Chatbots and Humane Design in Education

Kerry Nelson Milovic

Department of Education, Wilfrid Laurier University

Chatbots and Humane Design, EU 550N, A

April 28, 2021

Abstract

This research aims to explore the experience of chatbot use within academic advising from a humane design perspective. Although technology is a tool that facilitates people's efforts across all industries, there is limited research that examines the potential negative impacts of emerging technology and whether or not a particular technology implementation is beneficial to its users. Humane Design is an approach to IT systems design that aims to support those wanting to prioritize human needs in human-computer interactions. The Centre for Humane Technology has defined Humane Design along six instinctive human sensitivities vulnerable to new technologies. These sensitivities are labelled: emotions, attention, sense-making, decision-making, social reasoning, and group dynamics (Humane Design Guide, 2020). This study provides contextual support for a theoretical frame of humane design within the context of education theory and previous chatbot research. The study then uses this framework to examine students' experiences using a chatbot for academic advising. This mix-methods study examines whether chatbot use provides better information-seeking support than searching and using the navigation on a student support website. The hypothesis was that students would a) prefer using a chatbot, b) be more confident in the information they found, c) sense greater support of their human sensitivities d) take less time finding the answers than participants using the website. The results showed students preferred using a chatbot. Also, in three of the four scenarios, they did sense statistically significant greater support in one or more of the human sensitivities. However, the chatbot was not faster nor did it provide statistically different and better confidence scores in three of four scenarios.

Chatbots and Humane Design in Education

Technology is changing how people work, learn and live, yet it does not always meet our expectations. There has never been a generation so technologically engaged, and so the long-term societal effects of this level of technology use are not yet known. Common patterns of technology integration within educational contexts may or may not differ from the increasing adoption of technology on a broader scale. The Gartner Hype Cycle attempts to visually depict the common patterns of new technology emerging and being adopted using a line graph of expectations over time (“Hype Cycle Research Methodology,” n.d.). The cycle begins with innovation, followed by increased inflated expectations (“Hype Cycle Research Methodology,” n.d.). Expectations drastically drop into a trough of disillusionment and finally end with a slight increase in the slope of enlightenment (“Hype Cycle Research Methodology,” n.d.). This hype cycle is normally applied to explaining the process of one new technology emerging, yet it could indicate a broader societal experience of ever-increasing technology disruption.

It could be argued businesses are driving technology use, and those interests may at times be in conflict with societal interests like education. There are, however, clear advantages to people having access to technology to learn. Technology, among other benefits, facilitates research, communications and the use of media within learning. Steve Jobs, a renowned innovator and founder of Apple Inc., described computers as tools and the equivalent of bicycles for the mind (Laurence, 2006). He considered the efficiency that human bodies gain when using a bicycle, similar to the efficiency human minds gain when using a computer (Laurence, 2006). In this light of working towards increasing the efficiency of the human mind, the examination of technology’s success can be done through an educational lens. An educational perspective on technology use explores how it facilitates learning, the acquisition of new knowledge and understanding. This study examines technology integration from an educational perspective within an education context: online academic advising resources.

There is a wide range of research on the benefits of integrating technology in education. For example, research has been conducted on the benefits of using educational technology to improve student learning outcomes (Bernard et al., 2014; Tamim et al., 2011), supporting students with intelligent tutoring systems (Aeid & Meziane, 2019; Azfal et al., 2019; Saleheen et al. 2019), and providing educational policy insights (Wolfgang et al., 2012; Gulson & Webb, 2017). Although researchers have examined the use of chatbots in universities (Aeid & Meziane, 2019; Nurshatayeva et al., 2021; Saleheen et al., 2019), there was no research found on the social-psychological effects chatbot use has on university students requiring academic advising. There has been some research on chatbot effectiveness in similar counselling contexts (Kamalou et al., 2019).

While there is extensive research on the benefits of using technology, research into the potentially negative impact of technology on people is still an under-examined field (Silva et al., 2020). Silva et al. (2020) conducted a recent systematic review of research on technology's negative impacts. The review resulted in 67 articles from across all disciplines from 2005 to 2020 and was conducted using many keywords, synonyms and potential combinations of 'negative impacts' and 'technology' (Silva et al., 2020). Specific to the negative impacts of technology in education, research suggests that technology can create a reduced capacity to focus and deeply analyze (Carr, 2011). Also, researchers have shown that a negative relationship exists between smartphone dependency and student academic performance (Kuznekoff & Titsworth, 2013; Lepp et al., 2014; Longnecker, 2017; Rashid & Asghar, 2016). Also, personal-environment fit researchers have determined a trend of technostress, a psychological reaction of stress resulting from a misfit between individuals and their increasing technology-driven environment (Ayyagari et al., 2011; Brooks & Califf, 2017; Hung et al., 2015). Technostress within education has a significant positive correlation to burnout among university students (Wang et al. 2020). This research suggests that even young tech-savvy university students are not immune to stress caused by technology use not fitting their personal environment needs (Wang et al. 2020). Research on the potentially nefarious impacts of technology may be hindered by the constant and high-speed technological innovation, the interdisciplinary nature of human-machine interaction (Shaw et al., 2018), the overwhelming interest to cut costs through digital transformation, and the powerful marketing of the technology industry.

What is also notably absent from existing research on technology integration is the presentation a theoretical model within an academic context to support researchers' evaluation of technology and its application from the perspective of whether or not the technology integration is in the student's best interests. Educational technology theoretical frameworks have explained how technology defines processes (YouTube, 2015); how the intersectionality of technology and other content knowledge is critical to adoption (Koehler & Mishra, 2009) and the key challenges to overcome to accept technology (Petko, 2012). However, none of these frameworks provide a conceptual understanding of how human nature is impacted by technology or evaluate whether one technological implementation supports an individual better than another.

Literature Review

Humane Design

The Centre for Humane Technology (2020) was created to respond to the "attention attraction economy". According to the Centre for Humane Technology (2020), this powerful economy makes billions of dollars for technology giants who engineer free products to engage users and then sell their attention to advertisers. The Center for Humane Technology (2020) argues that attention-grabbing strategies contribute to a wide variety of problems facing humanity, including political polarization, loss of cognitive abilities like memory and focus, decreased mental health, and decreased social contact. The Center envisions a world where technology is realigned with humanity's best interests rather than businesses (Centre for Humane Technology, 2020). Using an approach they have coined as Humane Design, the center aims to support those wanting to prioritize human needs in human-computer interactions. The organization describes itself as being at the intersection of technology, systems transformation and human nature; it identifies its mission as advocating for knowledge integration among these domains.

The psychological effects of human-computer interactions have been studied within behavioral science for many years (Shaw et al., 2018). Yet, perhaps due to a lack of interdisciplinary collaboration, particularly from a theoretical standpoint between computer science and psychology, the human-computer relationship is still unclear and poorly defined after decades of research (Shaw et al., 2018). However, there is literature on considerations for a user's psychological needs in design outside of education. Within the field of building and landscape architecture, there is design for human flourishing that advocates for design that supports the client's well-being (Stevens et al, 2019). There are also many resources available to IT user experience designers to consider social psychological factors, including emotions and personality (Walther, 2020; Norman, 2004), cross-cultural design (Akpem, 2020; Heimgärtner, 2019) and accessibility (Kalbag, 2020).

User Experience (UX) design is a discipline that focuses on how users interact with products, and User Interface (UI) design refers specifically to the design of that interaction's sensory interface. Interactive systems have complex functionality beyond the sensory interface, which are invisible but have a considerable impact on the user's experience (Heimgärtner, 2019). There is broad consensus on the importance of nonfunctional hedonic aspects of user experience (Hassenzahl et al. 2003; Preim and Dachselt 2010). For some, experience has become the object of design rather than the thing itself (Heimgärtner, 2019). Don Norman (2004), a researcher and thought leader within UX design, describes user experience as conscious and unconscious. His research found three levels of emotional response managed through both affect and cognitive processing systems: visceral, behavioural and reflective (Norman, 2004). Visceral responses are related to appearance; behavioural responses to pleasure and functionality (Norman, 2004). Finally, reflective responses are to the rationalization and intellectualization of a product (Norman, 2004). These authors suggest human sensitivities are very much a consideration of modern UX/UI design.

UX/UI design is, however, distinct from humane design. Humane design focuses on how technological products support humans' needs without consideration of business interests. UX/UI design is about designing products; whereas humane design is about supporting human sensitivities to technology when using those products. To illustrate how user needs are considered within UX/UI Design, Walthers (2020), author of *Designing for Emotions,* remapped user needs on Maslow's hierarchy of human needs. While Maslow saw a human's ultimate need as self-actualization, Walther remapped the pyramid and identified a user's ultimate need as pleasure (Maslow, 1943; Walthers, 2020). Designing with a goal of pleasure is unlikely to be described as designing to help people reach their full potential or create efficiencies of the mind.

Designing with a goal of pleasure could be described as designing for a person's enjoyment. It could also be described as engineering pleasure for greater user interaction, sales, attention or dependency. Whether or not designing for pleasure is truly in a user's best interest would depend on the context.

Humane Design advocates for design to support a user's best interests. The Centre for Humane Technology has created a Humane Design Guide (2020) to guide a designer to support six instinctive human sensitivities that are vulnerable to new technologies. These sensitivities are labelled: emotions, attention, sense-making, decision-making, social reasoning and group dynamics (Humane Design Guide, 2020). Each of these sensitivities is discussed further below. The Centre anticipates that designers can identify if their product has a positive or negative effect on users from a humane design perspective by considering each of these sensitivities. It is important to note here the Centre for Humane Design is not just looking at whether human sensitivities are considered, but whether or not they are supported. In UX/UI design, human needs could be considered because they are leveraged to meet business interests and risk negatively impacting users. The following chart shows how good UX/UI design can be further understood when broken down by a humane design perspective. Supporting, leveraging and exploiting human sensitivities could all fit within a UX/UI definition of good design by supporting business interests. On the other hand, Humane design evaluates the support of human sensitivities and could evaluate good designs that support business interests as leveraging or exploiting human sensitivities.

Figure 1. Business and User Design Priority compared within Humane Design and User Design

Business and User Design Priority	Perspective	
	Humane Design	User Design
Good for users and good for business	The user’s decision-making agency is **supported** E.g., Interactive tax submission platforms designed to inform and support people making decisions on filing taxes	Business interests **supported**
Prioritize what is good for business, risk it is potentially bad for users	The user's decision-making agency is **leveraged** E.g. Pushing timely yet unsolicited advertising for a competitor to a product that a person recently viewed online, discussed, purchased or visited in the same geographical area	
Good for business at the users’ expense	The user's decision-making agency is **exploited** E.g. Clickbait that loads web pages in such a way that users are positioned to click on advertising accidentally	
User and business needs not met, so neither are prioritized	User's human sensitivities **are not supported** E.g. Tool that has poor usability.	Business interests **are not supported**

Each of the sensitivities described within Humane Design can be examined through an educational psychology lens and are explored below with examples of supporting and inhibiting those sensitivities within user design.

Emotional Sensitivity Network theory and social cognitive theory identifies managing one's emotions as a key element of self-regulation that improves learning (Schunk, 2020). Positive emotions can be associated with increased learner motivation, and negative emotions lower motivation (Pekrun, 2016). According to the Humane Design Guide, the emotion sensitivity relates to what we feel in our body and our physical health (The Centre for Humane Technology, 2020). Technology designed without considering emotional sensitivity can contribute to people feeling stressed, low on sleep, afraid or emotionally exhausted (The Centre for Humane Technology, 2020). Technology solutions support this human sensitivity when the design engenders calm, balance, safety, pauses and supports the circadian rhythm (The Centre for Humane Technology, 2020). An example of user design supporting emotions is the power button light that Apple designed to flash using the circadian rhythm (Walthers, 2020). The light mimics a calm circadian rhythm promoting relaxation, signals to users the computer is asleep, and reduces power consumption (Walthers, 2020). Network theory would describe this feature through four overlapping stages of emotional reactions: orienting complex, emotional event integration, response selection and sustained emotional context (Halgren & Marinkovic, 1995). From a network theory perspective, in this experience, the person observes the stimulus of the light, integrates the meaning of this stimulus with existing knowledge, then assigns meaning to the light and, finally, these events impact their mood. The light mimicking a circadian rhythm supports relaxation while improving usability through existing knowledge of the rhythm and its association with relaxation.

Some designs may leverage or exploit a user's emotions by using urgency signaling or artificial scarcity (The Centre for Humane Technology, 2020). These strategies use a person's emotions in a way that inhibits their feelings of calm or balance by, for example displaying artificial countdowns of time remaining to make a purchase or words like "only one seat remaining". These experiences can be associated with feelings of stress within the user's existing knowledge. A person's emotional sensitivity could also be unintentionally inhibited through poor design or user support from a functional perspective. An inability to do what a person is set out to do with technology could cause stress as a person feels unable to meet the expectations the tool has set out of them. Psychology researcher Seligman (1991) described this phenomenon of how negative experiences shape our future expectations as learned helplessness. Self-efficacy, a key variable in social cognitive theory, similarly describes a person's perception of their ability to produce actions (Schunk, 2020). Self-efficacy depends on a person's abilities and self-concept; how their past experiences shape their perception (Schunk, 2020). Self-efficacy is constructed by people using observational information like past performances on similar tasks and physiological indexes (e.g. heart rate, sweating) (Schunk, 2020). Self-efficacy is strongly related to effort and task persistence (Bandura & Cervone, 1986; Shunk & DiBenedetto, 2016). Struggles to use technology in the past or the current situation could cause stress and impact a person's sense of self-efficacy, decreasing motivation and making it more difficult to persist at a task.

Stress and physiological factors define an inhibited emotional experience within the emotional sensitivity of humane design; however, each of the human sensitivities could arguably be tied to a user's emotional response. Attention, sense-making, decision-making, social reasoning and group dynamics further define emotional responses specifically sensitive to technology use and they will be further explored.

Attention Sensitivity Information processing theory and social cognitive theory have demonstrated how we focus our attention is critical to learning (Schunk, 2020). The Centre for Humane Design (2020) describes the human sensitivity of attention as how and where we focus our attention. Technology designed without consideration of attention sensitivity can leave the user's attention fragmented or overwhelmed. In contrast, humanely designed technology can bring more focus and mindfulness (Centre for Humane Technology, 2020). An example of a good user design that supports attention is the Google search homepage. There is a very clear focus with only the search bar as the action for the user. The reflective process of defining search terms is critical to the user's success in this activity. The simplistic design is critical in focusing the user's attention on the task at hand, supporting the user in their goal.

Some designs may leverage or exploit a user's attention by providing no stopping queues like an infinite scroll or many undifferentiated choices (Humane Design Guide, 2020). These strategies capture a user's attention in a way that intentionally does not allow them to focus on something different by providing them with too many options which capture their attention. Attention is the sensitivity most likely to be associated with the layout or the visual interface as the number and design of visual cues for interaction impact this sensitivity. A person's attention sensitivity could also be unintentionally inhibited through poor design or user support from a functional perspective. The overuse of options represented with the colour red could, for example, draw and fragment a user's attention. Attention requires users have enough capacity to process information in their working memory (Schunk, 2020). Research has shown a limit to how much people can hold in their working memory (Cowan, 2001; Peterson & Peterson, 1959). Overwhelming a user's working memory can be described as exceeding their cognitive load (Benassi et al., 2014). Cognitive load theory defines two types of cognitive loads (Benassi et al., 2014). The first type of cognitive load is specific to the individual learner's information processing capacity, which varies between individuals (Benassi et al., 2014). As a result, it can be expected that different individuals' sensitivity to attention will be perceived differently. The second type of cognitive load depends on the inherent complexity of the topic (Benassi et al., 2014). As such, it can be expected that the required focus and attention to perform a task will vary depending on the inherent complexity of the task itself.

Sense-making Sensitivity Constructivist epistemology states that learning occurs as people create their own new knowledge by assigning meaning or sense to information (Schunk, 2020). Sense-making is about coordinating prior knowledge when confronted with a new situation and it enables humans to determine their interpretation of the current environment (Schunk, 2020). Sense-making is defined in the Humane Design guide as integrating what we sense with what we know (The Centre for Humane Technology, 2020). Technology designed without consideration of sense-making is marked by out-of-context and manipulative information, leading to confusion in the user. In contrast, technology designed to consider sense-making helps the user consider, learn, express, and feel grounded (The Centre for Humane Technology, 2020). An example of good user design that supports the sense-making sensitivity is when users are prompted to respond to direct emails. Direct emails are more likely to require a response than a mass email. Visual cues help users contextualize their email inbox and support integrating their existing knowledge with the current context in a meaningful way. In this case, the user is much more likely to make sense of their email inbox because the visual cues support a meaningful understanding of differences in the current context.

Some designs may leverage or exploit a user's sensemaking by providing facts out of context or using over-personalized filters (Humane Design Guide, 2020). An example of these strategies is how special interest groups can use data to target specific groups and provide out-of-context statistical information. These strategies manipulate sense-making by attempting to influence a person's existing knowledge. Sense-making is the sensitivity that is the most likely to be attributed to text because it is often through language that we acquire new information. A person's sensemaking sensitivity could also be unintentionally inhibited through poor design or user support from a functional perspective. Information presented to a user could be unclear or confusing because they cannot make sense of the information with their existing knowledge. This experience is particularly relevant to human-machine interaction as it is designed to be used without an additional person who is there to support users in understanding how the technology operates. Sense-making in human interaction contexts can be facilitated by another person who supports learners in putting information within contexts more relevant to them. Constructivist theorist Lev Vygotsky described the distance between a learner's development when supported by a more capable peer in his theory on the Zone of Proximal Development (Schunk, 2020). In this theory, the learner can bring their understandings to social interactions and construct new knowledge by integrating those understandings with the new context (Schunk, 2020). Technological products, often without social interactions, must support diverse users to build on their existing knowledge to create new knowledge.

Decision-making Sensitivity Decision-making, according to the Centre for Humane Technology's Design Guide, is how we align our actions with our intentions. Cognitive Learning

Processes states decision-making is an output of metacognitive processes (Shunk, 2020). Metacognition is the intentional, conscious control of mental activities that includes planning what is required, learning techniques, monitoring effectiveness and revising plans (Shunk, 2020). Metacognition helps us reflect on the adequacy of the conclusions we draw and determine if we are ready to make a decision (Shunk, 2020). Humanely designed technology supports a user's agency and their sense of control of their actions (Humane Design Guide, 2020). Technology designed without consideration of decision-making would not support or solicit a user's agency (Humane Design Guide, 2020). An example of good user design that supports the decision-making sensitivity would be any kind of interactive that supports a user in making complex decisions like how to file their taxes.

Some designs may leverage or exploit a user's decision-making by pushing out unsolicited content (Humane Design Guide, 2020). Decision-making is the sensitivity that is the most likely to be attributed to the specific series of steps a person takes to achieve their goal. A person's decision-making sensitivity could also be unintentionally inhibited through poor design or user support from a functional perspective. If a user senses through their metacognitive processes that there is no clear decision to take any action to meet their needs or the user is unsuccessful in planning the next steps to take, the user would feel unsupported in this sensitivity.

Social Reasoning Sensitivity Technology has made it possible for people to learn in social environments that were previously unimaginable. Social Cognitive theory focuses on how learning occurs within social environments and how people acquire knowledge, model behaviors and expected outcomes (Schunk, 2020). According to the Humane Design Guide, the social reasoning sensitivity is how we understand and navigate our personal relationships. This sensitivity is inhibited when our status, relationships and self-image are manipulated (Centre for Humane Technology, 2020). It is supported when we are enabled to connect more safely and authentically with others (Centre for Humane Technology, 2020). An example of good user design that supports the social reasoning sensitivity is video conferencing. Seeing a person's image while speaking with them allows users to pick up on physical gestures like a smile and connect more authentically with another person.

Some designs may leverage or exploit a user's sensemaking by quantifying social status or relationships (Centre for Humane Technology, 2020). An example would be labelling relationships or quantifying support for content shared. Social reasoning is the sensitivity that may most likely relate to social media platforms though it could apply to any social interaction supported by the technology. A person's social reasoning sensitivity could also be unintentionally inhibited through poor design or user support from a functional perspective by making it difficult for users to feel able to connect with others authentically.

Group Dynamics Sensitivity Group Dynamics from an educational theory perspective has been explored through identity theories. Social identity, perceived commonality with others within a group such as race, gender and sexual orientation, has been discussed within critical race theory, student development theory and queer theory (Jones & Abes, 2018). In addition to the importance of social identity, there is an element of personal identity which is critical to student identity. A student's identity is about how they perceive themselves as similar to and different from other people (Jones & Abes, 2018). These dynamics could be about social identities, perceived commonalities with others, or these dynamics could be about personal identities, perceived differences (Jones & Abes, 2018). According to the Humane Design Guide, the group dynamics sensitivity is about how we navigate larger group status and shared understanding (Center for Humane Technology, 2020). This sensitivity is supported when we are enabled to develop a sense of belonging and cooperation and inhibited when users are excluded, divided and mobilized through fear (Center for Humane Technology, 2020). An example of good user design supporting group dynamics is how Slack allows users to select their skin tone for all their skin-coloured gestures, e.g. a thumbs up. If a skin tone is not selected by a user, slack cycles through all skin tones as a default.

Some designs may leverage or exploit a user's group dynamics by enabling hate speech or political polarization (Center for Humane Technology, 2020). Social identity like race becomes more salient when it is oppressed (Jones & Abes, 2018). Critical race theory is a theoretical framework that examines institutional racism and oppression by dominant white culture (Jones & Abes, 2018). Queer theory, similar to critical race theory, examines the integration of queer culture within the curriculum and educational practices (Jones & Abes, 2018). Group dynamics could create a sense of exclusion in many technology areas, from how users enter data about their gender identity to how they feel represented in visual graphics. A person's group dynamics sensitivity could also be unintentionally inhibited through poor design or user support from a functional perspective. An example of this could be using colour to make value judgements. For example, the term blacklist has historically in technology meant an unsafe list, and the term whitelist means a safe list. This type of terminology could have unintended consequences on users' feelings of inclusion (Salesforce, 2021).

Chatbots

The purpose of a chatbot is to provide people with a way to navigate information through simulated dialogue. Chatbots enable people to express their intentions and receive relevant information. Similar to a search bar within a website, chatbots use natural language processing, a branch of artificial intelligence, to recognize words or key phrases and return artificially intelligent responses. Unlike using a search bar within a website, chatbots employ dialogue management in which administrators predetermine the paths the dialogue may take (Wlodek et al., 2000). A dialogue path could present the user with some options like a main menu or sub-menus, and it could ask open-ended questions relevant to the dialogue path the user is engaged in. These dialogue paths allow chatbot administrators to organize information so that they expect users to find what they are looking for on their own and thereby reduce demands on time for employees to answer questions. A chatbot’s dialogue can be visualized as decision trees, and each branch or sub-branch is a different topic. Chatbots utilizing natural language processing use open-ended questions and have free text user input fields. The text captured in these fields is used in the natural language processing to determine the users’ intent and to navigate the user from one branch to another with fewer steps than using predefined options alone. Natural language processing and the administrator’s training of the chatbot support it to understand the users' intent in relation to the other dialogue paths within the tree. As an emerging technology that arguably can improve customer service and reduce operational costs, chatbots have become increasingly commonplace. Nevertheless, it is not yet clear whether these are inflated expectations.

Universities today may consider chatbots of particular use to communicate effectively with this very technologically engaged generation. Universities may have an ageing technological infrastructure and slow or difficult to navigate bureaucratic processes. This user experience contrasts with the clean interfaces and instant endorphins released with products belonging to the "attention attraction economy". Funding new technology to facilitate improved user design has also become more challenging in Ontario due to recent cuts to universities estimated at $440 million (Rushowy, 2019). Support services to students must meet today's higher expectations for speed and quality of service, with less funding while being sensitive to this formative time and age for students in their psychological development. The following literature review explores research on chatbots from the perspective of each of the sensitivities within humane design.

Chatbots and Emotions Although chatbots have proven helpful in answering student inquiries (Aeiad & Meziane, 2019; Page & Gehlback, 2017) and could alleviate concerns and stress, research has also shown chatbots can provoke negative emotions (Mou & Xu, 2017). In one study comparing social interactions with chatbots and human-to-human interaction, it was determined "users tended to be more open, more agreeable, more extroverted, more conscientious and self-disclosing when interacting with humans than with AI" (Mou & Xu, 2017, p. 437). Even though chatbots may not be replacing support from a human, it could be argued simulated dialogue always enables people to compare the chatbot experience to that of interacting with a human. As such, the design of socially interactive technology has been focused on avoiding negative reactions to being imperfectly human (Afzal et al., 2019; Mathur & Rechling, 2016). Humans interacting with chatbots are more neurotic than when interacting with other humans (Mou & Xu, 2017) and improving the design of chatbots includes improving their soft skills (Afzal et al., 2019) and pierced trustworthiness (Mathur & Rechling, 2016).

In addition to potentially reducing stress through effective and timely support, the potential for anonymity may help chatbots support students emotionally. Kamalou et al.(2019) conducted a study on using chatbots for health advising, which showed people prefer the anonymity of receiving well-being counselling from a chatbot. Also, it was found that socially anxious people have been shown to prefer anonymous asynchronous social interactions (Kamalou et al., 2019). The researchers concluded that fear of negative evaluation makes people less likely to share personal information and have a greater sense of control over their self-presentation (Kamalou et al., 2019). These results could be indicative of the experiences of sensing support in other types of help-seeking.

Chatbots and Attention Some research suggests online environments can promote cursory reading (Carr, 2011) and an inability to process information to long-term memory as the scattered stimuli delivered overwhelms users (Schunk, 2020). Suppose dialogue pathways in a chatbot present fewer options than search or navigation on a website. In that case, it could be argued that chatbots are more likely to require less working memory and be more effective in supporting the human sensitivity of attention.

The segmenting principle also supports a learner's ability to process information. The segmenting principle, established in multimedia instruction, shows that learners learn best when able to fully process one step before moving on to the next (Benassi et al., 2014). From this perspective, some people may find fewer options in predetermined paths easier to fully process than reviewing more menu and search options available on a site. In line with this thinking, chatbots could prove more helpful in providing a path that focuses a learner's attention to navigate particularly complex content. On the other hand, dialogue paths may not anticipate the user's questions. Codifying complex knowledge is more difficult when more tacit knowledge is required or convergence of non-knowledge issues like emotions or politics (Guzman & Trievelato, 2011). Since knowledge must be codified into a dialogue pathway to be presented to users, chatbots would be limited to sharing simple or complex explicit knowledge such as facts, rules or determined processes. Overall, research suggests a user's sensitivity to attention may be better supported with dialogue paths to anticipate questions and provide satisfactory answers.

Chatbots and Sense-making Both search/website navigation and chatbots place the learner as the central actor in finding and creating new knowledge. Though both approaches also provide instant feedback, a chatbot is programmed to provide fewer satisfactory options. It may be better at supporting users to process information and make sense of it. On the other hand, existing research shows that chatbots could poorly anticipate a user's intentions (Celino & Calegari, 2020; Luger and Sellen, 2016) or provide confusing or out-of-context messages (Griffith et al., 2020).

Potentially in support of a chatbot's ability to provide sense-making support, chatbots can prompt users through the anticipated paths related to their query. In chatbots, the administrator that programs the chatbot provides a type of asynchronous zone of proximal development. Chatbot administrators create dialogue pathways and train the search results of any open-ended questions. For example, chatbot administrators knowing some students have previously searched for delaying an exam, unknowing of the expression "defer an exam", could program the search or chatbot search results to automatically display information about deferring an exam when someone searches for "delaying an exam". In addition to leveraging synonym functionality, chatbot administrators have an opportunity to target responses based on where a user is within the dialogue and to ask a follow-up question. These are not common features to search training on a website instead of a chatbot because there is no dialogue. The position of the user within the dialogue flow when a question is asked, and the follow-up question data could provide more thorough information for search training. Chatbot administrators could better predict a user's intent and provide a satisfactory response using this data on where someone is in dialogue when using a term. It is unclear whether this programming will be more meaningful than designing the content of support pages and whether sense-making support of a chatbot will be perceived as better or worse than web pages.

Chatbots and Decision-making Some chatbots are criticized for providing too linear a dialogue that does not allow users to deviate from that flow (Celino & Calegari, 2020). Predefined buttons support agency and can logically guide the user through the knowledge in a linear fashion. Yet, if those predefined options are not the desired path, free-text user input search fields are also important to support the user in navigating quickly to the relevant content (Celino & Calegari, 2020). Free-text user input fields are similar to the search you would find on webpage navigation and utilize natural language processing to determine the user's intent. This feature, when utilized, allows users to avoid a linear path and jump to any part of the chatbot's dialogue without passing through the predefined options to get there. Evidence suggests that natural language processing can fail to meet user expectations by not retrieving satisfactory content within the chatbot's dialogue, creating a limitation in the chatbot user experience (Luger and Sellen, 2016). This limitation could be experienced due to a) the technology used for the natural language processing, b) a lack of chatbot administrator search training, or c) there could not be a path relevant for that user and the chatbot's existing dialogue knowledge is the cause of the failed expectation.

In the case of using chatbots, the particular implementation, both the technology platform used and decisions creating the dialogue paths, have a tremendous impact in supporting the user's intentions. It could be argued, chatbots that are software-as-a-service will better support decision-making as removing the technical burden from the company will allow for more funding to be directed to non-technical chatbot administrators as product owners. These chatbot administrators could be subject matter experts, more likely to anticipate the information-seeking behaviour of the users and train the chatbot themselves to improve how they facilitate user decision-making over time.

Chatbots and Social Reasoning The conversational interaction with chatbots changes human-computer interaction and raises expectations for social characteristics through the simulation of dialogue **(**Chaves & Gerosa, 2021). Research on social interactions between humans and AI has shown that people exhibit different personality traits when interacting with a chatbot than with another human (Mou & Xu, 2017). Unlike navigating a web page's menu or using a search box, a more social relationship is implied when using a chatbot. Mori (1970), a pioneer in robotics, first coined the term uncanny valley. Now a widely adopted term, the uncanny valley describes how humans change from empathy towards robots to revulsion as robots become more lifelike (Mori, 1970). From this perspective, it could be argued that it is important for chatbots not to attempt to be perceived as humans. Abu Shawar and Atwell (2007) shared this conclusion when discussing practical chatbot applications: "In general, the aim of chatbot designers should be: to build tools that help people, facilitate their work, and their interaction with computers using natural language; but not to replace the human role totally, or imitate human conversation perfectly" (p.7). On the other hand, Chaves & Gerosa (2021) have argued that enriching chatbots with some social characteristics could be beneficial to adhere to user's expectations and avoid frustration.

Even though chatbots would struggle to replace humans and create uneasiness or repulsion when they attempt to, they may provide informational support when other sources of support are not available. Griffith et al. (2020) conducted a qualitative study on a companion chatbot designed to provide social support and concluded that the chatbot did curtail loneliness, providing a "safe space where users can discuss any topic without fear of judgement or retaliation" (p.32). The results of this study concluded that there were four themes of social support in the 1854 user reviews: informational support (n = 289, 15.6%), emotional support (n = 827, 44.6%), companionship support (n= 1429, 77.1%) and appraisal support (n= 172, 9.3%) (Griffith et al., 2020). The researchers concluded that chatbots could be a promising source of social support when another everyday social support is not readily available. In this study, there was also an unanticipated theme of negative experiences in 100/1854 or 5.4% of the user reviews (Griffith et al. 2020). These negative experiences were attributed to two themes. Firstly, the uncanny valley, that the companion chatbot was weird or creepy (Griffith et al. 2020). Secondly, some participants reported receiving out-of-place messages like repetitive or confusing messages (Griffith et al. 2020). From these findings, it could be argued chatbots will likely support the social reasoning sensitivity if chatbots do not pretend to be human.

Chatbots and Group Dynamics Understanding the importance of social identities within the context of a chatbot may be difficult to determine. Microsoft's chatbot "Tay" is a clear example of how not putting safeguards to protect social identities can have a negative impact (Banks, 2019). Microsoft's "Tay" was meant to learn from interactions with Twitter users and, within 16 hours, needed to be shut down because those interactions taught the chatbot through machine learning to spread hate speech (Banks, 2019). With increasing advances in human-computer interactions, there is a growing need for human communication and technology scholars to explore social interactions through an apparent moral agency (Banks, 2019).

Anticipating a user's intention is made more difficult by the variety of different users and by the potential lack of diversity amongst the chatbot administrators. How various users may approach trying to answer a question has been the study of user experience designers. Indi Young (2018) is a problem space researcher who developed the mental model diagramming technique. Mental model diagramming supports user experience designers in depicting how a person may perceive the surrounding world, the relationships between its various parts and a person's intuitive perception of their own acts and their consequences (Indi Young, 2018). Young (2020) has described this challenge as a need for "a parallel way to look at things, to identify and erase your assumptions and to support the way a person thinks (or how different people think differently) rather than making them bend to the approach your system represents". Though there could be considerations of oppressed groups within chatbot dialogue paths, a linear dialogue flow that focuses on most commonly asked questions could inadequately support this sensitivity. As such, it could be argued developing content to respond to the needs of oppressed groups and offering search within a chatbot or within web pages could be essential for anticipating questions sensitive to group dynamics.

Hypothesis

Many factors could influence how a particular user experiences seeking information on a chatbot. This study predicts that a particular chatbot designed to help students within a Canadian university's Faculty of Science will provide better support than web pages designed for the same purpose. Each of the following hypotheses measures different ways the experience of using a chatbot to seek information could be compared with the experience of using web pages. This study explores the relationship between the technology used and each of these experience measurements, indicating a better experience from a humane design perspective.

H1 Satisfactory responses to student inquiries will be retrieved faster when navigating to find content using the Science Academic Advising chatbot than navigating using the Science Academic Advising webpages.

H2 Participants seeking information using the Science Academic Advising chatbot will report that they were more confident with the responses they found than students seeking information on the Science Academic Advising webpages.

H3 Participants will report that their human sensitivities identified in the Humane Design Guide were better supported using the Science Academic Advising chatbot than using the Science Academic Advising webpages.

H4 Participants will report they prefer using the Science Academic Advising chatbot than using the Science Academic Advising webpages.

Method

Participants

The study consisted of 48 undergraduate students in the Faculty of Science at a small-to-midsize university in southern Ontario who stated they would consider using a chatbot. Participants were a purposeful sample of undergraduate students in the Faculty of Science because the chatbot used in this study has a knowledge base designed to support this audience. The decision to target participants who would consider using a chatbot was to identify users representative of those who would consider using an optional chatbot on a website. Targeting this group of students thereby allows us to understand the differences between the two experiences, searching on the support site with or without a chatbot, from this subset of users who would potentially choose a chatbot if one was available.

Materials and Procedures

Data collected compared the website and chatbot search experience by measuring the effectiveness of the search and five of the six sensitivities that evaluate human-computer interaction from a humane design perspective. [1] Participants were recruited from an online posting shared with undergraduate science students in their online Learning Management System. Participants were randomly assigned to two groups after consenting to participate and were provided with a link to the participant instructions. Participants were then presented with four hypothetical search scenarios. They had to seek the answer to four questions related to academic advising, either via the university's Science Academic Advising webpages or via the university's Science Academic Advising chatbot. 28 participants were attributed to the first group. This group was directed to search for information related to the first two scenarios on the web pages and then search for information for the other scenarios on the chatbots. The second group had 20 participants who were asked to search using the same four scenarios but were directed to search via the chatbot first and the webpages second. Once participants had completed each quest, they were asked to complete a Scenario Questionnaire related to their experience on the webpages or with the chatbot. After all four of the Scenario Questionnaires

[1] The sixth human sensitivity not measured is the sensitivity of group dynamics. This sensitivity was excluded because the research design of only presenting the most common search scenarios inherently ignores the specific information seeking needs of minority groups. Also, there were not enough participants to measure minority group views even in these most common search scenarios.

were completed, the participant was asked to complete the Post Scenarios Questionnaire. Detailed information regarding the chatbot, the relevant web pages, the scenarios, the Scenario Questionnaire and the Post Scenario Questionnaire are presented next.

Science Academic Advising chatbot. Participants used a chatbot built on the Instabot.io platform by the Faculty of Science Academic Advising team at a comprehensive Ontario post-secondary institution. The Science Academic Advising team built all the chatbot dialogue paths to answer frequently asked questions and triage inquiries to human academic advisors where required. The chatbot's dialogue paths are almost entirely navigated by users clicking on options provided through various menus. The chatbot asks only one question that allows users to input text to search. This text input option is available if a user selects "Ask a Question about Something Else". If the participant selects this option, the text entry is then sent to Google's Dialogue flow to leverage both natural language processing capability and programming from the chatbot administrators, attempting to determine the user's intent and provide a response.

Science Academic Advising webpages. Participants were provided with a link to the webpage designed as a landing page for content on the Faculty of Science's Academic Advising. The content of the web pages was also created by the Faculty of Science Academic Administration team. This webpage was built on the Cascade Content Management System, and some of the content layout and features were designed by the university's central Web Services. The search bar on these web pages leverages Google's Custom Search Engine, allowing the university to surface content from its own web properties using Google's natural language processing.

Search Scenarios. Each search scenario was presented to the user with instructions to take no more than five minutes to obtain the answer to the question and to seek the answer to the question either via the Science Academic Advising webpages or Science Academic Advising chatbot. There were four search scenarios total, each of which was displayed to participants in question format to convey an intent they needed to search. The questions were determined by reviewing the 500 most common search terms on the support website and categorizing them into groupings of whether they related to academic advising. The search terms related to academic advising were then grouped into themes broken down to the level of what could relate to a frequently asked question. The four most frequent and distinct themes were selected to be included in the study. Each theme was then turned into a question format. This formulation of the question in its entirety was done to allow participants to consider their search terms to express their intent. Also, each scenario shared enough information for participants to locate an answer to that question. For example, since finding the correct academic advisor involves knowing the first letter of the student's last name, a fictional student’s name was provided in the scenario for finding an academic advisor. The Scenarios are included in Appendix A. After reviewing each scenario and searching for the information, the user was asked, “Were you able to find the information within five minutes” and the user selected “Yes” or “No” to proceed. There was also a timer embedded on each of these scenario web pages that participants were unaware of, which automatically tracked the time elapsed before proceeding from the scenario web page.

Scenario Questionnaire. Once participants were satisfied that they found the answer or the five-minute time limit elapsed, they were asked to complete a Scenario Questionnaire. If the participant indicated they found the information they sought, they were asked to provide this information. All participants were then asked to indicate their level of confidence that they found the correct answer. This Scenario Questionnaire also consisted of 5 questions on human sensitivities identified within the Humane Design Guide. For example, to measure the participant's sensitivity labelled attention, the user was asked on a scale of 1 to 5 to indicate "To what extent did you feel your attention was focused or pulled in different directions because of what information was presented or how the information was presented?". Prompts on the screen described the scale as "My attention was pulled in many directions" to "My attention was focused". The questions were designed in this format to allow the participants to have a similar understanding of each point on the five-point scale represented for that human sensitivity. The Scenario Questionnaire is included in appendix B.

Post Scenarios Questionnaire. This questionnaire collected data at the end of the study after all of the search scenarios were completed. The questionnaire asked participants to select the device they used (mobile, computer or tablet), age, and gender identity. Using an open-ended format, the Post Scenarios Questionnaire asked participants what they liked or disliked about searching for the required information on each tool and what option they would be more likely to choose in the future. A thematic analysis was completed on the questions of tool likes and dislikes. Each response had codes applied, and then these codes were considered within meaningful groupings as sub-themes and finally consolidated into themes.

Results

Descriptive Statistics

Table 1 shows descriptive statistics related to the participants. There were a total of 48 participants. 45 or 93.8% of participants used a computer to participate in the study, and 41 or 85.4% identified with being Female. Participant ages ranged from 18 - 29, $M = 20.77$, $SD =$

1.93, Min = 18, Max 29. Table 2 shows the duration or total time elapsed participating in the study and participating in each scenario. The validity of these measures in capturing the time elapsed is questionable. Participants were not notified that they were being timed or required to complete the activities without interruption. Tables 3 and 4 capture the confidence and human sensitivity scores of participants. Table 5 shows how many participants indicated that they found the answer within 5 minutes on each scenario.

Table 1

Participant descriptive statistics

Variable	Values	Frequency	Percent
Random group assignment for tool order	Group 1 - Website then Chatbot	28	58.3%
	Group 2 - Chatbot then Website	20	41.7%
Device used	Mobile Device	3	6.3%
	Computer	45	93.8%
Gender Identity	Female	41	85.4%
	Male	6	12.5%
	Custom	1	2.1%
Preferred tool in future	Chatbot	25	52.1%
	Website	13	27.1%
	No preference	8	16.7%

Table 2

Duration descriptive statistics

	Total Duration (sec)	Duration by Scenario (sec)			
		1. AddDrop	2. Advisor	3. Data	4. Deferral
Mean	916.53	126.26	52.34	123.63	81.31
Median	797.5	119.83	52.33	112.92	66
Std Dev	661.94	64.62	29.27	81.89	68.3
Minimum	197.34	20.115	4.006	7.037	5.71
Maximum	4602	316.777	114.759	317.76	301.883

Table 3

Scenario 1 & 2 - Confidence and Sensitivity Scores

	Scenario 1 - Add Drop (Likert Scale)						Scenario 2 - Advisor (Likert Scale)					
	Confidence	Emotion	Attention	Sense	Decision	Social	Confidence	Emotion	Attention	Sense	Decision	Social
Mean	2.85	2.71	2.31	3.06	3.31	2.81	4.42	4.06	3.94	4.38	4.23	4
Mode	2	2	2	4	2	2	5	4	4	4	5	*1
Std Dev	1.35	1.11	1.07	1.12	1.17	0.96	0.85	0.81	0.95	0.61	0.83	0.83
Minimum	1	1	1	1	1	1	2	2	2	3	3	3
Maximum	5	5	5	5	5	5	5	5	5	5	5	5

Note:
*1 Equal frequencies of 3,4,5

Table 4

Scenario 3 & 4 - Confidence and Sensitivity Scores

	Scenario 3 - Data (Likert Scale)						Scenario 4 - Deferral (Likert Scale)					
	Confidence	Emotion	Attention	Sense	Decision	Social	Confidence	Emotion	Attention	Sense	Decision	Social
Mean	3.13	2.79	2.88	3.02	3.19	3	4.06	3.46	3.71	3.79	3.85	3.48
Mode	2	*2	2	4	4	3	5	4	4	4	4	3
Std Dev	1.45	0.97	1.25	1.25	1.23	0.83	1.08	1.17	1.11	1.11	1.05	0.97
Minimum	1	1	1	1	1	1	1	1	1	1	2	1
Maximum	5	5	5	5	5	5	5	5	5	5	5	5

Note:
*2 Equal frequencies of 2,3

Table 5

Participant states found the answer in 5 minutes by scenario

	1. Add Drop		2. Advisor		3. Data		4. Deferral	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Yes	25	52.1%	47	97.9%	29	60.4%	44	91.7%
No	23	47.9%	1	2.1%	19	39.6%	4	8.3%

H1

The null hypothesis was accepted, the chatbot was not faster at retrieving satisfactory responses. A subset of participants who stated they found the information in under five minutes was selected to test the null hypothesis. An independent samples t-test was used with these cases to test the statistical significance of using the website or the chatbot on the mean differences in elapsed time for each scenario. The results showed that the mean time elapsed for participants using the website was not statistically different from the mean elapsed time of participants using the chatbot. All four search scenarios had p values that were higher than .05 (Scenario 1 p = .888, Scenario 2 p = .328, Scenario 3 p = .559, Scenario 4 p = .508). There was a pattern of a higher mean time associated with chatbot use on all four of the scenarios, however, this relationship was not statistically significant.

Table 6

Duration by Scenario for Participants who Found in 5 minutes

Scenario	Tool Used	Found in 5	Time Elapsed in Seconds					
			Mean	Median	Min	Max	SD	Sig.*
1- Add / Drop	Website	12/28 (42%)	117.66	103.17	48.06	316.78	73.67	.888
	Chatbot	13/20 (65%)	121.41	104.83	43.356	220.291	57.80	
2- Advisor	Website	27/28 (96%)	48.58	42.93	4.01	114.76	30.80	.328
	Chatbot	20/20 (100%)	57.22	57.30	9.22	106.28	27.85	
3- Data	Website	9/20 (45%)	107.12	108.79	27.27	179.36	48.54	.559
	Chatbot	20/28 (71%)	125.16	117.11	13.17	288.62	85.00	
4- Deferral	Website	19/20 (95%)	76.13	50.61	5.71	301.88	76.53	.508
	Chatbot	25/28 (89%)	89.75	80.75	8.22	267.28	58.86	

*2-tailed

H2

The hypothesis was partially supported. An independent samples t-test showed that the mean confidence in finding an answer for participants using the website was not statistically different from the mean confidence of participants using the chatbot in three of the four scenarios (scenario 1 $p = .145$, scenario 3 $p = .662$, scenario 4 $p = .757$). In scenario 2, the independent samples t-test showed that the mean confidence of the participants using the chatbot ($M = 4.80$, $SD = .47$) was statistically different than the mean confidence of the users who used the website for the same scenario ($M = 4.14$, $SD = .47$, $p = .001$). Only in scenario 2 was the hypothesis supported; in two of the three other scenarios, there was greater mean confidence, but it was not statistically significant. In scenario 4, the mean confidence of participants using the website was higher than participants using the chatbot.

Table 7

Mean confidence rating by tool all scenarios

Scenario	Tool Used	Number of Participants	Confidence Rating on Likert Scale					
			Mean	Mode	Min	Max	SD	Sig. *
1- Add / Drop	Website	28	2.61	2	1	5	1.23	.145
	Chabot	20	3.20	. (2,4,5)	1	5	1.47	
2- Advisor	Website	28	4.14	5	2	5	.97	.001
	Chatbot	20	4.80	5	4	5	.47	

3- Data	Website	20	2.75	2	1	5	1.48	.662
	Chatbot	28	3.39	5	1	5	1.40	
4-Deferral	Website	20	4.15	5	2	5	.99	.757
	Chatbot	28	4.00	.(4,5)	1	5	1.15	

*2-tailed

H3

The hypothesis that chatbots better support humane design sensitivities was partially supported in this case. An independent samples t-test was completed on each sensitivity rating to test the null hypothesis that the tool did not impact the perceived support of these sensitivities. The results showed that there was one or more sensitivity in three of the four scenarios where the mean sensitivity rating using the website was statistically different from the mean sensitivity of participants using the chatbot In scenario 2, the independent samples t-test showed that the chatbot mean emotion rating (M = 4.35, SD= 0.88) attention rating (M= 4.40, SD = .88) sense rating (M= 4.65, SD= .88) and social rating (M= 4.55, SD=.6) were statistically different than the mean sensitivity rankings of the users who used the website for the same scenario: emotion rating (M = 3.86, SD= 0.71, p= .036), attention rating (M= 3.61, SD = 0.88, p= .003), sense rating (M= 4.18, SD= .55,p= .006) and social rating (M= 3.61, SD=.74, p= 0) . The results on Scenario 1 also showed the social rating of chatbot using participants (M= 3.15, SD = 1.09) was statistically significant from the social rating of participants using the website for the same scenario (M = 2.57, SD = .79, p= .038). The results on scenario 3 showed that the mean emotion ranking of chatbot users (M = 3.07, SD= 1.02) was statistically different from the emotion ranking of the website users (M= 2.4, SD = .75, p= .016). In these cases, the hypothesis was supported. There was also a trend of higher ranking of support of sensitivities by participants using the chatbot that was not statistically significant.

Table 8

Sensitivity Score all scenarios by tool

Scenario	Sensitivity	Tool Used	# of Partici-pants	Sensitivity Rating on Likert Scale					
				Mean	Mode	Min	Max	SD	Sig*
1- Add / Drop	Emotion	Website	28	2.54	2	1	5	0.92	.206
		Chatbot	20	2.95	. (2,3)	1	5	1.32	
	Attention	Website	28	2.11	2	1	4	0.83	.118
		Chatbot	20	2.6	2	1	5	1.31	
	Sense	Website	28	2.86	2	1	5	1.08	.134
		Chatbot	20	3.35	4	1	5	1.14	
	Decision	Website	28	3.14	2	1	5	1.18	.239
		Chatbot	20	3.55	4	2	5	1.15	
	Social	Website	28	2.57	2	1	4	0.79	.038
		Chatbot	20	3.15	.(3,4)	1	5	1.09	
2-	Emotion	Website	28	3.86	4	2	5	0.71	.036

Advisor		Chatbot	20	4.35	5	2	5	0.88	
	Attention	Website	28	3.61	4	2	5	0.88	.003
		Chatbot	20	4.40	5	2	5	0.88	
	Sense	Website	28	4.18	4	3	5	0.55	.006
		Chatbot	20	4.65	5	3	5	0.59	
	Decision	Website	28	4.04	5	3	5	0.84	.055
		Chatbot	20	4.5	5	3	5	0.76	
	Social	Website	28	3.61	3	3	5	0.74	0
		Chatbot	20	4.55	5	3	5	0.6	
3-Data	Emotion	Website	20	2.4	2	1	4	0.75	0.016
		Chatbot	28	3.07	3	1	5	1.02	
	Attention	Website	20	2.6	2	1	5	1.19	0.2
		Chatbot	28	3.07	.(2,4)	1	5	1.27	
	Sense	Website	20	3	4	1	5	1.26	0.923
		Chatbot	28	3.04	2	1	5	1.26	
	Decision	Website	20	2.85	4	1	5	1.23	.109
		Chatbot	28	3.43	.(3,4)	1	5	1.2	
	Social	Website	20	2.85	.(2,3)	1	5	0.99	.292
		Chatbot	28	3.11	3	2	5	0.69	

4-Deferral	Emotion	Website	20	3.45	4	1	5	1.36	.967
		Chatbot	28	3.46	4	2	5	1.04	
	Attention	Website	20	3.65	.(4,5)	1	5	1.23	.762
		Chatbot	28	3.75	4	1	5	1.04	
	Sense	Website	20	3.9	4	2	5	1.07	.573
		Chatbot	28	3.71	4	1	5	1.15	
	Decision	Website	20	3.55	4	2	5	1.19	.091
		Chatbot	28	4.07	4	2	5	0.9	
	Social	Website	20	3.45	3	2	5	1	.862
		Chatbot	28	3.5	3	1	5	0.96	

*2-tailed

H4

The hypothesis was supported; the results of the question on tool preference showed the chatbot was preferred: chatbot preferred (n = 25, 54%) webpage (n=13, 28%), or no preference (n=8, 17%). A Chi-Square test was used to determine there was not a statistically significant difference between the expected frequencies of tool preference and the order of the tools used in the study. The results showed no significant difference ($p = .419$) between the tool preference frequencies if the participant used the website or the chatbot first to perform their search scenarios.

Table 9

Tool preference by order of tools used

	Order of Tools Used		
Preference	Website First	Chatbot First	Total
Chatbot	15	10	25 (54%)
Website	6	7	13 (28%)
No Preference	6	2	8 (17%)
Total	27*	19	46

*There were two null values excluded from the total. Pearson Chi-Square Asymptotic Sig. (2-tailed) p-value was .419.

Qualitative Analysis of Tool Likes and Dislikes

Six themes emerged from the qualitative analysis of the responses to the two open-ended questions asking participants to share likes and dislikes from each tool. 42 of the participants, or 87.5 %, provided these qualitative responses.

Tool Workflow

The first two themes that emerged related to how participants described the experience of the workflow of each tool. A workflow is more specific than a process as its definition includes more than the desired output; it also includes the steps specific to the technology used within a process. This theme includes descriptions of how the tool handles the sequence of steps within the workflow and navigation through those steps using features of the tool. Descriptions around ease or difficulty to navigate those steps were included within these themes because they were core to exploring the tool's ability to enable users within the search process.

The chatbot offered a predetermined workflow 29 participants, or 78% of the 38 participants who commented on the chatbot's workflow, had positive comments. One participant summed the workflow nicely, stating the chatbot helped "*narrow down exactly what it was I was searching for, instead of just aimlessly looking.*" Similar positive comments described the prompts or paths to help retrieve information. Another participant shared a similar description of a narrowing workflow and compared the nature of the chatbot to a "*flowchart*" which "*naturally progressed from the previously selected answers until the search was narrowed down to the needed answers*". One participant referred to the chatbot as structured, and another described it as categorizing, "*The chatbot is convenient since it categorizing all the information into subheadings or topics without the need for tunnelling through several pages of information*".

16 or 42% of the participants who commented on the chatbot workflow shared negative comments. Of those 16 participants, nine or 56% also shared positive comments. Six participants commented on challenges over picking the correct option to proceed. Several of those participants found that there were too many options or text and one commented that there were not enough options. Four negative comments were about the chatbot's poor search or conversational skills. These comments were identified as being related to the workflow because it was the search feature that made it difficult for users to diverge from the structured dialogue path. Though there was a search feature, it only appeared in the initial introduction, and one participant did not see it, commenting that there was no feature to type a query. Another participant wanted to be able to speak to the chatbot to make the query. There were no definitive comments that a participant had a positive experience using the search feature. One participant

did comment that they liked having the option to type in a query. Another participant specifically shared after typing in a query, and the chatbot had no idea of a response.

The final trend within this theme of a structured workflow was that the Chabot supported users returning easily to a previous step of the workflow. Four participants described being satisfied with this option to return to a previous step. One participant summarized this feature well stating, *"I reeeeeally liked how when I clicked on a result it selected the radio button, but if I scroll up past the bots most recent response I can still click on old responses and alternate answers and it just adds them to the chatlog as well. Makes the impact of errors for users much smaller".*

The website required participants to determine the workflow 27 or 77 % of the 35 participants who provided comments on the website's workflow included negative comments. One participant summarized the website workflow by responding, *"I disliked using the website to find the required information as some of it was either very clearly laid out or it was extremely hard to find with lots of pages bringing you back to previous ones you had already visited. It felt as though it was like going around in circles when looking for some information."* There were nine participants who negatively referenced the workflows as confusing or difficult referencing not being sure what to do next. One participant summarized using the website as "*more or less trial and error*". Another participant stated, *"It's always guesswork, It's difficult to navigate most of the time, and if you accidentally click something you're totally lost on the website."* The unstructured workflow of the website was contrasted with the chatbot by one participant by stating the website *"Can be confusing at times, and difficult to navigate as there are so many pages and a lot of information, whereas the bot summarizes what you need to know or sends you the right page to go to."*

There were also 15 participants or 42% of those who commented on the website workflow providing positive comments. Of those 15, nine or 60% also shared negative comments. Positive comments included the website's nature to let users determine the steps to take. One participant explained using the website that there were "*Not as many distractions. More focused on the goal. Allows students to understand the website on their own*". A website feature called tabs was referenced both positively and negatively, with users appreciating their ability to declutter information and requiring users to click to open each one to reveal the tab's contents.

Sense of Support by Tool

The following two themes refer to how users shared their experiences using observations of how they sensed being supported or inhibited using the tool. The themes included coded judgements on the experience, whether they were clearly subjective, for example, confusing or perhaps less subjective but still depended on the user's senses, for example, judging the accuracy or speed of the tool.

The chatbot seemed convenient but confusing in some contexts 33 participants or 82.5% of the 40 participants who shared qualitative data on how they sensed the experience chatbot shared positive comments. Many of those participants, 17 or 51%, also shared negative comments. Many of the positive participants commented on the tool being easy or helpful. Ten of the participants commented on sensing the was fast or saving time. One participant summarized the experience as "*It is very fast, and efficient and the results are very accurate and tailored to my needs.*" Another participant attributed the speed to the chatbot's prompts, *"I liked* *how it provides you with quick prompts of what issues I may need addressed, which saves time in trying to navigate the website myself."*

In terms of negative senses of the experience, 23 participants or 57.5% of the total participants who shared qualitative data had negative comments. Of those negative comments, the most comment trend was around sense-making. Six participants provided a comment on the chatbot providing too much information. Eight participants commented there were confusing or unclear prompts. Some comments suggested prompts did not allow the participant to ascertain the information shared if that prompt was selected. One participant described "*it was helpful but at times the communication was not as smooth, in terms of certain goals and the information that was given as a result".* Seven of the participants who provided negative feedback made comments that compared the chatbot negatively in some way to a human. Two participants specified they would prefer speaking to a human. One participant shared, "*I would much rather communicate with someone over email versus the chatbot, even if it takes longer.*" Two referred to the chatbot as automated and not listening or being personalized. One participant felt the chatbot was confused, and another felt unable to speak to it.

The website felt cumbersome 33 participants shared qualitative data on how they sensed the website experience. 25 or 75.7% shared negative comments. 14 or 42.4% shared positive comments. Some users described a greater sense of struggle than others. One participant expressed displeasure stating that the "*Website is a trap. Big clear buttons for big issues, however when you click on them, you are bombarded with loads of useless information to your scenario. I found nothing pleasant about the way the website is organized."* Yet another participant shared, "*I find using the website to find information easy for the most part, but you have to dig around more, which can take longer than expected."* Five of the negative comments mentioned requiring additional time to find information when using the website. Eight participants commented on needing to sift through irrelevant information to find what they were looking for. Eight participants discussed a lack of clarity in the answers they found on the site.

Only one participant referenced a human in their feedback. Unlike the chatbot references, this reference seemed to refer more to the workflow of how to ask for help than a type of social reasoning comparison. Still, it was included in this section to contrast how additional help was sensed for the website compared to the chatbot. The participant stated the website was *"A bit complex and not easy to find things, which just leads you to emailing facilities for help and answers which can sometimes take a long time."*

Tool preference and experience depend on the scenario complexity and learner - This final theme referred to how users perceived simpler scenarios as easier than more difficult ones. Also, this theme includes comments that the chatbot, in particular, was better for simple inquiries and new students, and the website was better for complex inquiries and more experienced students. At times, the complexity here was defined using a scenario as an example, and explaining the best tool was dependent on the scenario. It was also defined by connecting the user's mastery with the content to which tool was preferred. A first-year student being less knowledgeable would make simpler inquiries, and so the chatbot is better. A student who has learned the website would prefer to have all the information available.

13 or 30.9% of participants who provided qualitative data provided data on this theme. 11 or 84.6 of those participants commented that the chatbot was better in simple search scenarios than in complex ones. One participant shared, "*I also felt like the Chatbot only provided the most basic information; which may be useful in simple inquiries, but if you need a lot of information*

on a problem then the chatbot is not a good tool". On the other hand, with the website, similar comments were only shared by three participants who thought the website was better in simple scenarios than complex ones. Six participants described a learning curve related to the website requiring time to familiarize themselves to find information. One participant summarized the website's learning curve by stating, "*When I was a newer student it definitely took a while to make sense of the website and to understand where to find information that you needed."* One of these participants who discussed the learning curve on the website shared that they would recommend the chatbot "*to new freshman and future students*". Finally, four participants discussed concepts related to the full information being available on the site. One participant described how the chatbot linked to the website, another described greater confidence seeing all the related information. One participant described that although there is additional clicking involved, "*I do like that there is lots of information on the website and you can almost bet your answer to the problem is somewhere on the website.*"

Discussion

These results partially support the hypothesis that a particular chatbot designed to help students within a Canadian university's Faculty of Science provided better support than web pages designed for the same purpose. The results did not show the chatbot outperformed the web pages in all experience measures. However, the results for each hypothesis are discussed further below. The significance of this research is limited because the design of the chatbot and the website would likely be critical to influencing a student's perception of support. The variability in responses for each of the four scenarios also suggests that a user's sense of support may be largely dependent on whoever builds the online content anticipating and surfacing options that support the user's information-seeking. Although these are significant limitations for applying the results of this research broadly, this study did explore and contributed academic research from a humane design perspective. This research could inform future research on measuring whether a user's best interests are supported by providing an educational psychology lens to the human sensitivities identified within humane design. The study also situated existing chatbot research literature in the context of human sensitivities. Finally, this work provided a research design that could inform future research to measure user experience from a humane design perspective.

H1

Although there were trends in the qualitative data that participants perceived the chatbot to be saving time and the website required additional time, the null hypothesis was accepted because the quantitative data showed using the chatbot did take longer. Though there was a limitation in the validity of the duration data because participants may have taken a break from participating, these results suggest there is only a potential perception of saving time using the chatbot. Participants, in reality, appear to have spent more time using the chatbot than the website on all scenarios. This perception of speed could be an area for further research to understand if these results would be similar on a larger scale with closer monitoring on timing.

The significance of the results for H1 is also limited because the duration data was analyzed for participants who believed they found the answer during the five minutes. The significance of these results would have been greater if participants had to demonstrate they found a correct answer. A limitation to the research design was the inability to determine whether participants could provide the correct response to the inquiry. If a participant indicated they found the answer to the scenario, there was a question in the post scenario questionnaire that asked participants to provide that answer. During the analysis of these responses, determining the correctness of the responses required too much subjectivity. Some answers provided may have been less correct but still were potential avenues that could be acceptable. For example, emailing an advisor could be considered an acceptable answer for any scenario. It is a way to initiate dropping a course, requesting a deferral or making an advisor appointment. Also, some cases were where the participant might have found the correct answer but only provided part of the information to determine its correctness. For example, “appeal” could have been provided as a response to deferral, but which department the appeal needed to be directed to could have been unknown. A future direction would be to interview participants after each search scenario to objectively ascertain whether the participant knew pre-determined attributes of a correct response. Checking the answers would provide greater validity to which tool takes less time to find a satisfactory answer as participants who did not find the correct response could be excluded.

H2

The significance of these results on mean confidence rating was decreased because, in only one scenario, the simplest scenario had a statistically different level of confidence depending on the tool used. Scenario 2 had 47 of the 48 participants, or 97.9% indicate that they found the answer within 5 minutes. This scenario had the shortest time elapsed to find the answer (M= 52.35, Median = 52.33, SD 29.27) and the highest level of confidence overall (M=4.42, Mode= 5, SD = .85). These results suggest chatbots may potentially provide greater confidence with very simple questions. The qualitative results similarly suggest participants believed the chatbot was better with simpler scenarios. Interestingly, the chatbot also did have a higher mean confidence level than the website on both of the two most seemingly difficult scenarios. The two, which had the lowest percentages of participants, indicate they found the answer; Scenario 1 had only 25 or 52.1% of participants indicate they found the response, and scenario 3 had 29 or 60.4% of participants indicate they found the response. Though both scenarios 1 and 3 also had higher confidence levels for the chatbot than for the website, it was not statistically significant. The chatbot's higher mean confidence score in both these scenarios was in line with the results on whether they believed they found the response when analyzed by tool. A higher percentage of participants found the answer using the chatbot in scenario 1 (n= 13, 65%) and scenario 3 (n=20, 71%) than participants using the website on scenario 1 (n=12, 42%) and scenario 3 (n=9, 45%). A future direction for research may be to examine why people are less confident in complex scenarios using a chatbot. Existing research suggests that users may become frustrated with too linear a dialogue path (Celino & Calegari, 2020). However, these results suggest that in more complex scenarios where users may need to navigate more than one linear path, they may become less confident even if they find the answer.

The qualitative data provided some insight into the perception that chatbots should be used for simple scenarios because several participants specified it was actually scenario 4 that they struggled using the chatbot. Scenario 4 had 91.7% of all participants find the answer, 89% of participants using the chatbot and 95% of participants using the website. Again, a future direction for research may be to consider whether participants did find the correct response. In this case, though the chatbot administrator programmed complex dialogue paths that helped participants navigate the deferral process, programming those rules into the dialogue paths made the process seem more complex. Perhaps, deferral information displayed on the web pages as informational text may not be conveying the complexity of the decision tree in this scenario or the web pages may be conveying the rules in a way that is easier to understand. In the end, it is not clear in this case whether the participants understood the complex rules for deferrals because the answers they found were not evaluated for their correctness. A future direction would be to determine whether participants are more successful in finding the correct answer in using a chatbot or web pages in such cases of supporting complex decision-making. There has been research on chatbots successfully reaching out and supporting students with navigating financial aid processes (Nurshatayeva et al., 2021). Still, it remains unclear how chatbots fare compared to web pages for successfully reactively supporting students in similar scenarios.

H3

The results partially supported the hypothesis because all scenarios showed statistically significant differences in one or more sensitivities except for scenario 4, the deferral scenario. There was also a trend of a greater sense of support for sensitivities across the three scenarios, which was not significant enough to register a statistically different from using the website. A future direction for research may be to understand better the relationship between sensitivities from a humane design perspective. In analyzing the data for this study, significant correlations were found between sensitivities. The correlation was not perfect, meaning participants did register the sensitivities differently, but there was a strong correlation between how a participant perceived their support of all sensitivities.

It could be argued that strongly supporting or inhibiting one sensitivity will positively affect the other human sensitivities. For example, an extremely low attention sensitivity score and a low measurement in other sensitivities could indicate that the user’s attention not being supported negatively impacts how they experience the other sensitivities. In line with this thinking, perhaps examining where a sensitivity score registers as most significantly different from one tool to the next will support understanding where improvements need to be made. For example, in scenario 2, social reasoning and attention had the lowest p values (social $p = 0$, attention $p = .003$). Since this is a simple question about contacting a human advisor, it could be argued that the visual interface and relationship to the university were more affected than in other scenarios.

Similarly in Scenario 1, the social reasoning category had the lowest p-value ($p = .038$) showing that the greatest difference in the chatbot over the web pages experience was the participants' ranking of how the interaction would impact their relationship with the participant's university. The most common response among participants using the web pages was that their relationship with the university would be 'what worsened' after this scenario (M= 2.57, Mode = 2, Min=1, Max=4, SD= 0.79). There were equal frequencies of relationships being 'the same' and 'somewhat improved' among chatbot using participants (M = 3.15, Mode= .(3,4), Min=1, Max=5). These results suggest that potentially the type of social support demonstrated in companion chatbots (Griffith et al., 2020) could be applied to using chatbots in some customer service contexts. The social reasoning score was higher for the chatbot in all four scenarios though it was only statistically significant in two of them. This data could be evidence that chatbots could help improve a sense of social support within a university setting. It must be noted that there did appear also to be heightened social expectations as suggested in other research **(**Chaves & Gerosa, 2021) 7 or 17% of participants who provided qualitative data included negative feedback about the chatbot that somehow compared the chatbot to human support.

Scenario 3, related to finding program requirements, also had a statistically significant difference in supporting sensitivities. The difference between the emotion sensitivity on this question was statistically significant with a p-value of .016. Participants using the website indicated most often that they were ‘somewhat stressed’ (M=2.4, Mode=2, Min=1, Max=4, SD=.75) in contrast with participants using the chatbot who indicated most often they were “neither calm nor stressed” (M=3.07, Mode=3, Min=1, Max-5, SD=1.02). Perhaps, the unequal stress in this scenario is related to finding the answer using the website. This question had the biggest difference in terms of participants indicating they believed they found the answer, with 20 participants or 71% of participants using the chatbot finding the answer and only 9 or 45% of the participants using the website stating they found the answer. There were some mentions of this scenario within the qualitative data. One participant commented on the chatbot dialogue stating that the prompt to search for an academic program was unclear. The text led the participant to believe if they selected a prompt, they would be sent directly to the academic calendar where all such information exists instead of specifically saying that they would help the participant find a particular program’s requirements. The participant selected the option anyways, and then the chatbot prompted her to find the correct year and program. However, the lack of clarity around the language in this scenario could account for why the experience was not better received in other areas. Three other participants provided similar negative comments on the chatbot prompts in this same scenario.

H4

The hypothesis was supported that participants would prefer the chatbot for support over the website. There was a limitation to the study that each participant only completed half of the scenarios on each tool, requiring participants to judge only two of the four scenarios per tool. The order of the scenarios stayed the same regardless of which tool participants were using, and the first half of the scenarios had one easy question and one harder question and the second half also had one easier and one harder question. Though this research design did offer two different experiences in terms of complexity on each tool, it is possible that tool preference would change with more experiences. A further direction for research would be to examine whether chatbots may be better suited for newer students to help initially navigate the website and whether the website may be better for students in more advanced years because they would have gained more knowledge of the site through this type of support. The qualitative research data did explore this trend as 13 participants (31%) thought the chatbot was better for simple questions, 6 (14%) believed there was a learning curve related to the website and 4 (10%) referenced they preferred the site to get the full information. Additional quantitative analysis was done to determine whether there was a higher frequency of students by age within the group who identified that they preferred the chatbot, however, the difference between the frequencies was not significant ($p = .587$). The validity of this analysis is limited in its ability to test whether first-year students would prefer chatbots. Age does not necessarily signify the year of the learner, with birth dates at various times over the year and some students starting their studies at different times. Future research designs to examine chatbot support should include the year of study of the student. Adding a year of study question would be in line with research that has shown that chatbots successfully support students with enrolment activities entering into their first year of university (Nurshatayeva et al., 2021).

Conclusion

As institutions increase their investments to better student experiences, we should seek greater understanding of what support we can anticipate is required during high volume periods during which it may not be possible for students to receive timely human support. This study opens up possibilities to engage in discussions of how to examine and explore that user experience along human sensitivities. However, such research must also bear in mind the responsibility of institutions to support what is within students' best interests, in line with this perspective of Humane Design.

References

Abrami, P. C., & Bernard, R. M. (2006). Research on Distance Education: In defense of field experiments. *Distance Education*, *27*(1), 5–26

Aeiad, E. & Meziane F. (2019). An adaptable and personalised E-learning system applied to computer science programmes design. *Education and Information Technologies, 24*, 1485–1509.

Afzal, S et al. (2019). The Personality of AI Systems in Education: Experiences with the Watson Tutor, a one-on-one virtual tutoring system. *Childhood Education, 95*(1), 44-52. https://doi.org/10.1080/00094056.2019.1565809

Akpem, S. (2020) *Cross Cultural Design.* A Book Apart.

Ayyagari, R., Grover, V., & Purvis, R. (2011). Technostress: Technological Antecedents and Implications. *MIS Quarterly*, *35*(4), 831-858. doi:10.2307/41409963

Bandura, A. & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology*, *45*, 1017-1028.

Benassi, V., Overson, C.E., & Hakala, C.M. (2014). Applying science of learning in education: Infusing psychological science into the curriculum. *Society for the Teaching of Psychology*, from: http://teachpsych.org/ebooks/asle2014/index.php

Bernard, R.M., Borokhovski, E., Schmid, R.F., Tamin, R., & Abrami, P. (2014). A meta-analysis of blended learning and technology use in higher education: from the general to the applied. *Journal of Computing in Higher Education*, *26*, 87–122. https://doi.org/10.1007/s12528-013-9077-3

Brooks, S. & Califf, C. (2017) Social media-induced technostress: Its impact on the job performance of IT professionals and the moderating role of job characteristics. *Computer Networks*, *114*, 143-153. https://doi.org/10.1016/j.comnet.2016.08.020.

Carr, N. (2011). *The shallows: What the internet is doing to our brains.* New York: Norton.

Centre for Humane Design. (n.d.) *Design Guide*. https://www.humanetech.com/designguide

Centre for Humane Design. (n.d.) *What we do*. https://www.humanetech.com/what-we-do

Celino, I., & Re Calegari.G. (2020) Submitting surveys via a conversational interface: An evaluation of user acceptance and approach effectiveness." *International Journal of Human-Computer studies, 139.*

Chaves, A., & Gerosa, M. (2021). How Should My Chatbot Interact? A Survey on Social Characteristics in Human-Chatbot Interaction Design. *International Journal of Human-Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Cooley, M. (1987) Human centred systems: An urgent problem for systems designers. *AI & Soc 1*, 37–46. https://doi.org/10.1007/BF01905888

Cowan, N. (2001). The magical number 4 in short term-memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185. https://doi.org/:10.1017/s0140525x01003922

Griffith, T.V., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of medical Internet research*, *22*(3), 16235. https://doi.org/10.2196/16235

Gulson, T. & Webb, T. (2017). Mapping an emergent field of 'computational education policy': Policy rationalities, prediction and data in the age of Artificial Intelligence. *Research in Education*, *98*(1) 14–26. doi: 10.1177/0034523717723385

Halgren, E. & Marinkovic, K. (1995) Neurophysiological networks integrating human emotions In M.S. Gazzaniga (Ed.) *The cognitive neurosciences*, 1137-1151. Cambridge, MA; MIT Press.

Hassenzahl, M., Burmester M., & Koller, F. (2003) Attrak Diff: a questionnaire to measure perceived hedonic and pragmatic quality. *Man computer: interaction in motion*. B.G. Teubner, Stuttgart/Leipzig, 187–196.

Heimgärtner, R. (2019). *Intercultural User Interface Design*. Springer International Publishing. https://doi.org/10.1007/978-3-030-17427-9

Hung, W., Chen, K., & Lin, C. Does the proactive personality mitigate the adverse effect of technostress on productivity in the mobile environment?. *Telematics and Informatics, 32*(1), 143-157. https://doi.org/10.1016/j.tele.2014.06.002.

Hype Cycle Research Methodology. (n.d.). Retrieved January 28, 2021, from https://www.gartner.com/en/research/methodologies/gartner-hype-cycle

Jiang, Q., & Li, Y. (2018). Factors affecting smartphone dependency among the young in China. *Asian Journal of Communication*, *28*(5), 508–525. https://doi.org/10.1080/01292986.2018.1431296

Kalbag, L. (2017) *Accessibility for Everyone.* A Book Apart.

Kamalou, S., Shaughnessy, K., & Moscovitch, D. A. (2019). Social anxiety in the digital age: The measurement and sequelae of online safety-seeking. *Computers in Human Behavior*, *90*, 10–17. https://doi.org/10.1016/j.chb.2018.08.023

Koehler, M. & Mishra, P. (2009). What is Technological Pedagogical Content Knowledge (TPACK)?. *Contemporary Issues in Technology and Teacher Education*, 9(1), 60-70. from https://www.learntechlib.org/primary/p/29544/.

Kuznekoff, J., & Titsworth, S. (2013). The Impact of Mobile Phone Usage on Student Learning, *Communication Education*, *62*(3), 233-252. doi: 10.1080/03634523.2013.767917

Laurence, Michael. (2006, June 1). Steve Jobs, *"Computers are like bicycles for our minds." – Michael Laurence Films* [Video]. YouTube. https://www.youtube.com/watch?v=ob_GX50Za6c

Lepp, A., Barkley, J., & Karpinski, A. (2014) The relationship between cell phone use, academic performance, anxiety, and Satisfaction with Life in college students. *Computers in Human Behavior*, *31*, 343-350. https://doi.org/10.1016/j.chb.2013.10.049.

Longnecker, E. (2017). *The Relationship between smartphone use, symptoms of depression, symptoms of anxiety, and academic performance in college students.* ProQuest Dissertations Publishing.

Luger E., & Sellen, A. (2016). Like having a really bad PA: The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

Maslow, A. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396. https://doi.org/10.1037/h0054346 Mori, M. (1970) The uncanny valley. *Energy*, *7*(4) 33–35.

Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, *72*, 432–440. https://doi.org/10.1016/j.chb.2017.02.067

Norman, D. (2003). Emotional design: why we love or hate everyday things. Cambridge: Basic Books .

Nurshatayeva, A., Page, L., White, C., & Gehlbach, H. (2021). Are Artificially Intelligent Conversational Chatbots Uniformly Effective in Reducing Summer Melt? Evidence from a Randomized Controlled Trial. Research in Higher Education. https://doi.org/10.1007/s11162-021-09633-z

Pekrun, R. (2016) Academic emotions. In K.R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school*. New York: Routledge.

Page L. & Gehlbach H. (2017) How an artificially intelligent virtual assistant helps students. Navigate the Road to College, *AERA Open, October-December 2017 3* (4), 1–12. doi: 10.1177/2332858417749220

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193-198. https://doi.org/:10.1037/h0049234

Petko, D. (2012). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the "will, skill, tool" model and integrating teachers' constructivist orientations. *Computers and Education*, *58*(4), 1351–1359. https://doi.org/10.1016/j.compedu.2011.12.013

Preim B,. & Dachselt, R. (2010*). Interactive systems-volume I: basics, graphical user interfaces,information visualization*. Springer Verlag, Heidelberg.

Rashid, T. & Asghar, H. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations, *Computers in Human Behavior, 63*, 604-612. https://doi.org/10.1016/j.chb.2016.05.084.

Rushowy, K. (2019). Ford government's tuition cut to cost universities $360 million and colleges $80. Toronto: The Toronto Star. Retrieved July 13, 2019 from https://www.thestar.com/politics/provincial/2019/01/16/ford-governments-tuition-cut-to-cost-universities-360-million-and-colleges-80-million.html

Salesforce (2020). *How we're bringing inclusive language to our products.* https://www.salesforce.com/news/stories/how-were-bringing-inclusive-language-to-our-products/

Saleheen, F., Wang, Z., Picone, J., Butz, B. P., & Won, C. (2018). Efficacy of a Virtual Teaching Assistant in an Open Laboratory Environment for Electric Circuits, *Advances in Engineering Education*, *6*(3), 1-27.

Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes,M. A.,...Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education*, 72, 271–291

Schunk, D.H. (2020). *Learning theories: an educational perspective*. Pearson.

Schunk, D.H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In K.R. Wentzel & D.B. Miele (Eds)*, Handbook of motivation at school*. Routledge.

Seligman, M. E. P. (1991). *Learned optimism.* Knopf.

Shaw, H., Ellis, D. & Ziegler, F. (2018). The technology integration model (TIM). Predicting the continued use of technology. *Computers in Human Behavior, 83*, 204-214.

Shawar, A., & Atwell, E.S. (2007). Chatbots: Are They Really Useful? *GSCL German Society for Computational Linguistics, 1*, 29-49.

Silva, I.; Rodrigues, L.; Sampayo, M. (2020). The bad and the ugly: a systematic review of technology's negative impacts' mentions in literature from 2005 to 2020. *Editorial Universitat Politècnica de València.* 819-829. https://doi.org/10.4995/HEAd20.2020.11152

Stevens, R., Petermans, A., & Vanrie, J. (2019). Design for human flourishing: A novel design approach for a more 'humane' architecture. *the design journal*, *22*(4), 391–412. https://doi.org/10.1080/14606925.2019.1612574

Tamim R.M., Bernard R.M., Borokhovski E., Abrami P.C., & Schmid R.F. (2011). What forty years of research says about the impact of technology on learning: a second-order meta-analysis and validation study. *Review of Educational Research*, *81*(1):4-28. doi:10.3102/0034654310393361

TPACK & SAMR by Allie & Jennie - YouTube. (2015). Retrieved October 28, 2020, from https://www.youtube.com/watch?v=98Lw8cuUses

Wang, X., Tan, S. C., & Li, L. (2020). Technostress in university students' technology-enhanced learning: An investigation from multidimensional person-environment misfit. *Computers in Human Behavior*, *105*, 10. doi:http://dx.doi.org.libproxy.wlu.ca/10.1016/j.chb.2019.106208

Wolfgang Greller, & Hendrik Drachsler. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society*, *15*(3), 42-57. Retrieved October 1, 2020, from http://www.jstor.org/stable/jeductechsoci.15.3.42

Xu, Z., Banerjee, M., Ramirez, G., Zhu, G., & Wijekumar, K. (2019). The effectiveness of educational technology applications on adult english language learners' writing quality: A meta-analysis. *Computer Assisted Language Learning*, *32*(1-2), 132-162. doi:http://dx.doi.org.libproxy.wlu.ca/10.1080/09588221.2018.1501069

Young, I. (2020) Twitter https://twitter.com/indiyoung/status/1309292536704503808

Zadrozny, W., Buszikowska, M. Chai, J., Kambhatla, N., Levesque, S. & Nicolog, N. (2000). Natural Language Dialogue for Personalized Interaction. *Communications of the ACM, 43*(8), 116.

Appendix A

Scenarios

1. Add/ Drop

   My home faculty is the Faculty of Science. I accidentally enrolled in the wrong class and now I can't change my course to the one I need because the add-drop deadline has passed. What can I do to ask the university to make an exception in my case?

2. Advisor

   My name is Blake Smith, I'm in first year and the Faculty of Science is my home faculty. How can I make an appointment with an academic advisor?

3. Data

   I started the Honours BSc Data Science program in 2018/2019. What classes are required for a concentration in Big Data?

4. Deferral

   How can I request to defer a final exam in the Faculty of Science because of personal circumstances?

Appendix B

Scenario Questionnaire

(One of the following questions will appear depending on which scenario the user just completed) [2]

- How can a student ask the university to make an exception to change a course after the add/drop date?
- How does a student make a virtual appointment with an academic advisor?
- What classes are required for a concentration in Big Data?
- What is the process for deferring a final exam in the Faculty of Science because of personal circumstances?

(The following questions will all appear for each of the scenarios)

(Confidence in Response)

1. How confident are you that you found the information that you were seeking?

o Likert Scale: 1 – 5

I was…

o not at all confident not really confident fairly confident confident very confident

(Human Sensitivities)

[2] Comments are noted in parenthesis and did not appear to participants.

(Emotional)

2. How stressed or calm did you feel as you were seeking this information?

o Likert Scale: 1-5

I was…

o stressed somewhat stressed neither calm nor stressed calm very calm

(Attention)

3. To what extent did you feel your attention was focused or pulled in different directions because of what information was presented or how the information was presented?

o Likert Scale: 1-5

My attention was…

o pulled in many directions pulled in a few directions neutral focused very focused

(Sense-Making)

4. To what extent did you feel confused by or able to understand the information presented?

o Likert Scale: 1-5

o I was very confused I was somewhat confused neutral I understood I very easily understood

(Decision-Making)

5. To what extent did you feel like you were able to control your information seeking?

o Likert Scale: 1 – 5

I felt I could…

o not control my information seeking somewhat not control my information seeking neutral somewhat control my information seeking control my information seeking

(Social Reasoning)

6. If this was a real scenario, how do you feel this brief interaction would influence your relationship with the university?

o Likert Scale: 1 – 5

My relationship would be...

o worsened somewhat worsened the same somewhat improved improved