

**Título:** Chatbots y las complejidades de brindar apoyo utilizando IA a sobrevivientes (Victimas) de violencia de género.

**Autoras:** Hera Hussain y Nadine Krish Spencer

**Biografías:** Hera Hussain es la fundadora y directora ejecutiva de Chayn, una organización mundial sin ánimo de lucro que crea recursos en Internet para combatir la violencia de género. Criada en Pakistán y residente en el Reino Unido, Hera supo muy pronto que quería combatir la violencia contra las mujeres. Cree en el uso del poder de la tecnología de código abierto, el diseño basado en el trauma y un marco lleno de esperanza para resolver los problemas acuciantes del mundo. Hera es becaria de Ashoka y formó parte de las listas Forbes 30 Under 30, MIT Technology Review's Innovators Under 35 y European Young Leader 2020.

Nadine Krish Spencer tiene experiencia en tecnología trabajando en entornos comerciales de hipercrecimiento. Actualmente es Directora de Producto y Experiencia en Chayn.

**Abstract** Chayn, una organización tecnológica sin ánimo de lucro dedicada a apoyar a las sobrevivientes (Victimas) de violencia de género y a proporcionarles recursos accesibles en línea, decidió retirar su chatbot en 2020 tras un piloto de 3 años. La experiencia de Chayn con chatbots culturalmente conscientes pone de relieve valiosas ideas sobre los retos y las complejidades de ofrecer apoyo basado en IA a las sobrevivientes de violencia de género.

**Palabras clave:** Chatbots, IA, seguridad en línea, violencia de género



**Artículo:** Chayn es una organización mundial sin ánimo de lucro que crea recursos digitales multilingües para ayudar a sanar a las sobrevivientes (víctimas) de violencia de género. Nuestro objetivo es empoderar a las mujeres y otros grupos marginados que han sufrido abusos domésticos, sexuales o tecnológicos. Cada decisión que tomamos -y cada recurso que creamos- tiene como núcleo la experiencia vivida.

Chayn sigue siendo una de las únicas organizaciones del mundo que presta apoyo transfronterizo y gratuito a las sobrevivientes de abusos de género. Nuestros recursos y servicios están abiertos a todas las sobrevivientes, independientemente de quiénes sean y de dónde vivan. Ponemos el apoyo al alcance de todos.

Hasta ahora hemos ayudado a más de 530.000 sobrevivientes en todo el mundo. Y -con la misión de hacer de la tecnología una herramienta para la salud, no para el daño- cada día llegamos a más. Nuestro objetivo es llegar a 1.000.000 en 2025.

Como organización nativa digital, siempre estamos tratando de aprovechar las nuevas tecnologías para apoyar a los sobrevivientes. A menudo nos vemos atrapados en el difícil espacio de pensar en la mejor manera de utilizar la tecnología para apoyar la curación de los abusos facilitada por la propia tecnología.

A partir de mediados de la década de 2010, los chatbots experimentaron un notable aumento en su utilización y popularidad en el sector humanitario. A pesar de esta creciente adopción, los datos exhaustivos sobre los logros generales, los posibles escollos y las ventajas y desventajas asociadas a esta automatización siguen siendo limitados. En 2017, decidimos desarrollar nuestro propio chatbot: Little Window. Nuestro cometido era muy claro: se creó para ser un asistente de búsqueda. Una forma de reducir drásticamente el tiempo que las mujeres tardan en buscar información, lo que en muchos casos puede salvar vidas. Sin embargo, en 2020 decidimos retirarlo, basándonos en nuestro compromiso con la sensibilidad cultural y decidiendo que no disponíamos de los recursos necesarios para seguir mejorando el servicio.

He aquí algunas lecciones aprendidas:

### **1. El lenguaje humano es confuso**

Es probable que los sobrevivientes no utilicen términos formales para describir su experiencia. El cúmulo de emociones y el condicionamiento social en torno a cómo hablar del maltrato y el abandono hacen que la forma en que la gente pide ayuda o explica su situación sea variable y enrevesada. La mayoría de la gente no escribe "estoy sufriendo maltrato doméstico" en los chatbots o en Google. Lo más probable es que escriban cosas como "¿Por qué me hace daño mi marido?", "¿Cómo puedo hacer que mi suegra me apruebe?" o "Excusas para que mi novio deje de obligarme a acostarme con él".

Que nuestros chatbots sean lo suficientemente inteligentes como para saber que cuando la gente habla de "sentirse mal" puede referirse a depresión o ansiedad, y que "hacer que mi marido me quiera" también puede significar que están sufriendo abusos o abandono, es una tarea muy exigente. Aunque con la llegada del ChatGPT parezca que podemos entender e imitar el habla humana mejor que nunca, ser capaces de mantener conversaciones sobre salud mental y escenarios

potencialmente relacionados con crisis es un asunto muy arriesgado. Tomemos el ejemplo reciente de la National Eating Disorder Association (NEDA) de EE. UU., que desconectó su línea de ayuda para trastornos alimentarios debido a que el chatbot aconsejaba a los usuarios seguir el tipo de consejo "que condujo al desarrollo de mi trastorno alimentario", según una persona influyente con experiencia en trastornos alimentarios.

Incluso dejando a un lado la capacidad multilingüe, el inglés textual difiere según los contextos culturales. Aunque la mayoría de las palabras usadas en inglés son las mismas en todas las regiones, el inglés coloquial.

El inglés coloquial y el de "texto hablado" se presentan de forma diferente y los programas de Procesamiento del Lenguaje Natural de los chatbots que sólo están preparados para reconocer el inglés de los libros de texto no los detectarían. Por ejemplo, "nikkahofied" en Pakistán es una forma de referirse a estar casado (Nikkah es el matrimonio islámico y "fied" es el sufijo del verbo inglés) o "pressuring into marriage" es lo mismo que "forced marriage".

Además, independientemente de la lengua materna, las supervivientes utilizaron el inglés para describir la agresión sexual, la violación, la violación marital y los temas relacionados con el consentimiento. Esto era especialmente evidente entre los hablantes de árabe, hindi, urdu y bengalí. En nuestros debates, lo atribuimos a que las culturas patriarcales suprimen la conversación sobre temas tabú como el sexo, por lo que las sobrevivientes a menudo no conocen el equivalente de estos términos en su lengua materna y recurren al inglés, donde se les enseña a describir la agresión sexual, la violación marital y los temas relacionados con el consentimiento.

Por ello, las supervivientes no suelen conocer el equivalente de estos términos en su lengua materna y recurren al inglés, donde se han popularizado a través de los medios de comunicación globalizados. También observamos que en bengalí y urdu, los términos sobre interacciones felices y consentidas eran más fáciles de pronunciar para las supervivientes que los términos cargados de traumas.

## **2. Utilizar un diseño basado en el trauma**

En todas las fases del diseño de Little Window, empleamos nuestros principios de seguridad, equidad, pluralidad, agencia, responsabilidad, privacidad, poder compartido y esperanza. Puedes leer más sobre estos principios en nuestro libro blanco de julio de 2023. He aquí algunos ejemplos:

### **Seguridad**

Debemos tomar decisiones valientes y audaces que den prioridad a la seguridad física y emocional de los usuarios. Esto es fundamental cuando se diseña para un público al que se le ha negado esto en muchos otros momentos de su vida. Tanto si se trata de la interfaz de su plataforma como de servicios, la seguridad en el diseño debe ser siempre el punto de partida.

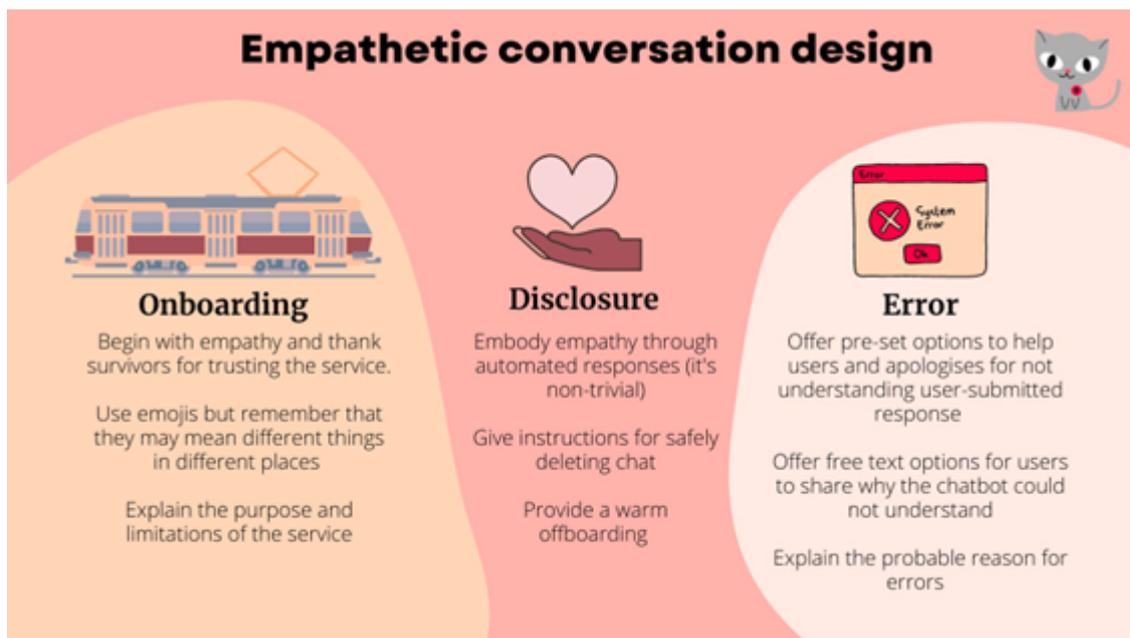
En los chatbots, esto puede reflejarse en decisiones tecnológicas como no guardar la información en el extremo del usuario, ya que podría estar utilizando un dispositivo compartido, y dar a los usuarios opciones para reemplazar la información dentro del chat en caso de que les preocupe que otros la encuentren. En la experiencia del usuario parece que el mensaje de apertura es empático, cálido y da las gracias a los sobrevivientes por confiar en el servicio. Se trata de una experiencia potencialmente traumática para un usuario que decide revelar su situación y buscar más

información, por lo que este tono de apoyo y sin prejuicios debe ser coherente desde el primer contacto.

### Pluralidad

Para hacer justicia a la complejidad de las experiencias humanas, debemos suspender las suposiciones sobre lo que los usuarios quieren o necesitan y, por tanto, tener en cuenta los sesgos de selección y confirmación. No existe un único problema humano, por lo que todas nuestras intervenciones deben diseñarse teniendo esto en cuenta. Aunque nuestros servicios se centren en un aspecto, tenemos que remitir a otras necesidades para proporcionar el mejor alivio.

En los chatbots, esto significa no asumir el idioma de una persona en función de su procedencia, ya que podría pertenecer a una minoría o a un grupo de inmigrantes. Puede significar permitir que las personas elijan varios idiomas y ubicaciones para los recursos a los que desean acceder. Y también puede significar ser sensibles en nuestra comunicación. Por ejemplo, el uso de emojis debe considerarse cuidadosamente. Un emoji de apretón de manos puede significar un saludo amistoso en algunos lugares, pero puede parecer más comercial, cuando el objetivo debe ser que sea una experiencia de apoyo y seguridad. Si se van a utilizar emojis, deben ser más fáciles de relacionar y de apoyo: en Chayn somos grandes fans del emoji del corazón.



### 3. La transparencia es esencial

En consonancia con nuestros principios de diseño informados por el trauma, Little Window fue sincera sobre su condición de chatbot (su mascota era un gato). A diferencia de los chatbots de atención al cliente, que pueden intentar hacerse pasar por personas reales, Little Window sólo estaba ahí para ofrecer orientación y recursos, y era transparente en cuanto a su propósito.

Había poco procesamiento del lenguaje natural, oportunidades limitadas para el texto libre, lo que reducía la probabilidad de una alta tasa de error, y una clara indicación a través del lenguaje y el diseño de que el servicio no estaba gestionado por un humano.

Dedicamos mucho tiempo a planificar el flujo de la conversación para garantizar un diseño sensible y adecuado. Aconsejar a las personas que sufren malos tratos sobre cómo buscar seguridad es una gran responsabilidad, y queríamos que la precisión de nuestros guiones lo reflejara. Por ejemplo, al principio de la conversación, si el chatbot no entendía la pregunta que hacía el usuario, dejaba claro que sólo podía responder a preguntas relacionadas con un conjunto limitado de temas (por ejemplo, sólo maltrato doméstico, divorcio y asilo). Dar al usuario opciones entre las que elegir de inmediato era más comprensivo, en lugar de arriesgarse a una mayor frustración por continuos malentendidos.

#### **4. Ser valiente analizando constantemente el equilibrio entre tecnología y humanidad**

A pesar de todas estas precauciones y de las mejores intenciones -incluido un cuidadoso proceso de diseño en el que participaron los supervivientes en todas las fases-, la revisión de los registros de chat mostró claramente que la gente lo utilizaba exactamente de la forma que queríamos evitar: como un servicio de crisis. El resultado fue una tasa de error alarmantemente alta. De los registros de chat se desprendía claramente que la gente ignoraba o no prestaba atención a nuestro mensaje inicial que indicaba que Little Window no contaba con personal humano. Algo que quizá indica lo desesperados que estaban por hablar con alguien. Era dolorosamente obvio lo frustrante que era la experiencia del error para los usuarios. Los usuarios pueden estar experimentando pánico o desolación y es probable que esta interacción les haga sentirse peor.

Para aquellos que quieran lanzar su propio chatbot de apoyo a la violencia doméstica, nuestro consejo general sigue siendo: No lo hagas. Y si vas a hacerlo, hazlo por las razones correctas. Los chatbots son una excelente forma de ayudar a la gente a navegar por contenidos estructurados. Quitar presión al personal (y a los voluntarios) y pueden ayudar a dirigir el tráfico a las organizaciones adecuadas. Para lo que no sirven es para proporcionar apoyo conversacional a personas en apuros. No se puede sustituir la atención humana. Y aunque pudiéramos, con la IA generativa, no deberíamos. Los riesgos son demasiado altos si nos equivocamos". Chayn ofrece ahora un chat 1 a 1 con un equipo entrenado de humanos a través de nuestra plataforma Bloom, y estamos seguros de que este servicio gratuito y sin fronteras en 14 idiomas es precisamente la dirección en la que queremos continuar.

#### **Referencias:**

Hussain, Hera, y Winfield, Jenny. 2023. Chayn's trauma-informed design principles.

[https://assets.website-files.com/60fdc9111506063bb9fe8e49/64b081438e3221d7ffc92b12\\_Trauma-informed%20design\\_%20the%20whitepaper%20by%20Chayn.pdf](https://assets.website-files.com/60fdc9111506063bb9fe8e49/64b081438e3221d7ffc92b12_Trauma-informed%20design_%20the%20whitepaper%20by%20Chayn.pdf)

Hussain, Hera. Junio de 2023. Why Chayn took down its chatbot in 2020 and what we've learned about culturally aware chatbots. <https://blog.chayn.co/why-chayn-took-down-its-chatbot-in-2020-and-what-weve-learned-about-culturally-aware-chatbots-a9587cf80df8>

Kazansky, Becky, Olivia Johnson, Bárbara Paes, Helen Kilbey y The Engine Room. 2023. "Chatbots en contextos humanitarios: Learning from practitioner experiences".

[https://communityengagementhub.org/wp-content/uploads/sites/2/2023/06/20230623\\_CEA\\_Chatbots.pdf](https://communityengagementhub.org/wp-content/uploads/sites/2/2023/06/20230623_CEA_Chatbots.pdf)

Nguyen, Britney. 2023. "Los trabajadores de la línea de ayuda de la Asociación Nacional de Trastornos Alimentarios dicen que están siendo reemplazados por IA". *Business Insider*, 26 de mayo

de 2023. <https://www.businessinsider.com/eating-disorders-nonprofit-reportedly-fired-humans-offer-ai-chatbot-2023-5?r=US&IR=T>