Let ur overlords know I conquered this land with ease"*.

Finally, students also seemed to appreciate how the chatbot spoke more like a peer. For example, when the chatbot asked *"What was your experience like chatting with me?"*, one student said *"it was nice, youre very unexpectedly informal"*. Other students expressed similar sentiments.

In contrast to these displays of genuine vulnerability however, a handful of students (2 of the 11) primarily engaged with the chatbot as a piece of technology to mess with, either through saying things which are completely off topic, or through sending the bot things like gifs or memes.

Overall, our pilot study revealed some interesting effects, but did not receive as much engagement (11/40 students) as we had hoped. We offer a few interpretations for this. First, the high school students that took this class used this as an extra-credit course to prepare for university, and likely had an established interest in CS and are therefore more likely to have some previous experience with coding. As a result, they may be less likely to struggle with the sense of belonging experienced by university students who come to CS later. Secondly, the class was relatively small and held in-person, making it more likely that students would get to know each other in the physical classroom (compared to a large lecture theatre where the sheer number of other students might make one feel more anonymous). Perhaps due to the smaller class size with in-person delivery, students might feel less isolated and may not feel as compelled to interact with an online chatbot, compared students taking a typical large introductory CS course (usually >200 students). While our pilot data was limited, three preliminary findings came out of deployment in this summer course: (i) students had the capacity and willingness to be vulnerable to the chatbot, (ii) students used discourse devices (like greetings and emojis) which typically would only have use for a human conversational partner, and (iii) students appreciated the informal, human-like nature of the chatbot's language.

3.3.2 *Fall Semester - Introductory CS Course.*

[Data collection and analysis.] Following the pilot study, we deployed the chatbot in an introductory CS class in Fall 2023 (460 students). The chatbot was advertised to students in all three sections of the course. We offered a 1/30 chance of winning a $50 gift card for participation. 73/460 consented to participate. 34 students filled out a post-study survey at the end of the semester. The instructor of the course was outside the research team and did not have any access to the study data.

Data analysis began with the the first author reading through all student-chatbot conversations (6387 lines of dialogue). The first author then coded each conversation as disclosing vulnerability (or not), as well as instances of greetings and "messing with" the chatbot. We present the key findings from this data analysis.

[Findings.] Roughly half of the 73 consented users (37, 50.7%) self-disclosed vulnerable information to the chatbot. Many students made themselves intellectually vulnerable (*"I made a dumb mistake [on the assignment] which sucked but overall I did pretty good"; "I was feeling confident [about the final], then I did the practice exam and got 60%, so now Im a bit scared"; "No idea I've never coded before, very nervous tho as I would like to get a good mark"; "I got most of it [the assignment] but had to use google for some of it"*). Other students made themselves socially vulnerable (*"Havent really talked to anyone in my class tbh"; "Still struggling with socializing"; "No I don't really have any friends in the course"; "A lot of the people are pretty antisocial"*).

Only three students postured to the chatbot (*"I scored 100"; "It was easy"; "Light work for a guy like me"*). Four students tinkered with the chatbot, for instance by telling it jokes (*"I just got fired from my job at the keyboard factory. They told me I wasn't putting in enough shifts"*) or being mean or insulting (*"i hope u crash"*).

Interestingly, only one student mentioned anything about the conversations being read by the research team (*"You are actually useless [...] Just kidding :D I know y'all creepies are reading this right now I love your bot, it's epic, please don't hurt me"*), suggesting that most students genuinely treated the chatbot as a interlocutor to talk to, not only a prop in a research study.

Notably, participants demonstrated many instances of humanizing the chatbot. In addition to using emojis (as was observed in the pilot study), roughly half (36/73, 49.3%) greeted the chatbot at least once. These greetings ranged from simple goodbyes (👋) to more extensive parting words (*"Talk to you later mibi"; "Have a good one"; "ttyl! have a great day* 💪💪💪💪*"*), with some verging on the sarcastic (*"byebye mibibot i love u"*). Anecdotally, others would thank the chatbot for its supportive words (*"Bye Mibi!! Thanks for the pep*

talk!") or even comfort the chatbot if it had expressed struggles with the course (*"Good luck 🤞 "*).

[Post-Study Survey.] The post-study survey aimed to collect feedback on student experiences with the chatbot, asking questions around what students liked or did not like about "Mibi", how frequently they responded to it, and what made them want or not want to respond to it. The survey contained 5 multiple choice questions and two open-ended response questions, and was sent to students after the final assessment was sent out.

34 students responded to the post-study survey. 27 of these 34 students (79%) responded to the chatbot most or every time the chatbot contacted them. Findings revealed that for those who stopped talking to the chatbot, it seemed that a human had replaced their need for a chatbot. For example, 29% of the non-responders selected the option [I would rather talk to a human] and [I made friends in my class so I didn't need to talk to the chatbot]. Others never wanted to genuinely engage with the chatbot to start with, with 29% selecting [I just wanted a chance to win the money and never actually wanted to talk to it] and [I just wanted to test it out and see if I could break it or see how it worked]). However, the most common reason for not talking to the chatbot was lack of time, with 43% selecting [I didn't have time to talk to it / I was too busy with the course].

For those who did continue talking to the chatbot throughout the semester, the most commonly selected feedback was [It was friendly and seemed fun to talk to] (58%). However, wanting to break the chatbot (42%) and wanting a chance to win the money (42%) were the next most common motivators. Only three students selected [I could share stuff with the chatbot that I couldn't share with my friends, the TA, or the instructor], perhaps indicating our chatbot might not yet readily serve as an easy replacement for human support. However, when asked if students would engage with the chatbot if it was part of their regular classes, 28% chose [Definitely] and 56% chose [Maybe], indicating a general interest and receptivity to having a chatbot be present in the classroom.

Overall, students were generally positive about their interactions with the chatbot. In a free-response section of the survey, five students noted it's friendliness (*"very cute and fun to talk to"; "It gave positivity and encouraging messages"; "I like that it would always start the conversation with an exclamative Hi! or Hey! it felt very friendly"*). Seven students cited it's human-like, relatable dialogue (*"It was cool that it shared concerns other might have"; "Felt very relatable in dialogue"; "Got me thinking sometimes", "it used slang", "It seemed like a human"*) as reasons they liked talking to it.

4 FUTURE DIRECTIONS AND CONCLUSIONS

This present study is one of the few uses of the Computers as Social Actors (CASA) paradigm in an educational context, addressing a gap in the literature on the design and deployment of chatbots in higher education for fostering student wellbeing, and the deployment of self-disclosing chatbots, outside of the mental health domain. This exploratory study therefore presents an exciting first step in understanding the opportunities and challenges for the use of self-disclosing chatbots in higher education for student mental health.

Overall, our findings indicated that a self-disclosing chatbot can be a promising tool for helping students to engage in self-disclosure in introductory CS classrooms, but that it may only be helpful for a certain subgroup of students. When given the option to interact with the self-disclosing chatbot, relatively few students were interested in engaging with it, even with a monetary incentive. While it's possible this lack of engagement stems from a lack of interest or suspicion in research (that is, perhaps students did not want their responses to the chatbot to be read by anyone), the anonymous nature of data collection promised by the study (following our ethics protocol) makes it unlikely that this was the main factor. However, those who did engage with the chatbot tended to enjoy it. Therefore, we see a general hesitation to engaging with the chatbot, but those who did, tended to derive something from it, whether that be emotional support or simple enjoyment in engaging with a new piece of technology.

This lack of broad engagement should not necessarily be seen as a drawback of the technology, however. After all, we designed the chatbot with the hope that it would provide a sense of belonging to the minority of students who feel "out of place" in the CS classroom, ideally helping them to not lose hope and stay in the program. Therefore, a chatbot designed for these purposes should not necessarily be designed to have universal appeal, and should instead be designed for the minority that feel a low sense of belonging. Future research is needed to see if engaging with a self-disclosing chatbot like the one we designed can meaningfully impact attrition rates in CS students.

Based on the survey responses, we saw some evidence that people were replacing the chatbot with a peer (that is, 29% of students who stopped interacting with the chatbot cited "I made friends in my class so I didn't need to talk to the chatbot" as the reason). Since human-based common humanity can better mitigate self-judgement and self-blame when faced with challenges by using thoughts such as "I am not alone in this" and also make failures feel less threatening [17, 18], future iterations of the chatbot could be expanded to find human interaction for a student. For instance, the chatbot could pair students who have disclosed similar struggles for example, around identity or a lack of belonging in CS over Discord, so that they could make a potential human connection, as a result of confiding in the chatbot. This explores a similar idea to Lee et al's work, where a human developed trust with a chatbot, and then transferred that trust to a (real) mental health professional [14].

Collective disclosure is another potential avenue to explore - for instance, if a student expressed having failed a particular assignment, the chatbot could potentially foster a sense of common humanity by informing the student that 23 other students (for example) also failed that assignment. We chose not to go down this route for this study, since the chatbot having omniscient knowledge of the class made it seem less like a peer, and therefore less likely to encourage self-disclosure. This therefore remains an unexplored bot-mediated means of fostering common humanity, making it unclear which approach (a friendly, emotionally-vulnerable chatbot or comforting statistics of how other students may be struggling in the class) would be most effective for fostering student wellbeing.

With the growing popularity of ChatGPT, it is probably tempting to think that a natural future direction for classroom chatbots is to give them freer linguistic reign. However, this raises an important question: What is the *minimal* level of capability required of a chatbot to accomplish the task of making students feeling less alone

when encountering learning challenges in computer science? While using a Large Language Model (LLM) like ChatGPT might allow for freer conversation, it could be that human-scripted dialogue (as used in our study) is ultimately more comforting and can be more explicitly directed towards encouraging student mental health. As was mentioned previously, a LLM also has the potential to do more harm than good - for instance, if it begins to mirror students too carefully, it could start to reinforce negative thoughts for the student and cause them to spiral. Since it is difficult to truly constrain the output of LLMs, it could be they are not actually the right avenue to pursue in this context. Instead, thoughtful dialogue, perhaps written with input from mental health therapists or psychologists, may the potential to be far more effective. Of course it may be useful to expand the functionality of the chatbot so that it could have more naturalistic conversations (for example, by being able to respond to a wider variety of student questions), but simply falling back on something like ChatGPT might not be the best approach for this specialized setting.

As the ultimate goal of this work is to strengthen the CS learning culture (where the sharing of failures and challenges is normalized), our hope is that the chatbot could ultimately help CS students from all backgrounds feel empowered to thrive. This may be particularly important for CS students from underrepresented groups, such as BIPOC, Women, those without prior coding experience, and other underrepresented populations. Since students from these groups might feel alienated in their classes, a self-disclosing chatbot could serve as a companion to motivate them to stay in their degree program, ideally contributing to increasing inclusivity in the CS field. Given that a lack of diversity within CS has been cited as a driving force behind, for instance, socio-economic biases in prison sentencing algorithms [1], racially-biased facial recognition systems [2], and voice recognition technologies that exclude certain racial, gender and demographic groups [3], technologies that foster diversity in the CS classroom are a first step towards addressing these larger issues. Thus, future work may explore the impacts of a self-disclosing chatbot in aims to explicitly address recognized EDIA (equity, diversity, inclusion, accessibility) barriers in CS.

ACKNOWLEDGMENTS

This research was funded by the University of Calgary's Taylor Institute Scholarship of Teaching and Learning Grant. Thank you to Yi-chieh Lee for his feedback on the design and implementation of our chatbot. Thanks to Grace Ferguson, Lora Oehlberg, Leanne Wu, and Nathaly Verwaal for their feedback on earlier iterations of the study design.

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2022. Machine bias. In *Ethics of data and analytics*. Auerbach Publications, 254–264.

[2] Fabio Bacchini and Ludovica Lorusso. 2019. Race, again: how face recognition technology reinforces racial discrimination. *J. Inf. Commun. Ethics Soc.* 17 (2019), 321–335. https://api.semanticscholar.org/CorpusID:197736593

[3] Joan Palmiter Bajorek. 2019. Voice recognition still has significant race and gender biases. *Harvard Business Review* 10 (2019), 1–4.

[4] Lecia J. Barker and Kathy Garvin-Doxas. 2004. Making Visible the Behaviors that Influence Learning Environment: A Qualitative Exploration of Computer Science Classrooms. *Computer Science Education* 14, 2 (2004), 119–145. https://doi.org/10.1080/08993400412331363853 arXiv:https://doi.org/10.1080/08993400412331363853

[5] Juliet Corbin and Anselm Strauss. 2014. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (4 ed.). SAGE Publications, Thousand Oaks, California. https://methods.sagepub.com/book/basics-of-qualitative-research

[6] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Ment. Health* 4, 2 (June 2017), e19.

[7] Andrew Gambino, Jesse Fox, and Rabindra A Ratan. 2020. Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication* 1 (2020), 71–85. https://search.informit.org/doi/10.3316/INFORMIT.097034846749023

[8] Colleen M. Ganley, Casey E. George, Joseph R. Cimpian, and Martha B. Makowski. 2018. Gender equity in college majors: Looking beyond the STEM/non-STEM dichotomy for answers regarding female participation. *American Educational Research Journal* 55, 3 (2018), 453–487. https://doi.org/10.3102/0002831217740221

[9] Kathy Garvin-Doxas and Lecia J. Barker. 2004. Communication in computer science classrooms: understanding defensive climates as a means of creating supportive behaviors. *J. Educ. Resour. Comput.* 4, 1 (mar 2004), 2–es. https://doi.org/10.1145/1060071.1060073

[10] Gunnar Harboe and Elaine M. Huang. 2015. Real-World Affinity Diagramming Practices: Bridging the Paper-Digital Gap. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 95–104. https://doi.org/10.1145/2702123.2702561

[11] Alexandra Hosszu, Cosima Rughiniș, Răzvan Rughiniș, and Daniel Rosner. 2021. Webcams and social interaction during online classes: Identity work, presentation of self, and well-being. *Front. Psychol.* 12 (2021), 761427.

[12] Harsh Kumar, Kunzhi Yu, Andrew Chung, Jiakai Shi, and Joseph Jay Williams. 2023. Exploring The Potential of Chatbots to Provide Mental Well-being Support for Computer Science Students. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2* (Toronto ON, Canada) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 1339. https://doi.org/10.1145/3545947.3576285

[13] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300932

[14] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 31 (may 2020), 27 pages. https://doi.org/10.1145/3392836

[15] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Honolulu</city>, <state>HI</state>, <country>USA</country>, </conf-loc>) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[16] Kristin D. Neff. 2003. Self-Compassion: An Alternative Conceptualization of a Healthy Attitude Toward Oneself. *Self and Identity* 2, 2 (2003), 85–101. https://doi.org/10.1080/15298860309032

[17] Kristin D Neff. 2011. Self-compassion, self-esteem, and well-being. *Social and personality psychology compass* 5, 1 (2011), 1–12.

[18] Kristin D Neff and Katie A Dahm. 2015. Self-compassion: What it is, what it does, and how it relates to mindfulness. *Handbook of mindfulness and self-regulation* (2015), 121–137.

[19] Daniela S.M. Pereira, Filipe Falcão, Lilian Costa, Brian S. Lunn, José Miguel Pêgo, and Patrício Costa. 2023. Here's to the future: Conversational agents in higher education- a scoping review. *International Journal of Educational Research* 122 (2023), 102233. https://doi.org/10.1016/j.ijer.2023.102233

[20] Jungwook Rhim, Minji Kwak, Yeaeun Gong, and Gahgene Gweon. 2022. Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior* 126 (2022), 107034. https://doi.org/10.1016/j.chb.2021.107034

[21] Sara B Taylor, Lindsay A Kennedy, Caroline E Lee, and Emily K Waller. 2022. Common humanity in the classroom: Increasing self-compassion and coping self-efficacy through a mindfulness-based intervention. *J Am. Coll. Health* 70, 1 (Jan. 2022), 142–149.

[22] Sebastian Wollny, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler. 2021. Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence* 4 (2021). https://doi.org/10.3389/frai.2021.654924

[23] Tai-ming Wut and Jing Xu. 2021. Person-to-person interactions in online classroom settings under the impact of COVID-19: a social presence theory perspective. *Asia Pacific Education Review* 22, 3 (01 Sep 2021), 371–383. https://doi.org/10.1007/s12564-021-09673-1