SHIHbot: A Facebook chatbot for Sexual Health Information on HIV/AIDS

**Jacqueline Brixey, Rens Hoegen, Wei Lan, Joshua Rusow, Karan Singla, Xusen Yin,
Ron Artstein, Anton Leuski**
University of Southern California
brixey, rhoegen, artstein, leuski at ict.usc.edu
weilan, rusow, singlak, xusenyin at usc.edu

Abstract

We present the implementation of an autonomous chatbot, SHIHbot, deployed on Facebook, which answers a wide variety of sexual health questions on HIV/AIDS. The chatbot's response database is compiled from professional medical and public health resources in order to provide reliable information to users. The system's backend is NPCEditor, a response selection platform trained on linked questions and answers; to our knowledge this is the first retrieval-based chatbot deployed on a large public social network.

1 Introduction

HIV (human immunodeficiency virus) is an incurable virus that leads to chronic illness and is the precursor for the potentially fatal disease, AIDS (acquired immunodeficiency syndrome). Approximately 73% of those infected are aware of their HIV-status[1]. Without diagnosis information, HIV-infected people are unable to access medication that improves their health, and reduces their risk of passing HIV on to their partners. However, stigma and discrimination, particularly in the form of homophobia, may prevent people from accessing providers for support.

Of the approximately 20 million per year sexually transmitted infections diagnosed, half occur among individuals between the ages of 15 and 24[2], thus sexual education for youths is vital. Yet youths indicate that sexual education does not meet their needs, as many report discomfort discussing sensitive topics in front of peers and with teachers who they feel might inform their parents about inquiries (DiCenso et al, 2001). However,

access to medically accurate information or confidential, individual counseling decreases negative consequences by increasing condom and contraception use (Kirby et al.,1994).

In the United States, overall technology usage (including Internet and mobile phone use) for those age 12 to 29 years was over 90% in 2014, and social media use was also high: 12-17 year olds: 81%; and 18-29 year olds: 89%[3]. Technology, particularly mobile technologies and social media, thus offers a powerful method to not only reach, but also engage and retain youth and young adults in HIV prevention and care. To inform our target group of youths, we implemented a chatbot, named SHIHbot, to dispense relevant, professionally vetted, and easy to access HIV/AIDS information on Facebook. Continuation of current work will strengthen the system before full public deployment. SHIHbot will undergo several rounds of testing and evaluation by cohorts of participants in the target demographic. Evaluation will include metrics for satisfying social work goals as well as dialogue system goals.

2 Motivation

Recent research has investigated the acceptability of sharing sexual health information via technology due to high technology use among youths, particularly vulnerable minorities in this demographic. One online survey distributed to over 5000 youth age 13 to 18 found that 19% of heterosexual youth versus 78% of gay/lesbian/queer youth used the Internet to search for sexual health information. The sizable difference in usage was attributed to sexual minority youth reporting lack of credible offline information sources (Mitchell et al., 2014).

---

[1] https://www.cdc.gov/hiv/group/msm/index.html
[2] http://www.cdc.gov/std/stats/

[3] http://www.pewinternet.org/three-technology-revolutions/

Type	Purpose	Example dialogue
Don't know	No relevant response in domain above threshold	User: What is the weather in LA? SHIHbot: I am not confident on an answer for that question.
Personal	Questions about the chatbot	User: are you a robot? SHIHbot: I am a computer, but I have lots of human colleagues who helped make me so smart.
Greeting	Greetings and pleasantries	User: hello SHIHbot: Hi there! Do you have any questions about HIV/AIDS?
Alternative	Suggests a new conversation topic after 3 off topic inputs from user	User: will it rain today? SHIHBot: I'm not sure on that. I would be happy to answer any questions you have about HIV/AIDS though.
Say more	Gives a random response with the same topic as the previous response	User: Should I get tested even if I do not have symptoms? SHIHBOT: If you think that you may have been exposed to HIV: you should get tested as soon as possible. You may have HIV and have no symptoms for many years. User: tell me more about this topic. SHIHBot: The CDC recommends that everyone between the ages of 13 and 64 get tested for HIV at least once as part of routine health care. About 1 in 8 people in the United States who have HIV don't know they have it.
Repeat	Repeats previous response	
QR	In-domain responses annotated with topics	

Table 1: dialogue-type description for all responses in SHIHbot domain.

We are aware of two chat services have been developed and evaluated to deliver health information to youths via social media. One, named MiChat, is a live chat intervention for 18 to 29 year olds delivered on Facebook and consists of eight one-hour motivational interviewing and cognitive behavioral skills-based online sessions designed to reduce condomless anal sex and substance use. In a pre-posttest design among 41 participants with no control group, investigators found that participation in at least one session of the intervention (n= 31) was associated with reductions in instances of condomless anal sex (Lelutiu-Weinberger et al., 2015). Another chatbot developed in 2011 was deployed on Windows Live Messenger and exclusively answered questions dealing with sex, drugs, and alcohol. The chatbot was rated highly by adolescent participants and demonstrated the potential to reach youths via chatbots (Crutzen et al., 2011). As of time of writing the chatbot is no longer accessible.

Facebook has over 1.23 billion daily active users[4], providing a platform to reach a large audience. Facebook also currently hosts over 30,000 unique automated chatbots, nonetheless users have reported disappointing experiences[5]. Our system, hence, aims to give a satisfying and informative experience to users by providing reliable information through a service that is always available on a ubiquitous social network.

3 Question-Answer Corpus

To build a system that answers a variety of personal and relevant questions about HIV/AIDS concerns, we created a corpus containing linked questions and answers (QA). Although a lot of information can be obtained from the web, the information might contain errors. To counteract this, our chatbot only provides information from reliable sources.

We extracted questions and their respective answers from the Center for Disease Control (CDC), New York State Department of Health HIV guide (NY), and i-Base, a treatment advocacy group that provides information vetted by medical professionals on an online forum.

The QA corpus forms the domain knowledge for SHIHbot and also provides the training data for the response classifier. The three sources pro-

provided over 3000 questions; some were curated by medical professionals but most submitted directly by users seeking information. The questions cover more than forty categories, encompassing questions dealing with transmission of HIV, potential drug interactions, and the history of the disease, among others. The inclusion of questions from i-Base provided real-world questions from users, rounding out the "frequently-asked questions" nature of the questions from the CDC and NY.

We also pulled all responses linked with their original questions from the three sources. All responses are either provided by medical professionals directly (CDC and NY) or approved by medical professionals (i-Base). Due to the larger variety in domain for questions from the forum on i-Base, the responses from this source also demonstrate a large variety. We reduced the total number of responses as more questions were provided from i-Base than answers. This was due to a new question being referred to the answer for a similar question previously responded to. In these cases, both questions were matched with the same answer in the corpus. In addition, when answers repeated the same information, only one answer was repeated where appropriate. For example, synonymous questions about a cure for HIV were all provided with the same response. An expert in social work manually annotated answers with topic tags in order to provide topic information to the dialogue manager.

4 Architecture

The architecture of SHIHbot comprises NPCEditor, a dialogue manager, and plugins to Facebook.

4.1 Dialogue Management

To drive our chatbot responses we used NPCEditor, a response classifier and dialogue management system (Leuski and Traum, 2011). NPCEditor employs a statistical classifier that is trained on linked questions and responses; for each new user utterance, the classifier ranks all the available responses. We train the classifier on our QA corpus, which is augmented by questions and responses about the chatbot itself and utterances that maintain dialogue flow such as greetings and closings (Table 1).

The dialogue manager functionality within NPCEditor chooses which response to return back to the user. Typically it will choose the response

that was ranked highest by the classifier, but it may choose a lower ranked response in order to avoid repetition. If the score of the top ranked response is below a predefined threshold (determined during training), the dialogue manager will instead select an off-topic response that indicates non-understanding (such as "please repeat that" or "I don't understand"). The classifier also has special tokens to recognize when a user asks the chatbot to repeat an answer or elaborate on a pervious answer, and when such a token is identified, the dialogue manager will repeat or elaborate, based on the topic annotation of the responses. A counter keeps track of the number of consecutive times the chatbot has failed to provide a direct answer, and on the 3rd instance, an "alternative" response is given to suggest returning to the HIV/AIDS domain. The counter restarts after giving an "alternative" response.

Previous applications of NPCEditor have been used to drive interactive characters in various domains such as interactive museum guides (Swartout et al., 2010), entertainment experiences (Hartholt et al., 2009), and interviews with Holocaust survivors (Artstein et al., 2016). NPCEditor was applied to the HIV/AIDS domain in the development of a virtual reality application designed for HIV positive young men who have sex with men (YMSM) to practice disclosing their status to intimate partners in an immersive, nonjudgmental environment (Knudtson et al., 2016). While NPCEditor has been used for custom chat applications, this is the first deployment of NPCEditor with Facebook Messenger.

4.2 Facebook API

Facebook launched the Messenger platform supporting chatbots, as well as sending and receiving APIs in 2016.[6] SHIHbot is the first deployment of a Facebook bot using NPCEditor, and to our knowledge, is the first Facebook bot to use information retrieval based response selection.

To create a chatbot on Messenger, the free Facebook API was used. This API was then connected with NPCEditor plugins, bridging NPCEditor and Facebook. When a message event occurs, it notifies our web-hook and calls a predefined function. Once all the actions of NPCEditor are completed and a response has been selected, the response is then sent to Facebook to deliver to the
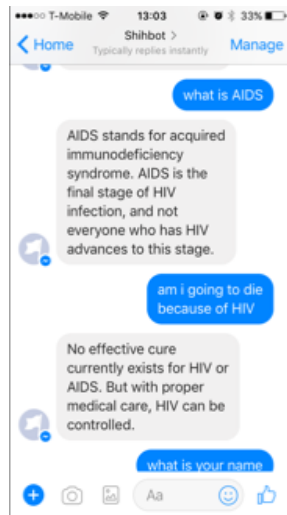
---

[6] http://newsroom.fb.com/news/2016/04/messenger-platform-at-f8

Figure 1: Interaction with mobile user on Facebook messenger**.**

user. A screenshot of an interaction with a mobile user is shown in Figure 1.

5 Demonstration outline

Participants will engage with SHIHbot via an open portal on Facebook Messenger available on a laptop at the demonstration. Participants will be invited to type input to the chatbot or welcome to provide suggestions for input. The live conversations will exhibit SHIHbot's ability to understand new questions, the chatbot's ability to cope with being asked questions outside of the domain knowledge, and the overall flow of dialogue. Participants will also be invited to view the real-time visualizations (Swartout et al., 2010) of how responses are selected based on user input.

Acknowledgments

References

Ron Artstein, Alesia Gainer, Kallirroi Georgila, Anton Leuski, Ari Shapiro, and David Traum. 2016. New dimensions in testimony demonstration. In *Proceedings of the 2016 Conference of the North American Chapter of the ACL: Demonstrations*, pages 32–36.

Rik Crutzen, Gjalt-Jorn Y. Peters, Sarah Dias Portugal, Erwin M. Fisser, and Jorne J. Grolleman. 2011. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *Journal of Adolescent Health*, 48(5):514-519.

Alba DiCenso, Vicki W. Borthwick, and Catia Creatura. 2011. Completing the picture: adolescents talk about what's missing in sexual health services. *Canadian Journal of Public Health.* 92(1):35.

rno Hartholt, Jonathan Gratch, and Lori Weiss. 2009. At the virtual frontier: Introducing Gunslinger, a multi-character, mixed-reality, story-driven experience. In *International Workshop on Intelligent Virtual Agents*. Springer Berlin Heidelberg, pages 500-501.

Kirby, D., et al.,1994. *School-based programs to reduce sexual risk behaviors: a review of effectiveness.* Public health reports, 109(3): p. 339.

Kelly Knudtson, Karina Soni, Kate Muessig, Margo Adams-Larsen, Ron Artstein, Anton Leuski, David Traum, Willa Dong, Donaldson Conserve, Lisa Hightow-Weidman. 2016. Tough Talks: Developing a virtual reality application to support HIV status disclosure among young MSM. In *21st International AIDS Conference (AIDS 2016) Abstract Book*, page 717.

Corina Lelutiu-Weinberger, John E. Pachankis, Kristi E. Gamarel, Anthony Surace, Sarit A. Golub, and Jeffrey T. Parsons. 2015. Feasibility, acceptability, and preliminary efficacy of a live-chat social media intervention to reduce HIV risk among young men who have sex with men. *AIDS and Behavior* ,19(7):1214-1227.

Anton Leuski and David Traum. 2011. NPCEditor: Creating virtual human dialogue using information retrieval techniques." *AI Magazine,* 32(2): 42-56.

Kimberly J. Mitchell, Michele L. Ybarra, Josephine D. Korchmaros, and Joseph G. Kosciw. 2014. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. *Health Education Research*, 29(1):147–57.

William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnekant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, Chad Lane, Jacquelyn Morie, Priti Aggrawal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu, and Kyle White. 2010. Ada and Grace: Toward realistic and engaging virtual museum guides. In *International Conference on Intelligent Virtual Agents.* Springer Berlin Heidelberg, pages 286-300.