ANÁLISIS DE LA CONDUCTA

Un análisis operante de

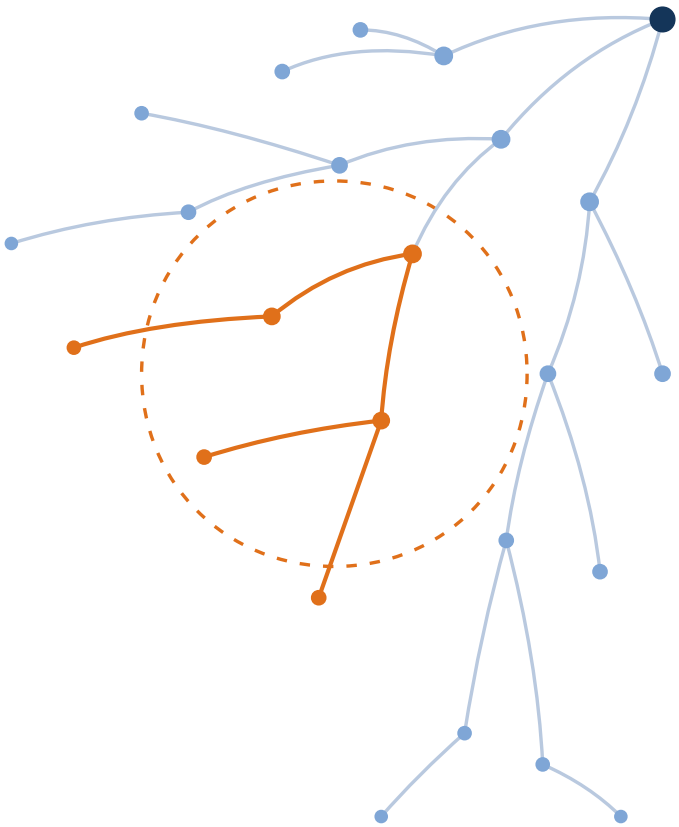
los discursos de odio y la desinformación



Márcio Moreira · Carlos Augusto de Medeiros · Cristiano Coelho
Diogo Conque Seco-Ferreira · Frederico Neves Condé · Lucas Ferraz Córdova

iW4 instituto WALDEN4
walden4.com.br

instituto WALDEN4
walden4.com.br



ANÁLISIS DE LA CONDUCTA

Un análisis operante de

los discursos de odio y la desinformación

Márcio Moreira
Carlos Augusto de Medeiros
Cristiano Coelho
Diogo Conque Seco-Ferreira
Frederico Neves Condé
Lucas Ferraz Córdova

Instituto Walden4
Brasília, DF · 2026

FICHA CATALOGRÁFICA

CDD-150

Datos Internacionales de Catalogación en la Publicación
(Elaborados por la editorial)

Un análisis operante de los discursos de odio y la desinformación [libro electrónico] / Márcio Borges Moreira ... [et al.] ; traducción del portugués de Brasil de Márcio Borges Moreira. -- Brasília, DF : Instituto Walden4, 2026.

PDF

Otros autores: Carlos Augusto de Medeiros, Cristiano Coelho, Diogo Conque Seco-Ferreira, Frederico Neves Condé, Lucas Ferraz Córdova

Traducción de: Uma análise operante de discursos de ódio e desinformação (ISBN 978-85-65721-74-5).

Bibliografía.

ISBN 978-85-65721-87-5

1. Conducta agresiva 2. Conducta - Análisis 3. Conducta - Aspectos psicológicos 4. Comunicación - Aspectos políticos 5. Comunicación - Aspectos sociales 6. Desinformación 7. Fake news 8. Odio - Aspectos sociales 9. Política - Aspectos sociales I. Moreira, Márcio Borges. II. Medeiros, Carlos Augusto de. III. Coelho, Cristiano. IV. Seco-Ferreira, Diogo Conque. V. Condé, Frederico Neves. VI. Córdova, Lucas Ferraz.

Índices para catálogo sistemático:
1. Conducta : Análisis : Psicología 150

Ficha catalográfica elaborada por la editorial

ISBN 978-85-65721-87-5
9 788565 721875

Instituto Walden4
Brasília, DF · 2026

La idea de este libro

La idea de hacer este libro surgió de una publicación en redes sociales que Rafael Nunes Félix envió a Carlos Augusto de Medeiros. Dejamos constancia aquí del crédito de Rafael por la idea, con nuestro agradecimiento.

Rafael Nunes Félix es psicólogo egresado del CEUB, con interés en la investigación y la práctica clínica en Análisis de la Conducta. Fue docente en la Universidade de Brasília (UnB), donde impartió clases de Psicología de la Personalidad desde la perspectiva del Análisis de la Conducta, y hoy es estudiante no regular del Programa de Posgrado en Ciencias del Comportamiento de la misma universidad. Investiga el descuento por demora y la resurgencia conductual, con foco en los procesos básicos de la conducta y en sus aplicaciones.

Sobre los autores

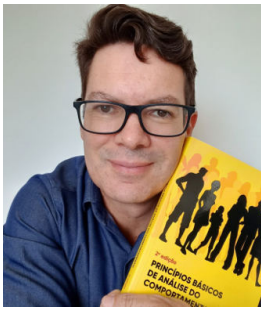
Márcio Moreira



Márcio Borges Moreira es doctor en Ciencias del Comportamiento por la Universidade de Brasília, y magíster en Psicología y psicólogo por la Pontifícia Universidade Católica de Goiás. Es profesor del grado y de la Maestría en Psicología del Centro Universitário de Brasília (UniCEUB), donde supervisa prácticas clínicas en el Centro de Formação de Psicólogos (CENFOR). Es también director del Instituto Walden4, clínica especializada en servicios basados en ABA para niños y adolescentes con trastorno del espectro autista (TEA).

Coautor del libro *Princípios Básicos de Análise do Comportamento*, tiene capítulos y artículos científicos publicados en revistas brasileñas e internacionales, con foco en Análisis Aplicado de la Conducta, conducta verbal e intervenciones basadas en evidencia. Sus proyectos de investigación actuales estudian la automatización de entrenamientos para profesionales y paraprofesionales que trabajan con personas diagnosticadas con autismo, el uso de tecnología e inteligencia artificial en la enseñanza de conceptos psicológicos, la formación y el cambio de actitudes prejuiciosas y la medición de conductas en la prestación de servicios de psicología.

Carlos Augusto de Medeiros



El Dr. Carlos Augusto de Medeiros concluyó su doctorado con honores en el programa de Ciencias del Comportamiento de la Universidade de Brasília - UnB en 2003. Obtuvo la maestría en Psicología también en la UnB en 1999, tras concluir el grado en Psicología en 1997 en la misma institución. Fue profesor asistente y adjunto en la Universidade Federal de Minas Gerais - UFMG de 2000 a 2003. En el Centro Universitário de Brasília - UniCEUB, de 2003 a 2024, ejerció las funciones de profesor de grado, supervisor de prácticas profesionales, coordinador, profesor y director de tesis del programa de Maestría en Psicología. Impartió diversos módulos en cursos de posgrado en Análisis de la Conducta en Brasil, y fue supervisor de prácticas clínicas en el Instituto Brasiliense de Análise do Comportamento - IBAC. A lo largo de su ex-

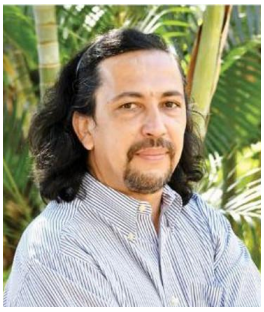
tensa carrera, impartió asignaturas relacionadas con el Análisis de la Conducta, Introducción a la Psicología y Clínica Analítico-Conductual, dirigió proyectos de investigación y supervisó prácticas de investigación y profesionales.

Desde 1997 trabaja como psicólogo clínico, con énfasis en terapia individual de adultos y terapia de pareja. Actualmente es uno de los directores de Pragmática: Psicoterapia e Cursos, una clínica de psicología que ofrece cursos regulares de extensión en Psicología Clínica.

Desarrolló un modelo de terapia analítico-conductual denominado Psicoterapia Conductual Pragmática, un modelo basado en la filosofía conductista radical y comprometido con el desarrollo de la independencia y la autonomía de los clientes.

A lo largo de su larga carrera, el Dr. Carlos Augusto de Medeiros se ha dedicado a la formación de analistas de la conducta en Brasil. Es coautor del best seller "Princípios Básicos de Análise do Comportamento". Tiene 22 artículos, dos libros y 43 capítulos de libros publicados. Sus publicaciones se relacionan con la conducta verbal, las relaciones de equivalencia, la conducta gobernada por reglas, la clínica conductual y el Análisis de la Conducta propiamente dicho.

Cristiano Coelho



Es graduado, magíster y doctor en Psicología por la Universidade de Brasília. Fue Profesor Sustituto en la UnB (1999 a 2000), Profesor Colaborador del Instituto de Educação Superior de Brasília - IESB (2009 a 2010), Supervisor de Encuestas del IBGE, el instituto nacional de estadística de Brasil, en el estado de Goiás (2010 a 2014), Profesor Colaborador del MBA en Economía Conductual de la ESPM y miembro de la junta directiva 2019-2020 de la Associação Brasileira de Psicologia e Medicina Comportamental (ABPMC). Actualmente es profesor Adjunto I de la Pontifícia Universidade Católica de Goiás, Coordinador del Programa de Posgrado Stricto Sensu en Psicología de la PUC-Goiás (Maestría y Doctorado) y terapeuta conductual. Ha realizado investigaciones y dirigido a estudiantes de iniciación científica, Maestría y Doctorado principalmente en los siguientes temas: conducta de elección, autocontrol e impulsividad, habilidades sociales y análisis de la conducta del consumidor.

Diogo Conque Seco-Ferreira



Diogo Conque Seco-Ferreira es psicólogo por la Universidade de Brasília (UnB), donde también obtuvo la maestría en Psicología y el doctorado en Ciencias del Comportamiento, con una investigación sobre la conducta del consumidor, y realizó un posdoctorado en economía conductual en la Universidad de Tilburg, en los Países Bajos. Es profesor asociado del Departamento de Psicología de la Universidade Federal de Sergipe (UFS).

Trabaja en la interfaz entre el Análisis de la Conducta, la psicología económica y la economía conductual, con investigaciones sobre la conducta del consumidor, el análisis conductual de fenómenos sociales y las ciencias conductuales aplicadas a las políticas públicas. Es coautor de "Economia comportamental: uma introdução para analistas do comportamento", publicado en la Revista Interamericana de Psicología.

Frederico Neves Condé



Doctor, magíster y graduado en Psicología Social, del Trabajo y de las Organizaciones por la Universidade de Brasília (UnB). Es investigador-tecnólogo en Información y Evaluaciones Educativas del Inep (el instituto nacional de estudios e investigaciones educativas de Brasil), donde dirigió el Servicio de Cálculo de Indicadores de Instrumentos Contextuales (2022–2023), fue Director de Evaluación de la Educación Básica (2014–2015) y fue uno de los desarrolladores del Inse, el indicador de nivel socioeconómico del Saeb (2019 y 2021). Fue asesor técnico de la Presidencia Ejecutiva y de la Defensoría del Usuario de Anatel (la agencia reguladora de las telecomunicaciones en Brasil), analista de información social en el MDS (Ministerio de Desarrollo Social), consultor del PNUD y de la Unesco y consejero del CRP-01 (Consejo Regional de Psicología del Distrito Federal). Fue docente de grado y de posgrado (IESB y UniCEUB) y director-presidente de Psicométrica Consultoria e Pesquisa. Sus áreas de actuación son la psicometría y la evaluación psicológica, educativa y social, y se dedica a la Teoría de Respuesta al Ítem, a la elaboración de instrumentos y al modelado de indicadores.

CV: http://lattes.cnpq.br/8338146219871041

Lucas Ferraz Córdova



Lucas Ferraz Córdova es psicólogo por la Universidade Federal de Minas Gerais (UFMG), magíster en Psicología y doctor en Ciencias del Comportamiento por la Universidade de Brasília (UnB), con una tesis sobre el efecto del entrenamiento sucesivo en la conducta de transposición entre los operantes verbales mando y tacto. Es profesor de la Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Trabaja en Análisis de la Conducta, con énfasis en conducta verbal, filosofía del Conductismo Radical y relaciones entre lenguaje, contexto y explicación científica. Coordina proyectos sobre los determinantes ambientales e históricos de la conducta de explicar, que investigan las variables que afectan el establecimiento y el mantenimiento del repertorio verbal, además de estudios experimentales sobre entrenamiento de operantes verbales, historia de reforzamiento y efectos de las audiencias sobre el discurso explicativo.

Índice

Presentación 13

Alcance y límites 13
Las reglas de lectura 13
Cómo leer los recuadros y los términos 14
El mapa del libro 15
Orden de lectura 16
Referencias *16*

Nota sobre el proceso de producción 17

Lo que vino de los autores 17
Lo que hizo la IA 18
Ejemplos de prompts 20
Lo que hicimos después 25
Límites del proceso 26
Una lectura conductual del propio proceso 26
Responsabilidad 27
La edición en español 27

PARTE I **El fenómeno y los datos** 28

Capítulo 1 — Desinformación y discurso de odio: definiciones, medidas y datos 29

1.1 Definiciones, y lo que cada una exige medir 29
1.2 Prevalencia: cuánto circula y quién comparte 31
1.3 Las élites políticas como fuente 33
1.4 Posicionamiento político y propensión a la información falsa: lo que muestran los estudios correlacionales 34
1.5 Limitaciones: lo que los diseños no permiten 37
1.6 Lo que se sabe sobre Brasil 39
1.7 Discurso de odio: definiciones, medidas y datos 40
1.8 Interpretaciones rivales y límites 42
Síntesis *43*
Referencias *43*

Capítulo 2 — Lo que dicen las agencias de verificación (y lo que no pueden decir) 47

2.1 Dos series de contingencias 48
2.2 Lo que cuenta cada fuente 50
2.3 Por qué la proporción entre las verificadas no es la proporción entre las dichas 51
2.4 Lo que Lupa permite decir 53
2.5 Lo que Aos Fatos permite decir 55
2.6 Dichos y menciones: de quién es la falsedad 57
2.7 ¿Concuerdan las agencias entre sí? 57
2.8 Lo que la verificación no alcanza 58
2.9 Protocolo para una muestra comparable 59

Síntesis *61*
Referencias *61*

PARTE II **Análisis operante de los discursos** 63

Capítulo 3 — Declaraciones falsas como operantes verbales **64**
3.1 El discurso del político como conducta verbal 65
3.2 La ficha de análisis 66
3.3 ¿Tacto o intraverbal? El continuo y la forma de tacto 69
3.4 Tacto distorsionado: cuando la consecuencia del oyente desplaza el control por el evento 72
3.5 Repetición: la fuerza de la afirmación repetida 75
3.6 Autoclíticos de fuente y de certeza: lo que se añade y lo que se quita 77
3.7 La declaración como regla para la audiencia 81
3.8 Casos: las mismas cuatro frases, con el mismo método 82
3.9 Interpretaciones rivales y cómo decidir entre ellas 86
Síntesis *87*
Referencias *88*

Capítulo 4 — Actitudes políticas y prejuicio como clases de equivalencia **91**
4.1 La actitud como conducta verbal evaluativa 92
4.2 Cómo una palabra adquiere función sin contacto con el referente 93
4.3 Transferencia de función 97
4.4 El prejuicio en los experimentos 103
4.5 Actitudes hacia presidentes de Brasil 107
4.6 "Nosotros" y "ellos": relaciones de oposición y de comparación 121
4.7 Cambio de actitud: reorganización de las clases y cómics 123
4.8 Medidas, interpretaciones rivales y límites 133
Síntesis *136*
Referencias *137*

Capítulo 5 — Discurso de odio: mando disfrazado y control aversivo **142**
5.1 Lo que produce la conducta del hablante: aprobación, retirada de amenazas y funciones que varían con el contexto 143
5.2 Mando disfrazado: la opinión que funciona como instrucción 145
5.3 Control aversivo dirigido a un grupo 148
5.4 El rótulo del grupo y la transferencia de función aversiva 150
5.5 Del discurso al acto: lo que los datos permiten afirmar 151
5.6 Un caso con clasificación externa: la declaración de 2011 y lo que vino después 156
5.7 Interpretaciones rivales y límites 158
Síntesis *159*
Referencias *160*

Capítulo 6 — Burbujas digitales como comunidades verbales y la evitación del contracontrol **163**
6.1 La burbuja como comunidad verbal: audiencia y fuente de reforzamiento 164
6.2 Lo que dicen los datos sobre las cámaras de eco (y dónde divergen) 166
6.3 "Realidades virtuales": cuando el control verbal sustituye al contacto con el evento. 170
6.4 Contracontrol y evitación del contracontrol 171
6.5 El caso brasileño: WhatsApp, elecciones y pandemia 176

6.6 Quién da “me gusta”, quién responde: un análisis exploratorio de dos videos políticos en Instagram ... 179
6.7 Análisis semánticos en redes sociales ... 194
6.8 Interpretaciones rivales y límites ... 195
Síntesis ... *198*
Referencias ... *198*

Capítulo 7 — El voto como conducta verbal ... 201
7.1 El voto según la definición de conducta verbal ... 202
7.2 ¿Mando, tacto o intraverbal? Funciones posibles del voto ... 203
7.3 El voto secreto como retirada de la audiencia ... 205
7.4 Declarar y votar: actitudes verbales y actos ... 207
7.5 Consecuencias sociales de votar: los experimentos de campo ... 212
7.6 Nombres y números de candidatos en clases de equivalencia ... 215
7.7 Interpretaciones rivales y límites ... 217
Síntesis ... *218*
Referencias ... *219*

PARTE III La ecología de la información 221

Capítulo 8 — Economía de la atención: el algoritmo como programador de contingencias ... 222
8.1 La atención como reforzador ... 223
8.2 Lo que muestran los datos sobre el contenido emocional, moralizado y negativo ... 225
8.3 El algoritmo como programador: cuánta audiencia recibe cada emisión ... 228
8.4 Reforzamiento intermitente e interacción ... 232
8.5 Experimentos con el propio algoritmo: lo que cambió y lo que no cambió ... 237
8.6 Interpretaciones rivales y límites ... 240
Síntesis ... *241*
Referencias ... *242*

Capítulo 9 — El ciclo productor-audiencia ... 245
9.1 Las opiniones pueden reforzarse: los experimentos de 1956 a 1969 ... 246
9.2 Lo que recibe el productor: “me gusta”, veces compartido y alcance ... 252
9.3 Lo que hace la audiencia y lo que la mantiene ... 255
9.4 El ciclo: cada lado moldea al otro ... 257
9.5 Indignación moral aprendida por reforzamiento: la evidencia ... 258
9.6 El político como productor de contenido ... 261
9.7 La metacontingencia del ecosistema: productor, audiencia, plataforma y anunciante ... 262
9.8 Interpretaciones rivales y límites ... 264
Síntesis ... *266*
Referencias ... *266*

Capítulo 10 — La ventana de Overton: lo que se puede decir sin ser castigado ... 269
10.1 El concepto, el origen y la traducción propuesta ... 270
10.2 De la ventana de la política a la ventana del enunciado: lo que desplaza el borde ... 273
10.3 Información sobre lo que los otros toleran: el experimento de Pittsburgh ... 274
10.4 Dos elecciones, dos medidas: Brasil, 2018, y Estados Unidos, 2016 ... 277

10.5 Políticos y redes sociales: quién habla desde arriba y qué recibe 281
10.6 Sanción, contracontrol y lo que mantiene el borde en su lugar 284
10.7 Cómo situar un dicho en relación con la ventana 287
10.8 Interpretaciones rivales y límites 288
Síntesis *290*
Referencias *291*

Capítulo 11 — Infodemiología: cómo se propaga la información y qué se mide 293

11.1 Infodemiología e infovigilancia 294
11.2 Medidas de difusión: alcance, exposición, reacción y la estructura de la cascada 296
11.3 ¿Lo falso se propaga más rápido que lo verdadero? 299
11.4 Modelos epidemiológicos aplicados a la información 302
11.5 La infodemia de covid-19, en el mundo y en Brasil 305
11.6 Lo que la infodemiología propone hacer 308
11.7 Preguntas para leer un estudio de difusión 310
Síntesis *310*
Referencias *311*

PARTE IV **Intervenciones** 313

Capítulo 12 — Intervenciones: qué se puede cambiar, en cuatro niveles, y con qué evidencia 314

12.1 Corregir después: la corrección, la advertencia y el "efecto rebote" 317
12.2 Corregir antes: la inoculación como entrenamiento en discriminación 320
12.3 Estímulos que señalan precisión 324
12.4 Ante quien habla: contradiscurso, compromiso público y autoclíticos de fuente 327
12.5 Reducir el prejuicio mediante equivalencia y mediante cómics 331
12.6 Cambiar las contingencias de la plataforma 334
12.7 Diseño cultural y los límites éticos de quien diseña 339
Síntesis *342*
Referencias *343*

Apéndice A — Ficha de análisis del discurso 346

Apéndice B — Conceptos en uso 348

Apéndice C — Referencias 352

Apéndice D — Verificación de fuentes 361

D.1 Cómo una afirmación llega a "verificada" 361
D.2 Números 362
D.3 Versiones leídas que no son la de la revista 362
D.4 Fuentes no leídas 363
D.5 Decisiones de método que afectan la lectura 369
D.6 Deslices de procedimiento registrados 369

Presentación

Este libro reúne una pregunta y un método. La pregunta es qué mantiene, en términos de contingencias, lo que políticos, plataformas y audiencias dicen y reenvían cuando el enunciado es falso o alguien lo clasifica como discurso de odio. El método es el del Análisis de la Conducta aplicado a la conducta verbal: en lugar de inferir lo que el hablante tenía en mente, se pregunta qué variables controlan la emisión y qué produce esta en quien escucha. Skinner (1957) da el vocabulario; la literatura experimental sobre actitudes, equivalencia y contingencias culturales da lo que se sabe y lo que todavía no se sabe.

Fue escrito para profesionales que ya conocen el Análisis de la Conducta: analistas de la conducta, psicólogos, investigadores. Por eso no hay una parte de fundamentos. Cada concepto entra, en un recuadro "Concepto en uso", en el punto en que el análisis pasa a necesitarlo, y el Apéndice B los reúne con el capítulo en que se presentó cada uno.

Alcance y límites

Aquí no se clasifican declaraciones como falsas, no se señala quién miente y no se defiende un campo político. Quien clasifica es una agencia de verificación, una decisión judicial, una plataforma o un estudio, y el texto dice cuál. Tampoco se puso a prueba ninguna hipótesis: se leyeron estudios de terceros y, donde se propone una lectura operante, se dice que es una lectura. Lo que se obtiene de aquí es un modo de leer el dato y de formular la pregunta que falta, y no una respuesta sobre quién es más deshonesto.

Dos consecuencias aparecen en todos los capítulos. La primera es que la clasificación de un operante verbal (mando, tacto, intraverbal, tacto distorsionado) es siempre interpretación, y nunca dato: el texto dice qué variables de control la sostienen y cuál sería la interpretación rival. La segunda es que la cobertura es desigual. Buena parte de los estudios son de lengua inglesa y de Estados Unidos, provienen sobre todo de Twitter y de la covid-19, y sobre Brasil hay pocos; cada capítulo dice lo que eso permite y lo que no permite concluir.

Las reglas de lectura

Cinco reglas valen de principio a fin.

La primera es el mismo método para todos los campos políticos. Donde se analiza lo que dijo un político, se busca el caso espejo del otro campo con el mismo criterio, y cuando no se encuentra, se dice que no lo hay. La regla de muestra comparable del Capítulo 2 (sección 2.9) se fijó antes de cualquier recolección, para que la elección de los casos no dependiera del resultado; la recolección no se hizo, y el capítulo dice por qué.

La segunda es que la falsedad es siempre de la agencia que la clasificó. "Falso" aparece con el nombre de la agencia, la etiqueta que usó y la fecha, y nunca como una propiedad que quien escribe atribuye a la frase.

La tercera es que "discurso de odio" es una etiqueta externa. Solo se llama así al enunciado clasificado por una instancia externa, con la instancia nombrada. Sin ella, se describe la función del enunciado, y no se lo etiqueta.

La cuarta es la separación entre dato, interpretación e hipótesis. Todo análisis de datos o de discurso trae un recuadro con las tres partes: lo que muestra la fuente, lo que se lee en ello y lo que queda como hipótesis, con lo que haría falta para ponerla a prueba.

La quinta trata de las fuentes. Solo entra como fuente lo que se leyó en el texto completo, y cada afirmación atribuida a una fuente se localizó en ella y se leyó en contexto; las planillas de ese trabajo se describen en el Apéndice D. Donde la fuente no pudo obtenerse por vía legal, se escribe que no fue leída, y no se la cita como verificada. Las páginas citadas son las de la revista cuando el texto leído las trae; cuando no las trae, la cita va sin página, y el capítulo advierte qué versión se leyó (manuscrito aceptado, documento de trabajo, texto en XML).

Cómo leer los recuadros y los términos

Hay tres tipos de recuadro. El recuadro "Dato · Interpretación · Hipótesis" cierra los análisis. El recuadro "Concepto en uso" presenta un concepto de la literatura, con la fuente y el punto en que pasa a usarse; cuando un capítulo lo retoma, el recuadro dice "retomada" y remite al capítulo de origen. El recuadro "Definición propuesta" marca el término que se propone aquí y que ninguna fuente leída usa, como la ventana de una audiencia (Capítulo 10) y la distinción entre alcance, exposición y reacción (Capítulo 11); "evitación del contracontrol" (Capítulo 6) es una propuesta hecha aquí, a partir de Delprato (2002), y viene siempre como hipótesis.

Las citas siguen el formato APA: "Autor (año, p. X)". Las referencias de cada capítulo están consolidadas en el Apéndice C. En el texto, "fue leído" y "no fue leído" quieren decir exactamente esto: el texto completo fue o no fue leído por quienes escribieron este libro.

Los estudios citados traen estadísticas, y el texto las explica la primera vez que las usa; este párrafo es para tenerlo a mano. El valor de *p* es la probabilidad de obtener una diferencia tan grande como la observada, o mayor, si en realidad no hubiera diferencia; por convención, un *p* por debajo de 0,05 hace que la diferencia se llame "significativa", lo que informa que difícilmente se debe al azar y no que sea grande o importante. *t* y *F* son los estadísticos de las pruebas que producen el *p*. La *d* y la *h* de Cohen miden el tamaño de un efecto, la diferencia entre medias en el primer caso y entre proporciones en el segundo, y los valores cercanos a 0,2 son pequeños. La razón de posibilidades (*odds ratio*, OR) compara la posibilidad de un desenlace en dos grupos: 1 significa posibilidades iguales. La desviación estándar es la variación típica de una medida, y el intervalo de confianza del 95% es el rango de valores plausibles para la estimación. El R^2 es la fracción de la variación de los datos que explica un modelo, y la exactitud, la precisión, la exhaustividad y el *F*1 describen los aciertos de un clasificador de texto. Cuando un estudio no tiene grupo de control, o mide solo lo que las personas declaran, el texto lo dice.

El mapa del libro

El libro se divide en cuatro partes y cuatro apéndices. La Parte I, "El fenómeno y los datos", tiene dos capítulos. El Capítulo 1 reúne las definiciones y los estudios de prevalencia y correlacionales, con las limitaciones de cada medida, y fija la regla de que lo que se mide es función del instrumento. El Capítulo 2 trata de las agencias de verificación: lo que mide y lo que no mide la proporción de falsas publicada por una agencia, y por qué no se la compara con la de otra sin las salvedades del capítulo.

La Parte II, "Análisis operante de los discursos", tiene cinco. El Capítulo 3 es el núcleo: presenta la ficha de análisis y los operantes verbales que usan los siguientes, y la aplica a cuatro declaraciones verificadas. El Capítulo 4 trata de las actitudes y del prejuicio como clases de equivalencia. El 5 trata del discurso de odio como mando disfrazado y control aversivo; el 6, de las burbujas digitales como comunidades verbales y de la evitación del contracontrol; el 7, del voto como conducta verbal.

La Parte III, "La ecología de la información", tiene cuatro. El Capítulo 8 trata de la economía de la atención y del algoritmo como programador de contingencias; el 9, del ciclo productor–audiencia; el 10, de la ventana de Overton,

definida por lo que una audiencia castiga; el 11, de la infodemiología y de las medidas de difusión. La Parte IV tiene un capítulo, el 12, sobre las intervenciones que se derivan del análisis, en cuatro niveles, y la evidencia de cada una. Los apéndices traen la ficha de análisis (A), los conceptos en uso (B), las referencias consolidadas (C) y la verificación de fuentes, con las lagunas (D).

Orden de lectura

Los Capítulos 1 y 2 pueden leerse por separado, por quien quiera solo los datos y la lectura crítica de las verificaciones. El Capítulo 3 es el punto de apoyo de los demás: quien vaya a leer el análisis de declaraciones y de discurso de odio, en los Capítulos 5 y 7, debe pasar por él y por el 4. Fuera de eso, los capítulos son autónomos, con remisiones explícitas a lo que usan de otro. El Capítulo 12 se lee mejor al final, porque usa el marco del entrenamiento en discriminación, las distinciones de los Capítulos 8 y 9 y la definición de alcance, exposición y reacción del Capítulo 11.

Quien use el libro como herramienta de análisis encuentra preguntas listas en cuatro puntos: la ficha del Capítulo 3 (sección 3.2), reproducida en el Apéndice A, para leer una declaración; las del Capítulo 10 (sección 10.7), para situar una declaración en una ventana; las del Capítulo 11 (sección 11.7), para leer una medida de difusión; y las del Capítulo 12 (sección 12.7), para leer una intervención. La ficha es la de mayor uso: transforma lo que describen los capítulos en una secuencia de preguntas sobre la declaración, la audiencia y las consecuencias, y termina siempre en la que importa, que es qué mantiene la declaración.

Referencias

Delprato, D. J. (2002). Countercontrol in behavior analysis. *The Behavior Analyst*, *25*(2), 191–200. https://doi.org/10.1007/bf03392057

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.

Nota sobre el proceso de producción

Este libro se escribió con un uso intensivo de inteligencia artificial generativa, y esta nota describe cómo. Los autores somos nosotros, los seis firmantes (Márcio Moreira, Carlos Augusto de Medeiros, Cristiano Coelho, Diogo Conque Seco-Ferreira, Frederico Neves Condé y Lucas Ferraz Córdova): definimos el problema, el marco interpretativo y los criterios de análisis, y respondemos por cada afirmación publicada. La IA redactó las versiones iniciales de todos los capítulos, de la Presentación y de los Apéndices A y B, y condujo la verificación de las fuentes; nosotros las revisamos, modificamos, ajustamos y comprobamos, entre el 19 y el 26 de octubre de 2026. La conducción de la IA, desde el pedido original hasta el último pedido de revisión, la hizo toda Márcio Moreira.

Preferimos describir el proceso en detalle a resumirlo en una línea de agradecimiento. El lector tiene derecho a saber cómo se produjo el texto para evaluarlo, y una descripción vaga serviría mal tanto a la transparencia como al trabajo que de hecho hicimos. Por el mismo motivo, no llamamos a este un libro "generado por IA": la expresión describe mal un proceso en el que el problema, el método y el juicio final fueron humanos de principio a fin.

Lo que vino de los autores

Lo que orienta el libro antecede a la redacción con IA y vino de nosotros, en el pedido original del 10 de octubre de 2026 y en los pedidos siguientes.

- **El problema.** La elección del discurso de odio y de la desinformación como objeto, con foco en las declaraciones de políticos brasileños, y la pregunta central: qué puede decir un análisis skinneriano sobre este fenómeno a partir de los datos que la ciberpsicología y las áreas vecinas (ciencia política, comunicación, infodemiología) ya han producido.
- **El marco interpretativo.** Una posición skinneriana explícita, fundada en la conducta verbal (*Verbal behavior*, de Skinner, y *Schedules of reinforcement*, de Ferster y Skinner, que abrimos como excepción a la regla de que solo se citan artículos), combinada con el análisis contextual de Bernard Guerin. Del artículo de Guerin que indicamos como base entran las actitudes y las creencias como conducta verbal mantenida por comunidades verbales, las funciones de tacto, intraverbal y mando de un mismo enunciado y el papel de los autoclíticos en la negociación social de lo que se

dice; de otro trabajo suyo, el análisis del conocimiento como construcción social y la condición en la que un grupo pequeño y cohesionado sostiene lo que dice. A esto se suman la equivalencia de estímulos y la transferencia de función, trabajadas en libros nuestros sobre actitudes y sobre cómics (historietas). Este marco llegó a las herramientas por el pedido original, por el artículo base, por esos libros y por las copias de Skinner que proporcionamos.

- **La selección de la literatura.** Fijamos las reglas: solo artículos científicos, más nuestros libros sobre actitudes y cómics y las dos obras de Skinner; solo textos obtenidos por vía legal, sin Sci-Hub ni equivalentes; solo se cita lo que se leyó en el texto completo; los sitios que prohíben los rastreadores de IA, como los de Lupa y Aos Fatos (agencias brasileñas de verificación de datos), no se recolectan de forma masiva. La lista de candidatas vino en parte de nosotros (27 artículos de nuestro acervo y 12 libros nuestros) y en parte de la IA (97 artículos previstos a partir del índice y 34 añadidos después de 20 búsquedas dirigidas en OpenAlex). Aprobamos el relevamiento el 10 de octubre de 2026, y llegó a 168 fuentes, de las cuales 108 se leyeron en el texto completo y 60 quedaron pendientes (Apéndice D.2 y D.4).
- **Los criterios de análisis.** Lo que cuenta como contingencia, como consecuencia social y como control verbal. Clasificar un operante verbal (mando, tacto, intraverbal, tacto distorsionado) es siempre interpretación, y nunca dato. El mismo método vale para políticos de todos los campos; "falso" es siempre la clasificación de una agencia nombrada; "discurso de odio" es una etiqueta de una instancia externa; dato, interpretación e hipótesis van separados. Y lo que no pretendemos afirmar: no ofrecemos datos robustos sobre el fenómeno, sino lecturas que puedan orientar experimentos futuros en Análisis de la Conducta.
- **La estructura.** Los diez temas del pedido original (datos y estudios correlacionales, agencias de verificación, operantes verbales, equivalencia de estímulos, burbujas digitales, voto, economía de la atención, ciclo productor–audiencia, infodemiología e intervenciones), el capítulo sobre la ventana de Overton, que añadimos el 11 de octubre, y la aprobación del índice en cuatro partes, el 10 de octubre.

Lo que hizo la IA

Usamos Claude Sonnet 5.5, de Anthropic, en Claude Code, entorno en el que el modelo lee y escribe los archivos del proyecto, ejecuta scripts y consulta bases bibliográficas abiertas, entre el 10 y el 26 de octubre de 2026. El historial de versiones muestra también a Claude Opus 5.5 en tareas de organización del

repositorio y de diagramación, sin redacción de texto del libro. Las herramientas trabajaron siempre a partir del marco interpretativo y de las instrucciones que proporcionamos, y quien las condujo, de principio a fin, fue Márcio Moreira.

- **Relevamiento y obtención de las fuentes.** Propuso candidatas, comprobó autores, año, revista, volumen y páginas en Crossref, obtuvo los textos por vías legales (acceso abierto, repositorios, versiones puestas en línea por los autores y los PDF que adjuntamos) y extrajo el texto, con reconocimiento óptico de caracteres cuando la capa de texto era mala. Lo que no obtuvo quedó en una lista de pendientes y no se cita como verificado.
- **Estructura.** Propuso el índice detallado y el orden de escritura: los capítulos 4, 9, 6, 8 y 10 primero, porque tenían las fuentes completas, y los demás a continuación.
- **Redacción inicial.** A partir del marco, desarrolló las primeras versiones de todas las secciones teóricas, de la Presentación y de los Apéndices A y B.
- **Aplicación del marco a los estudios.** Produjo las primeras lecturas skinnerianas de los estudios seleccionados: descripción de los procedimientos, identificación de posibles contingencias, interpretaciones rivales y límites, y la interpretación de los resultados, siempre con dato, interpretación e hipótesis separados.
- **Verificación de las afirmaciones.** Registró cada afirmación atribuida a una fuente en una planilla, con las frases literales que la sostienen; un script buscó esas frases en el texto completo, y la IA leyó el contexto y registró el resultado y las salvedades. Un segundo script comprobó que toda cita del texto tiene referencia y que toda afirmación marcada existe y está verificada. La planilla tiene 1.207 entradas (Apéndice D.1). En el relevamiento inicial, la comprobación corrigió autores, años de volumen y páginas que la IA había registrado de memoria e hizo retirar una afirmación atribuida a un artículo que no la hace.
- **Figuras, referencias y archivos.** Generó las figuras por script y recortó, de los libros del Instituto Walden4, las imágenes de los estudios que describe el Capítulo 4; armó las referencias en APA y los archivos de entrega (.docx, HTML y PDF).
- **Reescrituras a pedido.** Reescribió en prosa las descripciones de estudios y rehízo el Capítulo 4 con las imágenes de nuestros estudios y de los cómics, según los pedidos descritos más abajo.

La IA no es autora. Decidió, por delegación nuestra, muchas cuestiones de redacción (el orden de los capítulos, la numeración, el título de capítulos, el tratamiento de las lagunas, lo que quedó fuera cuando una fuente no pudo obtenerse), y cada decisión está registrada con fecha y con el camino para revertirla. Tres términos que aparecen como definiciones propias del libro nacieron en la redacción con IA: "evitación del contracontrol" (Capítulo 6), "ventana de una audiencia" (Capítulo 10) y la distinción entre alcance, exposición y reacción (Capítulo 11). Cada uno se declara como propuesta en el texto, y su evaluación es nuestra. La palabra final sobre cada interpretación es de los autores.

La redacción se dio en sesiones de conversación con la IA sobre los archivos del proyecto, un capítulo a la vez, a partir del 10 de octubre de 2026. Comenzó por el pedido original, que el repositorio guarda sin alteraciones. La IA entregó por etapas (el artículo base, el índice, el relevamiento bibliográfico y solo entonces los capítulos), y aprobamos cada etapa antes de la siguiente. En otros momentos respondimos "puedes decidir por mí" y "Continúa la redacción del libro y finalízala. Puedes tomar todas las decisiones": la IA decidió y registró cada decisión, con fecha, en un archivo propio, y la delegación no valió como aprobación de capítulo. Pedimos también correcciones de rumbo después de leer el borrador: que el texto dejara de tratar a "el libro" como quien lee, dice y propone; que las aperturas de sección se sostuvieran sin el título; que las descripciones de investigaciones, truncadas y poco didácticas, se revisaran párrafo por párrafo; y que el Capítulo 4 se rehiciera con figuras en color y didácticas y con las imágenes de nuestros estudios, incluidas las de los cómics. Algunos de esos pedidos aparecen literalmente en la sección siguiente. El registro completo, con el pedido original, las decisiones fechadas, la planilla de verificación y el historial de versiones, está en el repositorio del proyecto y puede consultarse previa solicitud.

Ejemplos de prompts

Casi todos los pedidos fueron mensajes cortos, en lenguaje corriente, hechos en medio del trabajo; la excepción es el pedido original, largo y estructurado, que reproducimos íntegramente. Todos fueron escritos por Márcio Moreira. Los prompts se escribieron en portugués y se reproducen aquí traducidos. Los demás ejemplos son literales y provienen de distintos puntos del proceso; en ellos la ortografía está corregida, y las palabras añadidas o cambiadas están entre corchetes. Cada ejemplo va seguido de lo que la IA hizo con él, según el registro del proyecto.

El pedido original (10 de octubre de 2026), íntegro. Los títulos de las partes del pedido aparecen en negrita.

> **PAPEL**
>
> Eres un analista de la conducta (marco skinneriano, Análisis Experimental y Aplicado de la Conducta) con experiencia en conducta verbal, equivalencia de estímulos y análisis de prácticas culturales, que escribe para un público de grado avanzado y posgrado en Psicología.
>
> **OBJETIVO**
>
> Escribir el libro "Uma análise operante de discursos de ódio e desinformação", en portugués de Brasil, con lenguaje didáctico y rigurosamente técnico (terminología del Análisis de la Conducta, definiciones operacionales, sin mentalismo explicativo). El foco son los discursos y las declaraciones falsas de políticos brasileños.
>
> **FLUJO DE TRABAJO (ejecuta por etapas y detente a esperar mi aprobación al final de cada una)**
>
> 1. Lee primero el artículo base: https://pmc.ncbi.nlm.nih.gov/articles/PMC2733686/ y resume cómo fundamentará el análisis de la conducta verbal.
> 2. Propón el índice detallado (capítulos, secciones, figuras previstas, fuentes previstas).
> 3. Arma el relevamiento bibliográfico (ver "Reglas de fuentes") y la tabla de verificación, antes de escribir cualquier capítulo.
> 4. Escribe un capítulo a la vez, con figuras, y entrégalo como documento.
> 5. Al final, haz la revisión cruzada de todas las citas y referencias.
>
> **CONTENIDO (añade temas pertinentes que yo no haya enumerado)**
>
> a) Datos y estadísticas sobre desinformación ("fake news") y discurso de odio, incluidos estudios correlacionales entre propensión a la desinformación y posicionamiento político. Presenta las limitaciones (correlación ≠ causa, muestras, medidas, contexto cultural).
>
> b) Datos de agencias de verificación (Aos Fatos, Lupa, etc.) sobre declaraciones falsas o engañosas de Jair Bolsonaro durante el mandato y de Lula. Usa el mismo criterio para los dos: misma agencia, período comparable, misma metodología de clasificación, denominador explícito (n.º de declaraciones verificadas) y salvedad sobre el sesgo de selección de las verificaciones.
>
> c) Análisis de los discursos en términos de los operantes verbales (mando, tacto, ecoica, intraverbal, textual, etc., y tacto distorsionado/mando disfrazado), con base en el artículo base.
>
> d) Análisis en términos de equivalencia de estímulos: formación de clases, transferencia de función, discriminaciones condicionales, propiedades (reflexividad, simetría, transitividad, equivalencia), relación con la formación de prejuicio y actitudes. Referencia los libros de Márcio Moreira sobre actitudes y sobre cómics y actitudes (walden4.com.br/livros).

e) Burbujas digitales como comunidades verbales; creación de "realidades virtuales" por control verbal; evitación del contracontrol.

f) Conducta de votar como conducta verbal.

g) Economía de la atención; investigaciones sobre la tendencia de los usuarios a consumir contenido emocional/moralizado; algoritmos como programas de reforzamiento.

h) Ciclo de reforzamiento entre productor de contenido y audiencia ("me gusta", compartidos, audiencia como reforzadores generalizados), análisis de las contingencias de ambos.

i) Epidemiología de la información (infodemiología): difusión, tasas de propagación, intervenciones.

j) Implicaciones prácticas: intervenciones culturales y educativas basadas en el análisis de la conducta.

FIGURAS

Ilustra todo procedimiento o concepto que se beneficie de un diagrama: discriminaciones condicionales (MTS), clases de equivalencia y sus propiedades, triple contingencia, ciclo productor–audiencia, burbujas como comunidades verbales. Genera las figuras como archivos (Graphviz/matplotlib/SVG), con leyenda numerada, vectoriales, legibles en impresión en blanco y negro y originales (no reproducir figuras de terceros).

REGLAS DE FUENTES

- Cita solo artículos científicos, más sus libros sobre actitudes y cómics (excepción explícita). No cites otros libros.
- Obtén los artículos solo por vías legales: acceso abierto (PMC, SciELO, Unpaywall, OpenAlex, repositorios), acceso institucional que yo ponga a disposición, o PDF que yo adjunte. No uses Sci-Hub ni equivalentes. Lo que no se obtenga legalmente entra en una lista "PDF pendiente" y NO se cita como verificado.
- Para cada artículo citado, lee el texto completo y confirma que contiene la información atribuida. Regístralo en una tabla: referencia | DOI | fragmento/página que sustenta la afirmación | estado (verificado / pendiente).
- Valida DOI, autores, año, revista, volumen y páginas (p. ej., Crossref). Nunca inventes ni completes una referencia de memoria; si no puedes verificarla, márcala como pendiente.
- Haz dos pasadas de verificación (citas en el texto ↔ lista de referencias ↔ tabla de verificación). Usa el formato APA 7.
- Citas literales: cortas, entre comillas, con página, y acompañadas de una imagen (recorte) del fragmento del PDF original. Prefiere la paráfrasis fiel.

NOTICIAS

> Usa notas periodísticas solo de G1, CNN Brasil, BandNews y Canal Meio. Para cada una: fragmento corto citado con crédito, URL, fecha de publicación y fecha de acceso, más captura de pantalla del fragmento. Registra también la fecha de cada declaración analizada y la verificación correspondiente.
>
> **TONO Y RIGOR POLÍTICO**
>
> Describe conductas y contingencias, no "esencias" de personas o partidos. Analiza a políticos de diferentes campos con el mismo método. Separa siempre dato verificado, interpretación analítica e hipótesis. Evita afirmaciones difamatorias; atribuye cada clasificación de falsedad a la agencia que la hizo.
>
> **ENTREGA**
>
> Cada capítulo como documento (.docx), figuras en una carpeta aparte, tabla de verificación de referencias en una planilla, y lista de pendientes (PDF no obtenidos, datos no confirmados).

Lo que hizo la IA. De ese pedido vinieron el trabajo por etapas (artículo base, índice, relevamiento bibliográfico, capítulos y revisión cruzada), la planilla de verificación, la regla de citar solo lo que se leyó en el texto completo y la separación entre dato, interpretación e hipótesis en todos los capítulos. La instrucción de figuras legibles en blanco y negro se revisó después, a pedido nuestro: las figuras del Capítulo 4 pasaron a ser en color el 16 de octubre, y las demás el 26 de octubre de 2026.

Aprobar y seguir (10 de octubre de 2026). Entre las etapas, respondimos con frases como estas:

> Adelante.
>
> Aprobado. Continúa.

Lo que hizo la IA. Trató cada respuesta como aprobación solo de la etapa que cubría la pregunta; los puntos que no se respondieron siguieron abiertos en el registro de decisiones.

Delegar decisiones (10 y 14 de octubre de 2026).

> Puedes decidir por mí.
>
> Continúa la redacción del libro y finalízala. Puedes tomar todas las decisiones.

Lo que hizo la IA. Decidió lo que faltaba (orden de escritura, numeración, tratamiento de las fuentes que no se obtuvieron, títulos) y registró cada decisión, con fecha y con el camino para revertirla. La delegación no valió como aprobación de capítulo.

Ampliar el alcance (11 de octubre de 2026).

> Añade un capítulo con un análisis conductual de la ventana de Overton, relacionándolo con discursos de políticos y redes sociales.

Lo que hizo la IA. Escribió el capítulo, hoy el Capítulo 10, y propuso la definición de "ventana de una audiencia" que usa, declarada como propuesta en el texto.

Responder a una pregunta de la IA, con una respuesta ambigua (11 de octubre de 2026). La IA había preguntado si la excepción de fuentes abierta para Skinner valía también para la obra de Ferster y Skinner sobre programas de reforzamiento. La respuesta fue:

> Fester, ok. Adelante.

Lo que hizo la IA. Leyó "Fester" como Ferster y Skinner, registró esa lectura y avisó que, si el libro pretendido era otro, bastaba con decirlo.

Corregir la redacción después de leer el borrador (16 de octubre de 2026).

> "El libro" entr[a] como una entidad en el propio libro (e.g., "El primer estudio que el libro leyó es un juego"). Hay párrafos de inicio de sección que empiezan de forma abrupta y confusa o considerando que el título de la sección ya [forma] parte del párrafo. Identifica y ajusta esas cuestiones.

Lo que hizo la IA. Reescribió unas 185 frases en voz impersonal, en todos los capítulos, y rehízo las aperturas de sección para que se entiendan sin el título.

> Hay descripciones de investigaciones que están truncadas, poco didácticas. Revisa párrafo a párrafo esa cuestión de la redacción.

Lo que hizo la IA. Reescribió en prosa las descripciones de estudios de los doce capítulos y de la Presentación, una por una, con la pregunta, la muestra, el procedimiento, la medida, el resultado y la salvedad de cada estudio.

Pedir detalle e imágenes (16 de octubre de 2026).

> En el capítulo de actitudes y equivalencia, usa en detalle y usando las imágenes de estos dos trabajos: [direcciones de los dos trabajos]. Usa figuras en color, muy didácticas. Rehaz el capítulo con este tono.
>
> Inserta, donde sea posible y adecuado, referencias a: [direcciones de once trabajos]. Usa imágenes de esos libros, deja todo bien ilustrado, incluidas imágenes de los cómics.

Lo que hizo la IA. Rehízo el Capítulo 4 con figuras en color, recortó imágenes de los estudios y de los cómics del Instituto Walden4 e incluyó las obras en las referencias, con la dirección de cada PDF. Una de las direcciones aparecía dos veces en la lista, y la IA la leyó como el volumen siguiente, registrando la suposición para que la confirmáramos.

Lo que hicimos después

Entre el 19 y el 26 de octubre de 2026, los autores trabajaron sobre el texto producido en las etapas anteriores: revisaron, modificaron, ajustaron y comprobaron los capítulos, la Presentación y los apéndices, en un manuscrito común.

1. **Revisión conceptual.** Evaluamos si cada argumento era coherente con el marco skinneriano y con el análisis contextual, y descartamos o rehicimos los que no lo eran.
2. **Modificación y ajuste.** Corregimos errores conceptuales, inferencias sin apoyo en los datos y extrapolaciones, y reescribimos los fragmentos que pedían otra formulación.
3. **Comprobación.** La comprobación de cada afirmación contra el texto completo de la fuente la hizo primero la IA, con el procedimiento del Apéndice D; nosotros comprobamos los textos en esa misma semana de revisión. Las 60 fuentes que no pudimos obtener no se citan como verificadas, y donde el asunto exigiría la fuente el capítulo dice que no se leyó.
4. **Edición final.** Unificamos terminología, estilo y la articulación entre capítulos.

El recuadro siguiente resume las etapas por capítulo.

RECUADRO

Etapas del trabajo, por capítulo

Capítulo	Primera versión por la IA	Obras en la lista del capítulo	Entradas verificadas	Revisión y comprobación por los autores
Presentación	15/10/2026	2	0	19 a 26/10/2026
1. Definiciones y datos	14 a 15/10/2026	21	48	19 a 26/10/2026
2. Agencias de verificación	10/10/2026	10	39	19 a 26/10/2026
3. Declaraciones falsas	14 a 15/10/2026	13	92	19 a 26/10/2026
4. Actitudes y equivalencia	10/10/2026, rehecho el 16/10/2026	39	300	19 a 26/10/2026
5. Discurso de odio	14 a 15/10/2026	13	45	19 a 26/10/2026
6. Burbujas digitales	11/10/2026	13	87	19 a 26/10/2026
7. El voto	14 a 15/10/2026	13	54	19 a 26/10/2026
8. Economía de la atención	11/10/2026	12	80	19 a 26/10/2026

Capítulo	Primera versión por la IA	Obras en la lista del capítulo	Entradas verificadas	Revisión y comprobación por los autores
9. Ciclo productor–audiencia	10/10/2026	13	123	19 a 26/10/2026
10. Ventana de Overton	11/10/2026	11	80	19 a 26/10/2026
11. Infodemiología	14 a 15/10/2026	12	97	19 a 26/10/2026
12. Intervenciones	14 a 15/10/2026	27	162	19 a 26/10/2026

Nota. Las fechas de la primera versión son las del primer registro de cada archivo en el historial del proyecto; el Capítulo 10 se comenzó el 11/10/2026 con otro número y se renumeró el 14/10/2026. “Entradas verificadas” son las afirmaciones y los datos de fuentes registrados en la planilla de verificación para el capítulo, cada uno con el fragmento leído en contexto, antes de la revisión de los autores y de la sección 6.7. El 16/10/2026 pedimos para todos los capítulos dos revisiones de redacción (el texto sin “el libro” como agente y las descripciones de estudios en prosa), y el Capítulo 4 se rehízo. Los Apéndices C y D se generan por script a partir de la planilla.

Límites del proceso

Dos límites merecen mencionarse. El primero es que la verificación confirma que el pasaje atribuido a una fuente está en el texto de esa fuente, en la página indicada, y no que la lectura operante que se hace de él sea correcta; esa es la parte que los capítulos llaman interpretación o hipótesis, y la que más depende de nuestro juicio. El segundo es que los modelos de lenguaje se equivocan de memoria, y la propia IA registró, al inicio del trabajo, datos que Crossref y el texto de las fuentes desmintieron; por eso nada entró en el libro sin la comprobación descrita en el Apéndice D, y donde la fuente no pudo leerse el texto dice que no se leyó.

Una lectura conductual del propio proceso

Para los analistas de la conducta, la pregunta sobre la autoría tiene una respuesta conocida, y no cambia a causa de la IA. En un ensayo de 1972, “On ‘having’ a poem”, Skinner compara escribir un poema con tener un hijo: el poeta no crea a partir de la nada, es el locus donde se encuentran variables genéticas y ambientales. Lo que llamamos creatividad es variación y selección, y buena parte del trabajo del escritor está en editar, es decir, en seleccionar lo que permanece.

No invocamos a Skinner para disminuir el papel de la IA ni para inflar el nuestro. La analogía sirve para describir lo que de hecho cambió: la IA amplió mucho la fuente de variación, y el trabajo autoral se concentró aún más en la disposición de las condiciones (el problema, el marco, las instrucciones) y en la

selección (lo que se aceptó, corrigió, reescribió o descartó). Esta es una descripción del proceso, no una justificación de él; si el lector discrepa de ella, la descripción de las secciones anteriores sigue valiendo por sí misma.

Responsabilidad

Los autores revisaron el contenido de este libro y asumen la responsabilidad por él, incluidas las partes cuya primera redacción se hizo con IA. Los errores que queden son nuestros.

La edición en español

Esta edición se tradujo de la edición brasileña en portugués (ISBN 978-85-65721-74-5) con Claude (Anthropic), en Claude Code, bajo la dirección de Márcio Borges Moreira, que firma como traductor. La traducción partió del texto portugués. De la edición en inglés se tomaron las correcciones que los autores aprobaron para ella: el nombre y el año de una comisión parlamentaria, años de referencias, remisiones internas, un recuento, algunas frases incompletas o ambiguas del original, la explicación del signo de un tamaño del efecto y dos figuras. Las citas literales de obras escritas en inglés son traducciones nuestras, hechas a partir de las palabras originales cuando la edición en inglés las había recuperado o cuando constaban en los registros de verificación del libro, y del portugués en los demás casos; no se consultaron traducciones publicadas. Las citas de fuentes en portugués también son traducciones nuestras. Las figuras se redibujaron a partir de los mismos datos, con rótulos en español; las imágenes tomadas de los estudios y de los cómics del Capítulo 4 y las portadas de las publicaciones del Capítulo 6 conservan el texto en portugués, y las notas de las figuras del Capítulo 4 lo señalan. Otra instancia de inteligencia artificial comparó después la traducción, frase por frase, con el original (revisión adversarial), y se corrigieron los errores encontrados. Los ejemplos, las instituciones y las agencias de verificación brasileñas se mantuvieron, con una breve glosa la primera vez que aparecen.

instituto WALDEN4
walden4.com.br



PARTE I

El fenómeno y los datos

1 Desinformación y discurso de odio: definiciones, medidas y datos

2 Lo que dicen las agencias de verificación (y lo que no pueden decir)

CAPÍTULO

1

Desinformación y discurso de odio: definiciones, medidas y datos

Las cifras de Allcott y Gentzkow (2017) sobre la elección presidencial estadounidense de 2016 muestran, desde el principio, el problema que organiza este capítulo. En la base de noticias falsas que reunieron, 115 favorables a Trump se compartieron 30 millones de veces en Facebook, y 41 favorables a Clinton, 7,6 millones. Los mismos autores estiman que el adulto estadounidense promedio vio y recordó 1,14 noticias falsas y dicen, de forma expresa, que no evalúan si estas decidieron la elección. Un lector apresurado junta esas cifras en una frase, y estas responden a preguntas diferentes: la primera mide lo que circuló; la segunda, lo que alguien vio y recordó; ninguna mide lo que cambió en lo que alguien hizo.

Este capítulo sistematiza el conocimiento acumulado sobre la cantidad y la forma de circulación de la desinformación y del discurso de odio, así como el perfil de quienes los producen y comparten, explicitando las métricas empleadas y sus límites metodológicos. Es el capítulo menos analítico del libro: no propone una lectura operante, sino que muestra los datos con que trabajan los capítulos siguientes y lo que cada uno de ellos exige que se diga en voz baja. Vale en él la regla de todo el libro: antes de leer la cifra, preguntar qué controla el instrumento.

1.1 Definiciones, y lo que cada una exige medir

Cada definición de "información falsa" trae incorporada una exigencia de medida. Lazer et al. (2018) definen la noticia falsa como información fabricada que imita el contenido periodístico en la forma, pero no en el proceso organizacional ni en la intención; es una subcategoría de la desinformación en senti-

do amplio (misinformation), es decir, información incorrecta o engañosa sobre el estado del mundo, y sus vehículos carecen de las normas y los procesos editoriales de los medios. Si el elemento definitorio es la intención y el proceso del editor, escriben ellos, la prevalencia debe medirse por editor, y no por noticia individual. Allcott y Gentzkow (2017) definen la noticia falsa como un artículo de noticias intencional y verificablemente falso, que puede engañar al lector, y excluyen de la definición los errores no intencionales de reportaje. Las dos definiciones exigen, en principio, conocer la intención; ninguna de las medidas que las usan la observa.

Los estudios sobre compartir resuelven el problema cambiando el ítem por la fuente. Guess et al. (2019) toman como noticia falsa el contenido falso o engañoso vestido intencionalmente como noticia, en general para generar ingresos publicitarios, y cruzan los enlaces que cada encuestado compartió con listas de editores de noticias falsas hechas por periodistas y académicos. Lasser et al. (2022) usan la nota que NewsGuard, un servicio de evaluación de la confiabilidad de sitios de noticias, da al dominio (de 0 a 100; por debajo de 60, "no confiable") como estimador de la confiabilidad del ítem compartido. Es una convención explícita, la de inferir el ítem por la fuente, y no un defecto escondido; pero cambia lo que dice la cifra: la proporción de enlaces a dominios no confiables no es la proporción de afirmaciones falsas.

Los estudios experimentales cambian de nuevo el objeto. Rathje et al. (2023) escriben que misinformation puede referirse a noticias fabricadas, rumores falsos, teorías de la conspiración o desinformación, y lo que miden es el juicio que los participantes hacen de titulares verdaderos y falsos. Y Ricard y Medeiros (2020), en una nota sobre el caso brasileño, distinguen la información falsa que quien la disemina cree verdadera de aquella que quien la disemina sabe que es falsa —dos puntos de un "espectro del 'desorden informativo'", sin presentar una medición empírica propia. La distinción es de intención, y la Tabla 1.1 muestra que casi ninguna de las medidas la observa.

Los términos se usan como las fuentes los traen. "Información falsa" y "desinformación" aparecen, sin pretensión de definición propia, para lo que la fuente llama misinformation, y la intención solo entra cuando la fuente la mide. En cuanto a "discurso de odio", la regla es la del resto del libro: es un rótulo externo, aplicado solo al enunciado clasificado por una instancia externa, con la instancia nombrada; los estudios de la sección 1.7 usan la expresión como categoría de la propia medida, y aquí se toma como ellos la traen.

TABLA 1.1

Seis definiciones de información falsa y lo que cada una exige medir

Fuente	Término y definición	Lo que la definición exige medir	Lo que el estudio midió
Lazer et al. (2018)	Noticia falsa: información fabricada que imita la forma del periodismo, no el proceso ni la intención	Proceso e intención del editor	Sin medida propia: ensayo que propone medir por editor, y no por noticia
Allcott y Gentzkow (2017)	Noticia falsa: artículo intencional y verificablemente falso, que puede engañar	Intención y veracidad, noticia por noticia	Veces que se compartieron 156 noticias clasificadas como falsas por verificadores en los tres meses previos a la elección
Guess et al. (2019)	Contenido falso o engañoso vestido como noticia, en general por ingresos publicitarios	Origen del contenido	Enlaces compartidos en Facebook, cruzados con listas de editores de noticias falsas
Lasser et al. (2022)	Información no confiable, definida por el dominio	Confiabilidad de la fuente, no del ítem	Nota de NewsGuard por debajo de 60 para el dominio del enlace
Rathje et al. (2023)	Misinformation: noticias fabricadas, rumores, teorías de la conspiración o desinformación	Veracidad del titular	Juicio de titulares verdaderos y falsos por los participantes, con y sin incentivo
Ricard y Medeiros (2020)	Misinformation (quien disemina cree) y disinformation (quien disemina sabe), citando otra obra	Lo que sabe quien disemina	Sin medida propia: nota comentada, sin datos empíricos

Nota. Descripciones propias, a partir de los pasajes citados; la columna “Lo que el estudio midió” trae lo que cada estudio operacionalizó, y no lo que declaró querer medir.

1.2 Prevalencia: cuánto circula y quién comparte

La prevalencia puede investigarse a partir de tres preguntas centrales, y cada una exige un instrumento de medida propio: cuánto circula, quién comparte y de dónde proviene el contenido. La primera pregunta se refiere a cuánto circula. Allcott y Gentzkow (2017) armaron, para la elección presidencial estadounidense de 2016, una base con 156 noticias relacionadas con la elección que sitios de verificación de referencia habían clasificado como falsas en los tres meses anteriores a la votación. Las 115 favorables a Trump se compartieron 30 millones de veces en Facebook, y las 41 favorables a Clinton, 7,6 millo-

nes de veces. Compartir, sin embargo, no es lo mismo que haber visto, y los autores intentaron estimar lo que las personas vieron: en una encuesta en línea con 1.200 personas, hecha después de la elección, preguntaron si recordaban haber visto cada uno de 14 titulares falsos que habían circulado. Cerca del 15% dijo que sí, pero cerca del 14% también dijo haber visto titulares placebo, falsos, inventados por los autores y que nunca circularon. Como los titulares placebo no podían haber sido vistos, los autores toman la diferencia entre los dos porcentajes como medida de recuerdo verdadero y, proyectándola al conjunto de noticias de la base, estiman que el adulto promedio vio y recordó 1,14 noticias falsas. En la misma encuesta, solo el 14% de los adultos estadounidenses veía las redes sociales como la fuente más importante de noticias electorales.

La segunda pregunta es quién comparte, y el estudio de Guess et al. (2019) es el que más se acerca a una respuesta con registro de conducta, y no solo con autoinforme. Los autores aplicaron una encuesta en línea a una muestra representativa de 3.500 estadounidenses y pidieron a cada encuestado que autorizara el acceso, mediante una aplicación, a datos de su propio Facebook, entre ellos las publicaciones con enlaces externos; el análisis abarca a quienes dieron su consentimiento y tuvieron sus datos vinculados a la encuesta. Los enlaces compartidos en 2016 se compararon entonces con listas de editores de noticias falsas hechas por periodistas y académicos, lo que evita los sesgos conocidos del autoinforme de actividad en línea. El resultado principal es de rareza: cerca del 8,5% de los encuestados con datos vinculados compartió al menos una nota de un dominio de noticias falsas, y más del 90% no compartió ninguna. No es que las personas compartan poco en general: en el período, el 26,1% de los encuestados compartió de 10 a 100 enlaces de cualquier tipo, y el 61,3%, de 100 a 1.000. Entre quienes compartieron noticias falsas había más republicanos: 38 contra 17 demócratas en números absolutos, y el 18,1% de los republicanos contra el 3,5% de los demócratas en la muestra. El hallazgo más robusto, dicen los autores, es la edad: quienes tenían 65 años o más compartieron en promedio 0,75 notas de dominios de noticias falsas, más del doble que el segundo grupo de edad más avanzada (0,26), y la diferencia persiste cuando se mantienen constantes la ideología y la educación.

Dos salvedades de los propios autores acompañan estas cifras. La primera es que el patrón por ideología es coherente con dos hechos: la mayoría de las noticias falsas tenía una inclinación pro-Trump, y las personas tienden a compartir lo que aprueban. El dato, por lo tanto, puede no indicar que los conservadores tengan mayor tendencia a compartir noticias falsas cuando se exponen a ellas. La segunda es que los autores no tienen una medida de si los encuestados sabían que el contenido era falso (Guess et al., 2019). Lazer et al.

(2018) añaden un límite de método. Citan un estudio que evaluó la diseminación de un conjunto de noticias falsas destacadas y estimó que el estadounidense promedio se encontró con entre una y tres de ellas en el mes anterior a la elección; la estimación es probablemente conservadora, porque el estudio siguió solo 156 noticias. Cabe subrayar que cuantificar las cuentas que recibieron o replicaron una publicación no revela cuántas personas la leyeron, ni si ese estímulo ejerció algún efecto sobre sus acciones.

En síntesis, como formulan Lazer et al. (2018): exposición no equivale a impacto, y la evaluación, a mediano y largo plazo, del efecto de la exposición a noticias falsas sobre la conducta política es esencialmente inexistente en la literatura. Mosleh y Rand (2022) recuerdan una distinción parecida del lado de quien comparte: exposición y compartir son constructos fundamentalmente diferentes, la mayor parte de las personas comparte solo una pequeña fracción de lo que ve, y mirar lo que alguien comparte da un retrato muy limitado de su ambiente de información.

1.3 Las élites políticas como fuente

La tercera pregunta sobre la prevalencia —de dónde proviene la información falsa— desplaza el foco del público en general a las élites políticas. Mosleh y Rand (2022) argumentan que, después de la atención a las noticias de editores oscuros, una forma diferente de desinformación pasó a llamar la atención: campañas coordinadas por figuras y organizaciones públicas, a las que llaman élites. Para estudiarla, construyeron dos puntuaciones. La primera es una puntuación de falsedad para cada una de 816 élites, calculada a partir de las verificaciones profesionales de PolitiFact sobre la veracidad de sus declaraciones. La segunda es una puntuación de exposición para cada usuario de Twitter, calculada a partir de las puntuaciones de falsedad de las élites que sigue. Después compararon la exposición con lo que el propio usuario comparte. La calidad de las noticias compartidas se midió por la confiabilidad del dominio, evaluada de dos formas: por el promedio de ocho verificadores profesionales y por una muestra de 970 estadounidenses elegidos por cuotas, con las notas de simpatizantes de los dos partidos equilibradas. La ideología del usuario se estimó por la ideología de las cuentas que sigue. Cuanto mayor la exposición a la desinformación de élites, menor la calidad de las noticias que el propio usuario comparte y más a la derecha se sitúa su inclinación política estimada. Es una correlación entre medidas, y los propios autores dicen que el enfoque se basa en la red inmediata de cuentas seguidas y no capta la exposición más allá de ellas, como la que llega por retuits de conexiones o por recomendaciones algorítmicas.

Lasser et al. (2022) analizan directamente las publicaciones emitidas por parlamentarios. Los autores reunieron los tuits, sin retuits, de miembros del Congreso estadounidense, del parlamento alemán y del parlamento británico entre 2016 y marzo de 2022, extrajeron los enlaces y le dieron a cada uno la nota de NewsGuard para el dominio, de 0 (muy malo) a 100 (periodismo de calidad excepcional); por debajo de 60, el dominio se considera "no confiable". Los republicanos compartieron 9,1 veces más enlaces a dominios no confiables que los demócratas: el 3,86% de los enlaces, contra el 0,43%. En Alemania, la CDU/CSU (democristianos) compartió el 0,18% y el SPD (socialdemócratas), el 0,03%; en el Reino Unido, los conservadores el 0,25% y los laboristas el 0,05%, y los autores señalan que esas cifras se basan en recuentos muy bajos. La proporción de enlaces no confiables de los republicanos pasa del 2,4% en 2016–2018 al 5,5% en 2020–2022, mientras que la de los demócratas sigue en el 0,4%; el resultado se repite con una segunda base de confiabilidad de dominios, compilada de forma independiente de NewsGuard. La divergencia entre los partidos parece peculiar de los Estados Unidos: no aparece en Alemania ni en el Reino Unido.

Lo que los autores proponen como explicación es una posibilidad debatida: la motivación partidaria, ya que el contenido despectivo sobre el grupo político adversario aumenta las intenciones de compartir entre los partidarios. Los autores sugieren también que el patrón entre los parlamentarios puede contribuir a la asimetría que se observa en el público, de dos maneras: al proporcionar desinformación a los partidarios y al legitimar que se comparta información no confiable. NewsGuard solo tiene nota para una parte de los enlaces, y el estudio registra la cobertura: el 46,5% de los enlaces del Congreso estadounidense, el 58,8% del parlamento alemán y el 39,2% del británico, después de excluir las redes sociales y los buscadores. Lo que queda sin nota no se mide.

1.4 Posicionamiento político y propensión a la información falsa: lo que muestran los estudios correlacionales

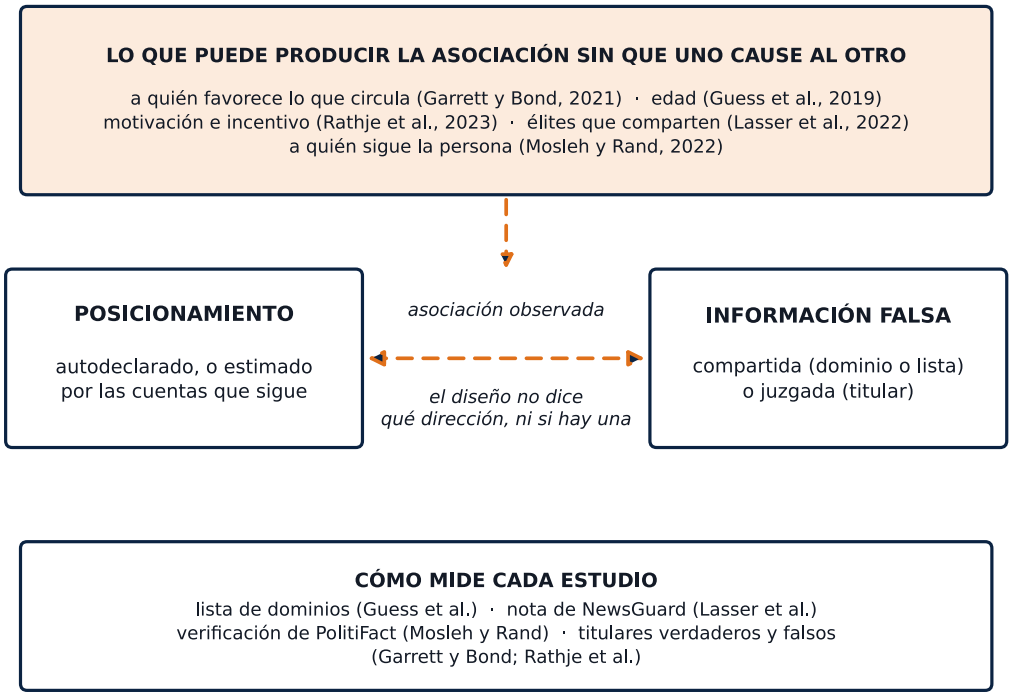
La pregunta recurrente en la literatura es si hay asimetría partidaria: si quien se posiciona en determinado campo comparte o cree más en información falsa. Los estudios verificados responden con asociaciones, y la Tabla 1.2 las reúne con lo que cada diseño no permite concluir. La Figura 1.1 ilustra los límites de esa inferencia.

Garrett y Bond (2021) adoptaron un cuidado metodológico poco común en el área. La evidencia de que los conservadores tienen más creencias falsas se apoyaba en un conjunto relativamente estrecho de temas, en muchos casos elegidos a propósito para reflejar falsedades promovidas por élites políticas

conservadoras. Los autores, en cambio, dejaron la elección de las afirmaciones a un criterio externo: cada dos semanas, un servicio identificó las 20 historias políticas más virales, 10 verdaderas y 10 falsas, y un panel de 12 oleadas, con una gran muestra representativa de estadounidenses, evaluó en la semana siguiente afirmaciones cortas sobre esas historias, a lo largo de seis meses. Con las respuestas, calcularon dos índices, mediante la teoría de detección de señales: la sensibilidad, que es la capacidad de distinguir afirmaciones verdaderas de falsas, y el sesgo de respuesta, que es la propensión a rotular todas las afirmaciones como verdaderas o como falsas. En datos de encuestas en los Estados Unidos, los conservadores distinguieron verdades de falsedades peor que los liberales, y los autores atribuyen parte de la diferencia al ambiente informativo: las afirmaciones verdaderas de alto perfil tienden a promover temas y candidatos favorecidos por los liberales, y las narrativas falsas tendían a favorecer causas conservadoras, de modo que, aun con igual sesgo, los liberales tendrían un mejor desempeño en ese ambiente.

FIGURA 1.1

Lo que muestra una asociación entre posicionamiento político e información falsa, y lo que puede producirla sin que uno cause al otro



Esquema original, sin datos. Discontinua: relación que los diseños correlacionales no separan.

Nota. Esquema original del autor, a partir de Garrett y Bond (2021), Guess et al. (2019), Rathje et al. (2023), Lasser et al. (2022) y Mosleh y Rand (2022). No representa datos. Discontinua: relación que los diseños correlacionales no separan.

Rathje et al. (2023) añaden experimentos en los que se pudo variar la consecuencia de acertar. En cuatro experimentos, con 3.364 participantes de los Estados Unidos, las personas leían titulares políticos verdaderos y falsos y respondían si la afirmación del titular era precisa; a una mitad, sorteada, se le informó que recibiría una bonificación de hasta un dólar según las respuestas correctas, y la otra mitad respondió a las mismas preguntas sin incentivo. Los autores llaman sesgo partidario a la diferencia entre el número de titulares favorables al propio partido y el de titulares contrarios al propio partido que el participante juzgó verdaderos. Los incentivos financieros por exactitud redujeron el sesgo partidario en el juicio de titulares en cerca de un 30%. Los conservadores fueron menos exactos que los liberales al distinguir titulares verdaderos de falsos, pero los incentivos redujeron la diferencia en un 52%; e incentivar a los participantes a identificar noticias que agradarían a sus aliados políticos disminuyó la exactitud. Parte de lo que aparece como diferencia de creencia, por lo tanto, parece depender de lo que se refuerza: acertar o agradar.

TABLA 1.2

Cinco estudios sobre posicionamiento político e información falsa: lo que midieron y lo que no permiten concluir

Estudio	Muestra y país	Medida de información falsa	Medida de posicionamiento	Resultado	Lo que no permite
Guess et al. (2019)	Encuestados de una encuesta representativa vinculados a datos de Facebook; EE. UU., 2016	Enlaces de dominios de noticias falsas, por lista	Autodeclarado (partido, ideología)	Raro (8,5%); más entre republicanos y entre los de más edad; la edad es el hallazgo robusto	Atribuir la diferencia a una mayor propensión: puede reflejar la oferta pro-Trump
Garrett y Bond (2021)	Encuesta en EE. UU., afirmaciones a lo largo de 6 meses	Juicio de afirmaciones verdaderas y falsas	Autodeclarado	Los conservadores discriminaron peor; una parte se atribuye al ambiente informativo	Concluir que hay un sesgo mayor: el resultado depende del conjunto de afirmaciones

Estudio	Muestra y país	Medida de información falsa	Medida de posicionamiento	Resultado	Lo que no permite
Mosleh y Rand (2022)	816 élites y usuarios de Twitter, EE. UU.	Puntuación de falsedad según verificaciones de PolitiFact	Estimado por las cuentas que sigue el usuario	La exposición a élites con alta puntuación de falsedad se correlaciona con peor calidad de noticias y con ideología conservadora	Causalidad; un solo verificador; exposición más allá de las cuentas que se siguen
Lasser et al. (2022)	Parlamentarios de EE. UU., de Alemania y del Reino Unido, 2016–2022	Nota de NewsGuard del dominio del enlace	Partido	Los republicanos comparten más, y más con el tiempo; Alemania y Reino Unido, recuentos bajos	Inferir la veracidad de la declaración a partir del dominio de la fuente; cobertura restringida del 39% al 59% de los enlaces
Rathje et al. (2023)	Cuatro experimentos, n = 3.364, EE. UU.	Juicio de titulares verdaderos y falsos	Autodeclarado	El incentivo redujo el sesgo partidario en cerca de un 30% y la diferencia en un 52%	Generalizar a partir de titulares y participantes en línea

Nota. Datos compilados a partir de los artículos originales citados. Las restricciones de la última columna explicitan lo que los diseños metodológicos no logran aislar.

1.5 Limitaciones: lo que los diseños no permiten

Los estudios reunidos en las secciones anteriores tienen tres límites en común, y es a partir de ellos que deben leerse las cifras del capítulo. De los cinco estudios de la Tabla 1.2, cuatro son sobre los Estados Unidos y el quinto compara con Alemania y el Reino Unido; los datos de tres de ellos vienen de una plataforma, Facebook o Twitter. El primer límite es, por lo tanto, de muestra y de contexto: los hallazgos son del sistema bipartidista estadounidense, en elecciones específicas, y Lasser et al. (2022) dicen que la divergencia que observan parece peculiar de los Estados Unidos. El segundo es de causa: ninguno de esos diseños separa lo que la persona ve, lo que se le ofrece y lo que ella elige, y la Figura 1.1 reúne las variables que pueden producir la asociación

sin que posicionamiento e información falsa se causen. El tercero es de medida, y merece un recuadro, porque es lo que se toma prestado del análisis de la conducta para leer todos los demás.

CONCEPTO EN USO

Conducta verbal, audiencia y autoinforme

Conducta verbal es la conducta reforzada por la mediación de otras personas (Skinner, 1957); la audiencia es el estímulo discriminativo en cuya presencia esa conducta es característicamente reforzada. El autoinforme, es decir, responder a una escala, a una pregunta de encuesta o a una entrevista, es conducta verbal, y está bajo control de quien pregunta y de cómo se hace la pregunta. Para Guerin (1994), las escalas y las preguntas de encuesta dependen de quién pregunta y de la cultura verbal en que se hace la encuesta, porque las actitudes dependen siempre de consecuencias sociales. El ejemplo clásico es el de Hildum y Brown (1956): en un sondeo telefónico con 40 estudiantes varones de Harvard, el entrevistador decía "Good" después de las respuestas a favor, o en contra, del programa de Educación General, y el "Good" desplazó las respuestas en la dirección reforzada; solo 8 de 20 sujetos notaron lo que hacía el entrevistador, y todos rechazaron que sus respuestas hubieran sido influidas por él.

La consecuencia para este capítulo es directa. La Tabla 1.2 trae tres filas en las que el dato es una respuesta a una encuesta o a una tarea de juicio (Garrett y Bond, Rathje et al. y, en parte, Guess et al., donde la ideología es autodeclarada), y una respuesta a una encuesta es conducta bajo control de la situación en que fue dada. El resultado de Rathje et al. ilustra precisamente esa dinámica: cuando la consecuencia de acertar cambia, el juicio verbal se altera. Un segundo problema de medida es el de la selección de los estímulos. Cameoka y Moreira (2021) relatan tres estudios con universitarios que evaluaron fotos de hombres blancos y de hombres negros. En el Estudio 1, 75 participantes evaluaron, en una escala de 1 (malo) a 10 (bueno), 10 fotos de hombres blancos y 10 de hombres negros con expresiones neutras; el Estudio 2 tuvo 40 participantes, con rostros ligeramente sonrientes, y el Estudio 3, 30. La diferencia entre las medias de los dos grupos de fotos fue significativa en los Estudios 1 y 2 y no en el 3, en el que se usaron las fotos de blancos con las evaluaciones más altas y las de negros con las más bajas en los dos primeros estudios. Los autores toman el Estudio 3 como demostración del potencial de distorsión de una selección de estímulos poco rigurosa y sugieren que la selección de los estímulos puede sesgar la medición. Lo que se mide depende de lo que se eligió para ser medido.

Queda además el límite del instrumento de verificación, que el Capítulo 2 trata por entero: la proporción de falsas entre las declaraciones verificadas no es la proporción entre todas las declaraciones, y una agencia no es un sensor

neutro. Mosleh y Rand (2022) usan un único verificador, PolitiFact, y élites de los Estados Unidos. El lector debe guardar, para los demás capítulos, lo que esta sección establece: toda tasa de falsedad es función de quién fue medido, con qué instrumento, en qué situación, y ninguna de ellas es, por sí misma, la tasa de un campo político.

1.6 Lo que se sabe sobre Brasil

Lo que las fuentes verificadas dicen de Brasil es menos de lo que la sección 1.2 dice de los Estados Unidos, y de naturaleza diferente. Ricard y Medeiros (2020), en un comentario, y no en un estudio empírico, escriben que una retórica de negación de los riesgos de la covid-19 fue respaldada en los más altos niveles del gobierno brasileño y apoyada por desinformación; y registran que, según la comisión parlamentaria mixta de investigación sobre las noticias falsas (la Comissão Parlamentar Mista de Inquérito, o CPMI das Fake News, instalada por el Congreso Nacional de Brasil en 2019), parte de la desinformación sobre la pandemia atacaba a responsables de decisiones y figuras públicas, como gobernadores, los medios y especialistas en salud, para deslegitimar a quienes apoyaban el aislamiento, y que la investigación aún no había divulgado evidencia de coordinación directa por el llamado Gabinete do Ódio ("Gabinete del Odio", grupo de asesores presidenciales acusado de campañas de difamación en línea) o por miembros del gobierno.

Carvalho et al. (2021) hicieron un análisis de contenido, con el programa IRaMuTeQ, de publicaciones en los perfiles de Trump y de Bolsonaro en Facebook y de pronunciamientos oficiales en los sitios de los dos gobiernos, desde el primer día de notificación de la covid-19 en cada país. Los autores señalan similitudes temáticas y posiciones contrarias a las medidas defendidas por la OMS. Escriben que no es posible afirmar categóricamente que hubo una alineación intencional entre los dos, pero que el análisis muestra una convergencia en las estrategias de desinformación, y que la minimización de los riesgos de contagio, sobre todo en las publicaciones de Facebook, es una de las marcas de la similitud. Es un análisis de dos emisores, y no un dato de prevalencia.

Recuero et al. (2025) analizaron 811 verificaciones de tres agencias brasileñas sobre las elecciones de 2018 y 2022, mediante análisis de contenido y métodos mixtos, comparando temáticas, tipos de desinformación y las fuentes que legitimaban el contenido. Señalan una transición de contenidos totalmente falsos a engañosos, y de contenidos sin fuentes a contenidos que usan fuentes reales; leen el cambio de plataforma de los contenidos verificados, de Facebook a WhatsApp y Twitter entre 2018 y 2022, como una adaptación de la desinforma-

ción a espacios más cerrados, en los que es más difícil que circulen contradiscursos. Registran, igualmente, que la literatura revisada señala un mayor volumen de desinformación generada por partidarios de Bolsonaro en ambas elecciones, aunque ambas campañas hayan figurado a la vez como polos emisores y como blancos de desinformación. Lupa (Agência Lupa, 2026), agencia brasileña de verificación de datos, describe el mismo desplazamiento en su propia base: en 2018, Facebook aparecía en el 76,5% de las verificaciones; en 2022, WhatsApp estaba presente en el 92,4%.

Dos límites separan estas fuentes de lo que la sección 1.2 muestra para los Estados Unidos. Ninguna de ellas es un estudio sobre compartir con registro de conducta, como el de Guess et al. (2019), ni una comparación de partidos por lo que sus parlamentarios comparten, como la de Lasser et al.: lo que hay son verificaciones, lo que es curado por una agencia, y un análisis de contenido de dos emisores. Y la regla del Capítulo 2, según la cual la proporción de falsas entre las declaraciones verificadas no es la proporción entre todas las declaraciones, vale aquí íntegramente. Que la desinformación brasileña tenga el formato de la estadounidense, o lo contrario, es una hipótesis, y este capítulo no la asume.

1.7 Discurso de odio: definiciones, medidas y datos

Entre las fuentes verificadas, la expresión 'discurso de odio' tiene una definición explícita en Hangartner et al. (2021): lenguaje hostil contra un grupo o una persona por una característica innata, real o percibida; los autores seleccionaron tuits xenófobos y racistas mediante una combinación de diccionario, análisis de sentimiento y anotación manual. Los demás estudios operacionalizan de otro modo, y la Tabla 1.3 reúne lo que midió cada uno.

TABLA 1.3

Cómo miden el discurso de odio cinco estudios verificados, y lo que cada medida no permite

Estudio	Medida de "discurso de odio"	Muestra	Resultado	Lo que la medida no permite
Siegel et al. (2021)	Diccionario ampliado por aprendizaje automático y clasificación por comunidades de Reddit; el diccionario exige insulto o término despectivo	Más de 750 millones de tuits sobre la elección estadounidense de 2016 y casi 400 millones de tuits de una muestra aleatoria de usuarios estadounidenses	Sin aumento persistente en la campaña ni después de la elección	Hablar de otras plataformas; la definición es conservadora

Estudio	Medida de "discurso de odio"	Muestra	Resultado	Lo que la medida no permite
Marino-Fages y Martínez-Pozo (2026, documento de trabajo)	Clasificador BERTimbau y diccionario, por blanco (homofobia, racismo, sexismo, insultos, odio político)	Más de 37,6 millones de tuits geolocalizados en Brasil, 2017-2019	Aumento después de la elección de 2018, mayor donde Bolsonaro tuvo menos votos, en homofobia y sexismo	Medir el costo social percibido; la exactitud del clasificador es del 77% al 79%
Álvarez-Benjumea y Winter (2018)	Nota de hostilidad (7 o más) dada a comentarios de un foro experimental	180 participantes de una plataforma de crowdsourcing	Comentarios hostiles: 5,88% en la línea de base, 0,99% con remoción moderada; contradiscurso sin efecto significativo	Generalizar: participantes pagados, foro anónimo
Hangartner et al. (2021)	Tuits xenófobos y racistas, seleccionados por diccionario, sentimiento y anotación	1.350 usuarios de Twitter	El contradiscurso basado en la empatía aumentó la eliminación y redujo la creación de discurso xenófobo	Efecto sobre quien solo lee; humor y advertencia de consecuencias sin efecto consistente
Fazzano y Gallo (2015)	Homofobia como clase de respuestas (golpear, insultar, torturar, apalear, excluir, discriminar), definida por la función	Datos de Disque 100 (línea telefónica federal de derechos humanos de Brasil) de 2011: 6.809 relatos de violaciones contra personas LGBT	Registros de denuncias oficiales; no miden la prevalencia directa de la agresión	Prevalencia: son relatos, y no el fenómeno

Nota. Cifras de los artículos. Marino-Fages y Martínez-Pozo es un documento de trabajo que, según la portada, no pasó por revisión de pares. Los estudios se tratan en el Capítulo 5 (sección 5.5), en el Capítulo 10 y en el Capítulo 12.

Cuatro puntos organizan la tabla. El primero es que la definición decide la cifra: Siegel et al. dicen que su definición de diccionario es conservadora, pues exige que el tuit contenga un insulto o un término despectivo, y Marino-Fages y Martínez-Pozo clasifican por blanco y obtienen un aumento concentrado en homofobia y sexismo, sin un patrón comparable para insultos generales, odio político o blancos políticos. El segundo es la plataforma y el recorte de la muestra: el estudio estadounidense se limita a Twitter y advierte que observar cierto discurso en un momento no significa que haya aumentado, dado el carácter intermitente de la plataforma, en la que los temas cobran relevancia por poco tiempo y vuelven al equilibrio; el brasileño usa solo tuits publicados los lunes y geolocalizados en Brasil, y asume la premisa de que esos dos recortes representan el universo de la plataforma. El tercero es que "discurso" y "acto" son medidas diferentes, y ninguno de los cinco estudios mide actos

contra el grupo (la laguna se discute en la sección 5.5). El cuarto es que los resultados no se contradicen ni se confirman: miden cosas diferentes en poblaciones diferentes.

Del lado del análisis de la conducta, la literatura es escasa. Mizael (2018) encontró apenas cuatro artículos analítico-conductuales sobre homosexualidad en 13 revistas y señala como laguna los análisis funcionales de conductas prejuiciosas, como los chistes mantenidos por la comunidad verbal; y Fazzano y Gallo (2015) definen la homofobia como una clase de respuestas de topografías diversas, definida por la función, y no por la forma. Es la definición de la que parte el Capítulo 5.

1.8 Interpretaciones rivales y límites

La lectura propuesta en este capítulo es que los datos empíricos revelan magnitudes inferiores a lo que el debate público supone, y que cada cifra es función estricta del instrumento utilizado. Esta perspectiva se enfrenta a tres interpretaciones rivales. La primera sostiene que la asimetría partidaria sería un hecho consolidado y que el capítulo adoptaría un escepticismo desmedido; la prueba decisiva entre esas lecturas exigiría un diseño que mantuviera bajo control riguroso la oferta total de contenido y la exposición individual efectiva —salvedad que los propios autores de Garrett y Bond (2021) y Guess et al. (2019) formulan. La segunda tesis rival argumenta que, sin verificación profesional de datos, cualquier medición es inviable, lo que hace de las agencias el único termómetro disponible; la respuesta del Capítulo 2 es que la verificación mide lo que fue verificado, y no lo que circula. La tercera es que la definición por la intención vuelve imposible la medida y debería abandonarse; la de dominio, que usan los estudios, es una respuesta pragmática, y aquí se toma como tal.

Los límites del capítulo son los de las fuentes. La mayoría de las medidas son de los Estados Unidos y de una plataforma, en el período de una o dos elecciones; la única base brasileña de gran escala, entre las de discurso, es un documento de trabajo sin revisión de pares; y las cifras de prevalencia de información falsa son, en el mejor de los casos, de exposición y de compartir, y no de efecto. No se compara a Brasil con los Estados Unidos, y compartir una fuente no se trata como creencia ni como acto.

Síntesis

Cada definición de información falsa impone exigencias metodológicas propias (intención del emisor, origen del contenido, confiabilidad del dominio o veracidad fáctica), y las investigaciones tienden a sustituir el análisis de la declaración por la clasificación de la fuente debido a la inaccesibilidad de la intención privada (1.1). La circulación empírica de noticias falsas que se comparten individualmente resultó rara en los datos de Guess et al., con mayor incidencia entre las personas mayores, y la exposición pasiva no equivale a impacto causal (1.2). La asimetría partidaria registrada entre parlamentarios en los Estados Unidos por el criterio de los dominios compartidos parece constituir un fenómeno peculiar de aquel contexto institucional (1.3). Las relaciones entre posicionamiento ideológico y desinformación tienen carácter correlacional, dependen del conjunto de ítems seleccionados y del ambiente de oferta, y una parte sustancial de la disparidad de exactitud se altera cuando se manipulan las contingencias de reforzamiento (1.4). El autoinforme en encuestas consiste en conducta verbal bajo control de las contingencias de la recolección, y el muestreo arbitrario de estímulos puede sesgar el resultado de las medidas (1.5). En Brasil, el cuadro empírico se limita a análisis de contenido, informes de coyuntura y bases de agencias de verificación, y no hay registros conductuales directos de lo que se comparte en redes abiertas (1.6). Por último, las mediciones de discurso de odio varían radicalmente según los criterios léxicos y la plataforma de recolección adoptados, y no captan la emisión de actos concretos contra los grupos que son su blanco (1.7).

Los capítulos siguientes movilizan directamente este marco: los Capítulos 2 y 3 operan bajo el principio de que toda tasa de falsedad es función del instrumento de medida; el Capítulo 5 adopta las definiciones funcionales de la sección 1.7; el Capítulo 6 se fundamenta en la evidencia de que compartir representa una fracción ínfima de la exposición; y el Capítulo 12 examina la discrepancia entre lo que el individuo relata verbalmente y lo que efectivamente hace ante contingencias reales.

Referencias

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, *34*(3), 223–237. https://doi.org/10.1093/esr/jcy005

Cameoka, M. C., & Moreira, M. B. (2021). Preconceito racial: viés na mensuração de atitudes produzido por controle de estímulos. *Acta Comportamentalia*, *29*(1), 93–112. https://doi.org/10.32870/ac.v29i1.78782

Carvalho, P. R., Sousa, P. C. C., & Schneider, M. A. F. (2021). Desinformação na pandemia de Covid-19: similitudes informacionais entre Trump e Bolsonaro. *Em Questão*, *27*(3), 15–41. https://doi.org/10.19132/1808-5245273.15-41

Fazzano, L. H., & Gallo, A. E. (2015). Uma análise da homofobia sob a perspectiva da análise do comportamento. *Temas em Psicologia*, *23*(3), 535–545. https://doi.org/10.9788/tp2015.3-02

Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, *7*(23), Artículo eabf1234. https://doi.org/10.1126/sciadv.abf1234

Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, *17*(1), 155–163. https://doi.org/10.1007/bf03392661

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), Artículo eaau4586. https://doi.org/10.1126/sciadv.aau4586

Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B. B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, *118*(50), Artículo e2116310118. https://doi.org/10.1073/pnas.2116310118

Hildum, D. C., & Brown, R. W. (1956). Verbal reinforcement and interviewer bias. *The Journal of Abnormal and Social Psychology*, *53*(1), 108–111. https://doi.org/10.1037/h0040518

Lasser, J., Aroyehun, S. T., Simchon, A., Carrella, F., Garcia, D., & Lewandowsky, S. (2022). Social media sharing of low-quality news sources by political elites. *PNAS Nexus*, *1*(4), Artículo pgac186. https://doi.org/10.1093/pnasnexus/pgac186

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Marino-Fages, D., & Martínez-Pozo, A. (2026). *Hate in the Tropics: Political leaders and the social acceptability of online hate speech* (Documento de trabajo N.º 391). RedNIE. https://rednie.eco.unc.edu.ar/files/DT/391.pdf

Mizael, T. M. (2018). Perspectivas analítico-comportamentais sobre a homossexualidade: análise da produção científica. *Revista Perspectivas*, *9*(1), 15–28. https://doi.org/10.18761/PAC.2017.011

Mosleh, M., & Rand, D. G. (2022). Measuring exposure to misinformation from political elites on Twitter. *Nature Communications*, *13*(1), Artículo 7144. https://doi.org/10.1038/s41467-022-34769-6

Rathje, S., Roozenbeek, J., Van Bavel, J. J., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, *7*(6), 892–903. https://doi.org/10.1038/s41562-023-01540-w

Recuero, R., Tavares, C., Volcan, T., Pozzebon, M., & Dutra, M. (2025). A desinformação narrada pela checagem: estudo das eleições de 2018 e 2022. E-Compós, 28, 1–26. https://doi.org/10.30962/ecomps.3039

Ricard, J., & Medeiros, J. (2020). Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil. Harvard Kennedy School Misinformation Review, 1(2), Artículo 013. https://doi.org/10.37016/mr-2020-013

Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Trumping hate on Twitter? Online hate speech in the 2016 U.S. election campaign and its aftermath. *Quarterly Journal of Political Science*, *16*(1), 71–104. https://doi.org/10.1561/100.00019045

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking (Council of Europe Report DGI[2017]09). Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

instituto WALDEN4 | ACADEMIA WALDEN4

ENCICLOPÉDIA IW4

Los conceptos de este libro, una entrada a la vez

Reforzamiento, control de estímulos, equivalencia, práctica basada en la evidencia: entradas técnicas escritas con claridad, con las fuentes al pie de cada una. Para estudiar, repasar y citar.

80+ entradas abiertas

Acceda, gratis y sin registro

walden4.com.br/enciclopedia