



Aplicación de conocimientos conductuales a la desinformación electoral en Chile

Propuesta al Programa de Naciones Unidas para el Desarrollo en Chile, agosto de 2021

Esta propuesta describe un programa de trabajo entre el Behavioural Insights Team (BIT) y el Programa de las Naciones Unidas para el Desarrollo (PNUD) para diseñar y probar intervenciones que puedan generar resiliencia a la desinformación. Esto se basará en el trabajo previo de TBI en el Reino Unido al probar intervenciones en el mundo real para disuadir el intercambio de desinformación en el contexto del proceso de reforma constitucional de Chile.

La intervención se realizará en Twitter, ya que es una de las plataformas de redes sociales más grandes y también la más abierta. Identificaremos a los usuarios de todo Chile que hayan compartido recientemente contenido de medios de comunicación identificados como fuentes de desinformación. A estos usuarios se les enviará una intervención conductual destinada a desalentar el intercambio de información falsa. Luego mediremos su posterior comportamiento de intercambio en el mundo real en Twitter. Esta metodología replica en gran medida lo que, en nuestra opinión, es la investigación de desinformación de vanguardia^{1 2} y proporcionará evidencia del mundo real para informar directamente la respuesta chilena a la desinformación.

BIT se encuentra en una posición única para ejecutar este proyecto. Nuestra investigación previa sobre desinformación en el Reino Unido junto con nuestra red de académicos y expertos en la materia interna nos permitirá entregar un proyecto impactante dentro del plazo.

El costo total de este programa de trabajo será de \$95,000 (sin IVA) y se llevará a cabo desde octubre de 2021 hasta marzo de 2022.

¹ Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. (2019). Comprender y reducir la difusión de información errónea en línea. Manuscrito no publicado: <https://psyarxiv.com/3n9u8>

² Yang, Qi, Khizar Qureshi, and Tauhid Zaman (2020). "Mitigating the Backfire Effect Using Pacing and Leading." arXiv preprint arXiv:2008.00049

Metodología

Esta sección de la propuesta describe nuestra metodología en detalle y se divide en las siguientes subsecciones:

- A. Identificación de fuentes y difusores de desinformación
- B. Diseñando las intervenciones
- C. Opciones para implementar el estudio
- D. Enfoque de análisis

A. Identificación de fuentes y difusores de desinformación

El público objetivo de nuestra intervención son los usuarios de Twitter que han compartido contenido de fuentes conocidas de desinformación. En primer lugar, identificaremos fuentes específicas de desinformación (por ejemplo, sitios web de noticias), que luego nos permitirán buscar cuentas de Twitter que compartan contenido de esas fuentes.

Identificación de fuentes de desinformación: Usaremos una combinación de bases de datos existentes, organizaciones respetadas de verificación de datos y anti-desinformación y expertos en el país para identificar una lista de fuentes de desinformación para Chile.

Identificación de los difusores de desinformación: Una vez que se hayan identificado los países y las fuentes de desinformación, utilizaremos la API de Twitter para buscar usuarios que hayan compartido contenido de estas fuentes recientemente. Twitter puede categorizar a los usuarios por país y por idioma, lo que nos permite adaptar el contenido de la intervención al contexto chileno. También filtraremos las cuentas que probablemente sean «bots». Esto proporcionará un conjunto de datos de usuarios que serán el público objetivo de nuestra intervención conductual. Ya nos hemos comunicado con la directora de políticas de Twitter, Kate Minshall, con quien seguiremos colaborando durante el proyecto.

B. Diseño de intervenciones

Las intervenciones conductuales dirigidas a las cuentas de difusión se basarán en la mejor evidencia académica disponible y se adaptarán al contexto específico del país. Haremos esto comprometiéndonos con académicos y la literatura actual, junto con expertos y profesionales locales en la lucha contra la desinformación. En esta sección, describimos algunas de las ideas más prometedoras de la literatura y cómo nos relacionaremos con expertos específicos de cada país para asegurarnos de crear intervenciones efectivas.

Es importante tener en cuenta que la cantidad de intervenciones que se pueden probar dependerá en última instancia del tamaño de la muestra disponible; sin embargo, esperamos poder probar al menos dos intervenciones contra un grupo de control que no recibirá ninguna intervención.

Enfoque para diseñar las intervenciones

Trabajaremos en estrecha colaboración con el PNUD al diseñar las intervenciones. También nos basaremos en los mismos profesionales y expertos regionales con los que nos hemos comprometido anteriormente para identificar las fuentes de desinformación.

Desarrollaremos intervenciones basadas en investigaciones previas de BIT y académicos líderes en el mundo, incluidos el profesor David Rand (MIT) y el Dr. Sander van der Linden (Cambridge), sobre lo que funciona para reducir el intercambio de desinformación. La siguiente tabla resume las perspectivas prometedoras en las que podríamos basar la intervención:

Tabla 2: Perspectivas prometedoras de la literatura	
Conocimiento	Detalle
Indicaciones de precisión	La investigación experimental en los EE. UU. indica que las personas no necesariamente quieren compartir información errónea en las redes sociales, pero aún pueden compartirla, ya que su atención no está en considerar la precisión del contenido. ³ El trabajo anterior de BIT con la FCDO también sugiere que las indicaciones de precisión pueden ser útiles para detener la propagación de la desinformación.
Juegos de inoculación	Los estudios han demostrado que una breve exposición a las tácticas utilizadas por los productores de noticias falsas puede actuar como una vacuna psicológica contra la desinformación. ⁴ Hemos trabajado con el Dr. Sander van der Linden para desarrollar esto. Esto podría adaptarse y probarse en un entorno del mundo real. Tenga en cuenta que el juego Malas noticias de Dr van der Linden ya se ha traducido a varios idiomas potencialmente relevantes, incluidos checo, alemán, polaco, rumano, serbio y esloveno.
Reglas empíricas	Los aspectos centrales de la alfabetización mediática se pueden sintetizar en simples reglas empíricas que aumentan la resistencia de las personas a la desinformación en un formato de fácil acceso. En nuestro trabajo anterior con FCDO, establecimos en un entorno simulado que las reglas generales pueden reducir la polarización y la susceptibilidad a la desinformación. Podríamos probar si las reglas generales pueden cambiar el comportamiento del mundo real.
Autoafirmación	La autoafirmación puede atenuar el razonamiento motivado en la difusión de desinformación. Esto se basa en la idea de que las personas que confían en sus creencias e identidad son menos defensivas y más abiertas a puntos de vista alternativos. ⁵ Reafirmar la imagen que tienen las personas de sí mismas, a través de simples ejercicios de pensamiento sobre los valores que son importantes para ellos, puede ayudar a defenderse contra el sesgo de confirmación y el pensamiento grupal. Podríamos probar esto en el mundo real para reducir la propagación de la desinformación.

C. Opciones para implementar el estudio

Esta intervención se entregará a través de Twitter a los usuarios identificados por haber compartido previamente contenido de fuentes conocidas de desinformación. Hay varias formas de realizar una intervención, cada una con sus pros y sus contras. Hay dos opciones principales que recomendamos considerar:

³ Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. (2019). Comprender y reducir la difusión de información errónea en línea. Manuscrito no publicado: <https://psyarxiv.com/3n9u8>

⁴ <https://www.cam.ac.uk/research/news/fake-news-vaccine-online-game-may-inoculate-by-simulating-propaganda-tactics>

⁵ Fein, S., & Spencer, S. J. (1997). El prejuicio como mantenimiento de la imagen de uno mismo: Afirmar el yo a través de la derogación de los demás. *Revista de personalidad y psicología social*, 73 (1), 31.

1. Mensajes directos enviados a usuarios específicos desde cuentas configuradas por BIT;
2. Anuncios en el suministro de noticias de usuarios específicos combinados con una intervención vinculada desde estos anuncios

Estos enfoques ya han sido utilizados con éxito por nuestros asesores académicos, quienes brindarán soporte técnico para realizar esta prueba.⁶ Proponemos que BIT, FCDO y Twitter acuerden en colaboración qué enfoque adoptar. BIT tiene un contacto existente en Twitter (Kate Minshall, directora de políticas) con quien nos relacionaríamos. Los factores clave a considerar al decidir el enfoque serán:

- **Costo:** Se debe pagar por colocar anuncios en Twitter, tanto para impresiones (personas que ven el anuncio) o clics en el enlace (estimamos que obtener una muestra de 2000 para hacer clic en un enlace costaría aproximadamente £ 6.000). Tenga en cuenta que el costo total de esta propuesta incluye £ 8,000 asignados a la colocación de anuncios y cualquier servicio de traducción necesario.
- **Políticas de Twitter:** Twitter tiene restricciones sobre cómo se pueden usar las cuentas «bot» en la plataforma en términos de seguimiento y envío de mensajes directos a otros usuarios. Por lo tanto, un enfoque que utilice tales cuentas debe cumplir con estas reglas.
- **Efectividad:** Es probable que una intervención más intensa, como un mensaje directo, sea más eficaz que una intervención de toque ligero, como la exposición a un mensaje en un anuncio.

Tabla 3: Opciones para la implementación

Opción	Metodología	Ventajas	Desventajas
1. Mensajes directos en Twitter	<p>BIT configura varias cuentas de Twitter. Luego, estas cuentas «siguen» a los usuarios específicos que recientemente han compartido desinformación.</p> <p>Un cierto porcentaje de ellos corresponderá (estimado del 3 al 10 %) y «seguirá» las cuentas de BIT, lo que nos permitirá enviarles mensajes directos.</p> <p>Los que le siguen se asignan para recibir el mensaje o mensajes de intervención o para el grupo de control.</p>	<ul style="list-style-type: none"> ● Replica fielmente el enfoque utilizado por Pennycook et al.⁸ que redujo con éxito el intercambio de información errónea. ● No requiere pagar por anuncios. ● Es más probable que el mensaje directo influya en el comportamiento que un anuncio. 	<ul style="list-style-type: none"> ● Depende de las políticas de Twitter. ● El tamaño de la muestra se restringe a quienes responden al «seguimiento».

⁶ Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. (2019). Comprender y reducir la difusión de información errónea en línea. Manuscrito no publicado: <https://psyarxiv.com/3n9u8>

⁷ Estimación basada en estudios previos utilizando una metodología similar

⁸ Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. (2019). Comprender y reducir la difusión de información errónea en línea. Manuscrito no publicado: <https://psyarxiv.com/3n9u8>

<p>2. Anuncios de Twitter e intervención vinculada</p>	<p>Los anuncios se envían a las fuentes de noticias de los usuarios específicos que recientemente han compartido desinformación. Los anuncios contendrían un mensaje de intervención sobre el intercambio de desinformación. El grupo de control recibiría un anuncio placebo.</p> <p>Estos anuncios también incitarían a las personas a hacer clic en un enlace a una intervención más atractiva (por ejemplo, un juego de inoculación).</p> <p>Mediríamos el impacto tanto en aquellos que solo ven los anuncios como en aquellos que interactúan con la intervención.</p>	<ul style="list-style-type: none"> ● La implementación sería técnicamente sencilla y estaría dentro de las pautas de Twitter. ● La muestra completa de usuarios objetivo estará expuesta al tratamiento de anuncios. ● El tratamiento de cantidad de clics es relativamente fuerte. ● Los aprendizajes de los anuncios se pueden transferir directamente a las campañas de comunicaciones. 	<ul style="list-style-type: none"> ● Los anuncios y los clics deberían pagarse. ● La mera exposición a un anuncio es una intervención débil, aunque el clic sería una intervención más fuerte. ● El tamaño de la muestra para el clic de destino se limitaría a los que se seleccionan automáticamente al hacer clic en el anuncio.
<p>Resultados conductuales medidos</p>	<p>En ambos enfoques anteriores, los resultados de comportamiento serían el número de enlaces compartidos posteriormente (directamente compartidos o retuiteados) al contenido de fuentes de desinformación conocidas. Esto se rastrearía a través de la API de Twitter.</p>		

Una vez que se haya finalizado el diseño del estudio, pasará por el proceso de revisión ética interna de BIT.

Enfoque de análisis

BIT tiene una amplia experiencia en la ejecución de experimentos de comportamiento en el mundo real, habiendo realizado más de 500 ensayos controlados aleatorios. Realizaremos este estudio como un ensayo controlado aleatorio, ampliamente reconocido como la metodología estándar de oro para sacar conclusiones sólidas sobre la efectividad de una intervención. Nuestro enfoque incluye la redacción de un protocolo de prueba que detalla el poder y los cálculos del tamaño de la muestra, las medidas de resultado, las técnicas de análisis (por ejemplo, análisis de regresión o similar) además de cómo se implementará el estudio.

Nuestro enfoque se puede aplicar a múltiples tipos de intervenciones en varios países, por lo que confiamos en que podemos lograr un tamaño de muestra suficiente para sacar conclusiones sólidas. Según la investigación previa de BIT y otros estudios similares, creemos que un tamaño de muestra de aproximadamente 2,000 participantes sería suficiente para este estudio, sin embargo, buscaremos maximizar el tamaño de muestra total (dentro de las limitaciones de recursos) para incrementar el poder estadístico del experimento y el número de intervenciones que podemos probar.

En el análisis podremos explorar si la intervención fue más o menos efectiva en diferentes tipos de usuarios. Por ejemplo, podremos ver el efecto en los usuarios que tienen un gran número de seguidores y, por lo tanto, son más influyentes. También podríamos comparar el efecto en diferentes países. Finalmente, como nuestros resultados se pueden rastrear a lo largo del tiempo utilizando la API de Twitter, podemos estimar los impactos dinámicos de la intervención e identificar si las intervenciones tienen un impacto a largo plazo en el comportamiento.

Actividades, entregables y cronograma

Objetivo y exploración | Octubre 2021

Actividades BIT	Entregables	Actividades PNUD
<p>BIT guiará al PNUD en Chile a través de un proceso de focalización para alinear los resultados de intereses y comportamientos prioritarios. Las actividades incluirán:</p> <ol style="list-style-type: none">1. Reunión de inicio de proyecto con PNUD Chile. Alineamiento en objetivos del proyecto y plan de trabajo. Durante esta reunión definiremos el enfoque metodológico a emplear, así como las potenciales organizaciones o bases de datos que nos puedan ayudar a identificar sitios web que son fuentes de noticias en Chile2. Coordinación con organizaciones trabajando en desinformación en Chile a fin de identificar fuentes de desinformación en Twitter	Plan de Trabajo	Apoyo en la identificación de organizaciones y/o bases de datos sobre desinformación en Chile

Solución y Prueba | Noviembre-Diciembre 2021

Actividades BIT	Entregables	Actividades PNUD
<p>BIT desarrollará un listado de potenciales ideas de intervención en colaboración con el PNUD y trabajará para refinar y adaptar el contenido. Asimismo, diseñará una estrategia de evaluación para medir si nuestras intervenciones son efectivas a la hora de combatir la desinformación. Las actividades incluirán:</p> <ol style="list-style-type: none"> 3. Desarrollar ideas de intervención informadas por las ciencias del comportamiento, en base a la literatura y a la investigación de organizaciones trabajando en desinformación en Chile. 4. Reunión de priorización con PNUD y Ministerio Público, a fin de identificar la(s) intervención(es) a implementar en función a potencial impacto. 5. Recoger las ideas de intervención en una nota de ideas, incluyendo análisis de potencial impacto vs. factibilidad (en formato Power Point). 6. Diseñar una estrategia de evaluación rigurosa para medir el impacto de las intervenciones informadas por las ciencias del comportamiento. 	<p>Presentación de ideas de intervención informadas por las ciencias del comportamiento y priorización (formato Power Point)</p> <p>Protocolo de diseño de evaluación (en inglés)</p>	<p><i>Previo a la reunión:</i> Convocar la reunión de priorización</p> <p><i>Durante la reunión:</i> Participación en la reunión de priorización Proveer retroalimentación sobre ideas de intervención</p>

Actividades BIT	Entregables	Actividades PNUD
<p>BIT analizará los resultados de la evaluación y proveerá al PNUD con recomendaciones de política:</p> <ol style="list-style-type: none">7. Recoger y analizar datos de resultados8. Recoger los resultados de la evaluación, así como las recomendaciones de política (por ejemplo para el escalamiento o adaptaciones de las intervenciones) en un informe final (en formato Power Point).	<p>Reporte Final recogiendo los resultados de la evaluación, así como recomendaciones de política (en formato Power Point).</p>	<p>Participación en la reunión de presentación de resultados Proveer retroalimentación sobre prioridades y posibilidades de escalamiento</p>

Propuesta financiera

Actividades y entregables	Entregables	Costo (USD)
Objetivo y exploración		
Reunión de inicio de proyecto	Plan de Trabajo	\$15.000
Solución y Prueba		
Desarrollo de ideas de intervención informadas por ciencias del comportamiento	Nota de ideas de intervención informadas por ciencias del comportamiento (en formato Power Point)	\$14.000
Reunión de priorización de intervenciones		
Diseño de estrategia de evaluación	Protocolo de diseño de evaluación (en inglés)	\$23.000
Costes de anuncios en Twitter y programación		\$8000
Escala		
Análisis de datos de resultado	Reporte Final recogiendo los resultados de la evaluación, así como recomendaciones de política (en formato Power Point).	\$35.000
Desarrollo de reporte final, recogiendo resultados y recomendaciones de política		
Total (excl. IVA)		\$95.000

Anexo

¿Quiénes somos?

El Behavioral Insights Team (BIT) fue creado en 2010 dentro de la oficina del Primer Ministro del Reino Unido, como la primera organización gubernamental a nivel mundial, dedicada a la aplicación de las ciencias del comportamiento a la política pública. Actualmente, somos una compañía de propósito social con el objetivo de apoyar organizaciones gubernamentales a introducir un modelo más realista del comportamiento humano al diseño de políticas públicas.






Parte fundamental de nuestra metodología es evaluar de manera rigurosa nuestras intervenciones y construir evidencia robusta sobre qué funciona en la práctica. Para esto, hemos diseñado e implementado, de manera exitosa, más de 500 ensayos controlados aleatorios en todo el espectro de política pública. Hemos, además, trabajado con organizaciones gubernamentales locales y nacionales en más de 80 países, incluido en 13 países en Latinoamérica, en temas tan diversos como digitación de servicios, salud, energía, empleo, crimen, transporte público, educación y recaudación de impuestos, entre otros.

¿Qué son «Behavioural Insights»?

En el 2010 acuñamos la expresión «Behavioural Insights» – o la perspectiva comportamental – para combinar ideas de varias disciplinas académicas interrelacionadas (incluyendo la economía comportamental, la psicología experimental y la antropología social), que buscan comprender cómo los individuos toman decisiones en la práctica y cómo suelen responder ante opciones o situaciones. Las ideas y conceptos de estas disciplinas nos permiten diseñar políticas que incorporan un modelo más realista del comportamiento humano y que apoyan, fomentan y alientan a los individuos a tomar mejores decisiones por y para sí mismos.

¿Cómo trabajamos?

BIT tiene una metodología de trabajo «a medida», llamada TEST, que adaptamos a las características específicas de cada proyecto y contexto. Nuestra metodología busca evaluar intervenciones comportamentales de una manera empírica, rigurosa, rápida y eficaz. Si bien cada proyecto se adapta a las necesidades de cada contexto, nuestra metodología sigue cuatro pasos de manera invariable:

 <p>Objetivo</p>	<p>Definir el problema: En la primera etapa, definimos de manera específica cuáles son los comportamientos por modificar. Para esto, utilizamos la información y datos administrativos disponibles y seguimos la orientación de nuestra contrapartida en los gobiernos u organizaciones.</p>
 <p>Exploración</p>	<p>Exploración y diagnóstico: Entender el contexto es fundamental, ya que no podemos asumir que lo que funcionó en un entorno particular funcionará en otros contextos. Esta es una etapa de inmersión en la que exploramos el contexto a profundidad por medio de herramientas cualitativas como entrevistas, grupos focales y herramientas del «pensamiento del diseño» («design thinking»). En lo posible tratamos de hacer el «viaje del usuario final» nosotros mismos, para así entender el proceso desde su perspectiva.</p>
 <p>Solución</p>	<p>Solución - Una vez hemos explorado el contexto, diseñamos la intervención. En esta etapa combinamos los aprendizajes del contexto con una revisión de la literatura de las ciencias del comportamiento y nuestro conocimiento institucional. Aquí también usamos nuestros marcos metodológicos EAST y MINDSPACE. En la etapa de Solución, producimos un listado de intervenciones potenciales que evaluamos de manera conjunta con nuestros socios de implementación.</p>
 <p>Prueba</p>	<p>Evaluar - El cuarto paso es Evaluar, Aprender, Adaptar. Una vez hemos decidido que intervención implementar, diseñamos una evaluación de impacto. Por medio de la aplicación de ensayos controlados aleatorios, evaluamos de manera científica nuestras intervenciones, aprendiendo qué funciona en la práctica y adaptando futuras intervenciones.</p>
 <p>Escala</p>	<p>Escalar - Desde el diseño de los proyectos y soluciones, el impacto social y la viabilidad de escalamiento toma prioridad. Después del análisis de datos descrito en la fase de Evaluación, escribiremos un informe que recoja los resultados de nuestros experimentos y daremos recomendaciones de política para el futuro. Asimismo, y dependiendo de lo que sea de interés y utilidad para nuestra contraparte, escribiremos notas de política para tomadores de decisiones y documentos de comunicación (blogs, op-eds, etc.). Con base en estos resultados y en la discusión con nuestros socios, haremos recomendaciones para institucionalizar las intervenciones exitosas.</p>