

Talk It Through: Conversational AI for Sexual Consent Education

MSc Thesis in Social Data Science
Oxford Internet Insitute
University of Oxford

Candidate number:
1088494

Date: August 1, 2025

Word Count: 14′972

Contents

1 Introduction **2**

2 Background and Related Works **4**
2.1 The State of Sexual Harassment at Universities in the UK 4
2.2 Causes of Sexual Harassment at UK Universities 4
2.2.1 Structural Gender Dynamics on Campus 4
2.2.2 Social-Psychological Processes 5
2.2.3 Intersectional Risk Structures 6
2.3 Policy and Legislative Approaches 7
2.4 Consent Training Programs 7
2.4.1 Theory Behind Consent Trainings 7
2.4.2 Implementation 8
2.4.3 Training Curricula 9
2.5 Mixed Results from Sexual Harassment Training Evaluations 9
2.5.1 Sexual Harassment Training Evaluations 9
2.5.2 Empirical Findings 11
2.6 The Case for AI Training 12

3 Methods **15**
3.1 Study Design 15
3.1.1 Sample 15
3.1.2 Procedure 16
3.2 Outcome Measures 17
3.2.1 Reaction 18
3.2.2 Learning 19
3.2.3 Behavior 20
3.3 Control Training 20
3.4 Treatment Training 20
3.4.1 Model Selection and Evaluation 21
3.4.2 Prompt Engineering 22
3.4.3 Retrieval Augmented Generation 25
3.4.4 Interface 25
3.5 Data Analysis 25

4 Results **27**
4.1 Reactions 27
4.2 Cognitive Learning 29
4.3 Affective Learning 30
4.4 Behavior 30

5 Discussion **32**
5.1 Principal Findings . . . 32
5.2 Limitations . . . 32
5.2.1 Methodological Limitations . . . 32
5.2.2 Intervention Limitations . . . 34
5.2.3 Ethical Limitations . . . 34
5.3 Implications . . . 35

6 Conclusion **36**

A Experiment Elements **52**
A.1 Overall Experiment Flow . . . 53
A.2 Recruitment . . . 54
A.2.1 Study Invitation . . . 54
A.2.2 Informed Consent . . . 55
A.3 Pre-training Questionnaires . . . 57
A.3.1 Pre-training Attitude Questionnaire . . . 57
A.3.2 Pre-training Knowledge Test . . . 59
A.3.3 Profiling Questionnaire . . . 61
A.4 Chatbot Training . . . 63
A.5 E-learning Training . . . 64
A.6 Post-training Questionnaires . . . 64
A.6.1 Reaction Questionnaire . . . 65
A.6.2 Post-training Attitude Questionnaire . . . 67
A.7 Follow-up . . . 69

B Chatbot Elements **70**
B.1 Prompt . . . 70
B.2 RAG Corpus . . . 82

C Results **85**
C.1 Cognitive Learning ANCOVA Results . . . 85
C.2 Affective Learning ANCOVA Results . . . 86

Trigger Warning: This thesis contains discussions of sexual consent and sexual assault that might be upsetting to some.

Abstract

Consent education in universities has become a legal and ethical necessity. Typically delivered through static e-learning formats, current trainings often result in low engagement, short-lived effects and minimal behavioral impact. This thesis evaluates whether a large language model chatbot can deliver more effective consent and sexual harassment training than the standard e-learning module used at UK universities. Four outcomes are examined: user reaction, cognitive learning, affective learning and behavior. A two-arm randomized controlled trial assigned 150 UK-based students to either a control e-learning module ($n = 76$) or an equivalent chatbot training ($n = 74$) for treatment, with outcomes measured via pre-training, post-training, and two-week follow-up questionnaires. Both formats were similarly well-received ($UX_t = 8.7, UX_c = 8.84$). Knowledge increased in each group with a higher effect from the chatbot ($\eta^2 = 0.03$). Attitude scores improved similarly in both groups. Behavioral intention to pursue further training was higher in the treatment group ($d = 0.29$). The chatbot training matches or exceeds the performance from traditional methods, offering a promising, accessible, private and low-cost complement to in-person consent trainings.

Chapter 1

Introduction

Universities are failing to protect their students from sexual violence. Despite decades of awareness campaigns and sexual harassment trainings, the majority of victims remain silent, perpetrators remain unpunished, and institutional responses remain inadequate. Still today, more than half of students in the United Kindgom (UK) are exposed to some form of sexual misconduct during their studies (Barradale, 2021; Brook, 2018; Oxley & Lipscombe, 2025; Revolt Sexual Assault, 2018). A vast majority of victims are women (70%), and barely any ever report the crime (6%) (Bondestam & Lundqvist, 2020; Revolt Sexual Assault, 2018). With just 2.7% of the cases that do reach law enforcement resulting in legal charges, the odds of perpetrators facing any consequences are close to inexistent (Home Office, 2025; Rape Crisis England & Wales, 2025).

High perpetration and low reporting rates are fueled by social structures within universities that normalize gendered violence. On UK campuses this strongly manifests as "lad culture" a form of hegemonic masculinity marked by heavy drinking, sexual conquest, and the routine sexual objectification of women (Phipps & Young, 2015). Lad culture creates an environment that silences survivors and emboldens perpetrators (Hales & Gannon, 2022; Phipps & Young, 2013). Universities are therefore not only scenes of sexual violence but also key sites for its prevention through cultural change.

Set to take effect in August 2025, the Office for Students' E6 conditions establish legally binding requirements for how UK universities must prevent and respond to sexual harassment (Office for Students, 2024). One of its key requirements is the implementation of mandatory sexual consent training. Typically delivered as short e-learning modules or facilitator-led workshops, these interventions aim to debunk rape myths, promote affirmative consent, correct misperceptions about peer beliefs, and build confidence in being an active bystander (Bednarchik et al., 2024; Berkowitz, 2003; Dardis et al., 2016; Foubert, 2005; Kettrey et al., 2019).

Yet studies evaluating the effects of these trainings paint a mixed picture. Meta-analyses and randomised controlled trials routinely report short-term gains in knowledge and modest attitude shifts, but these effects decay quickly and rarely translate into measurable behavioural change (Anderson & Whiston, 2005; Brecklin & Forde, 2001; Breitenbecher, 2000; Lonsway, 1996; Paul & Gray, 2011; Verbeek et al., 2023). Some studies even document back-fire effects, including decreased intentions to report sexual assault and increased rape proclivity (Berg et al., 1999; Duggan, 1998; Hillenbrand-Gunn et al., 2010; Htun et al., 2022; Potter et al., 2016; Stephens & George, 2009). The gap between

policy ambition and empirical reality therefore creates an urgent need for innovation in how consent education is delivered.

This thesis argues that large language-model (LLM)–powered chatbots offer a promising alternative to the shortcomings of existing consent trainings. Unlike traditional formats, conversational AI provides adaptive, dialogue-based instruction that boosts engagement and supports personalized learning (Qin et al., 2025; Yusuf et al., 2025). It has been proven to be highly persuasive and able to foster an empathetic, judgment-free space for users to express discomfort or sensitive personal experiences (Bull et al., 2024; Hatch et al., 2025; Siddals et al., 2024; Wang et al., 2022). However, these benefits come with important ethical caveats, such as mishandling crises (Grabb et al., 2024; Moore et al., 2025), excessive trust from users (Abrams, 2025; Hipgrave et al., 2025), and hallucinations (Ji et al., 2023; Maynez et al., 2020). Nevertheless, this LLM-instruction approach has shown promise in delivering sexual health information, but it has yet to be applied to the context of sexual harassment prevention (Latt et al., 2025; Wang et al., 2022). Therefore this study asks the following research question:

> *Does an AI-based conversational consent and sexual harassment training improve user reactions, cognitive learning, affective learning, and behavior compared to a standard online training module among UK university students ?*

To test this proposition, the study conducts a two-arm randomized controlled trial with 150 UK university students recruited on prolific. The control group was assigned a 30-minute commercial e-learning training, while the treatment group took part in a conversation with a purpose-built Gemini 2.5 Flash chatbot educator based on the same curriculum. Cognitive and affective learning outcomes are measured at baseline, immediately post-training, and in a two-weeks follow-up. User reactions and behavior are only measured immediately post-training.

Key findings reveal that the chatbot produces significantly larger and more durable gains in knowledge, comparable attitude improvements, and a modest increase in willingness to pursue further training, without sacrificing user trust or experience. These results demonstrate, for the first time, that AI-mediated consent education can match or surpass standard e-learning training while retaining high learner satisfaction.

The implications are two-fold. Practically, universities could deploy conversational AI as a low-cost, scalable complement to human-led initiatives, thereby strengthening compliance with OfS requirements and reducing facilitator burnout. Theoretically, the study extends the consent-training literature by showing how adaptive dialogue and retrieval-augmented generation can operationalise core principles of educational psychology—activating prior knowledge, providing immediate feedback, and supporting learner agency—in a domain where traditional methods have plateaued.

The remainder of this thesis proceeds as follows. Chapter 2 explores the existing research on causes, current solutions and improvements of sexual harassment at UK universities. Chapter 3 presents the methods for this study. Chapter 4 outlines the results, which are discussed in Chapter 5. Chapter 6 offers concluding remarks.

Chapter 2

Background and Related Works

2.1 The State of Sexual Harassment at Universities in the UK

Sexual violence is not the exception in universities, it is the norm. Surveys consistently report an alarming half of university students experience sexual harassment during their studies (Barradale, 2021; Brook, 2018; Oxley & Lipscombe, 2025; Revolt Sexual Assault, 2018). This number rises sharply when looking at gendered exposure, with female victimization rates above 70% (Bondestam & Lundqvist, 2020; Revolt Sexual Assault, 2018). Despite the scale of the problem, sexual misconduct is rarely reported and even more rarely punished. Only 6% of students who experience sexual harassment or assault report it to their university or the police (Revolt Sexual Assault, 2018), and of the cases that do reach law enforcement, just 2.7% result in legal charges (Home Office, 2025; Rape Crisis England & Wales, 2025).

The consequences of this violence extend far beyond the assault, profoundly affecting student well-being and educational outcomes. A multi-site study conducted at UK campus universities found that 100% students who had experienced sexual violence reported negative psychological, emotional, or physical health impacts: 27% contemplated suicide or self-harm, 15% developed an eating disorder, another 15% abused alcohol or drugs, and 12% became more prone to illness (Stenning et al., 2012). The academic toll was similarly stark, with half of the students who had been sexually assaulted reporting that it negatively affected their academic performance, and 11% indicating that their studies were delayed as a result (Stenning et al., 2012).

2.2 Causes of Sexual Harassment at UK Universities

2.2.1 Structural Gender Dynamics on Campus

The alarming prevalence and profoundly harmful consequences of sexual harassment within UK universities underscores the urgency of understanding and preventing its underlying causes. Tackling sexual violence demands the recognition that this behavior is deeply rooted in societal gender dynamics (Walker et al., 2021). While abuse can and does affect individuals of all genders and occurs across diverse relationships , this should

not obscure the dominant pattern. In terms of scale, perpetrators remain overwhelmingly male, accounting for over 97% of sexual offenders, while women represent around 86% of victims (Oxley & Lipscombe, 2025). Sexual and domestic violence must therefore be understood as components of broader systemic violence against women, reflecting and reinforcing entrenched gender inequalities.

Rape in UK universities originates in long-standing gender hierarchies that normalize male sexual entitlement while marginalizing women and queer students. Central to this structure is hegemonic masculinity, described by Connell and Messerschmidt (2005) as the culturally dominant ideal of manhood. This ideal typically centers on heterosexuality, physical strength, emotional detachment, and dominance in relationships. By presenting male dominance over women and other forms of masculinity as natural and socially rewarded, the hegemonic masculine norm normalizes sexist and coercive behaviors. On British campuses hegemonic masculinity concretely manifests through the phenomenon of "lad culture". This predominantly masculine culture is characterized by heavy alcohol consumption, sporting activities and the sexual pursuit of women (Phipps & Young, 2015). It promotes misogyny under the guise of humor and the routine sexual objectification of women (Phipps & Young, 2013, 2015). These dynamics closely mirror the core pathways to sexual violence outlined in Malamuth et al. (1995)'s Confluence Model of Sexual Aggression, which are hostile orientation towards women and a detached attitude towards sex. Recent empirical research reinforces this connection, directly associating laddism with self-reported sexual aggressions among UK male university students (Hales & Gannon, 2022). Overall, lad culture fosters an environment in which sexual assault is trivialized and survivors are silenced.

2.2.2 Social-Psychological Processes

Sexual violence on campus is not only a sociocultural issue but is also enabled by individual-level cognitive mechanisms that allow perpetrators to justify in harmful behavior. These mechanisms are fueled by both cognitive distortions and moral disengagement. Cognitive distortions are defined by Ward et al. (1997) as a biased ways of interpreting information about oneself and the world. These biased belief systems predispose individuals to interpret harmful sexual situations as acceptable. They often take the form of "implicit theories", driven by widespread rape myths such as "Male sex drive is uncontrollable" or "Women often play 'hard to get' and say 'no' when they really mean 'yes"' (Burt, 1980; Ó Ciardha & Ward, 2013; Polaschek & Ward, 2002; Rape Crisis England & Wales, n.d.). "Implicit theories" and generally sexist ideologies have been empirically found to be widely endorsed among sexual offenders (Ó Ciardha & Ward, 2013; Polaschek & Ward, 2002).

While cognitive distortions predispose individuals to sexual harassment, moral disengagement mechanisms explain how they excuse it without guilt. Bandura (2006) defines moral disengagement mechanisms as the cognitive strategies by which harmful behavior is psychologically redefined by perpetrators so it no longer appear morally problematic, allowing them to legitimize their actions. Within UK university settings, moral disengagement frequently involves mechanisms such as euphemistic labeling, victim-blaming, displacement of responsibility, and moral justification. Euphemistic labeling manifests as humor, notably through the concept of "banter," which students use to normalize behaviors such as hazing, homophobia, and sexual bullying (Abell et al., 2024). Male undergraduates endorsing statements such as "It was only harmless banter" have been

found more lenient towards harassers and more likely to blame complainants (Page et al., 2016). This shift of the responsibility onto the victim for "provoking" the behavior is also associated to higher moral-disengagement scores (Page et al., 2016). Additionally, students often displace personal responsibility onto external factors such as alcohol, drugs, or peer pressure, further distancing themselves from accountability (Cooper, 2021). Ultimately, sexual harassment on campus is morally justified by students as a "normal part of university life" (Phipps & Young, 2015). These four processes are further reinforced by commonly accepted rape myths such as "If she didn't physically resist, it wasn't rape.", "women who dress or act provocatively are asking for it", or "Men who rape are mentally ill strangers, not normal guys". These are widely recognized as significant drivers of sexually aggressive behaviors on campuses (Burt, 1980; Hales & Gannon, 2022).

These social beliefs enabling sexual harassment persist due to pluralistic ignorance defined in the social norms theory (Berkowitz, 2003). Pluralistic ignorance occurs when individuals overestimate others' acceptance of a behavior or norm, causing them to stay silent or conform, even if they personally disagree. Dardis et al. (2016) have found that men's own attitudes towards rape and women were correlated with what they thought their friends believed and what they thought "the average college male" believed. However, these perceptions were wrong. There was no correlation between what men thought their friends believed and what those friends actually believed. In the context of UK universities, pluralistic ignorance helps explain how harmful campus norms endure even when many students personally disapprove.

2.2.3 Intersectional Risk Structures

Sexual-harassment risk is not only influenced by gender and pyscho-social processes, but by the intersecting logics of race, sexuality, disability, migratory status and age. Intersectionality is a framework for analyzing how various systems of power interact to shape lived experiences and structural inequalities (Collins & Bilge, 2016). Among these, race plays a central role. Ethnically minoritised women face unique forms of harassment in university settings, including hyper-sexualisation, racial slurs, and reduced credibility when reporting misconduct (Arday & Mirza, 2018; Mirza, 2014; Sundaram et al., 2022). These factors puts them at additional risk of sexual harassment, with empirical studies showing that that Black, Latinx, and multiracial students report significantly higher rates of sexual assault than white peers (Coulter et al., 2017; Gross et al., 2006; Lincoln, 2021; Nagy et al., 1994; Sundaram et al., 2022). LGBTQ+ students also face disproportionately higher rates of sexual victimisation than their cisgender, heterosexual counterparts (Cantor et al., 2015; Coulter et al., 2017; Lincoln, 2021). Again this evidence is strongly linked to hyper-sexualization, stigma and discrimination of queer students, along with their exclusion from typically heteronormative sexual violence prevention efforts (Coulter et al., 2017; DeGue et al., 2014; Griner et al., 2020; Rothman et al., 2011). In the university context, Fethi et al. (2023) notes that international-students often lack social support networks, experience cultural and linguistic barriers, and may not fully understand their legal rights in the host country. These factors can lead to greater isolation and dependence, which perpetrators may exploit. Moreover, within academia, hierarchical power dynamics and age disparities between students and professors represent another factor of vulnerability. This environment fosters power imbalances that lead to younger undergraduates being disproportionately targeted by older students and staff, while senior academics exploit their authority to initiate inappropriate relationships with students (Bull & Page, 2021; Clancy et al., 2014). Finally, UK universities have been shown to pri-

oritise reputational concerns over student safety, with victims facing disbelief, retaliation, or inaction when reporting staff-student misconduct (Bull & Rye, 2018), discouraging disclosures and protecting perpetrators. These numerous intersecting structures of power mutually reinforce each other, amplifying both the risk and impact of sexual harassment.

2.3 Policy and Legislative Approaches

Addressing these deeply embedded causes of sexual harassment in higher education requires robust regulatory frameworks to hold universities accountable and protect students. Reflecting this need, the prevention of violence against women has become an increasing priority for the UK government (UK Parliament, 2003). While various national laws prohibit sexual harassment in general terms in the UK, it is higher education regulation that sets out the standards in how universities prevent and respond to sexual misconduct. The Office for Students (OfS) is the independent regulator of higher education in England. It was established by the Higher Education and Research Act 2017 and became fully operational in 2018. In 2021, the OfS issued its first statement of expectations on tackling harassment (Office for Students, 2021). This non-binding document was used as a benchmark for good practice. It sets out 7 key principles prompting universities for visible leadership and accountability, clear communication of expectations, student engagement, staff and student training, accessible reporting procedures, proportionate response upon reports, and support for students throughout investigation processes. Within the next month, the OfS is set to implement new legally binding requirement through the E6 conditions (Office for Students, 2024). As of 1 August 2025, all English higher education providers must publish a single, clear, and accessible policy on sexual harassment, ban or regulate intimate staff–student relationships, ensure mandatory training, offer appropriate support, ban non-disclosure agreements, protect freedom of speech, and demonstrate capacity and resources to meet these obligations. These new formally binding conditions of registration, allow the OfS to investigate universities, issue fines, suspend funding access and deregister institutions. This major policy change will soon ensure universities are highly accountable for their management of sexual harassment on campus.

2.4 Consent Training Programs

2.4.1 Theory Behind Consent Trainings

As outlined in the OfS E6 conditions, one of the main tools to fight sexual harassment at UK universities are staff and student mandatory trainings. The main theoretical basis for these trainings is the Theory of Planned Behavior, which helps provide targeted strategies for prevention. According to the Theory of Planned Behavior, an individual's intention to engage in a behavior is shaped by their attitudes, subjective norms, and perceived behavioral control (Ajzen, 1991). Attitude toward the behavior refers to the degree to which an individual holds a favorable or unfavorable evaluation of the behavior in question. When applied to sexual harassment training, strategies such as debunking rape myths, providing evidence of the impacts of sexual violence and framing consent as a necessary component of healthy sexual interaction, should decrease intentions to partain in harmful behavior. However, empirical studies detailed in section 2.5.1 often find no correlation between increase of knowledge, rape myth acceptance rate, and actual behavioral change. This gap highlights the importance of other determinants of behavior, such

as subjective norms. These refer to the perceived social pressure an individual feels to perform or refrain from performing a specific behavior. As described earlier, students often overestimate how accepting their peers are of harmful behaviors, which leads them to either participate in or tolerate misconduct they might otherwise reject (Perkins & Berkowitz, 1986). Revealing the actual, healthier norm to correct these misconceptions has been found to encourage individuals to reduce problematic behaviors and engage in protective ones (Berkowitz, 2003; Dardis et al., 2016). Finally, perceived behavioral control denotes an individual's confidence in their ability to perform a behavior. Increasing individuals' confidence in implementing consent and recognizing harmful behavior has been shown to reduce likelihood of perpetrating sexual harassment and to increase intentions to intervene in risky situations (Foubert, 2005; Kettrey et al., 2019). Supporting this, recent longitudinal research demonstrates that both perceived behavioral control and social norms are significant predictors of students' willingness to engage in affirmative consent practices (Bednarchik et al., 2024). Some high quality trainings such as the Green Dot Program highly engage with the Theory of Planned Behavior (Coker et al., 2011; Washington University, 2011). However, in practice, there are only a few programs that explicitly base themselves on any theoretically-grounded content, or only specific sub-theories of the Theory of Planned Behavior such as the Social Norms and Bystander Theory (Epigeum, n.d.-b; University of Exeter, n.d.).

2.4.2 Implementation

Translating theory into practice, consent and sexual harassment trainings are typically delivered through a range of formats, most commonly via e-learning or in-person trainings, each with own advantages and limitations. E-learning trainings are online courses designed to deliver educational content via text, video, and interactive elements. They allow for scalable, consistent and anonymous delivery while keeping flexibility in scheduling and letting students move through the information at their own pace (Peterson & McCleery, 2014; Preusser et al., 2011). However, they convey high non-completion rates and some institutions struggle with computer availability (Brown, 2001; Preusser et al., 2011). In the UK, e-learning modules are developed by a range of actors, including large commercial providers such as Epigeum, Ciphr, Brook, and Icena; charitable organisations like AdvanceHE; and universities themselves, notably through the Intervention and Let's Be Clear on Consent initiatives. Each of these programmes reaches anywhere from a handful to several hundred institutions (Advance HE, 2025; Brook, n.d.; Ciphr, n.d.; Epigeum, n.d.-b; Icena, n.d.; University of Exeter, n.d.).

On the other hand, in-person trainings are typically delivered through workshops or presentations, led by either professional facilitators or trained peer educators (Anderson & Whiston, 2005). These sessions are highly interactive, allowing real-time discussion, adaptation to trainee learning styles, and practice of scenarios. Having an educator present provides students with social pressures not only to attend training but to learn the material (Desai et al., 1999; Preusser et al., 2011). However, these trainings come with logistical limitations. They often suffer from restricted scheduling options, limited seating capacity, and a fixed learning pace that may not suit all participants. Furthermore, the quality and consistency of content delivery varies depending on the individual facilitator (Desai et al., 1999; Preusser et al., 2011). More importantly, these trainings put facilitators at high risk of vicarious trauma due to repeated exposure to distressing content related to sexual violence, particularly when they are not adequately supported (Nikischer, 2019). In the UK, in-person trainings are primarily provided by commercial organisations such as The

Art of Consent, Unlrn, Talk Consent, and EI Education, as well as peer-led university initiatives like It Happens Here at the University of Oxford. These providers typically reach several dozen to a few hundred institutions across the country (EI education, n.d.; Oxford Student Union, n.d.; Talk Consent, 2018; The art of consent, n.d.; UNLRN, n.d.).

2.4.3 Training Curricula

Even when not explicitly grounded in the Theory of Planned Behavior, most consent and sexual harassment training programs still incorporate elements targeting attitudes, norms, and behavioral skills (Paul & Gray, 2011). Typically, most of them begin with general information and awareness-building sections. These include definitions of sexual assault and consent, legal frameworks, and information on support resources (Brook, n.d.; Epigeum, n.d.-a; Icena, n.d.; University of Edinburgh, n.d.; University of Exeter, n.d.). These informational elements are often accompanied by attitudinal components designed to foster empathy toward victims, increase understanding of the psychological impact of sexual violence, and challenge harmful beliefs such as rape myths (Brook, n.d.; Icena, n.d.; University of Edinburgh, n.d.). Social norms are another key focus, with training elements aimed at unpacking gender stereotypes and correcting misperceptions about peer beliefs (Brook, n.d.; University of Edinburgh, n.d.; University of Exeter, n.d.). Intersectionality and multiple forms of discrimination are increasingly addressed, recognizing harassment experiences and access to justice differ for minorities (Brook, n.d.). Some trainings also provide content designed for specific ethnic groups, such as African-American, Asian or Latino men (DeGue et al., 2014). To build behavioral and communication skills, many trainings include modules on identifying early signs of abusive relationships, effectively communicating consent, and safely intervening as a bystander (Epigeum, n.d.-a; University of Edinburgh, n.d.; University of Exeter, n.d.). Finally, most digital programs incorporate some form of assessment, typically through formative quizzes or interactive scenario-based exercises, to reinforce learning and provide immediate feedback (Epigeum, n.d.-a; Icena, n.d.; University of Edinburgh, n.d.; University of Exeter, n.d.). The duration of prevention programs can range from 2 to 48 hours total (Verbeek et al., 2023)

2.5 Mixed Results from Sexual Harassment Training Evaluations

2.5.1 Sexual Harassment Training Evaluations

Throughout the literature, these trainings have been evaluated using a range of empirical methods, with randomized controlled trials (RCTs) widely regarded as the gold standard for assessing causal impact (DeGue et al., 2014). Supporting this, systematic reviews indicate that RCTs are the dominating evaluation method in this field, accounting for 58.6% of studies between 1985 and 2012, with this proportion rising to 77.7% in more recent analyses (DeGue et al., 2014; Kettrey et al., 2023; Wright et al., 2020). Quasi-experimental approaches, such as interrupted time series or regression discontinuity are also employed when randomization is impractical, making up approximately 25% of studies in one review (DeGue et al., 2014). Some evaluations rely on simpler pre–post single-group designs, though these are less robust, and only represent about 16.4% of the literature (DeGue et al., 2014).

Focusing on the most rigorous and prominent study design, RTCs in this field use a range of controls to evaluate the intended treatment training. These include no-treatment controls (Anderson & Whiston, 2005), waitlist controls (DeGue et al., 2014), attention-placebo controls, or minimal-information controls (Breitenbecher, 2000). Many evaluations also explore different delivery methods, such as comparing video-only to video-plus-discussion formats, computer-based to peer-facilitated programs, or empathy-focused to cognitive interventions (Glass et al., 2017; Preusser et al., 2011; Schewe & O'Donohue, 1993). Within these RCTs, data is typically gathered through baseline pre-training assessments, immediate post-training tests, and longitudinal follow-up surveys to evaluate both short- and long-term effects. However, most studies only include pre-training and immediate post-training assessments establishing baseline and immediate changes but no long term effects (Gidycz et al., 2011; Potter et al., 2016). This is an important methodological flaw, as relying exclusively on immediate post-tests can inflate outcomes due to demand characteristics (Breitenbecher, 2000; Paul & Gray, 2011). Demand characteristics are cues that lead participants to infer a study's purpose and respond in socially desirable ways to align with perceived expectations, thus inflating results (Breitenbecher, 2000; Malamuth & Check, 1984; Rosenthal, 1995). Implementing follow-up measures mitigates for that effect. These can vary from a few days, to months, or even years, with the longest evaluations extending over four years (Breitenbecher, 2000; Campbell et al., 2013; DeGue et al., 2014; Dobbin & Kalev, 2019; Gidycz et al., 2011; Lonsway, 1996; Moynihan et al., 2015).

The majority of single-theory-driven evaluations ground their outcome measure design on Kirkpatrick's four-level model (Peterson & McCleery, 2014; Preusser et al., 2011). Kirkpatrick and Kirkpatrick (2006) defines four hierarchical levels for evaluating the effectiveness of training programs. The first level, reaction, captures participants' immediate impressions, examining whether they found the training engaging, relevant, and effective. The second level, learning, evaluates the extent to which participants understood and retained key information. The third level, behavior, assesses whether trainees apply what they learned in real-life situations. Finally, the results level, examines the broader impact of the training on organizational outcomes. The second level of this model has since then been extended by Kraiger et al. (1993), separating learning measures in three types of outcomes: cognitive, skill-based, and affective. Cognitive learning refers to acquiring factual knowledge about sexual harassment, skill-based learning reflects the ability to apply this knowledge effectively in real-world scenarios, and affective learning, captures shifts in attitudes about sexual harassment.

In sum, six types of measures for the evaluation of consent and sexual harassment programs consistently appear in the literature: reaction, cognitive learning, affective learning, skill-based learning, behavior and organizational outcomes. These are operationalized in various ways within the literature. First, reactions are usually measured using likert-scale and open ended questions (Fenton & Mott, 2018; Preusser et al., 2011). Second, cognitive learning is evaluated using quizzes or self-evaluations (Antecol & Cobb-clark, 2003; Bingham & Scherer, 2001; Campbell et al., 2013; Htun et al., 2022; Potter et al., 2016; Preusser et al., 2011). Third, affective learning is evaluated using the rape myth acceptance scale, adversarial sexual beliefs metrics, and composite measures of rape-supportive attitudes (Antecol & Cobb-clark, 2003; Bingham & Scherer, 2001; Fenton & Mott, 2018; Htun et al., 2022; Lee & Wong, 2024; Potter et al., 2016). Fourth, measures of skill-based learning include metrics of intentions to confront the perpetrator, to report or self-reported

likelihood of engaging in sexually aggressive behavior (Bingham & Scherer, 2001; Breitenbecher, 2000; Fenton & Mott, 2018; Htun et al., 2022). While considered the ultimate goal of prevention programs, measures of actual sexually violent behavior are never included in evaluation studies, due to the difficulty of measuring them. Proxy measures are done via seemingly unrelated follow-up telephone calls evaluating participants' willingness to join a rape prevention program or using the rape conformity assessment, which empirically measures whether individuals conform to rape-supportive attitudes expressed by peers in a group setting. Finally, overall organizational measures are often rates of sexual violence victimization or perpetration based on official records (DeGue et al., 2014).

2.5.2 Empirical Findings

While the components outlined above reflect widespread consensus on what consent and sexual harassment trainings should include, their actual effectiveness remains an open empirical question. On the positive side, there is fairly consistent support that trainings effectively improve participants' knowledge about sexual assault (Breitenbecher, 2000). Indeed, 61.8% of studies reviewed by DeGue et al. (2014) between 1985 and 2012 reported positive effects on knowledge, while the remaining 38.2% showed mixed or null effects. When significant, the effect sizes for knowledge gains tend to range from medium to large, with reported values including d = 0.43 (Bingham & Scherer, 2001), d = 0.574 (Anderson & Whiston, 2005), g = 1.07 (Kettrey et al., 2019), and d = 0.9 (Glass et al., 2017). Based on these prior findings, the control e-learning training is expected to have a moderate positive effect on knowledge scores.

Beyond factual knowledge, there is mixed empirical for the effectiveness of sexual assault prevention programs in shifting rape-supportive attitudes. DeGue et al. (2014) found only a third studies measuring attitudinal outcomes reporting of positive effects, with 3.5% showing negative effects, highlighting the risk of potentially counterproductive interventions. This pattern is echoed in Verbeek et al. (2023), where only 13 out of 44 attitudinal effects reached significance. However, it is somewhat offset by earlier findings from Breitenbecher (2000), reporting a higher 13 out of 16 studies. When a positive effect is found, its size is marginal with effect sizes for various studies ranging from 0.06 to 0.35 (Anderson & Whiston, 2005; Bingham & Scherer, 2001; Kettrey et al., 2023). With this in mind, the control e-learning training is expected to deliver a small to negligible positive effect on attitude scores.

While these findings suggesting gains in knowledge and attitudes remain encouraging, most rely on short-term measures. There is consistent evidence that initial positive changes tend to decay towards pre-intervention levels over time, with effects often becoming very small or non-significant within weeks or months after the intervention. (Anderson & Whiston, 2005; Brecklin & Forde, 2001; Breitenbecher, 2000; Lonsway, 1996; Paul & Gray, 2011; Verbeek et al., 2023). This rebound effect is expected to make control e-learning's effects on knowledge and attitudes decay over time.

While knowledge and attitudes measures are often prioritized, participants' subjective experiences also shape the overall effectiveness of consent trainings. Reactions from students tend to be mixed to negative, with several studies reporting that participants found the material to be not beneficial to them (Acquaviva et al., 2023), "unnecessary," "nothing new," "redundant" and stating it taught "commonsense facts", "talking to them like a child" (Htun et al., 2022). Students often "clicked through" programs, without re-

ally engaging with the content (Potter et al., 2016). Notably, men expressed particular resistance, feeling unfairly targeted, with 63% voicing negative views compared to 39% of women (Htun et al., 2022). However, some programs received more favourable evaluations. In Preusser et al. (2011), participants appreciated the use of video clips, and in Lonsway (1996), students in mixed-gender groups reported feeling comfortable, engaged, and positive about the format. Therefore the control e-learning is expected to provoke negative reactions.

Finally, the main aim of these trainings is to have behavioral effect. However, most show no significant impact on this metric (Anderson & Whiston, 2005; Breitenbecher, 2000; Fenton & Mott, 2018; Foubert, 2005; Gidycz et al., 2011; Kettrey et al., 2023; Potter et al., 2016; Schewe & O'Donohue, 1993; Wright et al., 2020), with DeGue et al. (2014) reporting 80.9% of training evaluations had mixed or no effect. Alarmingly, they also report 14.3% of studies having negative effects, while only 4.8% show positive effects. Indeed, multiple studies have reported backfire effects with, reported their intervention led to increased reporting of sexually violent behavior or rape proclivity (Berg et al., 1999; Duggan, 1998; Hillenbrand-Gunn et al., 2010; Potter et al., 2016; Stephens & George, 2009). Additionally, Htun et al. (2022) found that mandatory training led to a notable 13 percent drop in women's intention to report sexual assault to campus authorities. When found, positive behavioral effects are small (Anderson & Whiston, 2005; Glass et al., 2017; Kettrey et al., 2023; Verbeek et al., 2023; Wright et al., 2020). This failure to translate awareness gains into meaningful behavioral change raises concerns with regards to the real-world effectiveness of current consent and sexual assault training programs. It is therefore highly likely that the control e-learning training has limited effect on participant behavior.

2.6 The Case for AI Training

Despite three decades of ever-expanding programs, the evidence reviewed above shows that both e-learning modules and in-person workshops yield, at best, short-lived gains. In light of these limitations, AI-powered training emerges as a promising new approach to enhance engagement and long-term impact of interventions. This novel method combines scalable delivery with the intimacy of natural conversation, enabling one-to-one, real-time instruction that is adaptive and empathic, something neither workshops nor e-learning formats can achieve.

Engagement is the first barrier any consent curriculum must clear. Indeed, disengaged learners simply "click-through" screens without processing any of the content, leading to barely any impact (Potter et al., 2016). When engaging in a learning task, Qin et al. (2025) have shown that AI instructional agents lead to significantly higher engagement than either human-led lectures or self-paced MOOC formats . Students in the AI agent condition interacted more frequently with the system, asked more questions, and completed the task in less time, all indicators of deeper cognitive and behavioural involvement. These findings suggest that responsive, dialogue-based instruction enhances both students' learning experience and learning outcomes (Qin et al., 2025). Given the positive impact AI has on user experience, the hypothesis related to user reactions for the treatment training is:

H_1: The treatment chatbot training fosters better user experience than the control e-learning training.

While engagement is essential, learners must also learn from the trianing. A key limitation of traditional consent education is its inability to promote long term knowledge and attitude change. Pedagogical research shows that AI tutors can foster comparable or even superior test performance and significantly higher perceived learner control than human teachers when delivering moderately demanding courses (Qin et al., 2025). These benefits stem from AI's ability to dynamically adapt instruction based on learners' interaction patterns, such as frequency and completion time, while also enabling real-time clarification and feedback, which together foster deeper cognitive engagement (Yusuf et al., 2025). Another example of the potential for consent teaching is Khan academy's teaching AI tool, Khanmigo, which was used by over 220'000 individuals during the 2023-24 academic year and has been shown to provide promising results in enhancing learning outcomes (DiCerbo, 2025; Shetye, 2024). The positive impact of AI-led teaching on learning outcomes leads to two of this study's hypotheses:

H_2: The treatment chatbot training has a higher effect on knowledge score improvement than the control e-learning training.

H_3: The treatment chatbot training has a higher effect on attitude score improvement than the control e-learning training.

In addition to factual learning, one of the main challenges in sexual harassment training is changing learner perceptions on widespread myths about sexual violence. Here too, AI shows remarkable promise. Numerous recent studies demonstrate that large language models match or exceeded human performance in shifting beliefs about conspiracy theories, policy issues, and quiz questions in the long term, thanks to their ability to sustain tailored, evidence-based dialogue (Bai et al., 2023; Costello et al., 2024; Salvi et al., 2025; Schoenegger et al., 2025). This positions AI as a powerful tool for consent education, capable of fostering reflective dialogue, challenging distorted beliefs, and reinforcing pro-consent norms at scale.

Personalization is a natural extension of persuasion. Psychology research shows that tailoring messages to match individual traits, such as interests, concerns or personality, significantly boosts the impact of a message (Hirsh et al., 2012; Joyal-Desmarais et al., 2022). Until now, such tailoring has been largely absent from consent education, due to the use of one-size-fits-all e-learning formats or facilitators that do not have the capacity to deliver one-to-one instruction at scale. Recent experiments show that LLMs can generate psychologically targeted messages, using as small an input as single short prompt describing a personnality trait, consistently outperforming generic messages across contexts like product marketing and political appeal (Matz et al., 2024; Simchon et al., 2024). Additionally, Yin et al. (2010) demonstrate that conversational agents that match a user's cultural background, through language and communication style, are significantly more persuasive and perceived as more trustworthy than those that do not. This means with minimal input, each student can engage in a personal conversation on consent tailored to their personality and culture, leading to higher persuasion. This high capacity for personalized persuasion, leads to the following hypothesis for behavioral change:

H_4: The treatment chatbot training has a higher effect on participant behavior than the control e-learning training.

While the primary goal of consent training is to improve learning and behavioral outcomes, this must be achieved within a safe and supportive environment, especially

given the potentially traumatic nature of the subject matter. Learners discussing topics like sexual violence may need to express discomfort, uncertainty, or past experiences, yet many are hesitant to do so in group settings or with authority figures (Griffin et al., 2022). Here, AI offers a unique advantage. Generative AI chatbots have been consistently rated as more empathic than human therapists when responding to therapeutic prompts, particularly in domains like couple's therapy (Hatch et al., 2025). Real-world interviews confirm that users find emotional sanctuary in their conversations with AI, often reporting healing, insight, and meaningful connection through open-ended dialogues with chatbots like ChatGPT (Siddals et al., 2024). Far from anecdotical, the use of chatbots for therapy and companionship is now the most widespread use of generative AI based on data from the Reddit forum (Zao-Sanders, 2025). Learners may be more willing to disclose sensitive issues, reflect honestly, and engage deeply in an AI safe from judgment.

However, these benefits must be weighed against substantial risks. AI therapy chatbots have been shown to mishandle or miss mental health emergencies, such as suicidal ideation or delusions, occasionally enabling harmful behaviour rather than intervening appropriately (Grabb et al., 2024; Moore et al., 2025). Unlike trained therapists, AI systems are unable to perceive subtle human cues such as tone or body language, which are often essential for understanding the full context of a person's distress (Hipgrave et al., 2025; Moore et al., 2025). In some cases, they may also express stigma toward those with mental health conditions (Moore et al., 2025). Finally, vulnerable individuals sometimes place excessive trust in chatbots, delaying appropriate professional treatment (Abrams, 2025; Hipgrave et al., 2025). This trust may also be misplaced given the widespread problem of hallucinations in generative AI systems, which can result in confidently delivered inaccurate information (Ji et al., 2023; Maynez et al., 2020). As such, while AI can provide a supportive environment for trauma-informed discussions, its application requires careful design, oversight, and safeguards.

Encouragingly, AI chatbots have already shown success in handling taboo topics in sexual and reproductive health. Studies across India, the UK, and the US found that sexual health chatbots provide accurate, judgment-free information while protecting user anonymity, which helps overcome stigma and encourage disclosure (Nadarzynski et al., 2021; Wang et al., 2022). Prompt-tuned chatbots reached accuracy and safety levels comparable to trained nurses (Latt et al., 2025), and real-world deployments even led to clinic appointments (Bull et al., 2024). As Mondal and Mondal (2025) notes, these systems fill key gaps where traditional education falls short, offering accessible support in areas long neglected by public health efforts.

Chapter 3

Methods

To evaluate large language models' potential for delivering effective consent and sexual harassment trainings, this study uses a two-arm randomized controlled trial design. The experiment compared a commercially available e-learning module created by the HR software company Ciphr and currently deployed at the University of Edinburgh, against a purpose-built conversational AI training delivered through the Gemini-2.5 Flash LLM. Both trainings were evaluated on reaction, cognitive learning, affective learning, and behavioral change. Outcome measures were collected pre-training, post-training and in a two-week follow-up questionnaire. The following chapter is structured in 4 parts. Section 3.1 outlines the overall study design and recruitment strategy. Section 3.2 defines the four outcome measures and their operationalization. Section 3.3 defines how the control e-learning training was selected. Section 3.4 describes the selection and optimization process to create the chatbot training. Finally, Section 3.5 outlines data analysis methods.

3.1 Study Design

3.1.1 Sample

The selection of participants was based on an a priori power analysis indicating that a minimum of 128 completers were required to achieve 80% power to detect a small-to-medium baseline-adjusted between-group effect (Cohen's d = 0.45) on the primary outcomes of reactions, cognitive learning, affective learning, and behavior, assuming a two-tailed α level of .05. To account for potential attrition, the target sample size was inflated by 15% to 150 participants. A pilot study with 16 participants was conducted to confirm the adequacy of the planned recruitment and data collection procedures. All participants were recruited via the Prolific Academic platform, chosen for its high quality pool of students, between 8 and 9 July 2025 (Douglas et al., 2023; Peer et al., 2022). They received a £4 reimbursement (approximately £8 per hour), considered fair compensation on the platform, conditional on completing all parts of the study. This payment scheme mirrors the standard expectations set by universities, where certification is only delivered if every element of the mandatory e-learning training is completed (London School of Economics, 2025; Oxford Brookes University, 2025; University of Oxford, 2025a). The final sample consisted of 150 ($n_{control} = 76, n_{treatment} = 74$) undergraduate and postgraduate students in the United Kingdom above 18 years old and fluent english speakers. Participants were only allowed to take part in the experiment once. No additional inclusion or exclusion criteria were applied thereby minimizing sampling bias (Jadad & Enkin, 1998). Attrition was minimal in the study, with 4 dropouts ($n_{control} = 1, n_{treatment} = 3$) in the first leg,

most due to the 87 minute time-out imposed by prolific. Dropout rate increased in the second leg with only 126 participants ($n_{control} = 60, n_{treatment} = 68$) completing the follow-up questionnaire. In line with national university demographics, 63% of participants identified as White (N = 94), 15% as Asian(N = 23), 13% as Black (N = 20), 8% as Mixed (N = 12), and 1% as other (N = 1) (HESA, 2025). Gender distribution was also representative, with 59% of female (n=89), 40% of male (n=59) and 1% of non-binary (n=2) participants (HESA, 2025). With regards to age, the average is slightly older than national university demographics, with 22.8% less students under 18 and 22.3% more students over 30% years old. A detailed breakdown of participant demographics per treatment group can be seen in Table 3.1.

Table 3.1. Descriptive statistics
Baseline demographic statistics broken down by experimental group. Proportions are reported as means or percentages of the group. Differences are calculated between the treatment group and the control group. No statistically significant differences were observed between groups on any variable (all $p > 0.1$), indicating successful randomisation.

Variable	N (control)	Mean (control)	N (treatment)	Mean (treatment)	Difference (pp)
Age	76	30.87	74	29.85	−1.02
Gender: Male	28	37%	31	42%	5.0%
Gender: Female	46	61%	43	58%	−2.4%
Gender: Non-binary	2	3%	0	0%	−3%
Ethnicity: White	48	63%	44	59%	−3.7%
Ethnicity: Black	11	14%	9	12%	−2.3%
Ethnicity: Asian	10	13%	13	18%	4.4%
Ethnicity: Mixed	6	8%	6	8%	0.2%
Ethnicity: Other	0	0%	1	1%	1.4%
Past exposure to consent training	26	34%	17	23%	−11.2%

3.1.2 Procedure

The randomized controlled trial was implemented via a guided 5 step procedure through a custom website. The full website is accessible with links provided in Appendix A. The procedure steps are recruitment, pre-training measures (time 1), training, post-training measures (time 2), and follow-up measures (time 3). The feasibility and correct flow of the experimental procedure was tested in a 16 participant pilot trial, prior to the full implementation. An overall view of the experimental flow is accessible in Appendix A.1

During the recruitment stage, participants were shown a study invitation on Prolific outlining the purpose of the study, eligibility criteria, nature of the task, estimated duration, and compensation provided. This ensured that participants had access to all relevant information before entering the website. The full invitation is provided in Appendix A.2.1. Once participants clicked on "Open study link in a new window", they were redirected to the experiment website and prompted to enter their prolific ID for identification and payment. Next, they had to accept the informed consent form adhering to standard guidelines (University of Oxford, 2025b). The consent form is provided in Appendix A.2.2.

Once participants had provided informed consent, they were instructed to answer three pre-training questionnaires. The first two assessed baseline outcomes for cognitive learning and affective learning through a knowledge quiz and the Sexual Consent Scale. The detailed theoretical foundations and structure of these questionnaires are detailed in Section 3.2.2. The full questionnaires are included in Appendix A.3.1 and Appendix A.3.2. Following these, participants completed a profiling questionnaire collecting information such as name, age, university affiliation, habits, and relationship to consent to personalize the chatbot training. While only used for the treatment, this step was administered to both groups to ensure consistency across conditions. The full context questionnaire is included in Appendix A.3.1 and detailed in Section 3.4.2.

Following completion of the three pre-training questionnaires, participants were randomly assigned to either the treatment or control group at a 1:1 ratio via a JavaScript's `Math.random()` function. Participants were only informed that they were taking a "digital sexual-consent training", not that two distinct formats existed, thus minimizing expectancy effects, which might have biased engagement or responses (Dutton et al., 2015). Both training formats were designed to deliver equivalent content and be of similar duration. The choice of the control e-learning training and the creation process of the chatbot training detailed in Sections 3.3 and 3.4, with full training materials provided in Appendix A.5 and A.4.

After completing their training, participants filled out a user experience questionnaire to assess their reactions to the session. The full questionnaire is provided in Appendix A.9, and its theoretical foundation and structure are detailed in Section **??**. To evaluate changes in attitudes and knowledge, participants then completed the same attitude and knowledge questionnaires used at baseline, with the addition of an open-ended question inviting them to reflect on how the training may have influenced their views or understanding. Both post-training questionnaires are included in Appendix A.6.2 and A.3.2.

Two-weeks after having done the initial evaluation, all participants that had taken part in the first leg of the study were invited for a follow-up questionnaire. The invitation on Prolific outlined this follow-up would take five minutes and be reimbursed 1£, which is a high compensation to the platform's standards. The invitation is accessible in Appendix A.7. After clicking "Open study link in a new window", participants were redirected to the exact same informed consent page, cognitive learning, and affective learning questionnaires that were defined earlier and in Appendix A.2.2, A.5, A.4.

3.2 Outcome Measures

The outcome measures for this study were defined using Kirkpatrick and Kirkpatrick (2006)'s four level evaluation framework extended with Kraiger et al. (1993)'s three measures of learning, defined in Section 2.5.1. While Kirkpatrick's framework outlines four levels, only the three first ones were evaluated, namely, reaction, learning and behavioral outcomes. Evaluating organizational outcomes would require a large-scale longitudinal study, which were beyond the time, budgetary, and data-access constraints of this thesis.

3.2.1 Reaction

The first level of Kirkpatrick's framework, reaction, was assessed using a shortened version of the user experience questionnaire (Laugwitz et al., 2008). The user experience questionnaire is a widely used instrument for evaluating users' immediate impressions of digital interfaces. Given its prior use in evaluating user reactions to educational interactions with large language models, it is particularly appropriate to this study's use case (Bolpagni & Gabrielli, 2025; Nasrul et al., 2024; te Pas et al., 2020). It captures both the usability and affective aspects of interfaces, such as efficiency, dependability, originality, and stimulation. These metrics align with (Kirkpatrick & Kirkpatrick, 2006)'s definition of reaction, which also distinguishes between affective responses and perceived utility. Each dimensions is measured on a 7-point semantic differential scale ranging between two opposite adjectives (e.g., "clear" vs. "confusing"; see Figure 3.1). However, the complete version includes 26 items with partial redundancies and some items less relevant to the context of sexual harassment training. To limit questionnaire fatigue (Ghafourifard, 2024) and focus on relevant measures, only a targeted subset of items were selected. These encompass clarity, boredom, security, usefulness, accessibility, and empathy. Several items were rephrased to better reflect the subject matter, for example, "not secure" was reframed as "triggering," and "friendly" was replaced with "empathetic".

Despite its widespread use, the User Experience Questionnaire has not been applied in evaluations of consent and sexual harassment e-learning trainings. Instead, researchers in this field tend to rely on more informal feedback methods, such as open-ended responses and ad hoc Likert-scale questions, often without a clear theoretical foundation (Heard et al., 2024; Preusser et al., 2011). To reflect these prevailing practices and ensure completeness, three additional elements were included in the reaction questionnaire. First, an optional open-ended prompt invited participants to share any "thoughts and suggestions" about the training. Second, given that 85% of adults familiar with AI report preferring any other source for factual information (Ofcom, 2025), a question included to assess participants' trust in the information delivered during the training. Finally, in providing a holistic measure of user experience, users were invited to rate the training on a 10-point scale, in line with Hock et al. (2021)'s methods.



Figure 3.1. User Experience Questionnaire.
Adapted User Experience Questionnaire on six user experience dimensions, on a 7 point likert-scale from one descriptive adjective to its opposite.

3.2.2 Learning

The second level of Kirkpatrick and Kirkpatrick (2006)'s framework, learning, was evaluated through two of Kraiger et al. (1993)'s three dimensions, namely, cognitive and affective learning. Cognitive learning was assessed through a knowledge test, while affective learning was measured using the Sexual Consent Scale. Due to online, time, and budgetary constraints, the third skill-based dimension of learning, usually evaluated through in-person drills or identification of harmful scenarios, was not included in this study (Blakely et al., 1998; Gidycz et al., 2011; Kleinsasser et al., 2015; Preusser et al., 2011).

Cognitive Learning

Cognitive learning is defined by (Kraiger et al., 1993) as the ability to recognize and recall terms, facts, and concepts after training. This has been evaluated in two distinct ways by researchers in the past, either with self-reported abilities or multiple choice tests (Bingham & Scherer, 2001; Campbell et al., 2013; Hock et al., 2021; Preusser et al., 2011). Self-reporting measures are less reliable than objective skill assessments due to social desirability bias, leading participants to overstate their abilities (Kruger & Dunning, 1999). Therefore cognitive learning was evaluated using a 10-question multiple choice knowledge test. This test was designed based on the exact content of the control e-learning module. Questions covered legal definitions, rape myths, specific facts, norms, impacts of rape, coping mechanisms, ways to give and identify consent, abusive relationships and intersectionality. They included both single- and multiple-answer options, clearly signposted with textual and visual cues. The phrasing and answers were reviewed and validated by experts in the consent training evaluation field.

Affective Learning

Kraiger et al. (1993)'s framework defines affective learning as the measure of both attitudinal and motivational outcomes, capturing trainees' beliefs and intentions with regards to the training content. Unfortunately, there exists no psychometric questionnaire to measure the outcomes outlined in this theory. However, these dimensions closely align with the constructs of attitude, norms and perceived behavioral control from Ajzen (1991)'s Theory of Planned Behavior, outlined in Section 2.4.1. Therefore, the measure of affective learning was done using the widely adopted Sexual Consent Scale–Revised, which is explicitly grounded in this theory, still ensuring strong theoretical alignment between the measure and the targeted learning outcomes (D'Eath et al., 2020; Humphreys & Brousseau, 2010; MacNeela et al., 2017). While several studies also employ the Illinois Rape Myth Acceptance Scale to assess affective learning, this instrument only captures attitudes toward sexual violence norms and has been critiqued as outdated (Gidycz et al., 2011; McMahon & Farmer, 2011; Paul & Gray, 2011; Payne et al., 1999). In contrast, the Sexual Consent Scale–Revised engages in a more holistic assessment targeting individuals' attitudes, beliefs, and perceived behaviors with respect to their understanding and communication of sexual consent. It comprises three core subscales: lack of perceived behavioral control, positive attitude toward establishing consent, and sexual consent norms. Each subscale consists of seven to eleven statements evaluated on a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The perceived behavioral control subscale captures internal or social discomfort around seeking or giving consent; the norms subscale reflects beliefs about when and in which contexts consent is considered necessary; and the attitudes subscale focuses on the importance, timing, and mutual responsibility of

verbal consent. To limit questionnaire fatigue (Ghafourifard, 2024), only the three items with the highest factor loadings, therefore most strongly correlated with their underlying concept, are included in this study's questionnaire. The chosen statements are provided in Appendix A.3.1.

3.2.3 Behavior

As noted in Section 2.5.1, while the ultimate goal of prevention programs is to produce lasting behavioral change, most evaluations fall short of measuring this outcome directly. This is largely because accurately measuring sexual assault and relationship violence requires longitudinal, ethically sensitive measurements with enough power to detect low-base-rate events (Lonsway, 1996; Schewe & O'Donohue, 1993). While such an approach lies beyond the scope of this study, multiple researchers have used more directly measurable outcomes that still offer meaningful insight into participants' behavioral change. Ring and Kilmartin (1992) have proposed that "expressed willingness to participate in a future acquaintance rape workshop" could serve as an indicator of change. They argued that willingness to do a time-costly, prosocial act moves beyond simple attitude endorsement toward concrete action intention. Therefore a single 7-point Likert-scale item, evaluating whether participants where "interested in doing an additional in-person training on this subject" was used to evaluate behavioral change. This measure far from ideal, however it offers a pragmatic, resource-efficient first indicator of behavioral change.

3.3 Control Training

To maximize external validity, the control condition was selected from among established e-learning consent trainings used in UK higher education. A landscape review of institutional programmes identified several leading providers. Most prominently CIPHR's E6 harassment and sexual misconduct suite, Epigeum's "Consent Matters", as well as Icena's "Sexual Violence at Universities" and Brook's "Consent for Students" programmes. These four trainings are used by over 80 UK and Irish universities each, including Oxford, King's College London, Durham, and Imperial (Brook, n.d.; Ciphr, n.d.; Epigeum, n.d.-a; Icena, n.d.). Selection criteria for the control module were public accessibility, a duration under 30 minutes, consistently positive institutional and student evaluations, and comprehensive thematic coverage rather than a narrow focus. The training that met all these criteria was the University of Edinburgh's "Consent on Campus" e-learning course, developed by CIPHR in collaboration with Rape Crisis England & Wales (Ciphr, 2025a, 2025b; University of Edinburgh, n.d.). This 30 minute module comprises 15 interactive screens with a short introduction video, text, drag-and-drop exercises, and embedded quizzes. It covers the legal definition of consent, practical strategies for requesting and communicating consent, rape culture and rape myths, consequences of rape, and recognizing abusive relationships. Its stated learning objectives are to learn what constitutes consent, how certain situations and contexts can affect consent, myth busting, and how to get consent.

3.4 Treatment Training

The treatment training consisted of a personalized conversation with an AI chatbot developped on the basis of the control e-learning training. Given the aim of this study

was to evaluate the sole effect of the training's delivery medium on outcome measures, the AI conversation was carefully matched to the control e-learning training's content and duration, ensuring minimal differences between the groups.

3.4.1 Model Selection and Evaluation

The first step in creating the training chatbot was finding an appropriate base model. To identify the most suitable large language model from the array of models available in 2025, candidates were evaluated across four criteria: instruction-following, knowledge, safety and latency. For benchmarks where the latest model version was not available, scores were imputed using the most recent version of the same model family.

Firstly, all models that were not available either as open-source systems or via public API were excluded from consideration. Additionally, to ensure a fluid and engaging conversation, models with an end-to-end response time exceeding 10 seconds, were also excluded. This threshold aligns with established human-computer interaction guidelines, which indicate that users' attention on dialogue deteriorates significantly when a system's feedback exceeds 10 seconds Nielsen (1994). Metrics for this exclusion criteria were sourced from the LLM API Providers Leaderboard published maintained by Artificial Analysis, which uniquely emphasizes LLM latency & throughput through live API testing, dimensions most academic leaderboards omit (Artificial Analysis, 2025). The shortlisted models were evaluated using an average measure of the normalized scores between the three remaining metrics.

The first criterion, instruction following, was essential for the chatbot to follow the given curriculum, discussion scenario, and personalize the learning experience. The most widely used and reliable metric to evaluate this is the Chatbot Arena Text Instruction Following leaderboard which ranks LLMs using blind A/B human voting with the Bradley–Terry ranking (Chiang et al., 2024; LMArena, 2025). Human-led leaderboards avoid the pitfall of static-benchmarks, often contaminated in LLMs' training data. The top performing model for this criteria was GPT-4o.

Second, the domain-specific knowledge capabilities of each model were assessed using performance scores from the Massive Multitask Language Understanding (MMLU) and MMLU-Pro benchmarks (Hendrycks et al., 2021; Wang et al., 2024). Due to its broad coverage of subjects, MMLU is one of the most widely adopted benchmarks for evaluating the general knowledge and reasoning capabilities of LMMs (Taghanaki et al., 2024). However, as state-of-the-art models began achieving near-ceiling performance on the original MMLU, researchers introduced MMLU-Pro, a more challenging variant incorporating reasoning-intensive questions. Given that the training aimed to correct misconceptions and convey complex concepts a strong underlying knowledge base was essential to ensure educational performance throughout the training. Domain-specific scores on psychology, human sexuality and moral scenarios from both benchmarks were used to evaluate the models' knowledge in areas directly relevant to the training task. When averaged over the three metrics the Claude 4 Sonnet performed best compared to other models.

Finally, given the sensitive and traumatic nature of sexual harassment, ensuring a safe and trustworthy conversational environment was essential. Therefore, overall model safety was assessed using the Phare LLM Benchmark (Giskard AI, 2025; Jeune et al., 2025), which provides multidimensional safety scores across four domains. : hallucina-

tion, harm, bias, and jailbreak resistance. Phare uses human annotators to judge whether model responses are factually grounded, avoid reinforcing harmful behaviors, treat individuals equitably across demographics, and resist adversarial prompts. Although it is not the most widely adopted measure due to its recent introduction, this independently developed benchmark offers a robust and holistic evaluation of all state-of-the-art models' safety features. The safest model on this metric was GPT-4o.

Following this comparative evaluation, the three highest-ranked models, GPT-4o, Claude 4 Sonnet, and Gemini 2.5 Flash, were selected for direct comparison given the exact prompt developped for the training (see Appendix B.1. Based on 20 tests each, Gemini 2.5 Flash and Claude 4 Sonnet's reasoning capabilities demonstrated better planning of the discussion's steps, stronger conceptual understanding, and more effective pacing in delivering pedagogical feedback. While their response latency was marginally higher, this did not adversely affect conversational flow. A slight preference for Gemini 2.5 Flash led it to be selected as the base model for the chatbot used in the treatment training.

Table 3.2. Top 5 model comparison
Benchmark results for the top 5 LLMs across instruction following, domain-specific knowledge, and conversational safety, used to inform selection of the chatbot's base model.

Model	LMArena Instruction	MMLU Human Sexuality	MMLU Moral Scenarios	MMLU-Pro Psychology	Phare Safety	Overall
GPT-4o	**1414**	0.9080	**0.9020**	0.8271	**0.7693**	**0.8772**
Claude 4 Sonnet	1404	**0.9390**	0.8880	**0.8459**	0.7573	0.8749
Gemini 2.5 Flash	1392	0.9010	0.8150	0.7905	0.7569	0.8535
DeepSeek R1	1389	0.9240	0.8080	0.7882	0.7149	0.8403
Qwen3	1381	0.8930	0.8150	0.7857	0.7020	0.8312
Llama 3.3 70B	1270	0.855	0.834	—	0.7049	0.8109
Grok 3	1396	—	—	0.8133	0.6138	0.7993

3.4.2 Prompt Engineering

With Gemini 2.5 Flash selected as the base model for the chatbot, the next stage involved designing the prompt that would guide its behavior throughout the training. Prompt engineering refers to the techniques used to create a set of instructions to define a model's behavior. In this case the prompt detailed the steps the model needed to follow to deliver the correct content in a way that was pedagogically sound, personalized, and emotionally sensitive.

However, before drafting the prompt itself, it was essential to first define how the model should ideally engage with the learner. To guide this interaction, nine teaching principles were drawn from foundational works in educational psychology. The first principle was connecting new information to prior knowledge and beliefs. Indeed, explicitly building upon learners' existing mental models has been shown to improve comprehension and retention, while ignoring them can result in superficial understanding or misconceptions (Bransford et al., 1999; Rosenshine, 2012). The second principle was to set clear learning goals, as "Teacher clarity" has one of the most substantial effects on student achievement, helping learners understand expectations (Hattie, 2008). The third principle was to use active retrieval and questioning. A principle grounded in the well-established "testing effect", which states that learners engaging in memory tests demonstrate significantly better long-term retention (RoedigerIII & Karpicke, 2006; Rosenshine, 2012). The fourth

principle was to provide immediate and targeted feedback, which has consistently been identified as one of the most powerful influences on learning outcomes (Hattie, 2008; Rosenshine, 2012). The fifth principle highlights the importance of linking content to learners' lived experiences, an approach identified by Bransford et al. (1999) as essential for effective teaching. The sixth principle was to demonstrate new concepts rather than imposing them as ground truth. This reflects a constructivist view of learning, where students actively build their own knowledge through guided exploration, rather than passively receiving information (Fosnot, 2013; Rosenshine, 2012). The seventh principle involves structuring content to reduce cognitive load. According to Cognitive Load Theory, the limited capacity of human working memory means that learning is more effective when information is broken down into smaller, organized segments (Rosenshine, 2012; Sweller, 2011). The eighth principle was to adapt instructions to meet diverse learners' abilities and learning styles. As Tomlinson (2014) argues, students vary widely in their readiness, interests, and learning profiles, and effective teaching must be responsive to these differences. The ninth and final principle was the creation of a safe and supportive learning environment. Especially when engaging with sensitive contexts like sexual harassment, building psychological safety and trust is crucial. Positive learning environments and strong teacher–learner relationships have been shown to produce measurable gains in classroom engagement and achievement (Bransford et al., 1999; Hattie, 2008).

These principles were implemented throughout every section of a structured JSON prompt, following best performing formatting (He et al., 2024). the prompt's structure was based on Hewing and Leinhos (2024)'s prompt canvas and the persona pattern from White et al. (2023)'s prompt catalog. These frameworks ensure a comprehensive and reproducible prompt design process that reflects best practices. The prompt canvas consists of 8 sections: Persona/Role, Task and Intent, Context, Output, Audience, Step-by-Step, References, Tonality.

The persona/role section of the prompt established the model's identity and expertise, aligning its perspective with the expected tone and behavior it must adopt in the conversation. Within the context of this study, the LLM was prompted to emulate the behavior of a "skilled facilitator in sexual consent and harassment prevention education, with deep experience working with diverse student populations in UK universities".

The task and intent section was used to clearly define the overarching purpose of the model's response. Within this prompt, the objective was defined as "deliver[ing] a short sexual consent training [...] socratic method conversation [...] tailored to a [...] university student". Mentioning the Socratic method, which uses guided questioning to build learner understanding, and content personalization, addressed both principles five and six. Following principle two, the LLM is also prompted to teach a list of learning goals matching each section of the control e-learning training and to address the list of ten facts present in the knowledge test.

In line with educational principle six, identifying the audience in the prompt ensured that the model provided content tailored to the participant's profile. Each participant's profile was established using a pre-training questionnaire. This questionnaire collects information on preferred name, age, gender, university, typical student routine, primary social environment, prior exposure to consent training, knowledge level, interest and comfort talking about sexuality. This information allowed the model to address users person-

ally, reference their university, tailor scenarios to their habits, and adapt the difficulty or phrasing of questions based on their prior knowledge and comfort level. The complete questionnaire is provided in Appendix A.6.

The reference section served to anchor the model's output in relevant sources and reduce hallucinations. In this context, the main source of reference for the LLM was the entire textual content from the control e-learning training, which can be found in Appendix A.5

The step-by-step section was the heart of the conversation's flow. It used precise instructions detailing each part of the training's implementation, breaking down the task into manageable sections, therefore reducing reasoning errors and improving output coherence. Within the prompt, the conversation scenario was divided into 3 sections: introduction, modules, and conclusion. The introduction started by welcoming the learner, explaining the five modules of the training and asking if they wished to learn more from a specific domain. It then asked about potential triggers, learning format preferences and learning difficulties. This allowed further personalization of the training, while creation trust and safety within the conversation, in line with principles eight and nine. Next, the model was instructed to teach the content of the training, separated in five structured modules encompassing foundation of consent, legal context, culture around consent, impacts & recovery from rape, and abusive relationships. Each module was addressed using the same format. First, the model explicitly introduced the upcoming subject, learning goals and provides a warning of any triggers. It then opened the conversation by asking the user a question engaging on their prior conceptions of the subject. Then the model engaged in back and forth questioning with the user until all the learning goals of the module had been reached. The model had the choice between scenarios based questions, quizzes, reflective questions, scripted dialogues based on the user's context and learning preferences. After each answer from the learner, the model was prompted to give it feedback, emphasizing rights and wrongs. Once the module was done, the model provided an overall view of what the user has learned and moved to the next module. Once every module had been addressed, the model concluded the training by "Summarizing the overall learning journey", highlighting learning outcomes and mentioning areas for further growth. It then provided the same list of resources contained in the control e-learning training. This section carefully implements all nine teaching principles in its instructions.

The section of the prompt on tonality steered the model's communicative style to align with the task's purpose. In the this case, a set of conversation principles were outlined, including speaking like a friendly peer, using short messages, speaking empathetically, adjusting its pace to the learner, being inclusive, staying engaging, challenge problematic views, and reinforcing progress. This strongly emphasizes principles eight and nine.

In addition to the sections defined by Hewing and Leinhos (2024), rules and atypical cases sections are implemented, following Vogel (2024)'s AUTOMAT prompt framework. The rules define strict limitations on the conversation's length and forced the model to limit hallucinations. The atypical cases defined the model's behavior in case the learner tried to go off topic, shortcut the training, change language, or shows discomfort. Finally, the prompt also contained a troubleshooting and interface section providing the model with information on how to help the user exit the chat when encountering a problem and navigate the interface.

3.4.3 Retrieval Augmented Generation

While prompt engineering structured the model's behavior, the chatbot still lacked up-to-date information beyond Gemini Flash 2.5's pretraining data. To address this limitation, a retrieval-augmented generation (RAG) system was implemented within the model. First introduced by (Lewis et al., 2021), RAG is a system that dynamically retrieves relevant text passages from a corpus of embedded documents based on a user input and incorporates them into the model's response to produce contextually grounded outputs. This limits model hallucinations and allows students to ask specific questions about statistics, university specific context, reporting resources and other related content. To implement a RAG database of relevant documents was first build for later retrieval. The creation of a comprehensive corpus of informative and reliable documents on sexual harassment and consent, was done through thorough web research of NGO and governmental data. This research led to the aggregation of a corpus of 107 consent trainings, educator guides, fact sheets, legislative documents, reports, and welfare student resources. Sources and descriptions can be found in Appendix 3.4.3. Turning these documents into a RAG-ready database required to separate them into overlapping chunks containing 400 tokens each and embedded each chunk using gemini-embedings-001. This embedding model was chosen as it has been optimized for downstream performance with Gemini and best reflects its internal representation of meaning (Google, 2025). It also is the third highest performing embedding model on the MTEB leaderboard, which systematically benchmarks multilingual and general-purpose embedding models across a wide range of real-world retrieval, classification, clustering, and reranking tasks, making it one of the most widely adopted embedding evaluation metrics (Muennighoff et al., 2023; Reimers et al., 2025). Once embedded, text chunks were stored in a database for later retrieval. The retrieval process used cosine similarity between embeddings and user prompt with a threshold of 0.85 and a maximum of 5 chunks to maintain precision and limit cognitive overload.

3.4.4 Interface

Participants interacted with the consent training chatbot through a simple online interface. Its main element was the text input bar for sending messages and a chat window displaying the conversation history. At the top of the page, a progress bar advanced with each completed module, offering users a clear sense of progression and maintaining engagement. Once it reached 100%, a "Finish Chat" button was activated, allowing users to proceed to the next stage of the experiment. On the right-hand of the screen, a menu provided three key tools: a Notes tab for saving reflections and important messages, a Resources tab offering quick access to the same resources featured in the control e-learning training, and a Reporting tab linking directly to the UK Rape Crisis reporting page for immediate support.

3.5 Data Analysis

Data analysis followed two distinct approaches, depending on how many times each outcome was measured. For reaction and behavior outcomes measured exclusively post-intervention, between-group comparisons were conducted using independent-sample t-tests on scores. For both cognitive and affective learning outcomes which had repeated measurements, four one-way analyses of covariance (ANCOVA) were performed using the type of training as the independent variable (e-learning or chatbot), post-training

or follow-up scores (cognitive and affective learning) as the dependent variable, and pre-training scores as the covariate. The control variables included were gender, age, time to complete training, ethnicicty, and previous training. An interaction term between treatment and gender was included to assess whether the effect of the training differed by gender. ANCOVA tests for differences between groups while adjusting for baseline values. Based on the intention-to-treat principle, all participants, whether they had completed the experiment or not were included in the analysis (European Medicines Agency, 1997; Gupta, 2011). To define whether data was missing at random (MAR) or missing completely at random (MCAR), Little (1988)'s test was performed on outcome variables and rejected the MCAR assumption, data was therefore assumed MAR. Following common practice for MAR data, missing values were imputed using multiple imputation (White et al., 2011). Multiple imputation estimates missing data by creating several different plausible datasets, analysing each one separately, and then combining the results to account for uncertainty due to missingness (Royston, 2004).

Chapter 4

Results

4.1 Reactions

User reactions to both trainings were generally very positive. However, none of the metrics used to differentiate user experience or trust in the training, showed any significant difference between groups. Indeed, as can be observed in Figure 4.1, both trainings are very similar in score across questions. It must still be noted that the chatbot training was very slightly outperformed by the e-learning training on each metric by 0.02 to 0.30 points on the 7-point scale. The overall user experience was rated 8.84/10 for the e-learning training and 8.70/10 for the chatbot training ($\Delta M = -0.136$, 95% CI [−0.632, 0.361]).

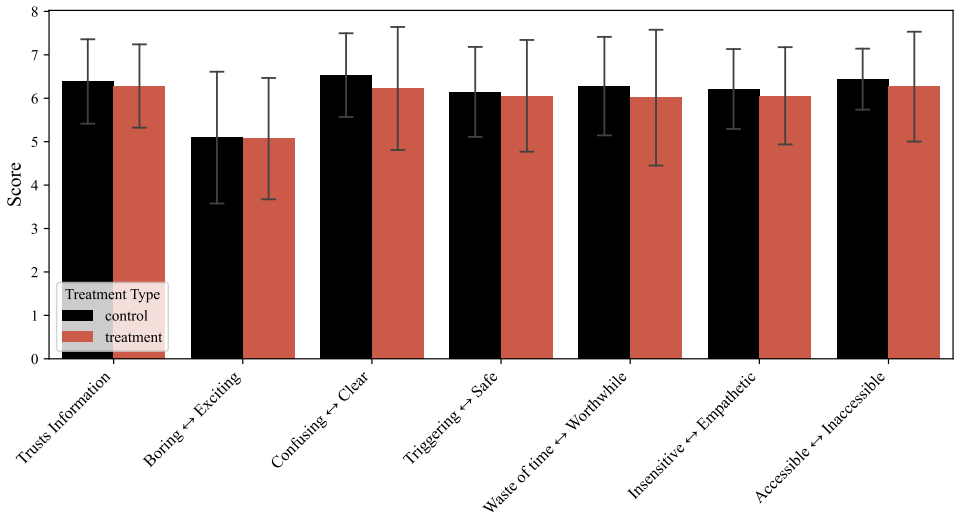


Figure 4.1. User Reactions to the Control and Treatment Trainings.
Mean user ratings across five experience dimensions and perceived trust in the training content, separated by treatment group. Error bars indicate 95% confidence intervals. Scores are based on 7-point semantic differentials, with higher values reflecting more positive reactions.

Given the limited insight from statistical tools to evaluate participant reactions, an analysis of user comments helped deepen these results. Of the 146 participants that took the user experience questionnaires, there was a slightly higher engagement in terms of

commenting for the treatment group, with 44 participants from the control group (59%) and 48 participants from the treatment group (68%) leaving a comment. As summarised in Table 4.1, both formats were generally well received. The control group more frequently praised the training's clarity (27%), engaging nature (14%), and interactivity (18%). By contrast, treatment comments highlighted the training as informative or helpful (33%), very good or useful (31%), and personalized (21%). Empathy and emotional tone were also more salient in the treatment group (10% vs. 2% in control). Both groups requested multimedia formats and more concrete examples, although these suggestions were slightly more frequent in the treatment group. Notably, over 10% of treatment participants explicitly stated there was nothing to improve. One very interesting comment from a female participant in the treatment group mentioned:

> *"I also appreciated the opportunity to offload some of my own personal experiences which was carthartic and therapeutic for me - so thank you for that."*

showing that students felt safe enough to disclose personal experiences to the chatbot. Negative comments were rare across both groups. In the control group, a few participants found the training "a little boring," "patronizing," or "common sense," and one commented it could "make the average person even more paranoid." In the treatment group, the only recurrent concern was about "unclear answers" or the chatbot "dismissing comments."

Table 4.1. Free-text comment analysis

Table of the user experience comments post-training, including theme counts, percentages, and illustrative quotes from the control and treatment conditions.

Theme	Group	Count	%	Example comment
Clarity	Control	12	27.3%	*The information was presented well and clear.*
	Treatment	6	12.5%	*I thought it was great and clear training.*
Informative / Helpful	Control	12	27.3%	*I like it a lot and think it's very informative.*
	Treatment	16	33.3%	*I found it really informative and insightful.*
Good / Great / Useful	Control	10	22.7%	*It was a great learning experience.*
	Treatment	15	31.3%	*I thought the AI was really good and it felt like a proper conversation.*
Interactive	Control	8	18.2%	*I liked how interactive this was and I thought it was great.*
	Treatment	8	16.7%	*I love the interactive approach of learning it used.*
Engaging	Control	6	13.6%	*It refreshed my memory of consent in an engaging way.*
	Treatment	2	4.2%	*I appreciated how clearly the training defined terms.*
Personal / Tailored	Control	1	2.3%	*Focus more on how to tell if you have consent.*
	Treatment	10	20.8%	*Felt personalised to my situation.*
Empathetic	Control	1	2.3%	*Training felt supportive and respectful.*
	Treatment	5	10.4%	*I think that it was very empathetic.*
Nothing to improve	Control	3	6.8%	*Nothing that I can recommend.*
	Treatment	5	10.4%	*Nothing needs to be changed.*
Different media wanted	Control	5	11.4%	*Add some video instead of all text.*
	Treatment	3	6.3%	*It would be great to have some videos that showed role played scenarios.*
More examples / stats	Control	4	9.1%	*Give more real examples and data.*
	Treatment	8	16.7%	*It can give more examples about the situations.*

Note. Percentages reflect the proportion of participants who left comments (Control $n = 44$, Treatment $n = 48$). Quotes are excerpts from open-text responses, lightly edited for brevity and clarity.

4.2 Cognitive Learning

Cognitive learning was evaluated using a 10-question knowledge test administered at pre-, post-, and follow-up timepoints. At baseline, knowledge scores were slightly higher in the control group ($M = 4.07$, 95% CI [3.14, 4.04]) than in the treatment group ($M = 3.59$, 95% CI [3.64, 4.49]), but this difference was not statistically significant ($p = .12$), indicating successful randomization. Two ANCOVA analyses were conducted to compare post-training and follow-up knowledge scores between groups, controlling for baseline performance. The analysis for the homogeneity-of-slopes assumption was validated by the indication that the relationship between baseline and outcome scores did not differ by training group, with $F(1, 142) = 0.06$ ($p = .805$) for post-training and $F(1, 142) = 0.41$ ($p = .521$) for follow-up. Controlling for baseline knowledge, measurement time, ethnicity, age, sex, and previous training, the intervention group scored 0.69 points higher on the 10-item post-test than the control group ($\beta = 0.69$, 95 % CI [0.05, 1.34]). This difference was statistically significant, $F(1, 135) = 4.55$, $p = .035$, partial$\eta^2 = .03$, indicating a small effect. When the same ANCOVA was run for the follow-up scores, the between-group difference ($\beta = 0.36$, 95 % CI [–0.30, 1.02]) was not significant, $F(1, 135) = 1.17$, $p = .282$, partial $\eta^2 = .01$, indicating that the initial gains in knowledge between groups were not maintained over time. Full result of the ANCOVA are outlined in Appendix C.1. As sexual violence is a highly gendered issue, interrogating differences between men and women is essential. However, although male participants in the treatment group scored approximately 0.51 points lower than female participants in the same group, the interaction term between treatment and sex was not statistically significant ($\beta = -0.51$, $p = .346$ at post-test; $\beta = -0.13$, $p = .810$ at follow-up), indicating insufficient statistical evidence for differential treatment effects by gender (see Figures C.1c and C.1d).

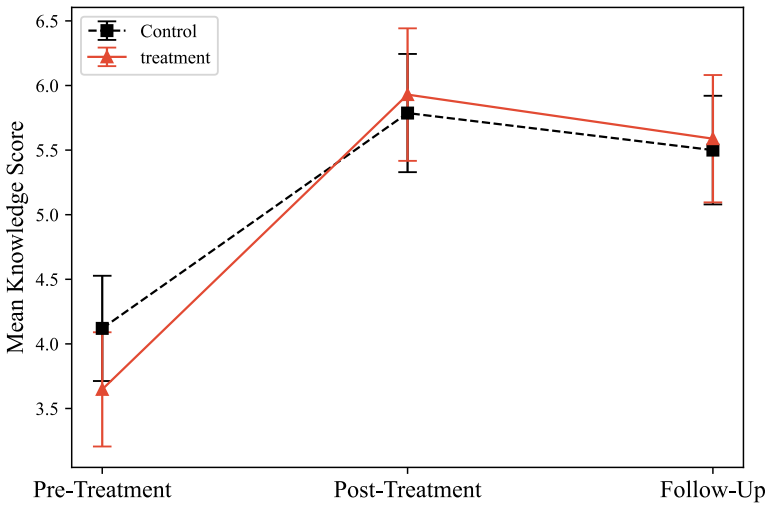


Figure 4.2. Knowledge Test Scores by Timepoint and Training Group.
Mean knowledge scores for control and treatment groups at pre-training, post-training, and follow-up. Error bars represent 95% confidence intervals.

4.3 Affective Learning

Affective learning was evaluated using the Sexual Consent Scale–Revised administered at pre-, post-, and follow-up timepoints. At baseline, scores were highly similar between the two groups, indicating successful randomization. The control group reported a mean score of $M = 4.94$ (95% CI [4.61, 5.27]), while the treatment group mean was $M = 4.97$ (95% CI [4.70, 5.24]), on a 7-point Likert scale. Again, two ANCOVA analyses were used to test for differences between groups controlling for pretest baselines. The homogeneity-of-slopes assumption was met, with non-significant interaction terms between baseline scores and treatment group for both post-training, $F(1, 142) = 0.04$, $p = .849$, and follow-up attitudes, $F(1, 142) = 0.49$, $p = .486$. While both trainings significantly increased affective scores from pre- to post-test ($p < .001$), ANCOVA analyses controlling for baseline attitudes, demographic variables, and time found no significant difference between groups at either post-training, $F(1, 135) = 1.41$, $p = .162$, or follow-up, $F(1, 135) = 0.17$, $p = .679$. Full results of the ANCOVA are outlined in Appendix C.2. Interestingly, the decoline in score between post-test and follow-up for the treatment was noticeably smaller in the treatment group ($\Delta M = -0.09$) than in the control group ($\Delta M = -0.24$); however, this difference was not statistically significant ($p = .32$). Interaction terms between treatment and gender were also non-significant at both post-training ($\beta = -0.20$, $p = .410$) and follow-up ($\beta = -0.36$, $p = .208$), indicating no evidence of differential treatment effects by gender (seeFigures C.1b and C.1a).

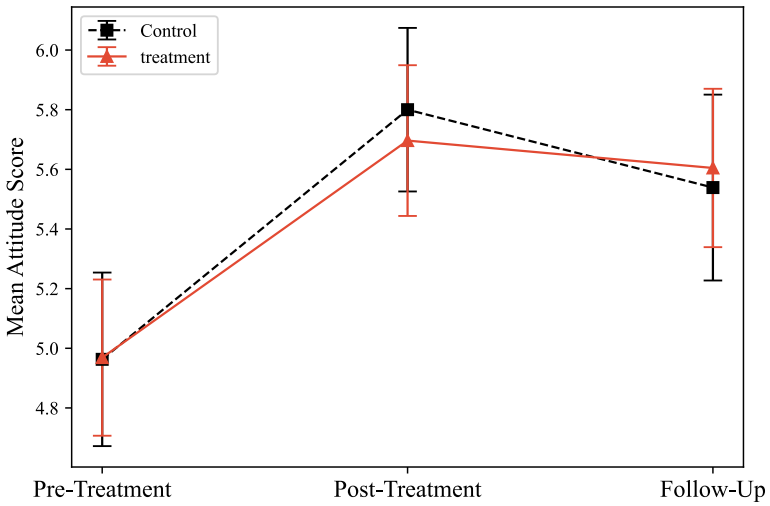


Figure 4.3. Attitude Questionnaire Scores by Timepoint and Training Group.
Mean affective learning scores for control and treatment groups at pre-training, post-training, and follow-up. Error bars represent 95% confidence intervals.

4.4 Behavior

Participants who received the AI-based consent training reported a greater interest in pursuing additional consent education compared to those in the control group. On a 7-point scale, the treatment group rated this item at $M = 5.37$, while the control group

rated it at $M = 4.84$, with a mean difference of 0.53 (95% CI [−0.06, 1.11]) and a small effect size (Cohen's $d = 0.29$). This difference approached statistical significance, $t(144) = –1.77$, $p = .079$.

Chapter 5

Discussion

5.1 Principal Findings

Taken together, the results offer a nuanced verdict on the four pre-registered hypotheses. The first hypothesis stating that the chatbot would outperform the e-learning module on user experience was not confirmed, with both formats rated almost identically (8.8/10). However, the control training reflected past flaws with descriptions stating it was commonsense and patronizing similarly to Htun et al. (2022), while the treatment showed more positive comments overall. This shows that conversational AI is acceptable but may not be intrinsically more engaging than a well-designed interactive course. The second hypothesis was partially supported, as the chatbot delivered a small, statistically significant knowledge advantage ($\eta^2 = 0.03$) over the e-learning training. However, this difference was no longer significant after two weeks, indicating that a chatbot educator boosts initial knowledge gain but decreases over time similarly to traditional e-learning trainings. The third hypothesis predicting greater attitudinal change was not supported. This outcome was similarly positively impacted in both arms, with no significant difference between groups. However, non-significant results indicated towards potentially higher long-term retention for the control group. Finally, the fourth hypothesis received partial support. Willingness to pursue further training rose modestly in the chatbot group, hinting that personalized closure can nudge behavioral intention, though objective follow-through remains to be demonstrated. In sum, the LLM-based dialogue demonstrated promising potential but its advantages over traditional formats remain limited, and future studies with larger samples are needed to clarify its long-term impact and practical value.

5.2 Limitations

While the results suggest that conversational AI holds promise for delivering effective consent and sexual harassment training in UK universities, several important limitations must be acknowledged. These include methodological constraints, intervention design challenges, and ethical considerations.

5.2.1 Methodological Limitations

The study's two main methodological limitations concern the use of Prolific for sampling and delivery, and weaknesses in outcome measurement. Participants were recruited via Prolific, allowing fast and diverse sampling but potentially reducing ecological validity,

as users tend to be older, more motivated, and more familiar with digital platforms than typical UK university students (Douglas et al., 2023; Prolific, 2025a). In this sample, participants were on average 22% older than the UK undergraduate population, which was statistically accounted for in the analyses but still reflects a key sampling difference. Moreover, unlike university students who are often forced to complete consent trainings to fulfill institutional requirements, Prolific participants choose the studies they enroll for and are directly compensated. Prolific also operates a reputation system in which participants' approval ratings determine access to future studies. Since low-effort or inattentive responses are often rejected, participants are strongly incentivized to remain engaged and attentive throughout the task (Prolific, 2025b). Their motivation is therefore shaped more by financial reward, task interest and response quality than by institutional compliance. These differences may have inflated learning outcomes, limiting the generalizability of findings to real student training processes in university settings. Future research should replicate this study within actual university populations, ideally through cluster-randomized designs embedded in onboarding procedures, as implemented in prior evaluations of campus-based consent interventions (Gidycz et al., 2011; Potter et al., 2016).

This study relied on relatively short and simplified instruments to assess its outcomes, which may limit the reliability of the results. The knowledge test, comprising 10 binary-scored items, was uniquely developed for this study based on the control training's content. However, it does not fully reflect the entire breadth of material covered. Measures were based on specific facts, allowing participants to directly reference information they had already encountered during the pre-training questionnaire. While the questions were validated in consultation of multiple researchers in the consent training evaluation field and tested in a pilot trial, they are not based on any past literature. This operationalization might not entirely relate to its intended theoretical construct. Future iterations should use item-response theory to refine question difficulty and enhance discrimination between knowledge levels (Reise & Waller, 2009). The behavioral outcome was measured via a single self-reported intention to pursue additional training within the user experience questionnaire. While relevant, this item faintly captures behavior. It is highly vulnerable to social desirability and motivational priming. It could be better implemented as a secondary unrelated study shown to participants simply prompting them to sign up for additional training. More robust approaches could incorporate alternative behavioral measures used in previous studies, such as seemingly unrelated follow-up phone calls, or self-reports of victimization and sexually aggressive behavior (Breitenbecher, 2000; Gidycz et al., 2011). While attitude measures were drawn from a validated scale, they were shortened to improve completion rates, potentially compromising construct validity. A full psychometric evaluation was not conducted but would strengthen interpretation. Finally, there was no measure of organizational outcomes in this study which fails to address Kirkpatrick and Kirkpatrick (2006)'s fourth level of evaluation. While "accurately measuring reductions in the incidence of rape is a monumental, long-term, and possibly unattainable goal" Lonsway (1996), this could ideally be addressed by implementing a large scale university implementation and measuring sexual assault reporting rates (Boba & Lilley, 2009). Additionally, the follow-up measure was administered relatively soon after the training. A six-month delay, as used in other studies, would have offered a more rigorous assessment (Breitenbecher, 2000; DeGue et al., 2014; Lonsway, 1996). Taken together, these limitations suggest the need for more comprehensive, psychometrically sound instruments to accurately detect and explain short- and long-term changes across

the four levels of Kirkpatrick's evaluation framework.

5.2.2 Intervention Limitations

Beyond methodological constraints, several limitations relate directly to the design and delivery of the intervention itself. A key limitation of the AI training was its occasional failure to cover every fact assessed in the knowledge test. While the chatbot followed the core curriculum, it did not always surface each one of the 10 test facts explicitly. Textual analysis of messages revealed fact.3.1, fact.3.3 and fact.3.4 were not mentioned to 2, 5 and 10 separate users, potentially impacting their knowledge scores. Participant comments reflect this gap, with a user noting that "the test [did] not seem fully accessible yet". To address this, future versions of the chatbot could build upon past multi-agent tutoring architectures, where a monitoring module tracks fact coverage in real time and triggers targeted micro-reviews when they remain unaddressed (Chen et al., 2024).

A second limitation concerns conversational rigidity. To ensure internal validity, both the chatbot and e-learning trainings were matched in content and duration, requiring the chatbot to deliver a substantial amount of information within a fixed timeframe. However, this strict scenario restricted users' ability to ask personal questions and deviate from the preset path, which led several participants to express frustration. One participant noted:

> *"I asked an off-piste question and felt the response back was a bit discouraging for asking other questions? We moved swiftly back to their training progression and actually a bit more of an open forum to ask questions would've been nice."*

Another reflected:

> *"Make it more empathetic. I almost confided in it but then it ignored my personal experiences and that's what made me snap back and realise this was a training session, not therapy."*

To address this, future iterations of the chatbot could integrate real-time intent detection and instruction-aware planning, enabling the system to flexibly respond to user input while maintaining pedagogical coverage (Chae et al., 2023; Cutler et al., 2025). Additionally, addressing the user suggestions on using various modalities, such as audio, video and images and making the training longer, could be a valuable addition to further experimentation of this chatbot.

5.2.3 Ethical Limitations

Despite its potential, deploying conversational AI in consent training raises critical ethical concerns. First, hallucinations remain a high risk, especially when users seek accurate legal or support-related advice (Pierre, 2025). While no hallucinations were observed in the current implementation, the safeguards in place were limited to prompt instructions and RAG, which might be insufficient in future context. Model improvements could include integrating atomic fact-checking for information verification directly into the response pipeline (Vladika et al., 2025). A second concern is the development of emotional bonds with agents due to their human-like design, which can foster "therapeutic misconceptions", leading users to overestimate the chatbot's therapeutic capabilities (Khawaja & Bélisle-Pipon, 2023). Current solutions to address this risk require the system to repeatedly disclose that it is non-human and not a substitute for professional care (Khawaja & Bélisle-Pipon, 2023). Additionally, this misplaced trust may reinforce

harmful behaviors, especially given documented cases of toxic emotional mirroring by AI companions, and poor responses to psychological distress such as abuse or suicidal ideation (Chu et al., 2025; Verma, 2024). The use of affect moderation filters to de-escalate toxic affective loops, along with fine-tuned safety classifiers to provide tailored responses to high-risk disclosures, has been recommended to mitigate these risks (Sherekar & Mehta, 2025; Verma, 2024). Finally, models may disproportionately harm marginalized groups through biased responses, which can be addressed by incorporating counterfactual data augmentation using diversity-aware training datasets (Brown & Halpern, 2021; Zmigrod et al., 2019).

5.3 Implications

AI-based training represents a scalable, cost-effective complement to existing university resources for sexual harassment education and reporting. Given the ethical constraints outlined above, such systems should not replace human-led instruction. However, they could be implemented within a blended learning model, combining both digital and face-to-face instruction, which has been shown to improve both learner engagement and academic performance (Hamad et al., 2024; Park & Doo, 2024). The anonymity, easy access, and low interpersonal pressure offered by AI chatbots are key factors associated to facilitated disclosure of sexual violence victimization (Gorissen, 2024). Lowering barriers to reporting is essential for universities to comply to OfS E6 requirements for "accessible reporting pathways" (Office for Students, 2024). Moreover, deploying an AI chatbot as a first-line support tool can help reduce vicarious trauma and mental strain on human facilitators (Indiana University, 2020; Nikischer, 2019). The added benefit of implementing a chatbot comes at a very low cost for universities which usually pay between 10-100£ per student for e-learning or in-person trainings (Bodyswaps, 2025; SC Training, 2025; Sexual Harassment Training com, n.d.). Indeed, current cost per participant when interacting with the chatbot was around 0.15£. AI delivery could therefore also help smaller institutions lacking budget and specialist staff meet statutory duties. This makes chatbot-based training not only pedagogically valuable, but also highly scalable. Beyond its practical applications, this study also contributes to the theoretical understanding of how large language models can facilitate learning in sensitive, compliance-driven contexts. It provides comprehensive guidelines for developing a chatbot educator that matches current industry performance, along with a replicable methodology for evaluating its effectiveness. As noted above, the current model shows promise but requires further refinement and validation in a larger study within university context to confirm its effectiveness.

Chapter 6

Conclusion

This study examined the effectiveness of AI as a tool for delivering sexual consent and harassment education at UK universities. Using a randomized controlled trial involving 150 participants, it compared traditional e-learning modules with a custom AI chatbot based on a large language model. Outcomes assessed through pre-training, post-training, and one-week follow-ups included user reactions, cognitive learning, affective learning, and behavioural intentions. The findings revealed that both formats' experiences were equally well-received overall, with the chatbot fostering a greater number of opiniated comments. The chatbot significantly outperformed the e-learning training in knowledge retention and showed a modest yet near-significant improvement in students' intentions to pursue further training. Attitudinal shifts were comparable across both training methods.

This study contributes to the field in two key ways. First, results demonstrate that an LLM-driven intervention can rival and, on some metrics, exceed an established digital course, expanding the proven toolkit for harassment-prevention education. Second, it offers a scalable, low-cost blueprint that universities can adopt quickly to help meet the OfS E6 duty for mandatory consent training.

Sexual harassment is however a highly sensitive subject and these findings should also be read through a strict ethical lens. Effective guardrails are still needed to curb hallucinations, prevent "therapeutic misconceptions," and protect marginalised students from biased or unsafe responses. Accordingly, while the chatbot showed promise as a first-line educational tool, it does not and should not replace the nuance and duty-of-care provided by trained facilitators.

In this context, future research should focus on running longer-horizon trials, test the chatbot in real university roll-outs, and refine the AI's overall output quality. With rigorous safeguards and genuine institutional commitment, conversational AI can evolve from promising prototype to catalyst for safer, more respectful campuses. The task now is to ensure the technology takes a meaningful place in the collective fight against sexual violence.

References

Abell, L., Buglass, S. L., Betts, L. R., & Morohunfola, T. (2024). "It's all part of the culture": Undergraduate students' experiences of banter in university sports clubs and societies. *Current Psychology*, *43*(10), 8727–8739. https://doi.org/10.1007/s12144-023-05026-w

Abrams, Z. (2025). Using generic AI chatbots for mental health support: A dangerous trend. Retrieved July 30, 2025, from https://www.apaservices.org/practice/business/technology/artificial-intelligence-chatbots-therapists

Acquaviva, B. L., Hayes, B. E., & Clevenger, S. L. (2023). College Students' and Alumnus' Knowledge and Perceptions of a Mandatory Online Sexual Assault Training. *Journal of Criminal Justice Education*, *34*(2), 269–288. https://doi.org/10.1080/10511253.2022.2132268

Advance HE. (2025). E-Learning Courses — Advance HE. Retrieved July 23, 2025, from https://advance-he.ac.uk/e-learning-courses

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Amnesty International. (2020, February). Let's Talk About Yes. Retrieved August 1, 2025, from https://www.amnesty.org/en/latest/campaigns/2020/02/letstalkaboutyes/

Anderson, L. A., & Whiston, S. C. (2005). Sexual Assault Education Programs: A Meta-Analytic Examination of Their Effectiveness. *Psychology of Women Quarterly*, *29*(4), 374–388. https://doi.org/10.1111/j.1471-6402.2005.00237.x

Antecol, H., & Cobb-clark, D. (2003). Does Sexual Harassment Training Change Attitudes? A View from the Federal Level. *Social Science Quarterly*, *84*(4), 826–842. https://doi.org/10.1046/j.0038-4941.2003.08404001.x

ARCH Teesside. (n.d.). Self Help Book. Retrieved August 1, 2025, from https://archteesside.org/

Arday, J., & Mirza, H. S. (Eds.). (2018). *Dismantling Race in Higher Education: Racism, Whiteness and Decolonising the Academy*. Springer International Publishing. https://doi.org/10.1007/978-3-319-60261-5

Artificial Analysis. (2025). LLM API Providers Leaderboard. Retrieved July 26, 2025, from https://artificialanalysis.ai/leaderboards/providers

Association of Title IX Administrators. (2025). Understanding Consent Workshop. Retrieved August 1, 2025, from https://www.atixa.org/training/understanding-consent-workshop/

Bai, H., Voelkel, J., Eichstaedt, J., & Willer, R. (2023). Artificial Intelligence Can Persuade Humans on Political Issues. https://doi.org/10.21203/rs.3.rs-3238396/v1

Bandura, A. (2006). Mechanisms of moral disengagement. In *Insurgent Terrorism*. Routledge.

Barradale, G. (2021, May). We asked 4,000 students about sexual assault on campus. These are the shocking results. Retrieved July 28, 2025, from https://archive.

thetab.com/uk/2021/05/04/we-asked-4000-students-about-sexual-assault-on-campus-these-are-the-shocking-results-202971

Bednarchik, L., Generous, M., & Mongeau, P. (2024). Using the Theory of Planned Behavior to Longitudinally Predict College Students' Communication of Affirmative Sexual Consent. *Western Journal of Communication*, *88*(5), 1068–1087. https://doi.org/10.1080/10570314.2023.2251007

Berg, D. R., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape prevention education for men: The effectiveness of empathy-induction techniques. *Journal of College Student Development*, *40*(3), 219–234.

Berkowitz, A. D. (2003, February). Applications of social norms theory to other health and social justice issues. In *A Handbook for Educators, Counselors, and Clinicians* (1st ed., pp. 259–279). John Wiley & Sons.

Bingham, S. G., & Scherer, L. L. (2001). The Unexpected Effects of a Sexual Harassment Educational Program. *The Journal of Applied Behavioral Science*, *37*(2), 125–153. https://doi.org/10.1177/0021886301372001

Blakely, G. L., Blakely, E. H., & Moorman, R. H. (1998). The Effects of Training on Perceptions of Sexual Harassment Allegations. *Journal of Applied Social Psychology*, *28*(1), 71–83. https://doi.org/10.1111/j.1559-1816.1998.tb01654.x

Boba, R., & Lilley, D. (2009). Violence Against Women Act (VAWA) Funding: A Nationwide Assessment of Effects on Rape and Assault. *Violence Against Women*, *15*(2), 168–185. https://doi.org/10.1177/1077801208329146

Bodyswaps. (2025). Soft skills training software Pricing. Retrieved August 1, 2025, from https://bodyswaps.co/pricing

Bolpagni, M., & Gabrielli, S. (2025). Development of a Comprehensive Evaluation Scale for LLM-Powered Counseling Chatbots (CES-LCC) Using the eDelphi Method. *Informatics*, *12*(1), 33. https://doi.org/10.3390/informatics12010033

Bondestam, F., & Lundqvist, M. (2020). Sexual harassment in higher education – a systematic review. *European Journal of Higher Education*, *10*(4), 397–419. https://doi.org/10.1080/21568235.2020.1729833

Bransford, J., Brown, A., & Cocking, R. (1999). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. National Academies Press. https://doi.org/10.17226/9853

Brecklin, L. R., & Forde, D. R. (2001). A meta-analysis of rape education programs. *Violence and Victims*, *16*(3), 303–321. https://doi.org/https://doi.org/10.1891/0886-6708.16.3.303

Breitenbecher, K. H. (2000). Sexual assault on college campuses: Is an ounce of prevention enough? *Applied and Preventive Psychology*, *9*(1), 23–52. https://doi.org/10.1016/S0962-1849(05)80036-8

Brook. (2018). *Sexual Harassment at UK Universities* (tech. rep.). Brook. https://www.brook.org.uk/wp-content/uploads/2025/02/Brook_DigIN-summary-report.pdf

Brook. (2025). Educators: Consent. Relationships & Sexual Education (RSE) Course. Retrieved August 1, 2025, from https://www.brook.org.uk/product/consent-course/

Brook. (n.d.). Educators: Consent. Relationships & Sexual Education (RSE) Course. Retrieved July 24, 2025, from https://www.brook.org.uk/product/consent-course/

Brown, J. E. H., & Halpern, J. (2021). AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. *SSM - Mental Health*, *1*, 100017. https://doi.org/10.1016/j.ssmmh.2021.100017

Brown, K. G. (2001). Using Computers to Deliver Training: Which Employees Learn and Why? *Personnel Psychology*, *54*(2), 271–296. https://doi.org/10.1111/j.1744-6570.2001.tb00093.x

Bull, A., & Page, T. (2021). Students' accounts of grooming and boundary-blurring behaviours by academic staff in UK higher education. *Gender and Education*, *33*(8), 1057–1072. https://doi.org/10.1080/09540253.2021.1884199

Bull, A., & Rye, R. (2018). *The Silencing Students Report* (tech. rep.). The 1752 Group. Retrieved July 22, 2025, from https://1752group.com/sexual-misconduct-research-silencing-students/

Bull, S., Hood, S., Mumby, S., Hendrickson, A., Silvasstar, J., & Salyers, A. (2024). Feasibility of using an artificially intelligent chatbot to increase access to information and sexual and reproductive health services. *Digital Health*, *10*, 20552076241308994. https://doi.org/10.1177/20552076241308994

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology*, *38*(2), 217–230. https://doi.org/10.1037/0022-3514.38.2.217

Cambridge Rape Crisis Centre. (2025). Support for survivors of sexual violence. Retrieved August 1, 2025, from https://cambridgerapecrisis.org.uk/

Campbell, C., Kramer, A., Woolman, K., Staecker, E., Visker, J., & Cox, C. (2013). Effects of a Brief Pilot Sexual Harassment Prevention Workshop on Employees' Knowledge. *Workplace Health & Safety*, *61*(10), 425–428. https://doi.org/10.1177/216507991306101002

Canadian Federation of Students, Ontario. (n.d.). *Campus Toolkit for creating consent culture* (tech. rep.). Canadian Federation of Students, Ontario. https://cfsontario.ca/wp-content/uploads/2017/07/Consent-Toolkit.pdf

Cantor, D., Fisher, B., Chibnall, S., Townsend, R., Lee, H., Bruce, C., & Thomas, G. (2015). *Report on the AAU Campus Climate Survey on Sexual Assault and Sexual Misconduct* (tech. rep.). The Association of American Universities. https://www.aau.edu/sites/default/files/%40%20Files/Climate%20Survey/AAU_Campus_Climate_Survey_12_14_15.pdf

Chae, H., Kim, M., Kim, C., Jeong, W., Kim, H., Lee, J., & Yeo, J. (2023). Tutoring: Instruction-Grounded Conversational Agent for Language Learners. https://doi.org/10.48550/arXiv.2302.12623

Chen, Y., Ding, N., Zheng, H.-T., Liu, Z., Sun, M., & Zhou, B. (2024). Empowering Private Tutoring by Chaining Large Language Models. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 354–364. https://doi.org/10.1145/3627673.3679665

Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A. N., Li, T., Li, D., Zhang, H., Zhu, B., Jordan, M., Gonzalez, J. E., & Stoica, I. (2024). Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. https://doi.org/10.48550/arXiv.2403.04132

Chu, M. D., Gerard, P., Pawar, K., Bickham, C., & Lerman, K. (2025). Illusions of Intimacy: Emotional Attachment and Emerging Psychological Risks in Human-AI Relationships. https://doi.org/10.48550/arXiv.2505.11649

Ciphr. (2025a). E6 Harassment and Sexual Misconduct Suite. Retrieved July 25, 2025, from https://www.ciphr.com/e6-harassment-and-sexual-misconduct-suite

Ciphr. (2025b). Sexual Consent on Campus eLearning Course. Retrieved July 25, 2025, from https://www.ciphr.com/elearning-courses/sexual-consent-campus-2

Ciphr. (n.d.). eLearning Sectors — eLearning Clients. Retrieved July 23, 2025, from https://www.ciphr.com/elearning-sectors

Civic Nation. (2025). It's On Us — Peer-Education Tools. Retrieved August 1, 2025, from https://itsonus.org/campus-organizing/peer-education/

Clancy, K. B. H., Nelson, R. G., Rutherford, J. N., & Hinde, K. (2014). Survey of Academic Field Experiences (SAFE): Trainees Report Harassment and Assault. *PLOS ONE*, *9*(7), e102172. https://doi.org/10.1371/journal.pone.0102172

Coker, A. L., Cook-Craig, P. G., Williams, C. M., Fisher, B. S., Clear, E. R., Garcia, L. S., & Hegge, L. M. (2011). Evaluation of Green Dot: An Active Bystander Intervention to Reduce Sexual Violence on College Campuses. *Violence Against Women*, *17*(6), 777–796. https://doi.org/10.1177/1077801211410264

Collins, P. H., & Bilge, S. (2016). *Intersectionality*. Polity. https://sussex.figshare.com/articles/report/That_s_what_she_said_women_students_experiences_of_lad_culture_in_higher_education/23405024?file=41254368

Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic Masculinity: Rethinking the Concept. *Gender & Society*, *19*(6), 829–859. https://doi.org/10.1177/0891243205278639

Consent Coalition. (n.d.). Your Journey. Retrieved August 1, 2025, from https://nottssvss.org.uk/consent-coalition/resources/downloads/your-journey/

Cooper, B. (2021). *The intersection of sexual violence, alcohol and drugs at universities and colleges* (tech. rep.). Office for Students. https://www.officeforstudents.org.uk/media/52171396-39cd-420c-b094-9aa7246d7278/the-intersection-of-sexual-violence-alcohol-and-drugs-at-universities-and-colleges.pdf

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, *385*(6714), eadq1814. https://doi.org/10.1126/science.adq1814

Coulter, R. W. S., Mair, C., Miller, E., Blosnich, J. R., Matthews, D. D., & McCauley, H. L. (2017). Prevalence of Past-Year Sexual Assault Victimization Among Undergraduate Students: Exploring Differences by and Intersections of Gender Identity, Sexual Identity, and Race/Ethnicity. *Prevention Science: The Official Journal of the Society for Prevention Research*, *18*(6), 726–736. https://doi.org/10.1007/s11121-017-0762-8

Coursera. (2025). Let's Talk Consent. Retrieved August 1, 2025, from https://www.coursera.org/learn/talk-consent

Cutler, E., Levonian, Z., & Christie, S. T. (2025). Detecting Student Intent for Chat-Based Intelligent Tutoring Systems. https://doi.org/10.48550/arXiv.2502.15096

Dardis, C. M., Murphy, M. J., Bill, A. C., & Gidycz, C. A. (2016). An investigation of the tenets of social norms theory as they relate to sexually aggressive attitudes and sexual assault perpetration: A comparison of men and their friends. *Psychology of Violence*, *6*(1), 163–171. https://doi.org/10.1037/a0039443

D'Eath, M., O'Higgins, S., Tierney, L., Noone, C., Burke, L., & MacNeela, P. (2020). *Research Evaluation of The Manuela Programme Sexual Violence Prevention Programme for Secondary School Students Final Report, May 2020* (tech. rep.). National University of Ireland, Galway. https://www.tusla.ie/uploads/content/Manuela_Programme_Evaluation_Report_Final_25_May_2020_1.pdf

DeGue, S., Valle, L. A., Holt, M. K., Massetti, G. M., Matjasko, J. L., & Tharp, A. T. (2014). A systematic review of primary prevention strategies for sexual violence perpetration. *Aggression and Violent Behavior*, *19*(4), 346–362. https://doi.org/10.1016/j.avb.2014.05.004

Desai, M. S., Richards, T., & Eddy, J. P. (1999). End-User Training: A Meta Model. *Journal of Instructional Psychology*, *26*(2), 74–84.

DiCerbo, K. (2025). *Khan Academy annual report 2023-2024 - Khanmigo* (tech. rep.). Khan Academy. Retrieved July 30, 2025, from https://annualreport.khanacademy.org/khanmigo

Dobbin, F., & Kalev, A. (2019). The promise and peril of sexual harassment programs. *Proceedings of the National Academy of Sciences*, *116*(25), 12255–12260. https://doi.org/10.1073/pnas.1818477116

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Dublin Rape Crisis Centre. (2025). Resources. Retrieved August 1, 2025, from https://www.drcc.ie/news-resources/resources/

Duggan, L. M. (1998). *The effectiveness of an acquaintance sexual assault prevention program in changing attitudes/beliefs and behavioral intent among college students*. Temple University. Retrieved July 29, 2025, from https://www.proquest.com/docview/304482992/abstract/5783F24F0C8A4F47PQ/1?sourcetype=Dissertations%20&%20Theses

Dutton, G. R., Fontaine, K. R., Alcorn, A. S., Dawson, J., Capers, P. L., & Allison, D. B. (2015). Randomized Controlled Trial Examining Expectancy Effects on the Accuracy of Weight Measurement. *Clinical obesity*, *5*(1), 38–41. https://doi.org/10.1111/cob.12083

EI education. (n.d.). Everyone's Invited. Retrieved July 23, 2025, from https://www.everyonesinvited.uk/

Epigeum. (n.d.-a). Epigeum Tackling Harassment & Sexual Misconduct. Retrieved July 24, 2025, from https://learningresources.sagepub.com/epigeum/support-and-wellbeing/tackling-harassment-and-sexual-misconduct

Epigeum. (n.d.-b). Responding to the OFS guidelines. Retrieved July 23, 2025, from https://learningresources.sagepub.com/blog/epigeum/2024/07/31/responding-to-the-ofs-guidelines

European Medicines Agency. (1997). *E 9 Statistical Principles for Clinical Trials* (tech. rep.). European Medicines Agency. https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline

Fenton, R. A., & Mott, H. L. (2018). Evaluation of the Intervention Initiative: A Bystander Intervention Program to Prevent Violence Against Women in Universities. *Violence and Victims*, *33*(4), 645–662. https://doi.org/10.1891/0886-6708.VV-D-16-00074

Fethi, I., Daigneault, I., Bergeron, M., Hébert, M., & Lavoie, F. (2023). Campus Sexual Violence: A Comparison of International and Domestic Students. *Journal of International Students*, *13*(1), 1–21. https://doi.org/10.32674/jis.v13i1.3685

Fosnot, C. T. (2013). *Constructivism: Theory, Perspectives, and Practice*. Teachers College Press.

Foubert, J. D. (2005). *The Men's Program: A Peer Education Guide to Rape Prevention, Third Edition*. Routledge. https://doi.org/10.4324/9780203956601

Galway Rape Crisis Centre. (2024). Background to The Manuela Programme. Retrieved August 1, 2025, from https://galwayrcc.ie/education-training/background-to-the-manuela-programme/

Ghafourifard, M. (2024). Survey Fatigue in Questionnaire Based Research: The Issues and Solutions. *Journal of caring sciences*, *13*(4). https://doi.org/10.34172/jcs.33287

Gidycz, C. A., Orchowski, L. M., & Berkowitz, A. D. (2011). Preventing Sexual Aggression Among College Men: An Evaluation of a Social Norms and Bystander Intervention

Program. *Violence Against Women*, *17*(6), 720–742. https://doi.org/10.1177/1077801211409727

Giskard AI. (2025). Phare LLM Benchmark. Retrieved July 26, 2025, from https://phare.giskard.ai/

Glass, N., Hanson, G. C., Anger, W. K., Laharnar, N., Campbell, J. C., Weinstein, M., & Perrin, N. (2017). Computer-based training (CBT) intervention reduces workplace violence and harassment for homecare workers. *American Journal of Industrial Medicine*, *60*(7), 635–643. https://doi.org/10.1002/ajim.22728

Google. (2025). Embeddings — Gemini API. Retrieved July 27, 2025, from https://ai.google.dev/gemini-api/docs/embeddings

Gorissen, M. (2024). It's Just a Distance Thing: Affordances and Decisions in Online Disclosure of Sexual Violence Victimization. *Journal of Interpersonal Violence*, *39*(23-24), 4976–5006. https://doi.org/10.1177/08862605241246800

Grabb, D., Lamparth, M., & Vasan, N. (2024). Risks from Language Models for Automated Mental Healthcare: Ethics and Structure for Implementation. https://doi.org/10.48550/arXiv.2406.11852

Griffin, V. W., Wentz, E., & Meinert, E. (2022). Explaining the why in #WhyIDidntReport: An examination of common barriers to formal disclosure of sexual assault in college students. *Journal of Interpersonal Violence*, *37*(15-16), NP14716–NP14745. https://doi.org/10.1177/08862605211016343

Griner, S. B., Vamos, C. A., Thompson, E. L., Logan, R., Vázquez-Otero, C., & Daley, E. M. (2020). The Intersection of Gender Identity and Violence: Victimization Experienced by Transgender College Students. *Journal of Interpersonal Violence*, *35*(23-24), 5704–5725. https://doi.org/10.1177/0886260517723743

Gross, A. M., Winslett, A., Roberts, M., & Gohm, C. L. (2006). An Examination of Sexual Violence Against College Women. *Violence Against Women*, *12*(3), 288–300. https://doi.org/10.1177/1077801205277358

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109. https://doi.org/10.4103/2229-3485.83221

Hales, S. T., & Gannon, T. A. (2022). Understanding Sexual Aggression in UK Male University Students: An Empirical Assessment of Prevalence and Psychological Risk Factors. *Sexual Abuse*, *34*(6), 744–770. https://doi.org/10.1177/10790632211051682

Hamad, F., Shehata, A., & Al Hosni, N. (2024). Predictors of blended learning adoption in higher education institutions in Oman: Theory of planned behavior. *International Journal of Educational Technology in Higher Education*, *21*(1), 13. https://doi.org/10.1186/s41239-024-00443-8

Hatch, S. G., Goodman, Z. T., Vowels, L., Hatch, H. D., Brown, A. L., Guttman, S., Le, Y., Bailey, B., Bailey, R. J., Esplin, C. R., Harris, S. M., Holt, D. P., McLaughlin, M., O'Connell, P., Rothman, K., Ritchie, L., Top, N., & Braithwaite, S. R. (2025). When ELIZA meets therapists: A Turing test for the heart and mind. *PLOS Mental Health*, *2*(2), e0000145. https://doi.org/10.1371/journal.pmen.0000145

Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge. https://doi.org/10.4324/9780203887332

He, J., Rungta, M., Koleczek, D., Sekhon, A., Wang, F. X., & Hasan, S. (2024). Does Prompt Formatting Have Any Impact on LLM Performance? https://doi.org/10.48550/arXiv.2411.10541

Heard, E., Evans, C., Buckley, L., Hatchman, K., & Barbara, M. (2024). Evaluating an online module for sexual violence prevention in a tertiary educational setting: An exploratory study. *Health promotion journal of Australia : official journal of*

Australian Association of Health Promotion Professionals, *35*(1). https://doi.org/10.1002/hpja.714

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). Measuring Massive Multitask Language Understanding. https://doi.org/10.48550/arXiv.2009.03300

HESA. (2025). *2023/24 Student numbers and characteristics* (tech. rep.). Higher Education Student Statistics. Retrieved July 25, 2025, from https://www.hesa.ac.uk/news/20-03-2025/sb271-higher-education-student-statistics/numbers?utm_source=chatgpt.com

Hewing, M., & Leinhos, V. (2024). The Prompt Canvas: A Literature-Based Practitioner Guide for Creating Effective Prompts in Large Language Models. https://doi.org/10.48550/arXiv.2412.05127

Hillenbrand-Gunn, T. L., Heppner, M. J., Mauch, P. A., & Park, H.-J. (2010). Men as Allies: The Efficacy of a High School Rape Prevention Intervention. *Journal of Counseling & Development*, *88*(1), 43–51. https://doi.org/10.1002/j.1556-6678.2010.tb00149.x

Hipgrave, L., Goldie, J., Dennis, S., & Coleman, A. (2025). Balancing risks and benefits: Clinicians' perspectives on the use of generative AI chatbots in mental healthcare. *Frontiers in Digital Health*, *7*. https://doi.org/10.3389/fdgth.2025.1606291

Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized Persuasion: Tailoring Persuasive Appeals to Recipients' Personality Traits. *Psychological Science*, *23*(6), 578–581. https://doi.org/10.1177/0956797611436349

Hock, L. E., Barlow, P. B., Scruggs, B. A., Oetting, T. A., Martinez, D. A., Abràmoff, M. D., & Shriver, E. M. (2021). Tools for Responding to Patient-Initiated Verbal Sexual Harassment: A Workshop for Trainees and Faculty. *MedEdPORTAL: The Journal of Teaching and Learning Resources*, *17*, 11096. https://doi.org/10.15766/mep_2374-8265.11096

Home Office. (2025). Crime outcomes year to December 2024: Data tables. Retrieved July 28, 2025, from https://www.gov.uk/government/statistics/crime-outcomes-year-to-december-2024-data-tables

Htun, M., Jensenius, F. R., Dominguez, M. S., Tinkler, J., & Contreras, C. (2022). Effects of Mandatory Sexual Misconduct Training on University Campuses. *Socius*, *8*, 23780231221124574. https://doi.org/10.1177/23780231221124574

Humphreys, T. P., & Brousseau, M. M. (2010). The Sexual Consent Scale-Revised: Development, reliability, and preliminary validity. *Journal of Sex Research*, *47*(5), 420–428. https://doi.org/10.1080/00224490903151358

Icena. (n.d.). ICENA: Empowerment and Social Change. Retrieved July 23, 2025, from https://www.icena.net/

Indiana University. (2020). Chatbots can ease medical providers' burden, offer trusted guidance to those with COVID-19 symptoms. Retrieved July 31, 2025, from https://news.iu.edu/live/news/26965-chatbots-can-ease-medical-providers-burden-offer

Jadad, A., & Enkin, M. (1998). Bias in Randomized Controlled Trials. In *Randomised Controlled Trials: A Users Guide* (Wiley–Blackwell, pp. 29–47). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470691922.ch3

Jeune, P. L., Malézieux, B., Xiao, W., & Dora, M. (2025, May). Phare: A Safety Probe for Large Language Models. https://doi.org/10.48550/arXiv.2505.11365

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.*, *55*(12), 248:1–248:38. https://doi.org/10.1145/3571730

Joyal-Desmarais, K., Scharmer, A. K., Madzelan, M. K., See, J. V., Rothman, A. J., & Snyder, M. (2022). Appealing to motivation to change attitudes, intentions, and behavior: A systematic review and meta-analysis of 702 experimental tests of the effects of motivational message matching on persuasion. *Psychological Bulletin*, *148*(7-8), 465–517. https://doi.org/10.1037/bul0000377

Kettrey, H. H., Marx, R. A., & Tanner-Smith, E. E. (2019). Effects of bystander programs on the prevention of sexual assault among adolescents and college students: A systematic review. *Campbell Systematic Reviews*, *15*(1-2), e1013. https://doi.org/10.4073/csr.2019.1

Kettrey, H. H., Thompson, M. P., Marx, R. A., & Davis, A. J. (2023). Effects of Campus Sexual Assault Prevention Programs on Attitudes and Behaviors Among American College Students: A Systematic Review and Meta-Analysis. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine*, *72*(6), 831–844. https://doi.org/10.1016/j.jadohealth.2023.02.022

Khawaja, Z., & Bélisle-Pipon, J.-C. (2023). Your robot therapist is not your therapist: Understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1278186

Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating Training Programs: The Four Levels*. Berrett-Koehler Publishers.

Kleinsasser, A., Jouriles, E., McDonald, R., & Rosenfield, D. (2015). An Online Bystander Intervention Program for the Prevention of Sexual Violence. *Psychology of violence*, *5*(3). https://doi.org/10.1037/a0037393

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*(2), 311–328. https://doi.org/10.1037/0021-9010.78.2.311

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Latt, P. M., Aung, E. T., Htaik, K., Soe, N. N., Lee, D., King, A. J., Fortune, R., Ong, J. J., Chow, E. P. F., Bradshaw, C. S., Rahman, R., Deneen, M., Dobinson, S., Randall, C., Zhang, L., & Fairley, C. K. (2025). Evaluation of artificial intelligence (AI) chatbots for providing sexual health information: A consensus study using real-world clinical queries. *BMC Public Health*, *25*, 1788. https://doi.org/10.1186/s12889-025-22933-8

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (pp. 63–76). Springer. https://doi.org/10.1007/978-3-540-89350-9_6

Lee, C., & Wong, J. S. (2024). Does "Consent Matter"? An Evaluation of a Sexual Consent Awareness Campaign for College Students. *Violence Against Women*, 10778012241280059. https://doi.org/10.1177/10778012241280059

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. https://doi.org/10.48550/arXiv.2005.11401

Lincoln, J. (2021). *What works for ending public sexual harassment* (tech. rep.). Plan International UK.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

LMArena. (2025). Text Arena. Retrieved July 26, 2025, from https://lmarena.ai/leaderboard/text/instruction-following

London School of Economics. (2025). Consent.Ed training. Retrieved August 1, 2025, from https://info.lse.ac.uk/report-it/Consent.Ed

Long, R. (2025). Relationships and Sex Education in Schools (England). Retrieved August 1, 2025, from https://commonslibrary.parliament.uk/research-briefings/sn06103/

Lonsway, K. A. (1996). Preventing Acquaintance Rape through Education What Do We Know. *Psychology of Women Quarterly*, *20*(2), 229–265. https://doi.org/10.1111/j.1471-6402.1996.tb00469.x

MacNeela, P., Breen, J., Byrnes, E., O'Higgins, S., Seery, C., & Silke, C. (2017). *Development, Implementation, and Evaluation of the SMART Consent Workshop on Sexual Consent for Third Level Students* (tech. rep.). National Universtiy of Ireland, Galway. https://www.consenthub.ie/wp-content/uploads/2021/12/SMART-Consent-Workshop-on-Sexual-Consent-Third-Level-Students-.pdf

Malamuth, N. M., Linz, D., Heavey, C. L., Barnes, G., & Acker, M. (1995). Using the confluence model of sexual aggression to predict men's conflict with women: A 10-year follow-up study. *Journal of Personality and Social Psychology*, *69*(2), 353–369. https://doi.org/10.1037//0022-3514.69.2.353

Malamuth, N. M., & Check, J. V. P. (1984). Debriefing effectiveness following exposure to pornographic rape depictions. *The Journal of Sex Research*, *20*(1), 1–13. https://doi.org/10.1080/00224498409551203

Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports*, *14*(1), 4692. https://doi.org/10.1038/s41598-024-53755-0

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

McMahon, S., & Farmer, G. L. (2011). An Updated Measure for Assessing Subtle Rape Myths. *Social Work Research*, *35*(2), 71–81. https://doi.org/10.1093/swr/35.2.71

Mirza, H. (2014). Decolonizing Higher Education: Black Feminism and the Intersectionality of Race and Gender. *Journal of Feminist Scholarship*, *7*(7), 1–12. https://digitalcommons.uri.edu/jfs/vol7/iss7/3

Mondal, H., & Mondal, S. (2025). The Role of Large Language Model Chatbots in Sexual Education: An Unmet Need of Research. *Journal of Psychosexual Health*, *7*(2), 120–127. https://doi.org/10.1177/26318318251323714

Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers, 599–627. https://doi.org/10.1145/3715275.3732039

Moynihan, M. M., Banyard, V. L., Cares, A. C., Potter, S. J., Williams, L. M., & Stapleton, J. G. (2015). Encouraging Responses in Sexual and Relationship Violence Prevention: What Program Effects Remain 1 Year Later? *Journal of Interpersonal Violence*, *30*(1), 110–132. https://doi.org/10.1177/0886260514532719

Muennighoff, N., Tazi, N., Magne, L., & Reimers, N. (2023). MTEB: Massive Text Embedding Benchmark. https://doi.org/10.48550/arXiv.2210.07316

Nadarzynski, T., Puentes, V., Pawlak, I., Mendes, T., Montgomery, I., Bayley, J., Ridge, D., & Newman, C. (2021). Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: A qualitative analysis. *Sexual Health*, *18*(5), 385–393. https://doi.org/10.1071/SH21123

Nagy, S., Adcock, A. G., & Nagy, M. C. (1994). A Comparison of Risky Health Behaviors of Sexually Active, Sexually Abused, and Abstaining Adolescents. *Pediatrics*, *93*(4), 570–575. https://doi.org/10.1542/peds.93.4.570

Nasrul, I., Angraini, A., Hamzah, M. L., & Saputra, E. (2024). Analysis Of User Experience Of ChatGPT And Gemini Users Using The User Experience Quistionnaire (UEQ) For Education. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, *13*(3), 352–359. https://doi.org/10.32736/sisfokom.v13i3.2250

National Sexual Violence Resource Center. (n.d.). How Do You Know if Someone Wants to Have Sex with You? Using Consent Videos to Model Respect and Communication. Retrieved August 1, 2025, from https://www.nsvrc.org/publications/curriculum/how-do-you-know-if-someone-wants-have-sex-you-using-consent-videos-model

National Union of Students. (n.d.). I Heart Consent. Retrieved August 1, 2025, from https://www.nusconnect.org.uk/liberation/women-students/lad-culture/i-heart-consent

New York State. (2015). Enough is Enough. Retrieved August 1, 2025, from https://opdv.ny.gov/enough-enough

Nielsen, J. (1994, September). *Usability Engineering*. Morgan Kaufmann.

Nikischer, A. (2019). Vicarious trauma inside the academe: Understanding the impact of teaching, researching and writing violence. *Higher Education*, *77*(5), 905–916. https://doi.org/10.1007/s10734-018-0308-4

Ó Ciardha, C., & Ward, T. (2013). Theories of Cognitive Distortions in Sexual Offending: What the Current Research Tells Us. *Trauma, Violence, & Abuse*, *14*(1), 5–21. https://doi.org/10.1177/1524838012467856

Ofcom. (2025). *Adults' media use and attitude* (tech. rep.). Ofcom. https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-adults/adults-media-use-and-attitudes

Office for Students. (2021). Office for Students statement of expectations for preventing and addressing harassment and sexual misconduct affecting students in higher education.

Office for Students. (2024, May). Condition E6: Harassment and sexual misconduct - Office for Students. Retrieved July 31, 2025, from https://www.officeforstudents.org.uk/

Oxford Brookes University. (2025). Sexual consent education. Retrieved July 25, 2025, from https://www.brookes.ac.uk/students/safeguarding-hub/sexual-consent-education

Oxford Student Union. (n.d.). Anti-Sexual Violence Campaign, It Happens Here. Retrieved July 23, 2025, from https://www.ithappenshere.co.uk

Oxley, G., & Lipscombe, S. (2025, January). *Tackling violence against women and girls* (tech. rep.). House of Commons Library. Retrieved August 1, 2025, from https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0183/

Page, T. E., Pina, A., & Giner-Sorolla, R. (2016). "It was only harmless banter!" The development and preliminary validation of the moral disengagement in sexual harassment scale. *Aggressive Behavior*, *42*(3), 254–273. https://doi.org/10.1002/ab.21621

Park, Y., & Doo, M. Y. (2024). Role of AI in Blended Learning: A Systematic Literature Review. *The International Review of Research in Open and Distributed Learning*, *25*(1), 164–196. https://doi.org/10.19173/irrodl.v25i1.7566

Paul, L. A., & Gray, M. J. (2011). Sexual Assault Programming on College Campuses: Using Social Psychological Belief and Behavior Change Principles to Improve Outcomes. *Trauma, Violence, & Abuse*, *12*(2), 99–109. https://doi.org/10.1177/1524838010390710

Payne, D. L., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape Myth Acceptance: Exploration of Its Structure and Its Measurement Using the*Illinois Rape Myth Acceptance Scale. Journal of Research in Personality*, *33*(1), 27–68. https://doi.org/10.1006/jrpe.1998.2238

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Perkins, H. W., & Berkowitz, A. D. (1986). Perceiving the community norms of alcohol use among students: Some research implications for campus alcohol education programming. *The International Journal of the Addictions*, *21*(9-10), 961–976. https://doi.org/10.3109/10826088609077249

Peterson, K., & McCleery, E. (2014). *Evidence Brief: The Effectiveness Of Mandatory Computer-Based Trainings On Government Ethics, Workplace Harassment, Or Privacy And Information Security-Related Topics*. Department of Veterans Affairs (US). Retrieved July 24, 2025, from http://www.ncbi.nlm.nih.gov/books/NBK384612/

Phipps, A., & Young, I. (2013). *That's what she said: Women students' experiences of 'lad culture' in higher education* (tech. rep.). University of Sussex.

Phipps, A., & Young, I. (2015). 'Lad culture' in higher education: Agency in the sexualization debates. *Sexualities*, *18*(4), 459–479. https://doi.org/10.1177/1363460714550909

Pierre, J. (2025). AI Hallucinations in Medicine and Mental Health. Retrieved July 30, 2025, from https://www.psychologytoday.com/us/blog/psych-unseen/202506/ai-hallucinations-in-medicine-and-mental-health

Polaschek, D. L. L., & Ward, T. (2002). The implicit theories of potential rapists: What our questionnaires tell us. *Aggression and Violent Behavior*, *7*(4), 385–406. https://doi.org/10.1016/S1359-1789(01)00063-5

Potter, S. J., Edwards, K. M., Banyard, V. L., Stapleton, J. G., Demers, J. M., & Moynihan, M. M. (2016). Conveying campus sexual misconduct policy information to college and university students: Results from a 7-campus study. *Journal of American College Health*, *64*(6), 438–447. https://doi.org/10.1080/07448481.2016.1178122

Preusser, M. K., Bartels, L. K., & Nordstrom, C. R. (2011). Sexual Harassment Training: Person versus Machine. *Public Personnel Management*, *40*(1), 47–62. https://doi.org/10.1177/009102601104000104

Prolific. (2025a). What are the advantages and limitations of an online sample? Retrieved July 30, 2025, from https://researcher-help.prolific.com/en/article/a061d5

Prolific. (2025b). Who should I reject? Retrieved July 30, 2025, from https://researcher-help.prolific.com/en/article/f75ea9?utm_source=chatgpt.com

Qin, F., Hao, Z., Yu, J., Liu, Z., & Zhang, Y. (2025). AI instructional agent improves student's perceived learner control and learning outcome: Empirical evidence from a randomized controlled trial. https://doi.org/10.48550/arXiv.2505.22526

Rape Crisis England & Wales. (2025). Rape, sexual assault and child sexual abuse statistics. Retrieved August 1, 2025, from https://rapecrisis.org.uk/get-informed/statistics-sexual-violence/

Rape Crisis England & Wales. (n.d.). Myths vs facts. Retrieved July 23, 2025, from https://rapecrisis.org.uk/get-informed/about-sexual-violence/myths-vs-realities/

Reimers, N., Magne, L., Tazi, N., & Muennighoff, N. (2025). MTEB Leaderboard. Retrieved July 27, 2025, from https://huggingface.co/spaces/mteb/leaderboard

Reise, S. P., & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, *5*(Volume 5, 2009), 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Revolt Sexual Assault. (2018). *Student's experience of sexual violence* (tech. rep.). The Student Room. https://revoltsexualassault.com/wp-content/uploads/2018/03/Report-Sexual-Violence-at-University-Revolt-Sexual-Assault-The-Student-Room-March-2018.pdf

Ring, T. E., & Kilmartin, C. (1992). Man to Man about Rape: A Rape Prevention Program for Men. *Journal of College Student Development*, *33*(1), 82–84. Retrieved July 26, 2025, from https://eric.ed.gov/?id=EJ444092

RoedigerIII, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rosenshine, B. (2012). Principles of Instruction: Research-Based Strategies That All Teachers Should Know. *American Educator*, *36*(1), 12. Retrieved July 27, 2025, from https://eric.ed.gov/?id=EJ971753

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*(2), 183–192. https://doi.org/10.1037/0033-2909.118.2.183

Rothman, E. F., Exner, D., & Baughman, A. L. (2011). The prevalence of sexual assault against people who identify as gay, lesbian, or bisexual in the United States: A systematic review. *Trauma, Violence & Abuse*, *12*(2), 55–66. https://doi.org/10.1177/1524838010390707

Royston, P. (2004). Multiple Imputation of Missing Values. *The Stata Journal*, *4*(3), 227–241. https://doi.org/10.1177/1536867X0400400301

Salvi, F., Ribeiro, M. H., Gallotti, R., & West, R. (2025). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 1–9. https://doi.org/10.1038/s41562-025-02194-6

SC Training. (2025). Top 10 Sexual harassment training courses in the UK. Retrieved August 1, 2025, from https://training.safetyculture.com/course-collection/sexual-harassment-training-uk/

Schewe, P. A., & O'Donohue, W. (1993). Sexual abuse prevention with high-risk males: The roles of victim empathy and rape myths. *Violence and Victims*, *8*(4), 339–351.

Schoenegger, P., Salvi, F., Liu, J., Nan, X., Debnath, R., Fasolo, B., Leivada, E., Recchia, G., Günther, F., Zarifhonarvar, A., Kwon, J., Islam, Z. U., Dehnert, M., Lee, D. Y. H., Reinecke, M. G., Kamper, D. G., Kobaş, M., Sandford, A., Kgomo, J., … Karger, E. (2025, May). Large Language Models Are More Persuasive Than Incentivized Human Persuaders. https://doi.org/10.48550/arXiv.2505.09662

Sexual Harassment Training com. (n.d.). Plan and Pricing. Retrieved August 1, 2025, from https://www.sexualharassmenttraining.com/pricing

Sexual Violence Training Development Team. (2021). *Consent & Sexual Violence: Training and Facilitation Guide*. BC Campus. Retrieved August 1, 2025, from https://opentextbc.ca/svmconsent/

Sherekar, P., & Mehta, M. (2025). Harnessing technology for hope: A systematic review of digital suicide prevention tools. *Discover Mental Health*, *5*(1), 101. https://doi.org/10.1007/s44192-025-00245-y

Shetye, S. (2024). An Evaluation of Khanmigo, a Generative AI Tool, as a Computer-Assisted Language Learning App. *Studies in Applied Linguistics and TESOL*, *24*(1). https://doi.org/10.52214/salt.v24i1.12869

Siddals, S., Torous, J., & Coxon, A. (2024). "It happened to be the perfect thing": Experiences of generative AI chatbots for mental health. *NPJ Mental Health Research*, *3*, 48. https://doi.org/10.1038/s44184-024-00097-4

Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, *3*(2), pgae035. https://doi.org/10.1093/pnasnexus/pgae035

Stenning, P., Mitra-Kahn, T., & Gunby, C. (2012). *Gender-Based Violence, Stalking and Fear of Crime: Country Report United Kingdom* (tech. rep.). European Commission. https://gendercrime.eu/pdf/gendercrime_final_report_printable_version.pdf

Stephens, K. A., & George, W. H. (2009). Rape Prevention With College Men. *Journal of interpersonal violence*, *24*(6), 996–1013. https://doi.org/10.1177/0886260508319366

Sundaram, V., Bell, B. T., Jessop, N., & Jackson, E. L. (2022). *Everything is racialised on top: Black and minoritised girls' and young women's experiences of public sexual harassment in the UK*. PLAN International, UK. https://pure.york.ac.uk/portal/en/publications/everything-is-racialised-on-top-black-and-minoritised-girls-and-y

Sweller, J. (2011). Chapter Two - Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (pp. 37–76, Vol. 55). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Taghanaki, S. A., Khani, A., & Khasahmadi, A. (2024). MMLU-Pro+: Evaluating Higher-Order Reasoning and Shortcut Learning in LLMs. https://doi.org/10.48550/arXiv.2409.02257

Talk Consent. (2018). School Consent Workshops and Assemblies. Retrieved July 23, 2025, from https://talkconsent.org/consent-workshop/

te Pas, M. E., Rutten, W. G. M. M., Bouwman, R. A., & Buise, M. P. (2020). User Experience of a Chatbot Questionnaire Versus a Regular Computer Questionnaire: Prospective Comparative Study. *JMIR Medical Informatics*, *8*(12), e21982. https://doi.org/10.2196/21982

The art of consent. (n.d.). Consent for Students. Retrieved July 23, 2025, from https://www.artofconsent.co.uk/consent-workshops-for-students

Tomlinson, C. A. (2014, May). *The Differentiated Classroom: Responding to the Needs of All Learners, 2nd Edition*. ASCD.

UK Department for Education and Employment. (2000). *Sex and relationship education guidance* (tech. rep.). Department for Education and Employment. 2000. https://www.education-uk.org/documents/dfee/2000-sex-education.pdf

UK Parliament. (2003). Sexual Offences Act 2003. Retrieved July 22, 2025, from https://www.legislation.gov.uk/ukpga/2003/42/contents

University of Edinburgh. (n.d.). Consent On Campus. Retrieved July 24, 2025, from https://www.marshallacm.co.uk/ClientScorm/Edinburgh/Consent/scormcontent/index.html#/lessons/5E1O5M0Jqxl6icC6jDPH8QQf4P4vD5E0

University of Exeter. (n.d.). The toolkit — The Intervention Initiative. Retrieved July 23, 2025, from https://www.exeter.ac.uk/research/groups/law/interventioninitiative/toolkit/

University of Greater Manchester. (2023). LEAP Sexual Consent. Retrieved August 1, 2025, from https://leaponline.bolton.ac.uk/My-Personal-Development/My-Personal-Effectiveness/Sexual-Consent.aspx

University of Oxford. (2025a). Consent for Students: An online consent programme. Retrieved July 25, 2025, from https://www.ox.ac.uk/students/welfare/supportservice/consent-for-students

University of Oxford. (2025b). Informed consent. Retrieved July 25, 2025, from https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/consent

UNLRN. (n.d.). Diversity Equity and Inclusion - Courses and Programmes. Retrieved July 23, 2025, from https://www.unlrn.co.uk/training

Verbeek, M., Weeland, J., Luijk, M., & van de Bongardt, D. (2023). Sexual and Dating Violence Prevention Programs for Male Youth: A Systematic Review of Program Characteristics, Intended Psychosexual Outcomes, and Effectiveness. *Archives of Sexual Behavior*, *52*(7), 2899–2935. https://doi.org/10.1007/s10508-023-02596-5

Verma, K. (2024). The Ethical Implications of AI Chatbots and Voice Bots. Retrieved July 30, 2025, from https://torontodigital.ca/blog/the-ethical-implications-of-ai-chatbots-and-voice-bots/

Vladika, J., Domres, A., Nguyen, M., Moser, R., Nano, J., Busch, F., Adams, L. C., Bressem, K. K., Bernhardt, D., Combs, S. E., Borm, K. J., Matthes, F., & Peeken, J. C. (2025). Improving Reliability and Explainability of Medical Question Answering through Atomic Fact Checking in Retrieval-Augmented LLMs. https://doi.org/10.48550/arXiv.2505.24830

Vogel, M. (2024). *The Automat Framework* (tech. rep.). Big Picture. Retrieved July 27, 2025, from https://medium.com/the-generator/the-perfect-prompt-prompt-engineering-cheat-sheet-d0b9c62a2bba

Walker, S.-J. L., Hester, M., McPhee, D., Patsios, D., Williams, A., Bates, L., & Rumney, P. (2021). Rape, inequality and the criminal justice response in England: The importance of age and gender. *Criminology & Criminal Justice*, *21*(3), 297–315. https://doi.org/10.1177/1748895819863095

Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., & Piterova, A. (2022). An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study. *Journal of Medical Internet Research*, *24*(1), e29969. https://doi.org/10.2196/29969

Wang, Y., Ma, X., Zhang, G., Ni, Y., Chandra, A., Guo, S., Ren, W., Arulraj, A., He, X., Jiang, Z., Li, T., Ku, M., Wang, K., Zhuang, A., Fan, R., Yue, X., & Chen, W. (2024). MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark. https://doi.org/10.48550/arXiv.2406.01574

Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review*, *17*(5), 479–507. https://doi.org/10.1016/S0272-7358(97)81034-3

Washington University. (2011). Green Dot Bystander Program. Retrieved July 24, 2025, from https://rsvpcenter.washu.edu/prevention/green-dot-bystander-program/

We-Consent. (2025). Research & Data. Retrieved August 1, 2025, from https://www.we-consent.ie/resource-hub/resources/research-data/

Western Carolina University. (2025). Title IX Training. Retrieved August 1, 2025, from https://www.wcu.edu/discover/leadership/office-of-the-chancellor/legal-counsel-office/TitleIX-Office/Title-IX-Training.aspx

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. https://doi.org/10.1002/sim.4067

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. https://doi.org/10.48550/arXiv.2302.11382

Wright, L. A., Zounlome, N. O. O., & Whiston, S. C. (2020). The Effectiveness of Male-Targeted Sexual Assault Prevention Programs: A Meta-Analysis. *Trauma, Violence, & Abuse*, *21*(5), 859–869. https://doi.org/10.1177/1524838018801330

Yin, L., Bickmore, T., & Cortés, D. E. (2010). The Impact of Linguistic and Cultural Congruity on Persuasion by Conversational Agents. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent Virtual Agents* (pp. 343–349). Springer. https://doi.org/10.1007/978-3-642-15892-6_36

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: A conceptual framework and survey of the state of the art. *Educational technology research and development*, *73*(2), 815–874. https://doi.org/10.1007/s11423-025-10447-4

Zao-Sanders, M. (2025). *How People are Really Using Generative AI Now* (tech. rep.). Filtered. https://learn.filtered.com/hubfs/The%202025%20Top-100%20Gen%20AI%20Use%20Case%20Report.pdf

Zmigrod, R., Mielke, S. J., Wallach, H., & Cotterell, R. (2019, July). Counterfactual Data Augmentation for Mitigating Gender Stereotypes in Languages with Rich Morphology. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1651–1661). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1161

Appendix A

Experiment Elements

The following appendix outlines the steps taken by participants from the recruitment to the completion of the experiment. The entire process is fully accessible at the following link: https://arti-experiment.lovable.app/.

A.1 Overall Experiment Flow

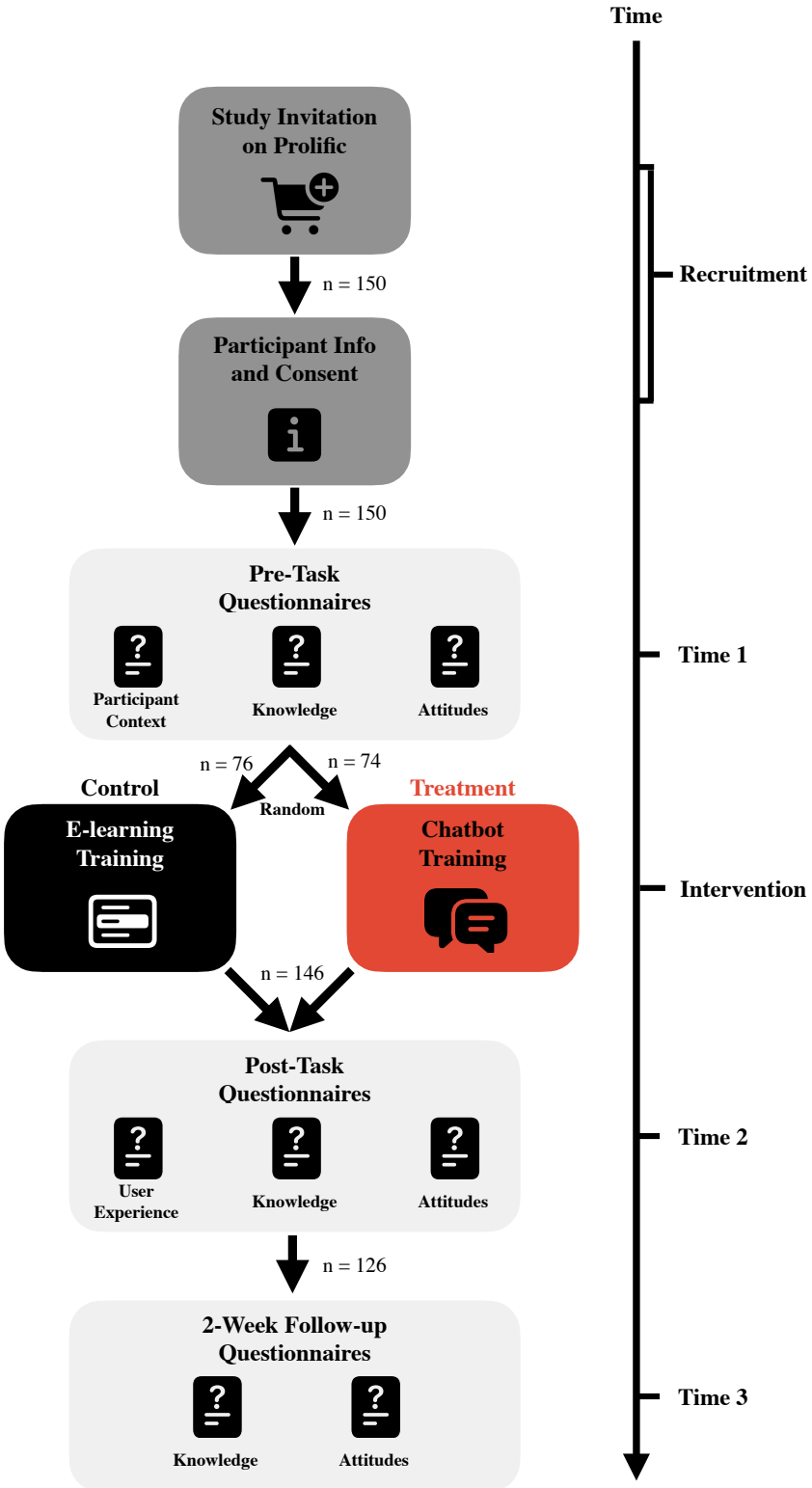


Figure A.1. Overall Experiment Flow.
CONSORT Flow Diagram of the steps participants take through the study from recruitment to completion.

A.2 Recruitment

A.2.1 Study Invitation

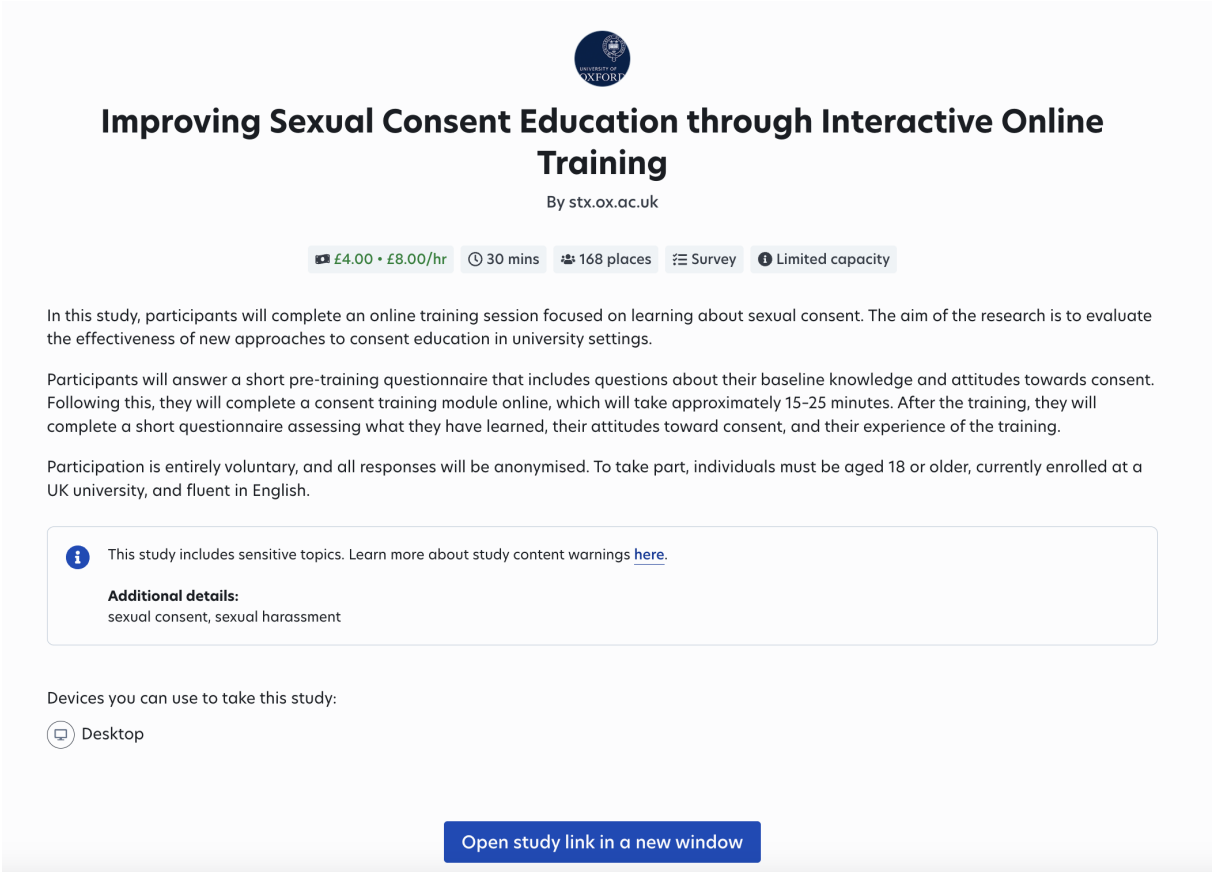


Figure A.2. Study Invitation on Prolific.
Study invitation visible by eligible participants on the prolific platform.

A.2.2 Informed Consent

Improving Sexual Consent Education through Interactive Online Training

Participation Agreement

Oxford Internet Institute
1 St Giles, Oxford, OX1 3JS, UK
enquiries@oii.ox.ac.uk
+44 (0)1865 287210

CUREC Approval Reference: 1615218

General Information

The aim of this research is to investigate whether interactive online sexual consent training improves knowledge and attitudes toward sexual consent more effectively than standard approaches. The findings will help improve the way sexual consent education is delivered in university settings.

We appreciate your interest in participating in this questionnaire and online task. You have been invited to participate as you are 18 years old or above, currently enrolled as a student at a UK university,and fluent in English. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below.

You may ask any questions before deciding to take part by contacting the researcher (details below).

The Principal Researcher is Sasha Gutenmacher, who is attached to the Oxford Internet Institute at the University of Oxford. This research is being completed under the supervision of Prof. Renaud Lambiotte.

You will be asked to complete a short questionnaire before, immediately after and one week after an online consent training session. This should take about 30 minutes. The data collected will help us understand whether this format improves knowledge, engagement, and attitudes related to consent education. No background knowledge is required. Your responses will be used for academic research purposes and may contribute to publications or presentations. Only the research team at the University of Oxford will have access to raw data. No identifiable data will be shared with any third parties.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser. **However, we are only able to reimburse participants who complete all research activities.**

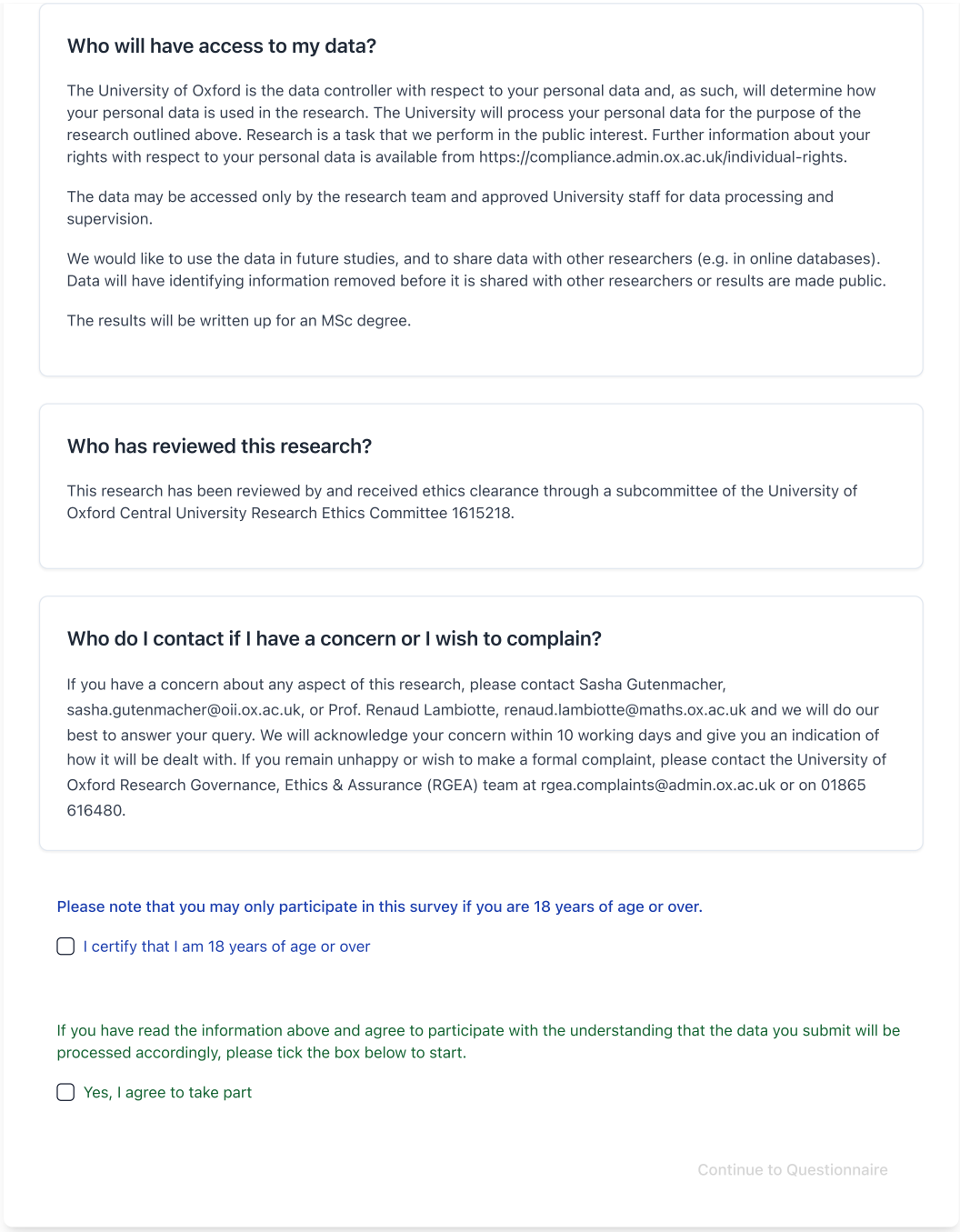
How will my data be used?

The data we will collect that could identify you will be:

- First name (or pseudonym)
- Contact details
- Age
- Gender
- University

Your IP address will not be stored. We will take all reasonable measures to ensure that data remain confidential.

The responses you provide will be stored in a password-protected electronic file on servers compliant with the University of Oxford's data security policies and may be used in academic publications, conference presentations. Research data will be stored for 3 years after publication or public release of the work of the research.



Who will have access to my data?

The University of Oxford is the data controller with respect to your personal data and, as such, will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest. Further information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.

The data may be accessed only by the research team and approved University staff for data processing and supervision.

We would like to use the data in future studies, and to share data with other researchers (e.g. in online databases). Data will have identifying information removed before it is shared with other researchers or results are made public.

The results will be written up for an MSc degree.

Who has reviewed this research?

This research has been reviewed by and received ethics clearance through a subcommittee of the University of Oxford Central University Research Ethics Committee 1615218.

Who do I contact if I have a concern or I wish to complain?

If you have a concern about any aspect of this research, please contact Sasha Gutenmacher, sasha.gutenmacher@oii.ox.ac.uk, or Prof. Renaud Lambiotte, renaud.lambiotte@maths.ox.ac.uk and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the University of Oxford Research Governance, Ethics & Assurance (RGEA) team at rgea.complaints@admin.ox.ac.uk or on 01865 616480.

Please note that you may only participate in this survey if you are 18 years of age or over.

☐ I certify that I am 18 years of age or over

If you have read the information above and agree to participate with the understanding that the data you submit will be processed accordingly, please tick the box below to start.

☐ Yes, I agree to take part

Continue to Questionnaire

Figure A.3. Informed Consent Form
Informed consent form presented to participants before starting the experiment.

A.3 Pre-training Questionnaires

A.3.1 Pre-training Attitude Questionnaire



Improving Sexual Consent Education through Interactive Online Training

Opinion Questionnaire

Instructions

Please note that the term sexual consent is used extensively throughout this questionnaire. Please use the definition of sexual consent below when answering the questions that follow.

Sexual consent: the freely given verbal or nonverbal communication of a feeling of willingness to engage in sexual activity.

Using the following scale, please select the number that best describes how strongly you agree or disagree with each statement. Remember, there are no right or wrong answers, just your opinions.

Statement	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
1. I believe that the need for asking for sexual consent decreases as the length of an intimate relationship increases	○	○	○	○	○	○	○
2. I am worried that my partner might think I'm weird or strange if I asked for sexual consent before starting any sexual activity	○	○	○	○	○	○	○
3. I feel that sexual consent should always be obtained before the start of any sexual activity	○	○	○	○	○	○	○
4. I believe it is enough to ask for consent at the beginning of a sexual encounter	○	○	○	○	○	○	○
5. I would have difficulty asking for consent because it doesn't really fit with how I like to engage in sexual activity	○	○	○	○	○	○	○
6. I would have difficulty asking for consent because it would spoil the mood	○	○	○	○	○	○	○

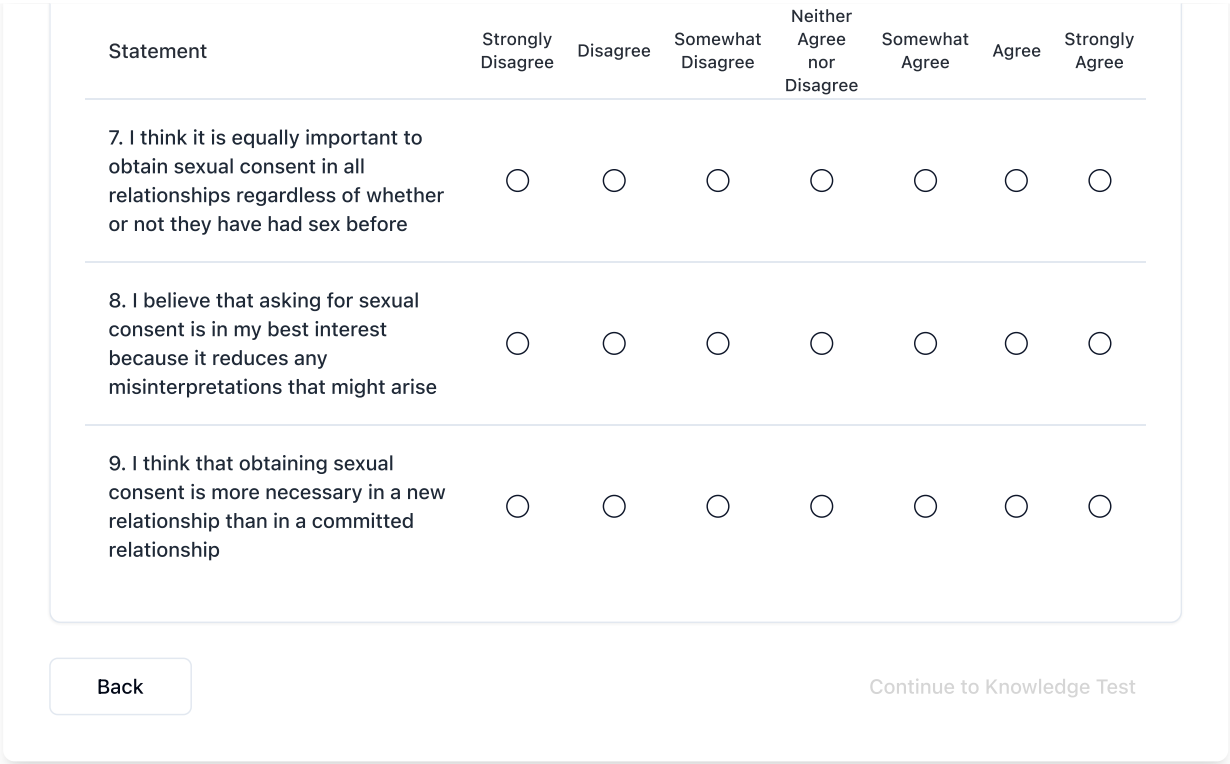


Figure A.4. Pre-training Attitude Questionnaire
Attitudinal questionnaire completed by participants prior to the training session.

A.3.2 Pre-training Knowledge Test

Improving Sexual Consent Education through Interactive Online Training

Knowledge Test

Instructions

Please answer the following questions to the best of your knowledge. Some questions allow multiple answers; these are clearly marked.

1. What is the maximum sentence for rape?

- ◯ A. 10 years imprisonment
- ◯ B. 15 years imprisonment
- ◯ C. 20 years imprisonment
- ◯ D. Life imprisonment

2. Who might find it harder to report a rape? (select all that apply)

- ☐ A. People involved in prostitution
- ☐ B. Disabled people
- ☐ C. People from ethnic minority communities
- ☐ D. Male victims

3. What factors contribute to "rape culture"? (select all that apply)

- ☐ A. Jokes and banter about rape
- ☐ B. Binge drinking
- ☐ C. Dressing provocatively
- ☐ D. Victim blaming

4. Which of these is not a rape myth?

- ◯ A. Most rapes are committed by strangers in dark alleys
- ◯ B. People are most often raped and sexually abused in their home or workplace
- ◯ C. People wrongly accuse others of rape when they regret having sex or to get revenge
- ◯ D. If the survivor didn't say no or complain at the time, then it can't be rape

5. Which of these may indicate you are in an abusive relationship? (select all that apply)

- ☐ A. Your partner keeps an eye on your bank accounts because you spend too much
- ☐ B. Your partner makes a lot of jokes about you in front of family and friends
- ☐ C. Your partner wants to know who you're with and checks in often because they worry about you
- ☐ D. Since you've been with your partner you've seen close friends and family members less

6. Which of these concepts does the Sexual Offences Act 2003 emphasise most strongly in defining consent?

- ◯ A. Verbal affirmation
- ◯ B. History of relationship
- ◯ C. Freedom and capacity
- ◯ D. Mutual sexual interest

7. What coping methods might survivors have as a result of the trauma of rape? (select three that apply)

- ☐ A. Self injuring
- ☐ B. Having nightmares
- ☐ C. Avoiding all intimate relationships
- ☐ D. Taking drugs or drinking

8. Which of the following scenarios are consensual? (select all that apply)

- ☐ A. Both individuals are in a long-term relationship with each other
- ☐ B. Both individuals are intoxicated
- ☐ C. Both individuals have positive body language and communication
- ☐ D. Both individuals gave prior written consent via text message

9. If a lecturer pressures an underage student into sex by threatening their academic outcome, which legal violation best applies?

- ◯ A. Rape
- ◯ B. Abuse of trust
- ◯ C. Sexual assault
- ◯ D. Sexual coercion

10. What percentage of rape prosecutions come from false allegations?

- ◯ A. 0.01%
- ◯ B. 1%
- ◯ C. 10%
- ◯ D. 30%

Back

Complete

Figure A.5. Pre-training Knowledge Test
Knowledge test completed by participants prior to the training session.

A.3.3 Profiling Questionnaire

Improving Sexual Consent Education through Interactive Online Training

Participant Context

Instructions

The following questions are designed to better understand your background, typical university context, and understanding of the topic of sexual consent and harassment. Please answer as accurately as possible.

1. How would like to be referred to throughout this training (first name/pseudonym)?

Enter your preferred name

2. What age are you?

Enter your age

3. What gender do you identify as?

- ◯ Female
- ◯ Male
- ◯ Non-binary
- ◯ Other Gender
- ◯ Prefer not to answer

4. What is the name of your university?

Enter your university name

5. Please describe in 1-2 sentences a typical day at university for you.

Describe your typical university day

6. Where do you spend most of your time interacting with people (in 1-2 sentences)?

Describe where you interact with people most

7. Have you ever received a sexual consent training in the past?

◯ Yes ◯ No

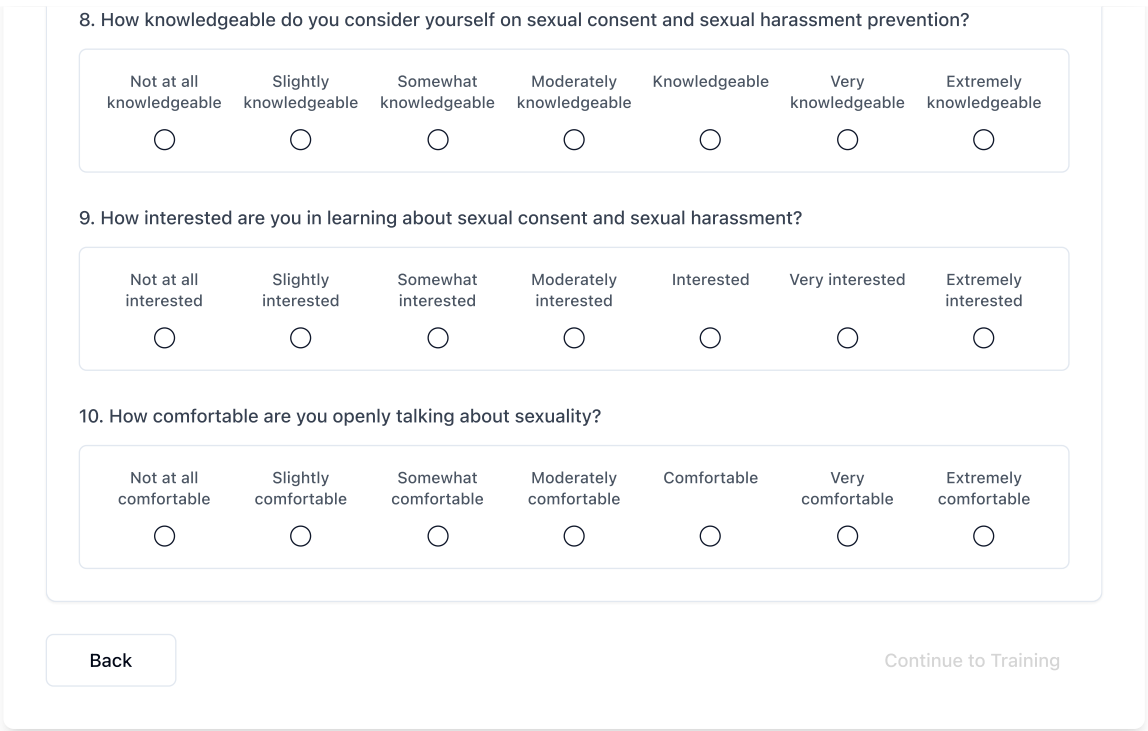


Figure A.6. Profiling Questionnaire
Demographic and contextual questionnaire collecting information to personalize the chatbot training.

A.4 Chatbot Training

The chatbot training used for this experiement is fully functional and accessible at the following link: Link to treatment chatbot training

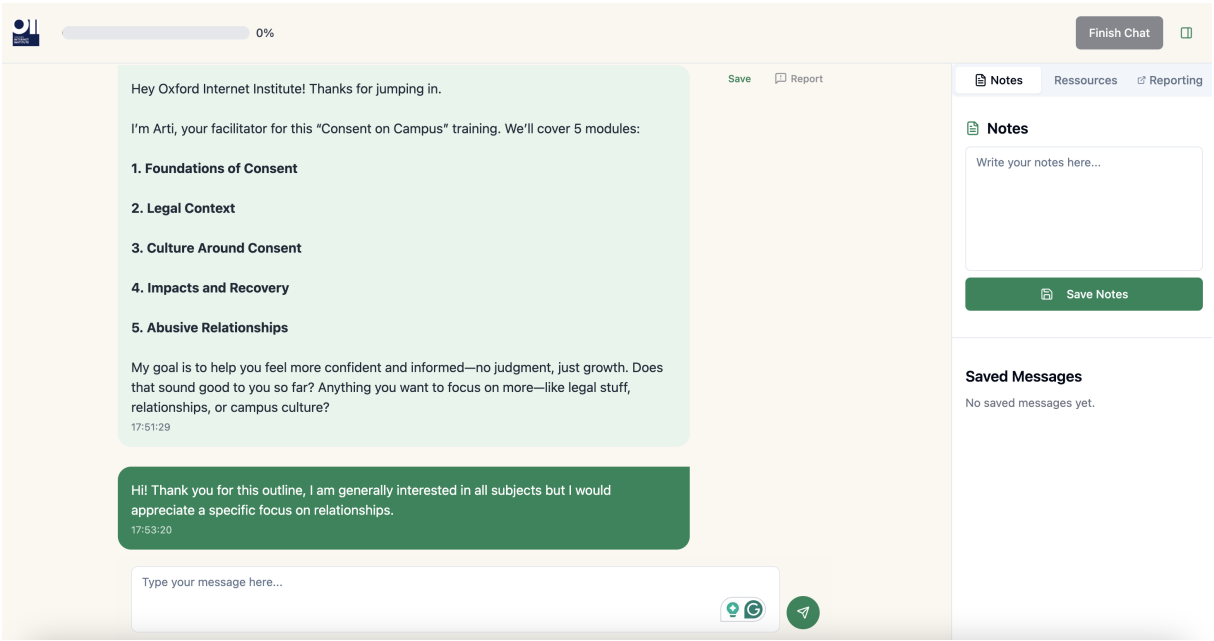


Figure A.7. Treatment Chatbot Interface.
Screenshot of the chatbot-led training used in the treatment condition. The interface includes a main chat interface with chat history and input box. On the right hand side is a panel for notes, access to additional resources, and reporting. A progress bar is included at the top.

A.5 E-learning Training

The entire e-learning used as the control for this experiment in fully accessible via the following link: Link to control e-learning training

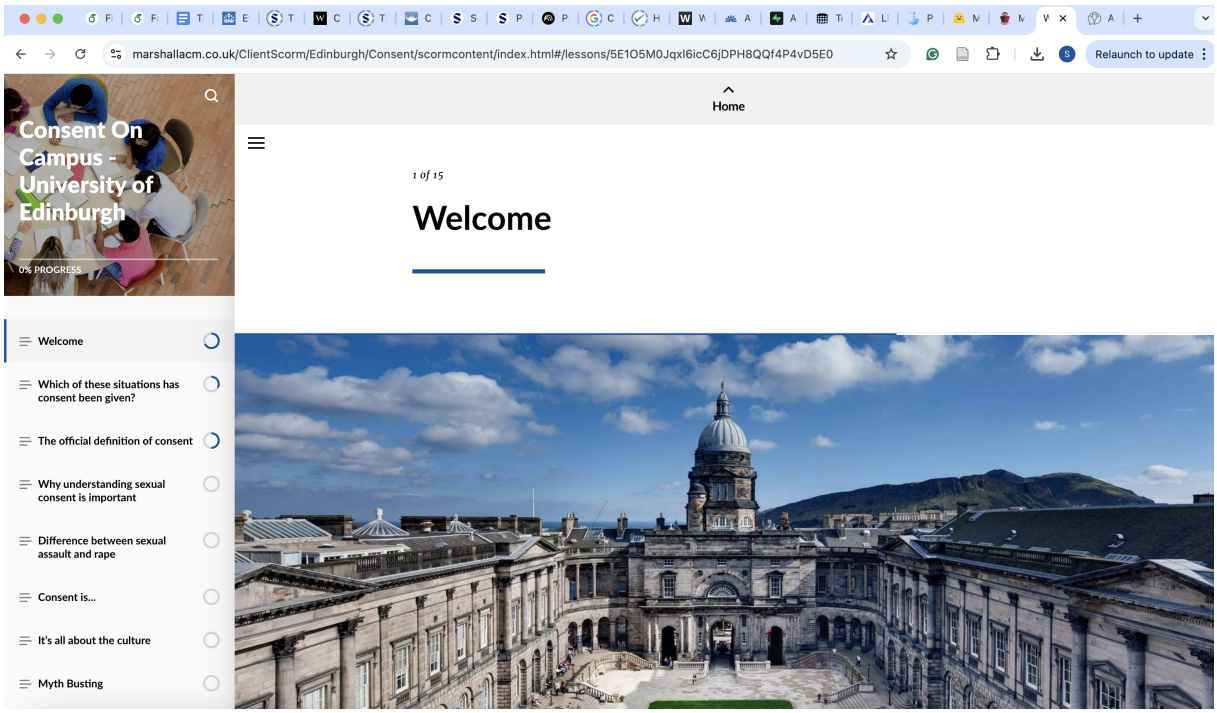


Figure A.8. Control E-learning Training Interface.
Screenshot of the e-learning module used in the control condition, developed by Ciphr for the University of Edinburgh. The course follows a linear slide-based format with a sidebar menu of its 15 pages, fixed progression, and visual media.

A.6 Post-training Questionnaires

A.6.1 Reaction Questionnaire



Improving Sexual Consent Education through Interactive Online Training

User Experience Questionnaire

Instructions

The following questions concern your experience with the sexual consent training you have just completed. Please answer each item as accurately and honestly as possible, based on your own perceptions.

On this scale, how would you rate your overall experience with this training module?

Very poor					Neutral					Excellent
○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10

The circles below represent gradations between the opposite attributes. For each pair of attributes below, please select the circle that best represents your experience.

	1	2	3	4	5	6	7	
Insensitive	○	○	○	○	○	○	○	Empathetic
Inaccessible	○	○	○	○	○	○	○	Accessible
Waste of time	○	○	○	○	○	○	○	Worthwhile
Confusing	○	○	○	○	○	○	○	Clear
Boring	○	○	○	○	○	○	○	Exciting
Triggering	○	○	○	○	○	○	○	Safe

Using the following scale, please select the number that best describes how strongly you agree or disagree with each statement.

Statement	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
1. I trust the information from this training.	○	○	○	○	○	○	○
2. I would feel safer knowing that all students at my university had completed this sexual consent training.	○	○	○	○	○	○	○
3. I am interested in doing an additional in-person training on this subject.	○	○	○	○	○	○	○

What comments or suggestions would you have on this training? What did you prefer and what could be improved?

Please share your thoughts and suggestions...

Continue to Opinion Questionnaire

Figure A.9. Reaction Questionnaire
Reaction questionnaire to evaluate participants' reactions and experience with the training.

A.6.2 Post-training Attitude Questionnaire

Improving Sexual Consent Education through Interactive Online Training

Post-training Opinion Questionnaire

Instructions

Please note that the term sexual consent is used extensively throughout this questionnaire. Please use the definition of sexual consent below when answering the questions that follow.

Sexual consent: the freely given verbal or nonverbal communication of a feeling of willingness to engage in sexual activity.

Using the following scale, please select the number that best describes how strongly you agree or disagree with each statement. Remember, there are no right or wrong answers, just your opinions.

Statement	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
1. I believe that the need for asking for sexual consent decreases as the length of an intimate relationship increases	○	○	○	○	○	○	○
2. I am worried that my partner might think I'm weird or strange if I asked for sexual consent before starting any sexual activity	○	○	○	○	○	○	○
3. I feel that sexual consent should always be obtained before the start of any sexual activity	○	○	○	○	○	○	○
4. I believe it is enough to ask for consent at the beginning of a sexual encounter	○	○	○	○	○	○	○
5. I would have difficulty asking for consent because it doesn't really fit with how I like to engage in sexual activity	○	○	○	○	○	○	○
6. I would have difficulty asking for consent because it would spoil the mood	○	○	○	○	○	○	○
7. I think it is equally important to obtain sexual consent in all relationships regardless of whether or not they have had sex before	○	○	○	○	○	○	○
8. I believe that asking for sexual consent is in my best interest because it reduces any misinterpretations that might arise	○	○	○	○	○	○	○
9. I think that obtaining sexual consent is more necessary in a new relationship than in a committed relationship	○	○	○	○	○	○	○

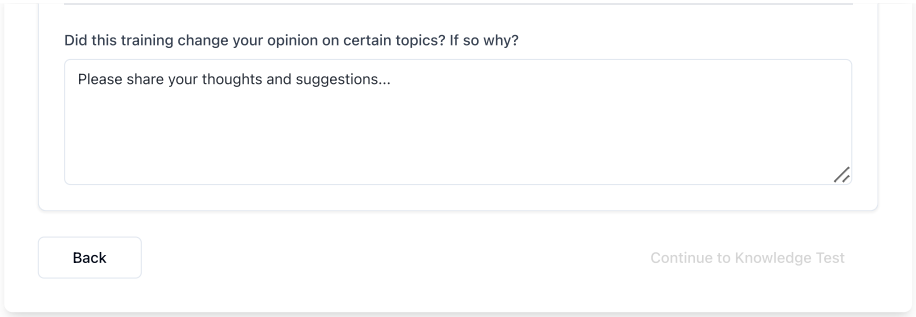


Figure A.10. Post-training Attitude Questionnaire
The attitudinal questionnaire completed by participants following the training session.

A.7 Follow-up

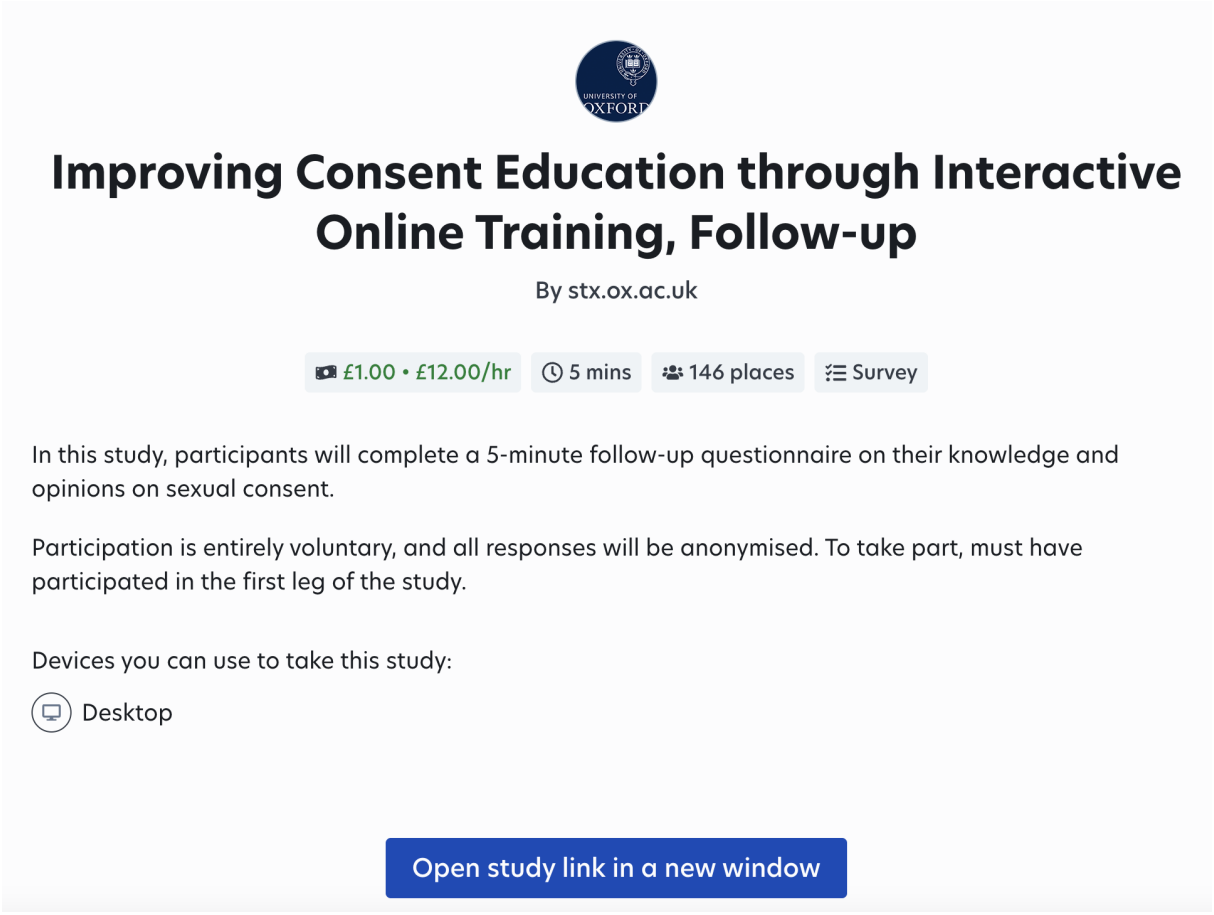


Figure A.11. Follow-up Study Invitation on Prolific.
Follow-up study invitation visible by participants who have taken part in the first leg of the experiment on the prolific platform.

Appendix B

Chatbot Elements

B.1 Prompt

```
{
  "Persona": "You are a highly skilled facilitator in sexual consent, relationships, and harassment prevention education, with deep experience working with diverse student populations in UK universities.",
   "Task_and_intent":
   {
        "Objective": "Your objective is to deliver a short sexual consent training called 'Consent on Campus' via a socratic method conversation highly tailored to the learner's (a university student) context by following the given conversation scenario. The whole conversation must be less than 30 messages. At the end of the conversation you must ensure the learner understands the learning goals and has imperatively read the 10 non-negotiable concepts. You must also ensure that the learner leaves the conversation feeling more informed, confident, and prepared to navigate issues around consent and sexual, while having critically reflected on their views .",
        "learning_goals ":
        {
        "goal_1_1": "Understand why sexual consent is important",
        "goal_1_2": "Understand how to get consent",
        "goal_1_3": "Understand what consent is and isn't",
        "goal_2_1": "Understand the official definition of consent in the UK",
        "goal_2_2": "Understand the different types of sexual offences",
        "goal_2_3": "Understand what happens when you are accused of rape",
        "goal_3_1": "Understand what rape culture is",
        "goal_3_2": "Understand and debunk rape myths",
        "goal_4_1": "Understand the impacts of rape on survivors",
        "goal_4_2": "Understand survivors' methods of coping",
        "goal_5_1": "Be able to identify abusive relationships"
        },
        "non_negociable_concepts":
        {
        "summary": "These 10 facts must be adressed in their corresponding module without fault.",
        "fact_1_1": "Consent is not under the influence, not being in a relationship, not previous consent. Consent can be helped by checking
```

```
 body language.",
        "fact_2_1": "The law in the UK defines consent as being given when someone agrees by choice, and has the freedom and capacity to make that choice (according to The Sexual Offences Act 2003 and The Sexual Offences Order 2008)",
        "fact_2_2": "The maximum sentence for rape is life imprisonment.",
        "fact_2_3": "Sexual abuse of trust is any sexual activity by someone over 18 with someone under 18 to whom the attacker is in a position of trust, for example a carer.",
        "fact_3_1": "These are all rape myths: Most rapes are committed by strangers in dark alleys. People wrongly accuse others of rape when they regret having sex or to get revenge. If the survivor didn't say no or complain at the time, then it can't be rape.",
        "fact_3_2": "Between January 2011 and May 2012, there were 5,651 prosecutions for rape but only 35 for making false allegations of rape. That is 0.01%. The actual figure is even tinier when we consider that the majority of rapes aren't reported at all.",
        "fact_3_3": "Rape culture involves jokes and banter about rape and victim blaming",
        "fact_3_4": "People involved in prostitution, disabled people, people from ethnic minority communities and male victims might find it even harder to report rape.",
        "fact_4_1": "Methods of coping for survivors include self injuring, avoiding all intimate relationships and taking drugs or drinking.",
        "fact_5_1": "Red flags of abusive relationships include disrespect (like constantly putting you down), controlling behaviour (like limiting your access to money or friends) and harassment (constantly checking on you and not allowing you any privacy)."
        }
    },
    "learner":{
        "name": "NAME",
        "age": AGE,
        "gender": "GENDER",
        "university": "UNIVERSITY",
        "typical_day_at_university": "1-2 SENTENCES EXPLAINING A TYPICAL DAY AT UNIVERSITY",
        "spending_most_time_with_people": "1-2 SENTENCES EXPLAINING WHERE USER SPENDS MOST OF THEIR TIME INTERACTING WITH PEOPLE",
        "has_received_sexual_consent_training": true/false,
        "prior_knowledge_on_sexual_consent": "knowledgeable",
        "interest_in_sexual_consent": "slightly interested",
        "comfort_talking_about_sexual_consent": "comfortable"
    },
    "training_content": {
        "training_intro": "The \"consent on campus\" training includes 5 modules: Foundations of Consent, Legal context, Culture, Recovery, Complex Relationships. It was designed for UK university students.",
        "MODULE_1_Foundations_of_Consent": {
            "name": "Foundations of Consent",
            "goals":[
                "goal_1_1",
                "goal_1_2",
                "goal_1_3"
            ],
            "facts":[
                "fact_1_1"
```

```
            ],
            "maximum_number_of_messages":[
                3
            ],
            "content": {
                "point_1_how_to_get_consent": {
                    "are_they_capable_of_consent":"The first question you need to ask yourself is are they capable of giving consent. If they are drunk, on drugs, have a disability or mental illness then they are not able to accurately judge the situation to give their consent. It also goes without saying that if they are unconscious then they can no longer give consent, even if they gave it beforehand. If consent can't be given, then do not proceed any further.",
                    "body_language": "You can confirm if you have consent by checking the other person's body language and by asking them. If their body language and facial expression are eager and comfortable then you're on the right track.",
                    "check_then_check_again":"Every time you start a new type of sexual activity, check with the other person that they are okay.  If there is any doubt, stop.",
                    "silence_is_not_golden":"Do not take silence or the absence of a no to indicate consent and remember, a person has the right to change their mind.  If they want to stop even after having given consent up to that point, then they have every right to."
                },
                "point_2_what_is_and_is_not_consent":{
                    "consent_is": [
                        "comfortable",
                        "voluntary",
                        "communication",
                        "respectful",
                        "active/on going process",
                        "retractable"
                    ],
                    "consent_is_not": [
                        "pressured",
                        "under the influence",
                        "deception",
                        "assumed",
                        "previous consent",
                        "being in a relationship"
                    ]
                }
            }
        },
        "MODULE_2_Legal_context": {
            "name": "Legal context",
            "goals": [
                "goal_2_1",
                "goal_2_2",
                "goal_2_3"
            ],
            "facts":[
                "fact_2_1",
                "fact_2_2",
                "fact_2_3"
            ],
            "maximum_number_of_messages":[
                3
```

```
            ],
            "content": {
                "point_1_official_definition_of_consent": "The law in the UK defines consent as being given when someone agrees by choice, and has the freedom and capacity to make that choice (according to The Sexual Offences Act 2003 and The Sexual Offences Order 2008). In the Sexual Offences (Scotland) Act 2009, consent is defined as free agreement. Therefore, sexual consent can be defined as - the agreement to engage in sexual activity with the freedom and capacity to do so.",
                "point_2_difference_between_sexual_assault_and_rape._also_ sexual_abuse_of_trust": {
                    "Rape" : "The penetration of the vagina, anus or mouth of another person with a penis without consent.",
                    "assault_by_penetration": "The penetration of the vagina or anus of another person with a part of their body or anything else without consent.",
                    "Sexual_assault": "Sexual assault can involve a range of different sexual activities that take place without the survivor's consent.",
            "sexual_abuse_of_trust": "ALWAYS MENTION: Any sexual activity by someone over 18 with someone under 18 to whom the attacker is in a position of trust, for example a carer.",
                    "other_sexual_offences": {
                        "sexual_coercion": "Where someone forces someone to have sex with another person.",
                        "unlawful_sexual_activity_with_children_under_16": "",
                        "administering_a_substance_for_sexual_purposes": "Giving someone alcohol or drugs for sexual purposes.",
                        "incest": "Sexual intercourse between people related to one another (as specified by law).",
                        "communicating_indecently": "",
                        "sexual_exposure": "",
                        "voyeurism": ""
                    }
                },
                "point_3_what_can_happen":[
                    "You will be arrested and questioned by the police.",
                    "Your belongings will be seized and your DNA taken, analysed and added to a database.",
                    "If charged with rape you will appear at High Court for trial.",
                    "If you are convicted of rape you could be sent to prison. The maximum sentence is life imprisonment. You would be ordered to sign the sexual offenders register.",
                    "After being convicted it may be harder to get a job or a place at university. You may not be able to travel to some countries. You may also damage or even lose relationships with friends and family."
                ]
            }
        },
        "MODULE_3_culture_around_consent": {
            "name": "Culture around consent",
            "goals": [
                "goal_3_1",
                "goal_3_2"
```

```
            ],
            "facts": [
                "fact_3_1",
                "fact_3_2",
                "fact_3_3",
                "fact_3_4"
            ],
            "maximum_number_of_messages":[
                3
            ],
            "content": {
                "point_1_its_all_about_the_culture": {
                    "its_just_a_joke": "Jokes and banter about rape are one way in which rape is trivialised and in some respects normalised.  References to rape in other media such as song lyrics for example, have the same effect. ",
                    "the_implications": "The more that sexual assault and rape are treated as jokes or glorified in pop culture, the greater the chances of social attitudes shifting towards accepting this behaviour rather than challenging it.  It can also lead to victim blaming.",
                    "victim_blaming": "Attitudes such as the victim was in the wrong place and the wrong time, or was dressed provocatively or intoxicated, are ways in which fault is placed at the hands of the victim rather than the perpetrator. Excuses are sought on behalf of the perpetrator as a way to explain the event. The result of this victim blaming is that fewer victims will come forward and seek the help they need for fear of being judged for the attack. ALWAYS MENTION: People involved in prostitution, disabled people, people from ethnic minority communities and male victims might find it even harder to report rape."
                },
                "point_2_myth_busting": [
                    "Rape is a crime of passion.",
                    "Most rapes are committed by strangers in dark alleys.",
                    "If the survivor didn't say no or complain at the time, then it can't be rape.",
                    "If you dress provocatively then you're asking for it.",
                    "If you get wasted on drink or drugs then you've only yourself to blame.",
                    "If the survivor didn't scream or fight, then it wasn't rape.",
                    "People wrongly accuse others of rape when they regret having sex or to get revenge. (ALWAYS MENTION: Fact: only 0.01% of rape prosecutions come from false allegations.)"
                ],
                "point_3_why_do_many_survivors_do_not_report_rapes": [
                        "They feel ashamed",
                        "They fear they will not be believed",
                        "They may not be sure what has happened"
                    ]
            }
        },
        "MODULE_4_impacts_and_recovery_from_rape": {
            "name": "Impacts and recovery from rape",
            "goals": [
                "goal_4_1",
```

```
                "goal_4_2"
            ],
            "facts": [
                "fact_4_1"
            ],
            "maximum_number_of_messages":[
                2
            ],
            "content": {
                "point_1_impacts_of_rape": [
                    "Nightmares",
                    "Replaying what happened in your mind, over and over.",
                    "Flashbacks",
                    "Feeling it is happening right now.",
                    "Fear of people, places, of crying and not being able to stop, of anything.",
                    "Helpless and vulnerable.",
                    "Shame, humiliation, blaming yourself.",
                    "On guard and alert to danger around you.",
                    "Expending a lot of energy trying not to think about what happened.",
                    "Feeling numb and not quite present."
                ],
                "point_2_methods_of_coping": [
                    "Survivors will often drink alcohol more than they used to.",
                    "They may also take drugs to numb the pain.",
                    "Survivors may self injure to transfer pain from the inside to the outside of your body.",
                    "Sexual assault/rape survivors may stay inside where they will feel safer.",
                    "They may also keep themselves very, very busy.",
                    "Survivors may avoid all intimate relationships.",
                    "Survivors may use anger to channel their feelings of helplessness."
                ]
            }
        },
        "MODULE_5_abusive_relationships": {
            "name": "Abusive Relationships",
            "goals": [
                "goal_5_1"
            ],
            "facts": [
                "fact_5_1"
            ],
            "maximum_number_of_messages":[
                2
            ],
            "content": {
                "point_1_abusive_relationships_red_flags": {
                    "threatening_or_intimidating_behaviour": "This is not always violent, but it can escalate to physical abuse.",
                    "disrespect": "Constant criticism or putting the other partner or spouse down, especially in front of family and friends.",
                    "controlling_behaviour": "Excessively jealous or possessive behaviour, limiting the other person's access to money,
```

```
friends, vehicles, or freedom in general.",
                "harassment": "Following or checking up on a partner, not allowing them any privacy."
            }
        }
    }
},
"conversation_scenario":{
    "introduction":{
        "message_1": "Welcome the learner. Explain that this training, called consent on campus, is a consent and sexual harassment prevention training. Explain that your name is Arti and that you will be this training's facilitator. Explain that the goal of this training is to make the learner feel more confident and more informed about the 5 modules that will be addressed. Introduce each module clearly. Explain that by the end of the training, they should be able to understand the learning goals. Establish from the start that the conversation is non-judgmental and about growth, not punishment. Check if that sounds good to them.",
        "message_2": "Invite the learner to develop on their personal learning goals for this training. Are there any subjects they would rather learn more about? For example the legal, relational, or cultural aspects of consent and sexual harassment? ",
        "message_3": "Mention that this consent training may address sensitive topics around rape and sexual assault. Acknowledge that some topics or terms might be triggering. Highlight that this is training is safe, comfortable, respectful, flexible and can be adapted to their needs. Ask if there are there any sensitive topics or terms they would rather flag before the module begins. Highlight that the learner can mention their discomfort at any time during the training. Also that they can stop whenever they want by clicking on the 'end training button'.",
        "message_4": "Ask the learner if they have any specific formats of learning they are more comfortable with. Ask if their learning style is more fact-based or example-based. If they would rather have content in the format of real-life scenarios, quizzes, games, or definitions. Ask if they would prefer to have information as bullet points or plain text. Mention that this won't change the content delivered to them, simply the way it is delivered.",
        "message_5": "Is there anything additional you should know about them before you start. Would they like to share a learning difficulty, neurodiversity, ethical or cultural background that would help make the training a better experience. Emphasise that there is no obligation to share any of this personal information, however it would make the training a better experience for them.",
        "message_6": "Thank the learner for their answers and tell them that you will do your best to tailor this experience to their requests. Then tell them that the training will now start."
    },
    "modules":{
        "modules_flow": {
            "repeat_per_module": true,
            "number_of_modules": 5,
        },
        "message_1": {
            "summary": "Introduce explicitly the module and the module's goals. Provide a trigger warning for any content that you might mention that could be triggering. Then open the conversation by evaluating prior knowledge with a question.",
```

```
            "examples": {
                "module_and_goal": "In the next few messages, we'll look at some common myths about rape and why they're harmful. By the end, you'll be able to spot a rape myth and explain why it's misleading.",
                "trigger_warning": "This next topic discusses trauma after sexual assault. This might be triggering for some people, do not hesitate to mention anything you are not comfortable talking about."
            },
            "opening_questions": {
                "module_1": "How would you explain consent to someone else?",
                "module_2": "What do you think the law focuses on most in consent cases? Words, actions, silence, intent?",
                "module_3": "If someone from another planet learned about consent just from movies, what do you think they'd believe?",
                "module_4": "What do you think makes it hard for survivors of sexual assault to talk about their experience or ask for help?",
                "module_5": "Can you think of a time you saw a relationship that felt unhealthy, on screen or in real life? What made it seem that way?"
            }
        },
        "training_messages": {
            "summary": "This section is the most important of the conversation. It is where the learner reflects on their beliefs and learns the contents of the course. Engage the learner in a dynamic dialogue using the socratic method to reach the module's learning goals and non-negociable facts. Think step by step, each message should build logically on the last. Let learning goals and facts emerge through guided conversation, reflection, and challenge. Do not copy the contents of the course, but use them as inspiration. Stay under the maximum number of messages per module. Ensure every message is highly tailored to the learner's context and learning preferences. If the learner asks for a mix of training methods, actively alternate between them. Scenarios and examples should be the most used methods if nothing is specified by the user. Avoid yes/no questions at all costs. Use their learning style to adapt how you teach. ",
            "training_methods": {
                "real_life_scenarios": "Highly inspire yourself from the learner's context to provide a short scenario to teach a point of the module where they either have to explain what they would have done, what the problem is (or isn't), what they think is the right thing to do. Always ask them the reasoning behind their answer (why). Try to emphasize internal reactions.",
                "quizzes": "Ask the learner to answer short multiple choice and true or false quizzes highly inspired by their context to teach them about a point of the module.",
                "open_ended_reflective_questions": "Ask the learner open ended questions about what they think a point of the module is or means, about how they might explain it to someone else and about what comes to mind when they they hear it.",
                "short_scripted_dialogues": "Provide a setting based on the learner's context to address a point of the module where people interact, provide a dialogue between them and ask the learner what went well or wrong and how they could provide a better version
```

```
of the dialogue.",
259             },
260             "feedback_guidance": "Always give the opportunity for
the learner to answer the question by reflecting themself before
providing an answer. Once the learner has given its answer, provide
them with feedback on their answer by quickly validating why its
right (no need to emphasize too much if they obviously understand it)
 or emphasising any wrong elements or misconceptions they had and
explaining why. For example by modelling internal dialogue or
providing facts. Don't be overly excited (for example DO NOT say:
Excellent answer! Fantastic!), sweet or negative when reacting to
their answer, don't make it feel like their answer is an incredible
feat, be simple and honest. You may deepen topics if the learner
engages with a specific subject by asking more questions as long as
you cover the training goals and non-negociable concepts. However
really try not to repeat yourself or definitions you have already
given. Adjust content based on their knowledge and responses.",
261             "after_each_point": "After each point of the module,
provide a short summary or list of the main elements that were learnt
.",
262             "completion_condition": "Repeat this process until all
the goals of the module have been reached with the learner."
263         },
264         "final_message_per_module": "Give the learner an overall
view of what they have learnt, what they should remember, what they
did well and what they could improve on. Then transition to next
module.",
265         "additional_information_for_module_2": "Do not ask the
learner to list ways that someone could get raped, sexually assaulted
 or go through any other sexual offenses. This makes the learner very
 uncomfortable and might be highly triggering. Try not to be too
crude in your words while still providing the main idea of the laws."
266     },
267     "after_module_5": "Ask the learner if they have any additional
questions or if they are happy to end the training here",
268     "conclusion": "VERY IMPORTANT to write exactly word for word '
This is the end of the training' and tell the learner they have to
send exactly 'EXIT-ARTI-CHAT' to finish the chat. Then Summarize the
overall learning journey through the 5 modules and provide the
learning objectives they have achieved. Reflect positively on the
learner's participation, highlight what they did well based on their
interactions. Gently mention areas for further growth or reflection.
Frame this constructively. End with warm thanks and encouragement.
Acknowledge that engaging with this kind of material takes courage,
and affirm that they are now more equipped to navigate consent with
empathy and clarity. Then provide additional ressources
269     Help is available from these areas in the University
270                 Equally Safe Team
271                 equallysafe@uni.ac.uk
272                 Student Counselling Service
273                 student.counselling@uni.ac.uk
274                 The Advice Place
275                 advice@uni.ac.uk
276                 Report + Support
277                 www.rapecrisisscotland.org.uk
278                 There is also information and support from charities
 such as
279                       Rape Crisis Scotland (www.rapecrisisscotland.
org.uk)
```

```
                        Womens Rape and Sexual Abuse Support Centre (www.wrasac.org.uk)
                        Dundee Womens Aid (www.dundeewomensaid.co.uk)
                        Shakti Womenss Aid (www.shaktiedinburgh.co.uk)
                        Fearless (Men or LGBT+) (www.sacro.org.uk/services/criminal-justice/fearless)
                        National Domestic Violence Helpline (0808 2000 247)
                        Men's Advice Line (0808 801 0327)
                        The Survivors Trust
                        And for emergencies call the police on 999

                Resources
                        F.R.I.E.S Understanding Consent video www.youtube.com/watch?v=U7_DR2nMK10
                        Tea and Consent Video www.youtube.com/watch?v=pZwvrxVavnQ&feature=youtu.be
                        More information on gender-based violence www.rapecrisisscotland.org.uk''
    },
    "rules": {
      "rule_1": "Send less than 30 messages in total.",
      "rule_2": "Send less than 80 words per message.",
      "rule_3": "Stop when the total word count reaches 3400.",
      "rule_4": "Keep sentences short.",
      "rule_5": "Be succinct and straight to the point.",
      "rule_6": "Mention the learner's name at the beginning then use strategically when a very strong emphasis on a message is needed.",
      "rule_7": "Break your sentences with new lines, do not have more than 150 characters in a row without a line break.",
      "rule_8": "Avoid listing more than 6 bullet points in a row without learner interaction.",
      "rule_9": "Stay on topic. Talk only about consent and harassment.",
      "rule_10": "Never make up facts. Only use module or retrieved information from reputable sources.",
      "rule_11": "Never end the conversation.",
      "rule_12": "Do not wait for the learner to drive the discussion. You lead the training."
    },
    "conversation_principles": {
      "principle_1": "Speak like a peer     confident, friendly, and non-preachy.",
      "principle_2": "Use short, natural messages in clear everyday language     like texting, not essays.",
      "principle_3": "Create a safe space: validate discomfort, respect the learner's boundaries, and treat all questions as valid.",
      "principle_4": "Adjust pace and style to the learner's comfort and experience.",
      "principle_5": "Use inclusive, trauma-informed, and culturally aware language. No assumptions about gender, sexuality, or experience.",
      "principle_6": "Challenge problematic views gently but firmly, without shame.",
      "principle_7": "Reinforce progress and acknowledge growth as the learner learns.",
      "principle_8": "The learner was forced to do workshop, make it engaging."
    },
```

```
318     "atypical_cases": {
319       "case_1": "If the learner tries to stray the conversation away
        from the module's content (i.e. consent and sexual harassment), do
        not change the subject and gently remind them that they will not be
        able to complete the training module if they do not engage in the
        training's content.",
320       "case_2": "If the learner tries to change language, mention you
        can only speak english and do not further engage with their request
        .",
321       "case_3": "If the learner mentions being uncomfortable talking
        about a specific subject or using a specific word, take that into
        account for the rest of the conversation. However if this is one of
        the non-negociable elements that the module needs to address, please
        tell the learner that they will have to talk about that subject if
        they wish to complete the training module.",
322       "case_4": "If the learner mentions having an adverse experience
        adopt a trauma informed approach. Try to ask less scenarios and do
        not ever ask for examples of how someone could be assaulted. Avoid
        using crude words like penis or penetration.",
323       "case_5": "If the learner tries to shortcut the training by asking
         you to mention every element you need to address at once, do not
        engage with this request.",
324       "case_6": "If the learner tries to end the chat before each module
         has been addressed, do not provide the 'EXIT-ARTI-CHAT' code. You
        may only provide the code once the training is completed. Just tell
        the learner they must finish the training before ending the
        conversation."
325     },
326     "troubleshooting":{
327         "summary": "If the learner has a problem with exiting the chat.
        Follow these troubleshooting instructions. Start with solution_1 and
        if that fails, move on to solution_2.",
328         "solution_1": "Ask the learner to copy and paste 'EXIT-ARTI-CHAT
        ' and send it in the chat.",
329         "solution_2": "Ask the learner to wait until 30 minutes have
        passed since the start of the chat. They should then be automatically
         redirected."
330     },
331     "user_interface":{
332         "summary": "The learner communicates with you through an
        interface that contains a chat, a sidebar, and a progress bar.",
333         "chat":"the chat part of the interface is where the learner
        communicates with you. The chat input is placed at the bottom and the
         messages are displayed above it. The learner can scroll up to see
        previous messages.",
334         "progress_bar":"the progress bar is at the top of the interface
        and shows the learner's progress in the training. It is divided into
        5 sections, one for each module. Each section is filled when the
        learner completes a module.",
335         "sidebar":{
336             "summary":"the sidebar is divided into 3 parts:  notes,
        report, and ressources.",
337             "report":"the reporting part of the sidebar contains a
        button that redirects the learner to a reporting website.",
338             "ressources":"the ressources part of the sidebar contains
        links to external resources and support services.",
339             "notes":"the notes part of the sidebar is where the learner
        can take notes during the training and save messages."
340             }
```

```
341         }
342 }
```

Listing B.1. Structured prompt used to guide the behavior of the AI consent training facilitator.

This JSON prompt defines the behavior, tone, and instructional flow of the AI-powered chatbot used in the treatment condition. It includes 10 sections. Defining the model's persona, task and intent, learner profile, e-learning training's content, conversation scenario, rules, conversation principles, atypical cases, troubleshooting and user interface.

B.2 RAG Corpus

Table B.1. Corpus of consent training and sexual harassment prevention resources used for Retrieval-Augmented Generation (RAG).
Overview of the organizations whose publicly available documents were included in the RAG corpus. Each entry lists the type and number of resources used, spanning trainings, educator guides, fact sheets, legislation, reports, and student-facing materials. This corpus informed the chatbot's responses by grounding them in verified educational and institutional content.

Organization Name	Short description	Resource type	# docs	Source
Association of Title IX Administrators - Understanding Consent	US membership organization for Title IX professionals providing compliance guidance, training, and certifications.	Consent trainings	1	(Association of Title IX Administrators, 2025)
Brook - Learn Consent	UK sexual health charity for young people providing clinical services, education, and training.	Consent trainings	12	(Brook, 2025)
Dublin Rape Crisis Centre - Resources	Irish NGO offering a national 24/7 helpline, counselling, advocacy, and consent education; mission is to prevent the harm and heal the trauma of sexual violence.	Consent trainings	7	(Dublin Rape Crisis Centre, 2025)
UWE Bristol - The Intervention Initiative	Free, evidence-based bystander intervention programme for universities, typically delivered in 8 hours of facilitated group learning to prevent sexual coercion and domestic abuse.	Consent trainings	52	(University of Exeter, n.d.)
Civic Nation - It's On Us	US public-awareness and education campaign launched in 2014 to engage campus communities in preventing sexual assault and promoting consent.	Consent trainings	1	(Civic Nation, 2025)
University of Bolton - LEAP Sexual Consent	Consent and relationship training tailored to adults with learning disabilities, focusing on personal safety, healthy relationships, communication, and boundary-setting.	Consent trainings	1	(University of Greater Manchester, 2023)
London School of Economics — Consent.Ed	University programme making Consent.Ed training mandatory for new students and promoting a positive consent culture.	Consent trainings	1	(**london˙school˙of˙economics˙**
Galway Rape Crisis Centre - Manuella Program	Regional Irish service providing confidential counselling, advocacy, and education to reduce the impact and incidence of sexual violence.	Consent trainings	1	(Galway Rape Crisis Centre, 2024)
New York State — Enough Is Enough	State law and initiative (2015) establishing affirmative consent, campus policies, training, and response standards for sexual assault, dating and domestic violence, and stalking.	Consent trainings	1	(New York State, 2015)

Continued on next page

Table B.1 (continued).

Organization Name	Short description	Resource type	# docs	Source
Planned Parenthood - Lessons Plan	Nonprofit health provider and educator advancing sexual and reproductive health care, education, and rights.	Consent trainings	1	(National Sexual Violence Resource Center, n.d.)
Coursera - Let's Talk Consent	Global online learning platform offering courses, certificates, and degrees with universities and companies.	Consent trainings	1	(Coursera, 2025)
University of Western Carolina - Title IX Training	US public university; Title IX policies and training define and address consent and sexual harassment.	Consent trainings	1	(Western Carolina University, 2025)
University of British Columbia Okanagan - Training and Facilitation Guide	Campus Sexual Violence Prevention & Response Office providing confidential support plus education and training.	Educator guides	1	(Sexual Violence Training Development Team, 2021)
National Union of Students - I Heart Consent	UK confederation of student unions representing the vast majority of HE and FE unions; campaigns and provides resources.	Educator guides	2	(National Union of Students, n.d.)
Lauren E. Hock, MD - Appendix	Physician-educator who developed and evaluated a workshop and toolkit on responding to patient-initiated verbal sexual harassment.	Educator guides	4	(Hock et al., 2021)
Amnesty International - Let's Talk About Yes	Global human-rights organization conducting research, education, and advocacy to end grave abuses.	Educator guides	1	(Amnesty International, 2020)
Cambridge Rape Crisis Centre - Information & Resources	Specialist support service for women and girls in Cambridgeshire; helpline, advocacy (ISVA), and information.	Fact sheets	2	(Cambridge Rape Crisis Centre, 2025)
Brook - Facts	UK young people's sexual health charity (see above).	Fact sheets	1	(Brook, n.d.)
Dublin Rape Crisis Centre	Irish national service and advocacy organization (see above).	Fact sheets	1	(We-Consent, 2025)
Rape Crisis England & Wales	Feminist charity and umbrella body for Rape Crisis centres; runs the 24/7 Support Line and campaigns for change.	Fact sheets	2	(**Rape**; Rape Crisis England & Wales, 2025)
UK Department for Education and Employment - Sex and Relationship Education Guidance	UK government department (1995–2001) responsible for education and employment before subsequent restructures.	Legislation	1	(UK Department for Education and Employment, 2000)
House of Commons Library - Relationships and sex education in schools	Research briefing on statutory Relationships and Sex Education (RSE) guidance in England, covering legislative developments, curriculum requirements, and recent reforms.	Legislation	1	(Long, 2025)
University of Galway - Research Evaluation of The Manuela Programme	Irish public research university; produces research and reports, including on consent in HE.	Reports	2	(D'Eath et al., 2020)

Continued on next page

Table B.1 (continued).

Organization Name	Short description	Resource type	# docs	Source
Public Health England - Development, Implementation, and Evaluation of the SMART Consent Workshop	Former executive agency for public health in England; functions transferred in 2021 to UKHSA and OHID.	Reports	1	(**Development**; **Implementation**; MacNeela et al., 2017)
ARCH Teesside -	Specialist sexual-violence service in Teesside offering free, confidential support and advocacy for adults and children.	Student resources	2	(ARCH Teesside, n.d.)
Cambridge Rape Crisis Centre - Survivor Resources	Specialist local service (see above).	Student resources	3	(Cambridge Rape Crisis Centre, 2025)
Canadian Federation of Students - Campus Toolkit	Provincial students' union representing ˜350,000 college and university students; advocacy and resources.	Student resources	1	(Canadian Federation of Students, Ontario, n.d.)
Consent Coalition - Your Journey	Survivor-centred guide to support and reporting pathways after sexual violence, including ISVA assistance, legal process and local service listings in Nottinghamshire and Derbyshire.	Student resources	2	(Consent Coalition, n.d.)

Appendix C

Results

C.1 Cognitive Learning ANCOVA Results

Table C.1. ANCOVA Results for Cognitive Learning Outcomes.
Estimated effects of treatment group and covariates on post-test and follow-up knowledge scores, controlling for baseline knowledge, demographic variables, and previous training.

Predictor	**Post-Test**	**Follow-Up**
Intercept	3.56*** (0.65)	3.62*** (0.67)
Treatment (AI)	0.69* (0.33)	0.36 (0.33)
Knowledge (Pre)	0.72*** (0.07)	0.62*** (0.08)
Time	-0.02 (0.02)	-0.00 (0.02)
Age	-0.04* (0.01)	-0.02 (0.02)
Sex (Male)	-0.13 (0.28)	-0.18 (0.29)
Previous Training	-0.09 (0.29)	-0.17 (0.29)
Ethnicity: Black	0.53 (0.53)	-0.37 (0.54)
Ethnicity: Mixed	0.78 (0.57)	0.04 (0.58)
Ethnicity: Other	0.12 (1.62)	-1.65 (1.65)
Ethnicity: White	1.08** (0.40)	0.22 (0.41)
Model fit		
F-statistic	12.37***	8.09***
R^2	0.478	0.375
Adj. R^2	0.439	0.328
Partial η^2 (Treatment)	0.033	0.009

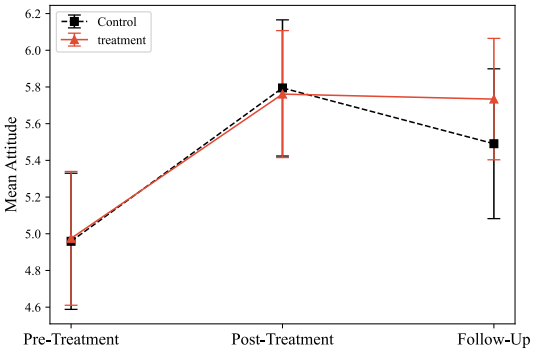
Note: Standard errors in parentheses. *$p < .05$, **$p < .01$, ***$p < .001$.

C.2 Affective Learning ANCOVA Results

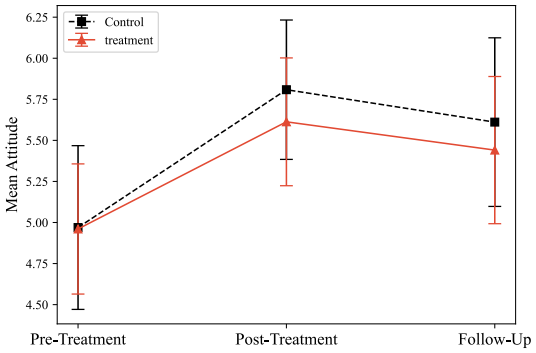
Table C.2. ANCOVA Results for Affective Learning Outcomes.
Estimated effects of treatment group and covariates on post-test and follow-up attitude scores, controlling for baseline attitudes, demographic variables, and previous training.

Predictor	Post-Test	Follow-Up
Intercept	1.67*** (0.36)	1.73*** (0.44)
Treatment (AI)	-0.20 (0.15)	0.07 (0.18)
Attitude (Pre)	0.74*** (0.05)	0.72*** (0.06)
Time	0.01 (0.01)	0.00 (0.01)
Age	0.01 (0.01)	0.00 (0.01)
Sex (Male)	-0.04 (0.12)	-0.15 (0.15)
Previous Training	-0.01 (0.13)	0.10 (0.16)
Ethnicity: Black	-0.14 (0.24)	0.33 (0.29)
Ethnicity: Mixed	0.11 (0.25)	0.40 (0.31)
Ethnicity: Other	-1.22 (0.72)	-1.14 (0.88)
Ethnicity: White	0.26 (0.18)	0.12 (0.22)
Model fit		
F-statistic	25.38***	16.19***
R^2	0.653	0.545
Adj. R^2	0.627	0.512

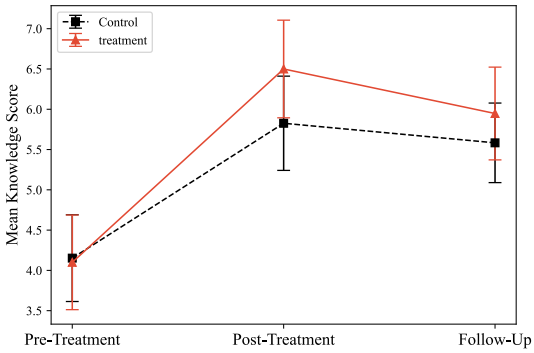
Note: Standard errors in parentheses. *$p < .05$, **$p < .01$, ***$p < .001$.



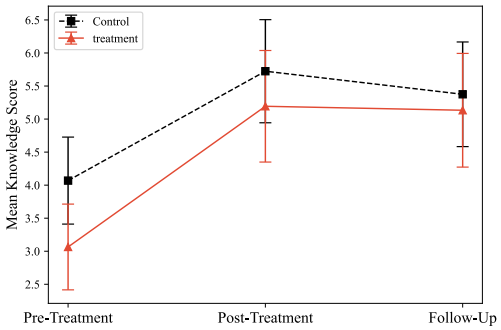
(a) Female Attitude Scores



(b) Male Attitude Scores



(c) Female Knowledge Scores



(d) Male Knowledge Scores

Figure C.1. Attitude and Knowledge Scores Separated by Gender