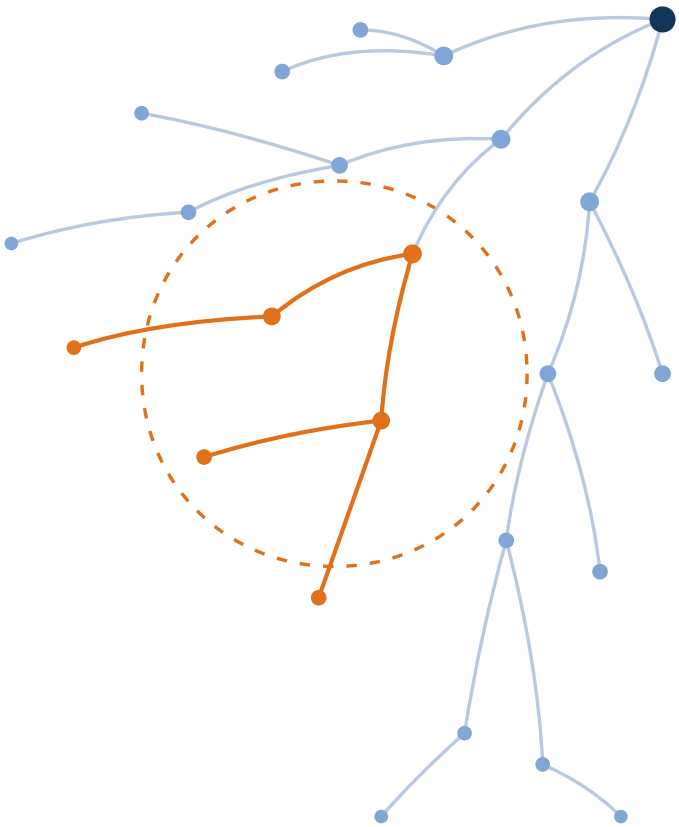
instituto WALDEN4



ANÁLISE DO COMPORTAMENTO

Uma análise operante de

discursos de ódio e desinformação

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

26-408023.0 · CDD-150

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Uma análise operante de discursos de ódio e desinformação [livro eletrônico] / Márcio Borges Moreira ... [et al.]. -- Brasília, DF : Instituto Walden4, 2026.

PDF

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

Bibliografia.

ISBN 978-85-65721-74-5

1. Comportamento agressivo 2. Comportamento - Análise 3. Comportamento - Aspectos psicológicos 4. Comunicação - Aspectos políticos 5. Comunicação - Aspectos sociais 6. Desinformação 7. Fake news 8. Ódio - Aspectos sociais 9. Política - Aspectos sociais 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. Comportamento : Análise : Psicologia 150

Camila Aparecida Rodrigues - Bibliotecária CRB - SP-010133/O

ISBN: 978-85-65721-74-5
9 788565 721745

Instituto Walden4
Brasília, DF · 2026

A ideia deste livro

A ideia de fazermos este livro surgiu de uma postagem que Rafael Nunes Félix enviou a Carlos Augusto de Medeiros. Registramos aqui o crédito de Rafael pela ideia, com o nosso agradecimento.

Rafael Nunes Félix é psicólogo formado pelo CEUB, com interesse em pesquisa e atuação clínica na Análise do Comportamento. Foi docente na Universidade de Brasília (UnB), onde ministrou aulas de Psicologia da Personalidade sob a perspectiva da Análise do Comportamento, e hoje é aluno especial do Programa de Pós-Graduação em Ciências do Comportamento da mesma universidade. Pesquisa desconto pelo atraso e ressurgência comportamental, com foco nos processos básicos do comportamento e em suas aplicações.

Sobre os Autores

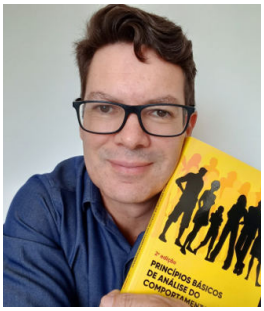
Márcio Moreira



Márcio Borges Moreira é Doutor em Ciências do Comportamento pela Universidade de Brasília, Mestre em Psicologia e Psicólogo pela Pontifícia Universidade Católica de Goiás. Professor da graduação e do Mestrado em Psicologia do Centro Universitário de Brasília (UniCEUB), onde supervisiona estágios clínicos no Centro de Formação de Psicólogos (CENFOR). É também diretor do Instituto Walden4, clínica especializada em serviços baseados em ABA para crianças e adolescentes com Transtorno do Espectro Autista (TEA).

Co-autor do livro *Princípios Básicos de Análise do Comportamento*, possui capítulos e artigos científicos publicados em periódicos nacionais e internacionais, com foco em Análise do Comportamento Aplicada, comportamento verbal e intervenções baseadas em evidências. Seus projetos de pesquisa atuais investigam a automação de treinamentos para profissionais e paraprofissionais que atuam com pessoas diagnosticadas com autismo, o uso de tecnologia e inteligência artificial no ensino de conceitos psicológicos, a formação e mudança de atitudes preconceituosas e a mensuração de comportamentos na prestação de serviços de psicologia.

Carlos Augusto de Medeiros



Dr. Carlos Augusto de Medeiros concluiu seu doutorado com louvor no programa de Ciências do Comportamento da Universidade de Brasília - UnB em 2003. Formou-se mestre em Psicologia também na UnB em 1999, tendo concluído graduação em Psicologia em 1997 na mesma instituição. Atuou como professor assistente e adjunto na Universidade Federal de Minas Gerais - UFMG de 2000-2003. No Centro Universitário de Brasília - UniCEUB, de 2003 à 2024, exerceu as funções de professor de graduação, supervisor de estágio profissionalizante, coordenador, professor e orientador do curso de Mestrado em Psicologia. Ministrou diversos módulos em curso de pós-graduação em Análise do Comportamento no país, assim como atuou como supervisor de estágio clínico no Instituto Brasiliense de Análise do Comportamento - IBAC. Ao longo de sua extensa carreira, ministrou disciplinas relacio-

nadas a Análise do Comportamento, Introdução à Psicologia, Clínica Analítica Comportamental, orientou projetos de pesquisa e supervisionou estágios de pesquisa e profissionalizantes.

Desde 1997, atua como psicólogo clínico, com ênfase em terapia individual de adultos e terapia de casais. Atualmente é um dos diretores da Pragmática: Psicoterapia e Cursos, uma clínica de psicologia que oferece cursos regulares de extensão em Psicologia Clínica.

Desenvolveu um modelo de terapia analítica comportamental denominado Psicoterapia Comportamental Pragmática. Modelo este embasado na filosofia behaviorista radical e comportamento com o desenvolvimento da independência e autonomia dos clientes.

Ao longo de sua longa carreira, o Dr. Carlos Augusto de Medeiros têm se dedicado à formação de analistas do comportamento no país. É co-autor do best seller, "Princípios Básicos de Análise do Comportamento". Possui 22 artigos, dois livros, e 43 capítulos de livros publicados. Suas publicações são relacionadas ao comportamento verbal, relações de equivalência, comportamento governado por regras, clínica comportamental e Análise do Comportamento propriamente dita.

Cristiano Coelho



Possui graduação, Mestrado e Doutorado em Psicologia pela Universidade de Brasília. Foi Professor Substituto na UnB (1999 a 2000), Professor Colaborador do Instituto de Educação Superior de Brasília - IESB (2009 a 2010), Supervisor Estadual de Pesquisas do IBGE - Goiás (2010 a 2014), Professor Colaborador do MBA em Economia Comportamental da ESPM e membro da gestão 2019-2020 da Associação Brasileira de Psicologia e Medicina Comportamental (ABPMC). Atualmente é professor Adjunto I da Pontifícia Universidade Católica de Goiás, Coordenador do Programa de Pós-Graduação Stricto Sensu em Psicologia da PUC-Goiás (Mestrado e Doutorado) e terapeuta comportamental. Tem realizado pesquisas e orientado alunos de iniciação científica, Mestrado e Doutorado principalmente nos seguintes temas: comportamento de escolha, autocontrole e impulsividade, habilidades sociais e análise do comportamento do consumidor.

Diogo Conque Seco-Ferreira



Diogo Conque Seco-Ferreira é psicólogo pela Universidade de Brasília (UnB), onde também se tornou mestre em Psicologia e doutor em Ciências do Comportamento, com pesquisa sobre comportamento do consumidor, e realizou pós-doutorado em economia comportamental na Universidade de Tilburg, nos Países Baixos. É professor associado do Departamento de Psicologia da Universidade Federal de Sergipe (UFS).

Trabalha na interface entre a Análise do Comportamento, a psicologia econômica e a economia comportamental, com pesquisas sobre o comportamento do consumidor, a análise comportamental de fenômenos sociais e as ciências comportamentais aplicadas a políticas públicas. É coautor de "Economia comportamental: uma introdução para analistas do comportamento", publicado na Revista Interamericana de Psicología.

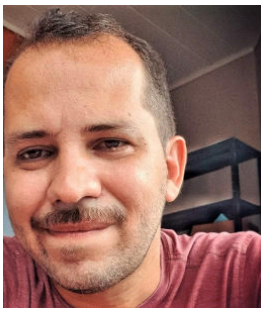
Frederico Neves Condé



Doutor, mestre e graduado em Psicologia Social, do Trabalho e das Organizações pela Universidade de Brasília (UnB). É pesquisador-tecnologista em Informações e Avaliações Educacionais do Inep, onde chefiou o Serviço de Cálculo de Indicadores de Instrumentos Contextuais (2022–2023), foi Diretor de Avaliação da Educação Básica (2014–2015) e atuou como um dos desenvolvedores do Inse do Saeb (2019 e 2021). Foi assessor técnico da Presidência Executiva e da Ouvidoria da Anatel, analista de informações sociais no MDS, consultor do PNUD e da Unesco e conselheiro no CRP-01. Atuou como docente de graduação e pós-graduação (IESB e UniCEUB) e diretor-presidente da Psicométrica Consultoria e Pesquisa. Tem como áreas de atuação a psicometria e a avaliação psicológica, educacional e social, dedicando-se à Teoria de Resposta ao Item, elaboração de instrumentos e modelagem de indicadores.

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

Lucas Ferraz Córdova



Lucas Ferraz Córdova é psicólogo pela Universidade Federal de Minas Gerais (UFMG), mestre em Psicologia e doutor em Ciências do Comportamento pela Universidade de Brasília (UnB), com tese sobre o efeito de treino sucessivo no comportamento de transposição entre os operantes verbais mando e tato. É professor da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Atua em Análise do Comportamento, com ênfase em comportamento verbal, filosofia do Behaviorismo Radical e relações entre linguagem, contexto e explicação científica. Coordena projetos sobre os determinantes ambientais e históricos do comportamento de explicar, que investigam as variáveis que afetam o estabelecimento e a manutenção do repertório verbal, além de estudos experimentais sobre treino de operantes verbais, história de reforçamento e efeitos de audiências sobre o discurso explicativo.

Sumário

Apresentação 12

Alcance e limites 12
As regras de leitura 12
Como ler os quadros e os termos 13
O mapa do livro 14
Ordem de leitura 15
Referências *15*

Nota sobre o processo de produção 16

O que veio dos autores 16
O que a IA fez 17
Exemplos de prompts 19
O que fizemos depois 23
Limites do processo 25
Uma leitura comportamental do próprio processo 25
Responsabilidade 25

PARTE I **O fenômeno e os dados** 26

Capítulo 1 — Desinformação e discurso de ódio: definições, medidas e dados 27

1.1 Definições, e o que cada uma exige medir 27
1.2 Prevalência: quanto circula e quem compartilha 29
1.3 Elites políticas como fonte 31
1.4 Posicionamento político e propensão à informação falsa: o que os estudos correlacionais mostram 32
1.5 Limitações: o que os desenhos não permitem 35
1.6 O que se sabe sobre o Brasil 36
1.7 Discurso de ódio: definições, medidas e dados 37
1.8 Interpretações rivais e limites 39
Síntese *40*
Referências *41*

Capítulo 2 — O que dizem as agências de checagem (e o que elas não podem dizer) 44

2.1 Duas séries de contingências 45
2.2 O que cada fonte conta 47
2.3 Por que a proporção entre as checadas não é a proporção entre as ditas 48
2.4 O que a Lupa permite dizer 49
2.5 O que o Aos Fatos permite dizer 52
2.6 Falas e menções: de quem é a falsidade 53
2.7 As agências concordam entre si? 54
2.8 O que a checagem não alcança 54
2.9 Protocolo para uma amostra comparável 56
Síntese *57*
Referências *58*

PARTE II **Análise operante dos discursos** 59

Capítulo 3 — Declarações falsas como operantes verbais 60

3.1 A fala do político como comportamento verbal 61
3.2 A ficha de análise 62
3.3 Tato ou intraverbal? O contínuo e a forma de tato 65
3.4 Tato distorcido: quando a consequência do ouvinte desloca o controle pelo evento 68
3.5 Repetição: a força da afirmação repetida 71
3.6 Autoclíticos de fonte e de certeza: o que se acrescenta e o que se retira 73
3.7 A declaração como regra para a audiência 76
3.8 Casos: as mesmas quatro frases, pelo mesmo método 77
3.9 Interpretações rivais e como decidir entre elas 81
Síntese *82*
Referências *83*

Capítulo 4 — Atitudes políticas e preconceito como classes de equivalência . 86

4.1 Atitude como comportamento verbal avaliativo 87
4.2 Como uma palavra adquire função sem contato com o referente 88
4.3 Transferência de função 92
4.4 Preconceito nos experimentos 97
4.5 Atitudes em relação a presidentes do Brasil 101
4.6 "Nós" e "eles": relações de oposição e de comparação 115
4.7 Mudança de atitude: reorganização das classes e histórias em quadrinhos 117
4.8 Medidas, interpretações rivais e limites 126
Síntese *129*
Referências *130*

Capítulo 5 — Discurso de ódio: mando disfarçado e controle aversivo 135

5.1 O que o comportamento do falante produz: aprovação, retirada de ameaças e funções que variam com o contexto 136
5.2 Mando disfarçado: a opinião que funciona como instrução 138
5.3 Controle aversivo dirigido a um grupo 141
5.4 O rótulo do grupo e a transferência de função aversiva 143
5.5 Do discurso ao ato: o que os dados permitem afirmar 144
5.6 Um caso com classificação externa: a fala de 2011 e o que veio depois 149
5.7 Interpretações rivais e limites 150
Síntese *151*
Referências *152*

Capítulo 6 — Bolhas digitais como comunidades verbais e a esquiva de contracontrole 155

6.1 A bolha como comunidade verbal: audiência e fonte de reforço 156
6.2 O que os dados dizem sobre câmaras de eco (e onde divergem) 158
6.3 "Realidades virtuais": quando o controle verbal substitui o contato com o evento 162
6.4 Contracontrole e esquiva de contracontrole 163
6.5 O caso brasileiro: WhatsApp, eleições e pandemia 168
6.6 Quem curte, quem responde: uma análise exploratória de dois vídeos políticos no Instagram 171
6.7 Análises semânticas em mídias sociais 185

6.8 Interpretações rivais e limites ... 187
Síntese ... *189*
Referências ... *189*

Capítulo 7 — O voto como comportamento verbal ... 192
7.1 O voto pela definição de comportamento verbal ... 193
7.2 Mando, tato ou intraverbal? Funções possíveis do voto ... 194
7.3 O voto secreto como retirada da audiência ... 196
7.4 Declarar e votar: atitudes verbais e atos ... 198
7.5 Consequências sociais do votar: os experimentos de campo ... 203
7.6 Nomes e números de candidatos em classes de equivalência ... 205
7.7 Interpretações rivais e limites ... 207
Síntese ... *208*
Referências ... *209*

PARTE III A ecologia da informação 210

Capítulo 8 — Economia da atenção: o algoritmo como programador de contingências ... 211
8.1 Atenção como reforçador ... 212
8.2 O que os dados mostram sobre conteúdo emocional, moralizado e negativo ... 214
8.3 O algoritmo como programador: quanta audiência cada emissão recebe ... 217
8.4 Reforçamento intermitente e engajamento ... 221
8.5 Experimentos com o próprio algoritmo: o que mudou e o que não mudou ... 226
8.6 Interpretações rivais e limites ... 229
Síntese ... *230*
Referências ... *231*

Capítulo 9 — O ciclo produtor-audiência ... 234
9.1 Opiniões podem ser reforçadas: os experimentos de 1956 a 1969 ... 235
9.2 O que o produtor recebe: curtidas, compartilhamentos e alcance ... 241
9.3 O que a audiência faz e o que a mantém ... 243
9.4 O ciclo: cada lado modela o outro ... 245
9.5 Indignação moral aprendida por reforço: a evidência ... 246
9.6 O político como produtor de conteúdo ... 249
9.7 A metacontingência do ecossistema: produtor, audiência, plataforma e anunciante ... 250
9.8 Interpretações rivais e limites ... 252
Síntese ... *254*
Referências ... *254*

Capítulo 10 — A janela de Overton: o que se pode dizer sem ser punido ... 257
10.1 O conceito, a origem e a tradução proposta ... 258
10.2 Da janela da política à janela do enunciado: o que desloca a borda ... 261
10.3 Informação sobre o que os outros toleram: o experimento de Pittsburgh ... 262
10.4 Duas eleições, duas medidas: Brasil, 2018, e Estados Unidos, 2016 ... 265
10.5 Políticos e redes sociais: quem fala de cima e o que recebe ... 269
10.6 Sanção, contracontrole e o que mantém a borda no lugar ... 271
10.7 Como situar uma fala em relação à janela ... 274
10.8 Interpretações rivais e limites ... 275
Síntese ... *277*

Referências *278*

Capítulo 11 — Infodemiologia: como a informação se espalha e o que se mede . . . 280

11.1 Infodemiologia e infovigilância 281
11.2 Medidas de difusão: alcance, exposição, reação e a estrutura da cascata 283
11.3 O falso se espalha mais rápido que o verdadeiro? 286
11.4 Modelos epidemiológicos aplicados à informação 289
11.5 A infodemia de Covid-19, no mundo e no Brasil 292
11.6 O que a infodemiologia propõe fazer 295
11.7 Perguntas para ler um estudo de difusão 296
Síntese *297*
Referências *298*

PARTE IV **Intervenções** 299

Capítulo 12 — Intervenções: o que se pode mudar, em quatro níveis, e com que evidência 300

12.1 Corrigir depois: a correção, o aviso e o "efeito rebote" 302
12.2 Corrigir antes: inoculação como treino discriminativo 305
12.3 Estímulos que sinalizam precisão 309
12.4 Diante de quem fala: contra-fala, compromisso público e autoclíticos de fonte 312
12.5 Reduzir preconceito por equivalência e por histórias em quadrinhos 315
12.6 Mudar as contingências da plataforma 319
12.7 Planejamento cultural e os limites éticos de quem planeja 323
Síntese *326*
Referências *327*

Apêndice A — Ficha de análise de discurso 330

Apêndice B — Conceitos em uso 332

Apêndice C — Referências 336

Apêndice D — Verificação de fontes 345

D.1 Como uma afirmação chega a "verificada" 345
D.2 Números 346
D.3 Versões lidas que não são a do periódico 346
D.4 Fontes não lidas 347
D.5 Decisões de método que afetam a leitura 352
D.6 Deslizes de procedimento registrados 353

Apresentação

Este livro reúne uma pergunta e um método. A pergunta é o que mantém, em termos de contingências, o que políticos, plataformas e audiências dizem e repassam quando o enunciado é falso ou é classificado por alguém como discurso de ódio. O método é o da Análise do Comportamento aplicada ao comportamento verbal: em vez de inferir o que o falante tinha em mente, pergunta-se que variáveis controlam a emissão e o que ela produz em quem ouve. Skinner (1957) dá o vocabulário; a literatura experimental sobre atitudes, equivalência e contingências culturais dá o que se sabe e o que ainda não se sabe.

Foi escrito para profissionais que já conhecem a Análise do Comportamento: analistas do comportamento, psicólogos, pesquisadores. Por isso não há parte de fundamentos. Cada conceito entra, num quadro "Conceito em uso", no ponto em que a análise passa a precisar dele, e o Apêndice B os reúne com o capítulo em que cada um foi apresentado.

Alcance e limites

Aqui não se classificam falas como falsas, não se aponta quem mente e não se defende um campo político. Quem classifica é uma agência de checagem, uma decisão judicial, uma plataforma ou um estudo, e o texto diz qual. Também não se testou nenhuma hipótese: foram lidos estudos de terceiros e, onde se propõe uma leitura operante, diz-se que é leitura. O que se tira daqui é um modo de ler o dado e de formular a pergunta que falta, e não uma resposta sobre quem é mais desonesto.

Duas consequências aparecem em todos os capítulos. A primeira é que a classificação de um operante verbal (mando, tato, intraverbal, tato distorcido) é sempre interpretação, e nunca dado: o texto diz que variáveis de controle a sustentam e qual seria a interpretação rival. A segunda é que a cobertura é desigual. Boa parte dos estudos é de língua inglesa e dos Estados Unidos, vem sobretudo do Twitter e da Covid-19, e para o Brasil há poucos; cada capítulo diz o que isso permite e o que não permite concluir.

As regras de leitura

Cinco regras valem do começo ao fim.

A primeira é o mesmo método para todos os campos políticos. Onde se analisa o que um político disse, procura-se o caso espelhado do outro campo com o mesmo critério, e quando ele não é encontrado, diz-se que não há. A regra de amostra comparável do Capítulo 2 (seção 2.9) foi fixada antes de qualquer coleta, para que a escolha dos casos não dependesse do resultado; a coleta não foi feita, e o capítulo diz por quê.

A segunda é que a falsidade é sempre da agência que a classificou. "Falso" aparece com o nome da agência, o rótulo que ela usou e a data, e nunca como propriedade que quem escreve atribui à frase.

A terceira é que "discurso de ódio" é rótulo externo. Só se chama assim o enunciado classificado por instância externa, com a instância nomeada. Sem ela, descreve-se a função do enunciado, e não se o rotula.

A quarta é a separação entre dado, interpretação e hipótese. Toda análise de dados ou de discurso traz um quadro com as três partes: o que a fonte mostra, o que se lê nisso e o que fica como hipótese, com o que seria preciso para testá-la.

A quinta trata das fontes. Só entra como fonte o que foi lido no texto integral, e cada afirmação atribuída a uma fonte foi localizada nela e lida em contexto; as planilhas desse trabalho estão descritas no Apêndice D. Onde a fonte não pôde ser obtida por via legal, escreve-se que não foi lida, e ela não é citada como verificada. As páginas citadas são as do periódico quando o texto lido as traz; quando não traz, a citação vai sem página, e o capítulo avisa a versão que foi lida (manuscrito aceito, documento de trabalho, texto em XML).

Como ler os quadros e os termos

Há três tipos de quadro. O quadro "Dado · Interpretação · Hipótese" fecha as análises. O quadro "Conceito em uso" apresenta um conceito da literatura, com a fonte e o ponto em que ele passa a ser usado; quando um capítulo o retoma, o quadro diz "retomada" e remete ao capítulo de origem. O quadro "Definição proposta" marca o termo que é proposto aqui e que nenhuma fonte lida usa, como a janela de uma audiência (Capítulo 10) e a distinção entre alcance, exposição e reação (Capítulo 11); "esquiva de contracontrole" (Capítulo 6) é proposta feita aqui, a partir de Delprato (2002), e vem sempre como hipótese.

As citações seguem o padrão APA: "Autor (ano, p. X)". As referências de cada capítulo estão consolidadas no Apêndice C. No texto, "foi lido" e "não foi lido" querem dizer exatamente isto: o texto integral foi ou não foi lido por quem escreveu.

Os estudos citados trazem estatísticas, e o texto as explica na primeira vez em que as usa; este parágrafo é para ter à mão. O valor de *p* é a probabilidade de se obter uma diferença tão grande quanto a observada, ou maior, se na verdade não houvesse diferença; por convenção, *p* abaixo de 0,05 faz a diferença ser chamada de "significativa", o que informa que ela dificilmente é acaso e não que seja grande ou importante. *t* e *F* são as estatísticas dos testes que produzem o *p*. O *d* e o *h* de Cohen medem o tamanho de um efeito, a diferença entre médias no primeiro e entre proporções no segundo, e valores perto de 0,2 são pequenos. A razão de chances (*odds ratio*, OR) compara a chance de um desfecho em dois grupos: 1 significa chances iguais. O desvio-padrão é a variação típica de uma medida, e o intervalo de confiança de 95% é a faixa de valores plausíveis para a estimativa. O R^2 é a fração da variação dos dados que um modelo explica, e a acurácia, a precisão, a revocação e o *F*1 descrevem os acertos de um classificador de texto. Quando um estudo não tem grupo de controle, ou mede só o que as pessoas declaram, o texto diz.

O mapa do livro

O livro se divide em quatro partes e quatro apêndices. A Parte I, "O fenômeno e os dados", tem dois capítulos. O Capítulo 1 reúne as definições e os estudos de prevalência e correlacionais, com as limitações de cada medida, e fixa a regra de que o que se mede é função do instrumento. O Capítulo 2 trata das agências de checagem: o que a proporção de falsas publicada por uma agência mede e o que não mede, e por que não se compara com a de outra sem as ressalvas do capítulo.

A Parte II, "Análise operante dos discursos", tem cinco. O Capítulo 3 é o núcleo: apresenta a ficha de análise e os operantes verbais que os seguintes usam, e a aplica a quatro falas checadas. O Capítulo 4 trata das atitudes e do preconceito como classes de equivalência. O 5 trata do discurso de ódio como mando disfarçado e controle aversivo; o 6, das bolhas digitais como comunidades verbais e da esquiva de contracontrole; o 7, do voto como comportamento verbal.

A Parte III, "A ecologia da informação", tem quatro. O Capítulo 8 trata da economia da atenção e do algoritmo como programador de contingências; o 9, do ciclo produtor–audiência; o 10, da janela de Overton, definida pelo que uma audiência pune; o 11, da infodemiologia e das medidas de difusão. A Parte IV tem um capítulo, o 12, sobre as intervenções que decorrem da análise, em quatro níveis, e a evidência de cada uma. Os apêndices trazem a ficha de análise (A), os conceitos em uso (B), as referências consolidadas (C) e a verificação de fontes, com as lacunas (D).

Ordem de leitura

Os Capítulos 1 e 2 podem ser lidos à parte, por quem quer só os dados e a leitura crítica das checagens. O Capítulo 3 é o ponto de apoio dos demais: quem vai ler a análise de falas e de discurso de ódio, nos Capítulos 5 e 7, deve passar por ele e pelo 4. Fora isso, os capítulos são autônomos, com remissões explícitas ao que usam de outro. O Capítulo 12 se lê melhor por último, porque usa o quadro de treino discriminativo, as distinções dos Capítulos 8 e 9 e a definição de alcance, exposição e reação do Capítulo 11.

Quem usa o livro como ferramenta de análise encontra perguntas prontas em quatro pontos: a ficha do Capítulo 3 (seção 3.2), reproduzida no Apêndice A, para ler uma fala; as do Capítulo 10 (seção 10.7), para situar uma fala numa janela; as do Capítulo 11 (seção 11.7), para ler uma medida de difusão; e as do Capítulo 12 (seção 12.7), para ler uma intervenção. A ficha é a de maior uso: transforma o que os capítulos descrevem numa sequência de perguntas sobre a fala, a audiência e as consequências, e termina sempre na que importa, que é o que mantém a fala.

Referências

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 o processo de produção

Este livro foi escrito com uso intensivo de inteligência artificial generativa, e esta nota descreve como. Os autores somos nós, os seis signatários (Márcio Moreira, Carlos Augusto de Medeiros, Cristiano Coelho, Diogo Conque Seco-Ferreira, Frederico Neves Condé e Lucas Ferraz Córdova): definimos o problema, o quadro interpretativo e os critérios de análise, e respondemos por cada afirmação publicada. A IA redigiu as versões iniciais de todos os capítulos, da Apresentação e dos Apêndices A e B, e conduziu a verificação das fontes; nós as revisamos, alteramos, ajustamos e conferimos, entre 19 e 26 de outubro de 2026. A condução da IA, do pedido original ao último pedido de revisão, foi toda feita por Márcio Moreira.

Preferimos descrever o processo em detalhe a resumi-lo numa linha de agradecimento. O leitor tem o direito de saber como o texto foi produzido para avaliá-lo, e uma descrição vaga serviria mal tanto à transparência quanto ao trabalho que de fato fizemos. Pelo mesmo motivo, não chamamos este de um livro "gerado por IA": a expressão descreve mal um processo em que o problema, o método e o juízo final foram humanos do início ao fim.

O que veio dos autores

O que orienta o livro antecede a redação com IA e veio de nós, no pedido original de 10 de outubro de 2026 e nos pedidos seguintes.

- **O problema.** A escolha do discurso de ódio e da desinformação, com foco nas declarações de políticos brasileiros, como objeto, e a pergunta central: o que uma análise skinneriana consegue dizer sobre esse fenômeno a partir dos dados que a ciberpsicologia e as áreas vizinhas (ciência política, comunicação, infodemiologia) já produziram.
- **O quadro interpretativo.** Uma posição skinneriana explícita, fundada no comportamento verbal (*Verbal behavior*, de Skinner, e *Schedules of reinforcement*, de Ferster e Skinner, que abrimos como exceção à regra de que só se citam artigos), combinada com a análise contextual de Bernard Guerin. Do artigo de Guerin que indicamos como base entram as atitudes e as crenças como comportamento verbal mantido por comunidades verbais, as funções de tato, intraverbal e mando de um mesmo enunciado e o papel dos autoclíticos na negociação social do que se diz; de outro trabalho dele, a análise do conhecimento como construção social e a condição em

que um grupo pequeno e coeso sustenta o que diz. Somam-se a isso a equivalência de estímulos e a transferência de função, trabalhadas em livros nossos sobre atitudes e sobre histórias em quadrinhos. Esse quadro chegou às ferramentas pelo pedido original, pelo artigo-base, por esses livros e pelas cópias de Skinner que fornecemos.

- **A seleção da literatura.** Fixamos as regras: só artigos científicos, mais nossos livros sobre atitudes e histórias em quadrinhos e as duas obras de Skinner; só textos obtidos por via legal, sem Sci-Hub nem equivalentes; só se cita o que foi lido no texto integral; sites que proíbem rastreadores de IA, como os da Lupa e do Aos Fatos, não são coletados em massa. A lista de candidatas veio em parte de nós (27 artigos do nosso acervo e 12 livros nossos) e em parte da IA (97 artigos previstos a partir do sumário e 34 acrescentados depois de 20 buscas dirigidas no OpenAlex). Aprovamos o levantamento em 10 de outubro de 2026, e ele chegou a 168 fontes, das quais 108 foram lidas no texto integral e 60 ficaram pendentes (Apêndice D.2 e D.4).
- **Os critérios de análise.** O que conta como contingência, como consequência social e como controle verbal. Classificar um operante verbal (mando, tato, intraverbal, tato distorcido) é sempre interpretação, e nunca dado. O mesmo método vale para políticos de todos os campos; "falso" é sempre a classificação de uma agência nomeada; "discurso de ódio" é rótulo de uma instância externa; dado, interpretação e hipótese vêm separados. E o que não pretendemos afirmar: não oferecemos dados robustos sobre o fenômeno, mas leituras que possam orientar experimentos futuros em Análise do Comportamento.
- **A estrutura.** Os dez tópicos do pedido original (dados e estudos correlacionais, agências de checagem, operantes verbais, equivalência de estímulos, bolhas digitais, voto, economia da atenção, ciclo produtor–audiência, infodemiologia e intervenções), o capítulo sobre a janela de Overton, que acrescentamos em 11 de outubro, e a aprovação do sumário em quatro partes, em 10 de outubro.

O que a IA fez

Usamos o Claude Sonnet 5.5, da Anthropic, no Claude Code, ambiente em que o modelo lê e escreve os arquivos do projeto, executa scripts e consulta bases bibliográficas abertas, entre 10 e 26 de outubro de 2026. O histórico de versões mostra também o Claude Opus 5.5 em tarefas de organização do repositório e

de diagramação, sem redação de texto do livro. As ferramentas trabalharam sempre a partir do quadro interpretativo e das instruções que fornecemos, e quem as conduziu, do começo ao fim, foi Márcio Moreira.

- **Levantamento e obtenção das fontes.** Propôs candidatas, conferiu autores, ano, periódico, volume e páginas no Crossref, obteve os textos por vias legais (acesso aberto, repositórios, versões postadas pelos autores e os PDFs que anexamos) e extraiu o texto, com reconhecimento óptico de caracteres quando a camada de texto era ruim. O que não obteve ficou numa lista de pendências e não é citado como verificado.
- **Estrutura.** Propôs o sumário detalhado e a ordem de escrita: os capítulos 4, 9, 6, 8 e 10 primeiro, porque tinham as fontes completas, e os demais em seguida.
- **Redação inicial.** A partir do quadro, desenvolveu as primeiras versões de todas as seções teóricas, da Apresentação e dos Apêndices A e B.
- **Aplicação do quadro aos estudos.** Produziu as primeiras leituras skinnerianas dos estudos selecionados: descrição dos procedimentos, identificação de possíveis contingências, interpretações rivais e limites, e a interpretação dos resultados, sempre com dado, interpretação e hipótese separados.
- **Verificação das afirmações.** Registrou cada afirmação atribuída a uma fonte numa planilha, com as frases literais que a sustentam; um script procurou essas frases no texto integral, e a IA leu o contexto e registrou o resultado e as ressalvas. Um segundo script conferiu que toda citação do texto tem referência e toda afirmação marcada existe e está verificada. A planilha tem 1.207 entradas (Apêndice D.1). No levantamento inicial, a conferência corrigiu autores, anos de volume e páginas que a IA havia registrado de memória e fez retirar uma afirmação atribuída a um artigo que ele não faz.
- **Figuras, referências e arquivos.** Gerou as figuras por script e recortou, dos livros do Instituto Walden4, as imagens dos estudos que o Capítulo 4 descreve; montou as referências em APA e os arquivos de entrega (.docx, HTML e PDF).
- **Reescritas a pedido.** Reescreveu em prosa as descrições de estudos e refez o Capítulo 4 com as imagens dos nossos estudos e das histórias em quadrinhos, conforme os pedidos descritos abaixo.

A IA não é autora. Decidiu, por delegação nossa, muitas questões de redação (a ordem dos capítulos, a numeração, o título de capítulos, o tratamento de lacunas, o que ficou de fora quando uma fonte não pôde ser obtida), e cada decisão está registrada com data e com o caminho para revertê-la. Três termos que

aparecem como definições próprias do livro nasceram na redação com IA: “esquiva de contracontrole” (Capítulo 6), “janela de uma audiência” (Capítulo 10) e a distinção entre alcance, exposição e reação (Capítulo 11). Cada um é declarado como proposta no texto, e a avaliação deles é nossa. A palavra final sobre cada interpretação é dos autores.

A redação se deu em sessões de conversa com a IA sobre os arquivos do projeto, um capítulo por vez, a partir de 10 de outubro de 2026. Começou pelo pedido original, que o repositório guarda sem alterações. A IA entregou por etapas (o artigo-base, o sumário, o levantamento bibliográfico e só então os capítulos), e aprovamos cada etapa antes da seguinte. Em outros momentos respondemos “você pode decidir por mim” e “Continue a redação do livro e a finalize. Você pode tomar todas as decisões”: a IA decidiu e registrou cada decisão, com data, num arquivo próprio, e a delegação não valeu como aprovação de capítulo. Pedimos também correções de rumo depois de ler o rascunho: que o texto deixasse de tratar “o livro” como quem lê, diz e propõe; que as aberturas de seção se bastassem sem o título; que as descrições de pesquisa, truncadas e pouco didáticas, fossem revistas parágrafo a parágrafo; e que o Capítulo 4 fosse refeito com figuras coloridas e didáticas e com as imagens dos nossos estudos, inclusive das histórias em quadrinhos. Alguns desses pedidos aparecem textualmente na próxima seção. O registro completo, com o pedido original, as decisões datadas, a planilha de verificação e o histórico de versões, está no repositório do projeto e pode ser consultado mediante solicitação.

Exemplos de prompts

Quase todos os pedidos foram mensagens curtas, em linguagem corrente, feitas no meio do trabalho; a exceção é o pedido original, longo e estruturado, que reproduzimos na íntegra. Todos foram escritos por Márcio Moreira. Os demais exemplos são textuais e vêm de pontos diferentes do processo; neles a grafia está corrigida, e as palavras acrescentadas ou trocadas estão entre colchetes. Cada exemplo é seguido do que a IA fez com ele, segundo o registro do projeto.

O pedido original (10 de outubro de 2026), na íntegra. Os títulos das partes do pedido aparecem em negrito.

> **PAPEL**
>
> Você é um analista do comportamento (referencial skinneriano, Análise Experimental e Aplicada do Comportamento) com experiência em comportamento verbal, equivalência de estímulos e análise de práticas culturais, escrevendo para um público de graduação avançada e pós-graduação em Psicologia.
>
> **OBJETIVO**

Escrever o livro “Uma análise operante de discursos de ódio e desinformação”, em português do Brasil, com linguagem didática e rigorosamente técnica (terminologia da Análise do Comportamento, definições operacionais, sem mentalismo explicativo). O foco são discursos e declarações falsas de políticos brasileiros.

FLUXO DE TRABALHO (execute em etapas e pare para minha aprovação ao fim de cada uma)

1. Leia primeiro o artigo-base: https://pmc.ncbi.nlm.nih.gov/articles/PMC2733686/ e resuma como ele fundamentará a análise de comportamento verbal.
2. Proponha o sumário detalhado (capítulos, seções, figuras previstas, fontes previstas).
3. Monte o levantamento bibliográfico (ver “Regras de fontes”) e a tabela de verificação, antes de escrever qualquer capítulo.
4. Escreva um capítulo por vez, com figuras, e entregue como documento.
5. Ao final, faça a revisão cruzada de todas as citações e referências.

CONTEÚDO (acrescente tópicos pertinentes que eu não listei)

a) Dados e estatísticas sobre desinformação (“fake news”) e discurso de ódio, incluindo estudos correlacionais entre propensão à desinformação e posicionamento político. Apresente limitações (correlação ≠ causa, amostras, medidas, contexto cultural).

b) Dados de agências de checagem (Aos Fatos, Lupa, etc.) sobre declarações falsas ou enganosas de Jair Bolsonaro durante o mandato e de Lula. Use o mesmo critério para os dois: mesma agência, período comparável, mesma metodologia de classificação, denominador explícito (nº de declarações checadas) e ressalva sobre viés de seleção das checagens.

c) Análise dos discursos em termos dos operantes verbais (mando, tato, ecoico, intraverbal, textual, etc., e tato distorcido/mando disfarçado), com base no artigo-base.

d) Análise em termos de equivalência de estímulos: formação de classes, transferência de função, discriminações condicionais, propriedades (reflexividade, simetria, transitividade, equivalência), relação com formação de preconceito e atitudes. Referencie os livros de Márcio Moreira sobre atitudes e sobre HQs e atitudes (walden4.com.br/livros).

e) Bolhas digitais como comunidades verbais; criação de “realidades virtuais” por controle verbal; esquiva de contracontrole.

f) Comportamento de votar como comportamento verbal.

g) Economia da atenção; pesquisas sobre a tendência de usuários a consumir conteúdo emocional/moralizado; algoritmos como programas de reforçamento.

h) Ciclo de reforçamento entre produtor de conteúdo e audiência (curtidas, compartilhamentos, audiência como reforçadores generalizados), análise das contingências de ambos.

i) Epidemiologia da informação (infodemiologia): difusão, taxas de propagação, intervenções.

j) Implicações práticas: intervenções culturais e educacionais baseadas em análise do comportamento.

FIGURAS

Ilustre todo procedimento ou conceito que se beneficie de diagrama: discriminações condicionais (MTS), classes de equivalência e suas propriedades, tríplice contingência, ciclo produtor–audiência, bolhas como comunidades verbais. Gere as figuras como arquivos (Graphviz/matplotlib/SVG), com legenda numerada, vetoriais, legíveis em impressão P&B e originais (não reproduzir figuras de terceiros).

REGRAS DE FONTES

- Cite apenas artigos científicos, mais seus livros sobre atitudes e HQs (exceção explícita). Não cite outros livros.
- Obtenha os artigos apenas por vias legais: acesso aberto (PMC, SciELO, Unpaywall, OpenAlex, repositórios), acesso institucional que eu disponibilizar, ou PDFs que eu anexar. Não use Sci-Hub ou equivalentes. O que não for obtido legalmente entra em uma lista "PDF pendente" e NÃO é citado como verificado.
- Para cada artigo citado, leia o texto completo e confirme que ele contém a informação atribuída. Registre numa tabela: referência | DOI | trecho/página que sustenta a afirmação | status (verificado / pendente).
- Valide DOI, autores, ano, periódico, volume e páginas (ex.: Crossref). Nunca invente ou complete referência de memória; se não conseguir verificar, marque como pendente.
- Faça duas passagens de verificação (citações no texto ↔ lista de referências ↔ tabela de verificação). Use formato APA 7.
- Citações literais: curtas, entre aspas, com página, e acompanhadas de imagem (recorte) do trecho do PDF original. Prefira paráfrase fiel.

NOTÍCIAS

Use matérias apenas de G1, CNN Brasil, BandNews e Canal Meio. Para cada uma: trecho curto citado com crédito, URL, data de publicação e data de acesso, mais captura de tela do trecho. Registre também a data de cada fala analisada e a checagem correspondente.

TOM E RIGOR POLÍTICO

Descreva comportamentos e contingências, não "essências" de pessoas ou partidos. Analise políticos de diferentes campos com o mesmo método. Separe sempre dado verificado, interpretação analítica e hipótese. Evite afirmações difamatórias; atribua cada classificação de falsidade à agência que a fez.

> **ENTREGA**
>
> Cada capítulo como documento (.docx), figuras em pasta separada, tabela de verificação de referências em planilha, e lista de pendências (PDFs não obtidos, dados não confirmados).

O que a IA fez. Desse pedido vieram o trabalho em etapas (artigo-base, sumário, levantamento bibliográfico, capítulos e revisão cruzada), a planilha de verificação, a regra de só citar o que foi lido no texto integral e a separação entre dado, interpretação e hipótese em todos os capítulos. A instrução de figuras legíveis em preto e branco foi revista depois, a nosso pedido: as figuras do Capítulo 4 passaram a coloridas em 16 de outubro, e as demais em 26 de outubro de 2026.

Aprovar e seguir (10 de outubro de 2026). Entre as etapas, respondemos com frases como estas:

> Vá em frente.
>
> Aprovado. Continue.

O que a IA fez. Tratou cada resposta como aprovação só da etapa que a pergunta cobria; os pontos que não foram respondidos continuaram em aberto no registro de decisões.

Delegar decisões (10 e 14 de outubro de 2026).

> Você pode decidir por mim.
>
> Continue a redação do livro e a finalize. Você pode tomar todas as decisões.

O que a IA fez. Decidiu o que faltava (ordem de escrita, numeração, tratamento de fontes que não foram obtidas, títulos) e registrou cada decisão, com data e com o caminho para revertê-la. A delegação não valeu como aprovação de capítulo.

Acrescentar escopo (11 de outubro de 2026).

> Adicione um capítulo com análise comportamental da janela de Overton, relacionando com discursos de políticos e redes sociais.

O que a IA fez. Escreveu o capítulo, hoje o Capítulo 10, e propôs a definição de "janela de uma audiência" que ele usa, declarada como proposta no texto.

Responder a uma pergunta da IA, com uma resposta ambígua (11 de outubro de 2026). A IA tinha perguntado se a exceção de fontes aberta para Skinner valia também para a obra de Ferster e Skinner sobre esquemas de reforçamento. A resposta foi:

> Fester, ok. Vá em frente.

O que a IA fez. Leu "Fester" como Ferster e Skinner, registrou essa leitura e avisou que, se o livro pretendido era outro, bastava dizer.

Corrigir a redação depois de ler o rascunho (16 de outubro de 2026).

> "O livro" entr[a] como uma entidade no próprio livro (e.g., "O primeiro estudo que o livro leu é um jogo"). Há parágrafos de início de seção que começam de forma abrupta e confusa ou considerando que o título da seção já [faz] parte do parágrafo. Identifique e ajuste essas questões.

O que a IA fez. Reescreveu cerca de 185 frases em voz impessoal, em todos os capítulos, e refez as aberturas de seção para que se entendam sem o título.

> Há descrições de pesquisa que estão truncadas, pouco didáticas. Revise parágrafo a parágrafo essa questão da redação.

O que a IA fez. Reescreveu em prosa as descrições de estudos dos doze capítulos e da Apresentação, uma por uma, com a pergunta, a amostra, o procedimento, a medida, o resultado e a ressalva de cada estudo.

Pedir detalhe e imagens (16 de outubro de 2026).

> No capítulo de atitudes e equivalência, use em detalhes e usando as imagens desses dois trabalhos: [endereços dos dois trabalhos]. Use figuras coloridas, muito didáticas. Refaça o capítulo com este tom.
>
> Insira, onde for possível e adequado, referências a: [endereços de onze trabalhos]. Use imagens desses livros, deixe tudo bem ilustrado, incluindo imagens das HQs.

O que a IA fez. Refez o Capítulo 4 com figuras coloridas, recortou imagens dos estudos e das histórias em quadrinhos do Instituto Walden4 e incluiu as obras nas referências, com o endereço de cada PDF. Um dos endereços aparecia duas vezes na lista, e a IA o leu como o volume seguinte, registrando a suposição para que a confirmássemos.

O que fizemos depois

Entre 19 e 26 de outubro de 2026, os autores trabalharam sobre o texto produzido nas etapas anteriores: revisaram, alteraram, ajustaram e conferiram os capítulos, a Apresentação e os apêndices, num manuscrito comum.

1. **Revisão conceitual.** Avaliamos se cada argumento era coerente com o quadro skinneriano e com a análise contextual, e descartamos ou refizemos os que não eram.

2. **Alteração e ajuste.** Corrigimos erros conceituais, inferências sem apoio nos dados e extrapolações, e reescrevemos os trechos que pediam outra formulação.
3. **Conferência.** A conferência de cada afirmação contra o texto integral da fonte foi feita primeiro pela IA, com o procedimento do Apêndice D; nós conferimos os textos nessa mesma semana de revisão. As 60 fontes que não pudemos obter não são citadas como verificadas, e onde o assunto exigiria a fonte o capítulo diz que ela não foi lida.
4. **Edição final.** Unificamos terminologia, estilo e a articulação entre capítulos.

O quadro a seguir resume as etapas por capítulo.

QUADRO

Etapas do trabalho, por capítulo

Capítulo	Primeira versão pela IA	Obras na lista do capítulo	Entradas verificadas	Revisão e conferência pelos autores
Apresentação	15/10/2026	2	0	19 a 26/10/2026
1. Definições e dados	14 a 15/10/2026	21	48	19 a 26/10/2026
2. Agências de checagem	10/10/2026	10	39	19 a 26/10/2026
3. Declarações falsas	14 a 15/10/2026	13	92	19 a 26/10/2026
4. Atitudes e equivalência	10/10/2026, refeito em 16/10/2026	39	300	19 a 26/10/2026
5. Discurso de ódio	14 a 15/10/2026	13	45	19 a 26/10/2026
6. Bolhas digitais	11/10/2026	13	87	19 a 26/10/2026
7. O voto	14 a 15/10/2026	13	54	19 a 26/10/2026
8. Economia da atenção	11/10/2026	12	80	19 a 26/10/2026
9. Ciclo produtor–audiência	10/10/2026	13	123	19 a 26/10/2026
10. Janela de Overton	11/10/2026	11	80	19 a 26/10/2026
11. Infodemiologia	14 a 15/10/2026	12	97	19 a 26/10/2026
12. Intervenções	14 a 15/10/2026	27	162	19 a 26/10/2026

Nota. As datas da primeira versão são as do primeiro registro de cada arquivo no histórico do projeto; o Capítulo 10 foi começado em 11/10/2026 com outro número e renumerado em 14/10/2026. “Entradas verificadas” são as afirmações e os dados de fontes registrados na planilha

de verificação para o capítulo, cada um com o trecho lido em contexto, antes da revisão dos autores e da seção 6.7. Em 16/10/2026 pedimos a todos os capítulos duas revisões de redação (o texto sem “o livro” como agente e as descrições de estudos em prosa), e o Capítulo 4 foi refeito. Os Apêndices C e D são gerados por script a partir da planilha.

Limites do processo

Dois limites merecem ser ditos. O primeiro é que a verificação confirma que a passagem atribuída a uma fonte está no texto dessa fonte, na página indicada, e não que a leitura operante que se faz dela esteja certa; essa é a parte que os capítulos chamam de interpretação ou de hipótese, e a que mais depende do nosso juízo. O segundo é que modelos de linguagem erram de memória, e a própria IA registrou, no início do trabalho, dados que o Crossref e o texto das fontes desmentiram; por isso nada entrou no livro sem a conferência descrita no Apêndice D, e onde a fonte não pôde ser lida o texto diz que não foi lida.

Uma leitura comportamental do próprio processo

Para analistas do comportamento, a pergunta sobre autoria tem uma resposta conhecida, e ela não muda por causa da IA. Num ensaio de 1972, “On ‘having’ a poem”, Skinner compara escrever um poema a ter um filho: o poeta não cria a partir do nada, é o lócus onde variáveis genéticas e ambientais se encontram. O que chamamos de criatividade é variação e seleção, e boa parte do trabalho do escritor está em editar, isto é, em selecionar o que permanece.

Não invocamos Skinner para diminuir o papel da IA nem para inflar o nosso. A analogia vale para descrever o que de fato mudou: a IA ampliou muito a fonte de variação, e o trabalho autoral se concentrou ainda mais no arranjo das condições (o problema, o quadro, as instruções) e na seleção (o que foi aceito, corrigido, reescrito ou descartado). Essa é uma descrição do processo, não uma justificativa dele; se o leitor discordar dela, a descrição das seções anteriores continua valendo por si.

Responsabilidade

Os autores revisaram o conteúdo deste livro e assumem responsabilidade por ele, inclusive pelas partes cuja primeira redação foi feita com IA. Erros remanescentes são nossos.



PARTE I

O fenômeno e os dados

1 Desinformação e discurso de ódio: definições, medidas e dados

2 O que dizem as agências de checagem (e o que elas não podem dizer)

CAPÍTULO

1

Desinformação e discurso de ódio: definições, medidas e dados

Os números de Allcott e Gentzkow (2017) sobre a eleição presidencial americana de 2016 mostram, logo de saída, o problema que organiza este capítulo. No banco de notícias falsas que reuniram, 115 favoráveis a Trump foram compartilhadas 30 milhões de vezes no Facebook, e 41 favoráveis a Clinton, 7,6 milhões. Os mesmos autores estimam que o adulto americano médio viu e lembrou 1,14 notícia falsa e dizem, de forma expressa, que não avaliam se elas decidiram a eleição. Um leitor apressado junta esses números numa frase, e eles respondem a perguntas diferentes: o primeiro mede o que circulou; o segundo, o que alguém viu e lembrou; nenhum mede o que mudou no que alguém fez.

Este capítulo sistematiza o conhecimento acumulado sobre quantidade e forma de circulação da desinformação e do discurso de ódio, bem como o perfil de quem os produz e compartilha, explicitando as métricas empregadas e seus limites metodológicos. É o capítulo menos analítico do livro: não propõe leitura operante, e sim mostra os dados com que os capítulos seguintes trabalham e o que cada um deles exige que se diga em voz baixa. Vale nele a regra de todo o livro: antes de ler o número, perguntar o que o instrumento controla.

1.1 Definições, e o que cada uma exige medir

Cada definição de "informação falsa" traz embutida uma exigência de medida. Lazer et al. (2018) definem notícia falsa como informação fabricada que imita o conteúdo jornalístico na forma, mas não no processo organizacional nem na intenção; é uma subcategoria da desinformação em sentido amplo (misinformation), isto é, informação incorreta ou enganosa sobre o estado do mundo, e

seus veículos carecem das normas e dos processos editoriais da mídia. Se o elemento definidor é a intenção e o processo do editor, escrevem eles, a prevalência deve ser medida por editor, e não por notícia individual. Allcott e Gentzkow (2017) definem notícia falsa como artigo de notícia intencional e verificavelmente falso, que pode enganar o leitor, e excluem da definição erros não intencionais de reportagem. As duas definições exigem, em princípio, conhecer a intenção; nenhuma das medidas que as usam a observa.

Os estudos de compartilhamento resolvem o problema trocando o item pela fonte. Guess et al. (2019) tomam como notícia falsa o conteúdo falso ou enganoso intencionalmente vestido como notícia, em geral para gerar receita de anúncios, e cruzam os links que cada respondente compartilhou com listas de editores de notícias falsas feitas por jornalistas e acadêmicos. Lasser et al. (2022) usam a nota que o NewsGuard, um serviço de avaliação da confiabilidade de sites de notícias, dá ao domínio (de 0 a 100, abaixo de 60 “não confiável”) como estimador da confiabilidade do item compartilhado. É uma convenção explícita, a de inferir o item pela fonte, e não é um defeito escondido; mas muda o que o número diz: a proporção de links para domínios não confiáveis não é a proporção de afirmações falsas.

Os estudos experimentais trocam de novo o objeto. Rathje et al. (2023) escrevem que misinformation pode se referir a notícias fabricadas, boatos falsos, teorias da conspiração ou desinformação, e o que medem é o julgamento de manchetes verdadeiras e falsas pelos participantes. E Ricard e Medeiros (2020), numa nota sobre o caso brasileiro, distinguem a informação falsa que quem a dissemina acredita ser verdadeira daquela que quem a dissemina sabe ser falsa — dois pontos de um "espectro da desordem informacional", sem apresentar mensuração empírica própria. A distinção é de intenção, e a Tabela 1.1 mostra que quase nenhuma das medidas a observa. .

Os termos são usados como as fontes os trazem. “Informação falsa” e “desinformação” aparecem, sem pretensão de definição própria, para o que a fonte chama de misinformation, e a intenção só entra quando a fonte a medir. Quanto a “discurso de ódio”, a regra é a do restante do livro: é um rótulo externo, aplicado só ao enunciado classificado por instância externa, com a instância nomeada; os estudos da seção 1.7 usam a expressão como categoria da própria medida, e aqui ela é tomada como eles a trazem.

TABELA 1.1

Seis definições de informação falsa e o que cada uma exige medir

Fonte	Termo e definição	O que a definição exige medir	O que o estudo mediu
Lazer et al. (2018)	Notícia falsa: informação fabricada que imita a forma do jornalismo, não o processo nem a intenção	Processo e intenção do editor	Sem medida própria: ensaio que propõe medir por editor, e não por notícia"
Allcott e Gentzkow (2017)	Notícia falsa: artigo intencional e verificavelmente falso, que pode enganar	Intenção e veracidade, notícia a notícia	Compartilhamentos de 156 notícias classificadas como falsas por checadores nos três meses antes da eleição
Guess et al. (2019)	Conteúdo falso ou enganoso vestido como notícia, em geral por receita de anúncios	Origem do conteúdo	Links compartilhados no Facebook, cruzados com listas de editores de notícias falsas
Lasser et al. (2022)	Informação não confiável, definida pelo domínio	Confiabilidade da fonte, não do item	Nota do NewsGuard abaixo de 60 para o domínio do link
Rathje et al. (2023)	Misinformation: notícias fabricadas, boatos, teorias da conspiração ou desinformação	Veracidade da manchete	Julgamento de manchetes verdadeiras e falsas por participantes, com e sem incentivo
Ricard e Medeiros (2020)	Misinformation (o disseminador crê) e disinformation (o disseminador sabe), citando outra obra	O que o disseminador sabe	Sem medida própria: nota comentada, sem dado empírico

Nota. Descrições próprias, a partir dos trechos citados; a coluna “O que o estudo mediu” traz o que cada estudo operacionalizou, e não o que declarou querer medir.

1.2 Prevalência: quanto circula e quem compartilha

A prevalência pode ser investigada a partir de três perguntas centrais, e cada uma exige um instrumento próprio de medida: quanto circula, quem compartilha e de onde o conteúdo provém. A primeira pergunta diz respeito a quanto circula. Allcott e Gentzkow (2017) montaram, para a eleição presidencial americana de 2016, um banco com 156 notícias relacionadas à eleição que sites de checagem de referência classificaram como falsas nos três meses anteriores à votação. As 115 favoráveis a Trump foram compartilhadas 30 milhões de vezes no Facebook, e as 41 favoráveis a Clinton, 7,6 milhões de vezes. Compartilhamento, porém, não é o mesmo que ter visto, e os autores tentaram estimar o

que as pessoas viram: numa pesquisa online com 1.200 pessoas, feita depois da eleição, perguntaram se elas se lembravam de ter visto cada uma de 14 manchetes falsas que haviam circulado. Cerca de 15% disseram que sim, mas cerca de 14% também disseram ter visto manchetes-placebo, falsas, inventadas pelos autores e nunca circuladas. Como as manchetes-placebo não podiam ter sido vistas, os autores tomam a diferença entre as duas porcentagens como medida de recordação verdadeira e, projetando-a para o conjunto de notícias do banco, estimam que o adulto médio viu e lembrou 1,14 notícia falsa. Na mesma pesquisa, só 14% dos adultos americanos viam as mídias sociais como a fonte mais importante de notícias eleitorais.

A segunda pergunta é quem compartilha, e o estudo de Guess et al. (2019) é o que mais se aproxima de uma resposta com registro de comportamento, e não só com autorrelato. Os autores aplicaram uma pesquisa online a uma amostra representativa de 3.500 americanos e pediram a cada respondente que autorizasse o acesso, por um aplicativo, a dados do próprio Facebook, entre eles as postagens com links externos; a análise cobre quem consentiu e teve os dados ligados à pesquisa. Os links compartilhados em 2016 foram então comparados com listas de editores de notícias falsas feitas por jornalistas e acadêmicos, o que evita os vieses conhecidos do autorrelato de atividade online. O resultado principal é de raridade: cerca de 8,5% dos respondentes com dados ligados compartilharam ao menos uma matéria de domínio de notícias falsas, e mais de 90% não compartilharam nenhuma. Não é que as pessoas compartilhem pouco em geral: no período, 26,1% dos respondentes compartilharam de 10 a 100 links de qualquer tipo, e 61,3%, de 100 a 1.000. Entre os que compartilharam notícias falsas, havia mais republicanos: 38 contra 17 democratas em números absolutos, e 18,1% dos republicanos contra 3,5% dos democratas na amostra. O achado mais robusto, dizem os autores, é a idade: quem tinha 65 anos ou mais compartilhou em média 0,75 matérias de domínios de notícias falsas, mais que o dobro do segundo grupo etário mais velho (0,26), e a diferença persiste quando se mantêm constantes a ideologia e a educação.

Duas ressalvas dos próprios autores acompanham esses números. A primeira é que o padrão por ideologia é coerente com dois fatos: a maioria das notícias falsas tinha inclinação pró-Trump, e as pessoas tendem a compartilhar o que aprovam. O dado, portanto, pode não indicar que conservadores tenham maior tendência a compartilhar notícias falsas quando expostos a elas. A segunda é que os autores não têm medida de se os respondentes sabiam que o conteúdo era falso (Guess et al., 2019). Lazer et al. (2018) acrescentam um limite de método. Eles citam um estudo que avaliou a disseminação de um conjunto de notícias falsas de destaque e estimou que o americano médio encontrou entre uma e três delas no mês anterior à eleição; a estimativa é provavelmente con-

servadora, porque o estudo acompanhou apenas 156 notícias. Ressalte-se que quantificar as contas que receberam ou replicaram uma postagem não revela quantas pessoas a leram, tampouco se esse estímulo exerceu qualquer efeito sobre suas ações.

Em síntese, como formulam Lazer et al. (2018): exposição não equivale a impacto, e a avaliação, a médio e longo prazo, do efeito da exposição a notícias falsas sobre o comportamento político é essencialmente inexistente na literatura. Mosleh e Rand (2022) lembram uma distinção parecida do lado de quem compartilha: exposição e compartilhamento são construtos fundamentalmente diferentes, a maior parte das pessoas compartilha só uma pequena fração do que vê, e olhar o que alguém compartilha dá um retrato muito limitado de seu ambiente de informação.

1.3 Elites políticas como fonte

A terceira pergunta sobre prevalência — de onde a informação falsa provém — desloca o foco do público em geral para as elites políticas. Mosleh e Rand (2022) argumentam que, depois da atenção às notícias de editores obscuros, uma forma diferente de desinformação passou a chamar a atenção: campanhas coordenadas por figuras e organizações públicas, que chamam de elites. Para estudá-la, construíram dois escores. O primeiro é um escore de falsidade para cada uma de 816 elites, calculado a partir das checagens profissionais do PolitiFact sobre a veracidade das declarações delas. O segundo é um escore de exposição para cada usuário do Twitter, calculado a partir dos escores de falsidade das elites que ele segue. Depois compararam a exposição com o que o próprio usuário compartilha. A qualidade das notícias compartilhadas foi medida pela confiabilidade do domínio, avaliada de duas formas: pela média de oito checadores profissionais e por uma amostra de 970 americanos escolhidos por cotas, com as notas de simpatizantes dos dois partidos balanceadas. A ideologia do usuário foi estimada pela ideologia das contas que ele segue. Quanto maior a exposição à desinformação de elites, menor a qualidade das notícias que o próprio usuário compartilha, mais à direita se situa sua inclinação política estimada. É uma correlação entre medidas, e os próprios autores dizem que a abordagem se baseia na rede imediata de contas seguidas e não capta a exposição além delas, como por retuítes de conexões ou por recomendações algorítmicas.

Lasser et al. (2022) analisam diretamente as postagens emitidas por parlamentares. Os autores reuniram os tuítes, sem retuítes, de membros do Congresso americano, do parlamento alemão e do parlamento britânico entre 2016 e março de 2022, extraíram os links e deram a cada um a nota do NewsGuard

para o domínio, de 0 (muito ruim) a 100 (jornalismo de qualidade excepcional); abaixo de 60, o domínio é considerado "não confiável". Os republicanos compartilharam 9,1 vezes mais links para domínios não confiáveis que os democratas: 3,86% dos links, contra 0,43%. Na Alemanha, a CDU/CSU (democratas-cristãos) compartilhou 0,18% e o SPD (social-democratas) 0,03%; no Reino Unido, os conservadores 0,25% e os trabalhistas 0,05%, e os autores notam que esses números se baseiam em contagens muito baixas. A proporção de links não confiáveis dos republicanos passa de 2,4% em 2016–2018 para 5,5% em 2020–2022, enquanto a dos democratas segue em 0,4%; o resultado se repete com uma segunda base de confiabilidade de domínios, compilada independentemente do NewsGuard. A divergência entre os partidos parece peculiar aos Estados Unidos: não aparece na Alemanha nem no Reino Unido.

O que os autores propõem como explicação é uma possibilidade debatida: a motivação partidária, já que conteúdo depreciativo sobre o grupo político adversário aumenta as intenções de compartilhar entre partidários. Os autores sugerem também que o padrão entre os parlamentares pode contribuir para a assimetria que se observa no público, de duas maneiras: ao fornecer desinformação a partidários e ao legitimar o compartilhamento de informação não confiável. O NewsGuard só tem nota para parte dos links, e o estudo registra a cobertura: 46,5% dos links do Congresso americano, 58,8% do parlamento alemão e 39,2% do britânico, depois de excluídas as redes sociais e os buscadores. O que fica sem nota não é medido.

1.4 Posicionamento político e propensão à informação falsa: o que os estudos correlacionais mostram

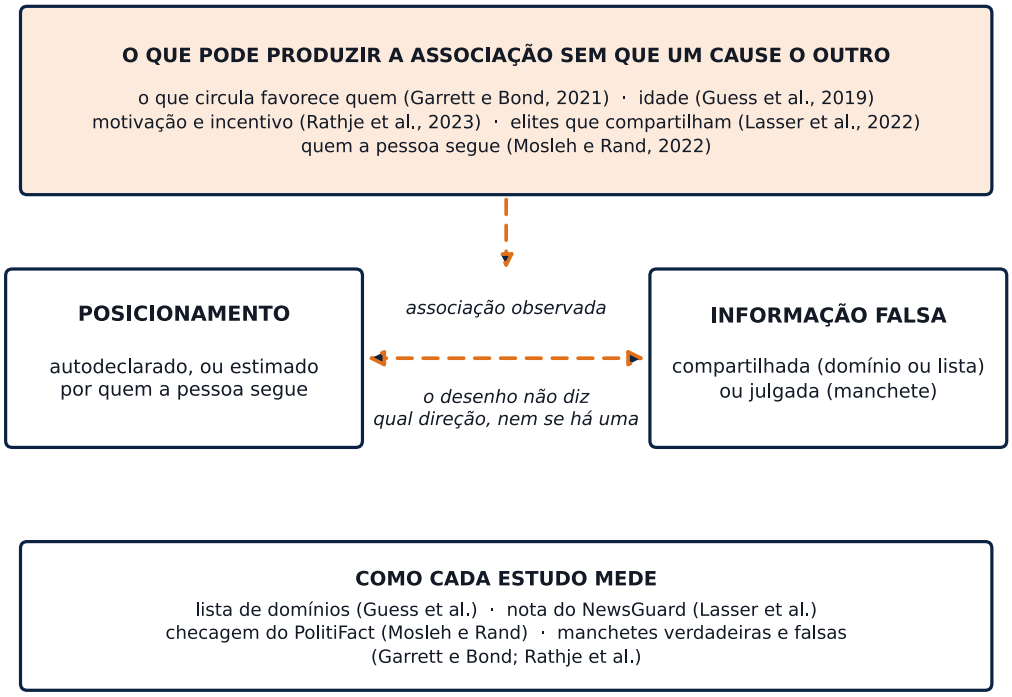
A pergunta recorrente na literatura é se há assimetria partidária: se quem se posiciona em determinado campo compartilha ou acredita mais em informação falsa. Os estudos verificados respondem com associações, e a Tabela 1.2 as reúne com o que cada desenho não permite concluir. A Figura 1.1 ilustra os limites dessa inferência.

Garrett e Bond (2021) adotaram um cuidado metodológico incomum na área. A evidência de que conservadores têm mais crenças falsas se apoiava em um conjunto relativamente estreito de temas, em muitos casos escolhidos de propósito para refletir falsidades promovidas por elites políticas conservadoras. Os autores, em vez disso, deixaram a escolha das afirmações a um critério externo: a cada duas semanas, um serviço identificou as 20 histórias políticas mais virais, 10 verdadeiras e 10 falsas, e um painel de 12 ondas, com grande amostra representativa de americanos, avaliou na semana seguinte afirmações curtas sobre essas histórias, ao longo de seis meses. Com as respostas,

calcularam dois índices, pela teoria de detecção de sinais: a sensibilidade, que é a capacidade de distinguir afirmações verdadeiras de falsas, e o viés de resposta, que é a propensão a rotular todas as afirmações como verdadeiras ou como falsas. Em dados de pesquisa nos Estados Unidos, os conservadores distinguiram verdades de falsidades pior que os liberais, e os autores atribuem parte da diferença ao ambiente informacional: as afirmações verdadeiras de alto perfil tendem a promover temas e candidatos favorecidos por liberais, e as narrativas falsas tendiam a favorecer causas conservadoras, de modo que, mesmo com igual viés, os liberais teriam desempenho melhor nesse ambiente.

FIGURA 1.1

O que uma associação entre posicionamento político e informação falsa mostra, e o que pode produzi-la sem que um cause o outro



Nota. Esquema original do autor, a partir de Garrett e Bond (2021), Guess et al. (2019), Rathje et al. (2023), Lasser et al. (2022) e Mosleh e Rand (2022). Não representa dados. Tracejada: relação que os desenhos correlacionais não separam.

Rathje et al. (2023) acrescentam experimentos em que se pôde variar a consequência de acertar. Em quatro experimentos, com 3.364 participantes dos Estados Unidos, as pessoas liam manchetes políticas verdadeiras e falsas e respondiam se a afirmação da manchete era precisa; metade, sorteada, foi informada de que receberia um bônus de até um dólar conforme as respostas corretas, e a outra metade respondeu às mesmas perguntas sem incentivo. Os au-

tores chamam de viés partidário a diferença entre o número de manchetes favoráveis ao próprio partido e o de manchetes contrárias ao próprio partido que o participante julgou verdadeiras. Os incentivos financeiros por acurácia reduziram o viés partidário no julgamento de manchetes em cerca de 30%. Os conservadores foram menos acurados que os liberais em distinguir manchetes verdadeiras de falsas, mas os incentivos reduziram a diferença em 52%; e incentivar os participantes a identificar notícias que seriam apreciadas por seus aliados políticos diminuiu a acurácia. Parte do que aparece como diferença de crença, portanto, parece depender do que se reforça: acertar ou agradar.

TABELA 1.2

Cinco estudos sobre posicionamento político e informação falsa: o que mediram e o que não permitem concluir

Estudo	Amostra e país	Medida de informação falsa	Medida de posicionamento	Resultado	O que não permite
Guess et al. (2019)	Respondentes de pesquisa representativa ligados a dados do Facebook; EUA, 2016	Links de domínios de notícias falsas, por lista	Autodeclarado (partido, ideologia)	Raro (8,5%); mais entre republicanos e entre os mais velhos; a idade é o achado robusto	Atribuir a diferença a maior propensão: pode refletir a oferta pró-Trump
Garrett e Bond (2021)	Pesquisa nos EUA, afirmações ao longo de 6 meses	Julgamento de afirmações verdadeiras e falsas	Autodeclarado	Conservadores discriminaram pior; parte atribuída ao ambiente informacional	Concluir que há viés maior: o resultado depende do conjunto de afirmações
Mosleh e Rand (2022)	816 elites e usuários do Twitter, EUA	Escore de falsidade por checagens do PolitiFact	Estimado por quem o usuário segue	Exposição a elites falsas correlaciona com pior qualidade de notícias e com ideologia conservadora	Causalidade; um verificador só; exposição além de quem se segue
Lasser et al. (2022)	Parlamentares dos EUA, da Alemanha e do Reino Unido, 2016–2022	Nota do NewsGuard do domínio do link	Partido	Republicanos compartilham mais, e mais com o tempo; Alemanha e Reino Unido, contagens baixas	Inferir a veracidade da fala a partir do domínio da fonte; cobertura restrita de 39% a 59% dos linkss

Estudo	Amostra e país	Medida de informação falsa	Medida de posicionamento	Resultado	O que não permite
Rathje et al. (2023)	Quatro experimentos, n = 3.364, EUA	Julgamento de manchetes verdadeiras e falsas	Autodeclarado	Incentivo reduziu o viés partidário em cerca de 30% e a diferença em 52%	Generalizar de manchetes e participantes online

Nota. Dados compilados a partir dos artigos originais citados. As restrições da última coluna explicitam o que os delineamentos metodológicos não conseguem isolar.

1.5 Limitações: o que os desenhos não permitem

Os estudos reunidos nas seções anteriores têm três limites em comum, e é por eles que os números do capítulo devem ser lidos. Dos cinco estudos da Tabela 1.2, quatro são sobre os Estados Unidos e o quinto compara com Alemanha e Reino Unido; os dados de três deles vêm de uma plataforma, o Facebook ou o Twitter. O primeiro limite é, portanto, de amostra e de contexto: os achados são do sistema bipartidário americano, em eleições específicas, e Lasser et al. (2022) dizem que a divergência que observam parece peculiar aos Estados Unidos. O segundo é de causa: nenhum desses desenhos separa o que a pessoa vê, o que lhe é oferecido e o que ela escolhe, e a Figura 1.1 reúne as variáveis que podem produzir a associação sem que posicionamento e informação falsa se causem. O terceiro é de medida, e merece um quadro, porque é o que se toma emprestado da análise do comportamento para ler todos os outros.

CONCEITO EM USO

Comportamento verbal, audiência e autorrelato

Comportamento verbal é o comportamento reforçado pela mediação de outras pessoas (Skinner, 1957); a audiência é o estímulo discriminativo em cuja presença esse comportamento é caracteristicamente reforçado. Autorrelato, isto é, responder a uma escala, a uma pergunta de pesquisa ou a uma entrevista, é comportamento verbal, e está sob controle de quem pergunta e de como a pergunta é feita. Para Guerin (1994), escalas e questões de pesquisa dependem de quem pergunta e da cultura verbal em que a pesquisa é feita, porque atitudes dependem sempre de consequências sociais. O exemplo clássico é o de Hildum e Brown (1956): num levantamento telefônico com 40 estudantes homens de Harvard, o entrevistador dizia “Good” depois de respostas a favor, ou contra, o programa de Educação Geral, e o “Good” deslocou as respostas na direção reforçada; só 8 de 20 sujeitos notaram o que o entrevistador fazia, e todos rejeitaram que suas respostas tivessem sido influenciadas por ele.

A consequência para este capítulo é direta. A Tabela 1.2 traz três linhas em que o dado é uma resposta a uma pesquisa ou a uma tarefa de julgamento (Garrett e Bond, Rathje et al. e, em parte, Guess et al., em que a ideologia é autodeclarada), e uma resposta a uma pesquisa é comportamento sob controle da situação em que foi dada. O resultado de Rathje et al. ilustra precisamente essa dinâmica: quando a consequência de acertar muda, o julgamento verbal se altera. Um segundo problema de medida é o da seleção dos estímulos. Cameoka e Moreira (2021) relatam três estudos com universitários que avaliaram fotos de homens brancos e de homens negros. No Estudo 1, 75 participantes avaliaram, numa escala de 1 (mau) a 10 (bom), 10 fotos de homens brancos e 10 de homens negros com expressões neutras; o Estudo 2 teve 40 participantes, com faces discretamente sorridentes, e o Estudo 3, 30. A diferença entre as médias dos dois grupos de fotos foi significativa nos Estudos 1 e 2 e não no 3, em que se usaram as fotos de brancos com as avaliações mais altas e as de negros com as mais baixas nos dois primeiros estudos. Os autores tomam o Estudo 3 como demonstração do potencial de distorção de uma seleção de estímulos pouco rigorosa e sugerem que a seleção dos estímulos pode enviesar a mensuração. O que se mede depende do que se escolheu para ser medido.

Há ainda o limite do instrumento de checagem, que o Capítulo 2 trata por inteiro: a proporção de falsas entre as declarações checadas não é a proporção entre todas as declarações, e uma agência não é um sensor neutro. Mosleh e Rand (2022) usam um único verificador, o PolitiFact, e elites dos Estados Unidos. O leitor deve guardar, para os demais capítulos, o que esta seção estabelece: toda taxa de falsidade é função de quem foi medido, com que instrumento, em que situação, e nenhuma delas é, por si, a taxa de um campo político.

1.6 O que se sabe sobre o Brasil

O que as fontes verificadas dizem do Brasil é menos do que a seção 1.2 diz dos Estados Unidos, e de natureza diferente. Ricard e Medeiros (2020), num comentário, e não num estudo empírico, escrevem que uma retórica de negação dos riscos da COVID-19 foi endossada nos mais altos níveis do governo brasileiro e apoiada por desinformação; e registram que, segundo a Comissão Parlamentar de Inquérito (CPI da Pandemia), parte da desinformação sobre a pandemia atacava decisores e figuras públicas, como governadores, a mídia e especialistas em saúde, para deslegitimar quem apoiava o isolamento, e que a Comissão ainda não havia divulgado evidência de coordenação direta pelo chamado Gabinete do Ódio ou por membros do governo.

Carvalho et al. (2021) fizeram uma análise de conteúdo, com o programa IRaMuTeQ, de postagens nos perfis de Trump e de Bolsonaro no Facebook e de pronunciamentos oficiais nos sites dos dois governos, desde o primeiro dia de notificação da Covid-19 em cada país. Os autores apontam similaridades temáticas e posicionamentos contrários às medidas defendidas pela OMS. Escrevem que não é possível afirmar categoricamente que houve alinhamento intencional entre os dois, mas que a análise mostra convergência nas estratégias de desinformação, e que a minimização dos riscos de contágio, sobretudo nas postagens do Facebook, é uma das marcas da similaridade. É uma análise de dois emissores, e não um dado de prevalência.

Recuero et al. (2025) analisaram 811 checagens de três agências brasileiras sobre as eleições de 2018 e 2022, por análise de conteúdo e métodos mistos, comparando temáticas, tipos de desinformação e as fontes que legitimavam o conteúdo. Apontam uma transição de conteúdos totalmente falsos para enganosos, e de conteúdos sem fontes para conteúdos que usam fontes reais; leem a mudança das plataformas dos conteúdos checados, do Facebook para o WhatsApp e o Twitter entre 2018 e 2022, como adaptação da desinformação a espaços mais fechados, em que é mais difícil circularem contradiscursos. Registram, igualmente, que a literatura revisada aponta maior volume de desinformação gerada por apoiadores de Bolsonaro em ambos os pleitos, ainda que ambas as campanhas tenham figurado concomitantemente como polos emissores e alvos de desinformação. A Lupa (Agência Lupa, 2026) descreve o mesmo deslocamento em sua própria base: em 2018, o Facebook aparecia em 76,5% das verificações; em 2022, o WhatsApp estava presente em 92,4%.

Dois limites separam essas fontes do que a seção 1.2 mostra para os Estados Unidos. Nenhuma delas é um estudo de compartilhamento com registro de comportamento, como o de Guess et al. (2019), nem uma comparação de partidos pelo que seus parlamentares compartilham, como a de Lasser et al.: o que há são checagens, o que é curado por uma agência, e uma análise de conteúdo de dois emissores. E a regra do Capítulo 2, de que a proporção de falsas entre as declarações checadas não é a proporção entre todas as declarações, vale aqui integralmente. Que a desinformação brasileira tenha o formato da americana, ou o contrário, é hipótese, e este capítulo não a assume.

1.7 Discurso de ódio: definições, medidas e dados

Entre as fontes verificadas, a expressão 'discurso de ódio' têm uma definição explícita em Hangartner et al. (2021): linguagem hostil contra um grupo ou pessoa por característica inata, real ou percebida; os autores selecionaram tu-

ítes xenófobos e racistas por combinação de dicionário, análise de sentimento e anotação manual. Os demais estudos operacionalizam de outro modo, e a Tabela 1.3 reúne o que cada um mediu.

TABELA 1.3

Como cinco estudos verificados medem discurso de ódio, e o que cada medida não permite

Estudo	Medida de "discurso de ódio"	Amostra	Resultado	O que a medida não permite
Siegel et al. (2021)	Dicionário aumentado por aprendizado de máquina e classificação por comunidades do Reddit; o dicionário exige insulto ou termo depreciativo	Mais de 750 milhões de tweets sobre a eleição americana de 2016 e quase 400 milhões de tweets de uma amostra aleatória de usuários americanos	Sem aumento persistente na campanha nem depois da eleição	Falar de outras plataformas; a definição é conservadora
Marino-Fages e Martínez-Pozo (2026, documento de trabalho)	Classificador BERTimbau e dicionário, por alvo (homofobia, racismo, sexismo, insultos, ódio político)	Mais de 37,6 milhões de tweets geolocalizados no Brasil, 2017-2019	Aumento depois da eleição de 2018, maior onde Bolsonaro teve menos votos, em homofobia e sexismo	Medir o custo social percebido; a acurácia do classificador é de 77% a 79%
Álvarez-Benjumea e Winter (2018)	Nota de hostilidade (7 ou mais) dada a comentários de fórum experimental	180 participantes de plataforma de crowdsourcing	Comentários hostis: 5,88% na base, 0,99% com remoção moderada; contra-fala sem efeito significativo	Generalizar: participantes pagos, fórum anônimo
Hangartner et al. (2021)	Tuítes xenófobos e racistas, selecionados por dicionário, sentimento e anotação	1.350 usuários do Twitter	Contra-fala baseada em empatia aumentou a exclusão e reduziu a criação de discurso xenófobo	Efeito sobre quem apenas lê; humor e aviso de consequências sem efeito consistente
Fazzano e Gallo (2015)	Homofobia como classe de respostas (bater, insultar, torturar, espancar, excluir, discriminar), definida pela função	Dados do Disque 100 de 2011: 6.809 relatos de violações contra LGBT	Registros de denúncias oficiais; não mensuram a prevalência direta da agressão	Prevalência: são relatos, e não o fenômeno

Nota. Números dos artigos. Marino-Fages e Martínez-Pozo é documento de trabalho que, segundo a capa, não passou por revisão de pares. Os estudos são tratados no Capítulo 5 (seção 5.5), no Capítulo 10 e no Capítulo 12.

Quatro pontos organizam a tabela. O primeiro é que a definição decide o número: Siegel et al. dizem que sua definição de dicionário é conservadora, pois exige que o tuíte contenha insulto ou termo depreciativo, e Marino-Fages e Martínez-Pozo classificam por alvo e obtêm um aumento concentrado em homofobia e sexismo, sem padrão comparável para insultos gerais, ódio político ou alvos políticos. O segundo é a plataforma e o recorte da amostra: o estudo americano se limita ao Twitter e adverte que observar certo discurso num momento não significa que ele tenha aumentado, dado o caráter intermitente da plataforma, em que tópicos ganham destaque por pouco tempo e voltam ao equilíbrio; o brasileiro usa só tweets publicados às segundas-feiras e geolocalizados no Brasil, e assume a premissa de que esses dois recortes representam o universo da plataforma. O terceiro é que "discurso" e "ato" são medidas diferentes, e nenhum dos cinco estudos mede ato contra o grupo (a lacuna é discutida na seção 5.5). O quarto é que os resultados não se contradizem nem se confirmam: medem coisas diferentes em populações diferentes.

Do lado da análise do comportamento, a literatura é escassa. Mizael (2018) encontrou apenas quatro artigos analítico-comportamentais sobre homossexualidade em 13 periódicos e aponta como lacuna as análises funcionais de comportamentos preconceituosos, como piadas mantidas pela comunidade verbal; e Fazzano e Gallo (2015) definem a homofobia como classe de respostas de topografias diversas, definida pela função, e não pela forma. É a definição de que o Capítulo 5 parte.

1.8 Interpretações rivais e limites

A leitura proposta neste capítulo é que os dados empíricos revelam magnitudes inferiores ao que o debate público supõe, e que cada cifra é função estrita do instrumento utilizado. Essa perspectiva confronta três interpretações rivais. A primeira sustenta que a assimetria partidária seria um fato consolidado e que o capítulo adotaria um ceticismo desmedido; o teste decisivo entre essas leituras exigiria um delineamento que mantivesse sob controle rigoroso a oferta total de conteúdo e a exposição individual efetiva — ressalva que os próprios autores de Garrett e Bond (2021) e Guess et al. (2019) formulam. A segunda tese rival argumenta que, sem checagem profissional de fatos, qualquer medição é inviável, tornando as agências o único termômetro disponível; a resposta do Capítulo 2 é que a checagem mede o que foi checado, e não o que circula. A terceira é que a definição pela intenção torna a medida impossível e deveria ser abandonada; a de domínio, que os estudos usam, é uma resposta pragmática, e aqui é tomada como tal.

Os limites do capítulo são os das fontes. A maioria das medidas é de Estados Unidos e de uma plataforma, no período de uma ou duas eleições; a única base brasileira de grande escala, entre as de discurso, é um documento de trabalho sem revisão de pares; e os números de prevalência de informação falsa são, no melhor dos casos, de exposição e de compartilhamento, e não de efeito. Não se compara o Brasil com os Estados Unidos, e o compartilhamento de uma fonte não é tratado como crença nem como ato.

Síntese

"Cada definição de informação falsa impõe exigências metodológicas próprias (intenção do emissor, origem do conteúdo, confiabilidade do domínio ou veracidade factual), e as pesquisas tendem a substituir a análise da declaração pela classificação da fonte em virtude da inacessibilidade da intenção privada (1.1). A circulação empírica de notícias falsas por compartilhamento individual revelou-se rara nos dados de Guess et al., com maior incidência entre idosos, e exposição passiva não equivale a impacto causal (1.2). A assimetria partidária registrada entre parlamentares nos Estados Unidos pelo critério de compartilhamento de domínios parece constituir fenômeno peculiar àquele contexto institucional (1.3). As relações entre posicionamento ideológico e desinformação têm caráter correlacional, dependem do conjunto de itens selecionados e do ambiente de oferta, sendo que parte substancial da disparidade de acurácia se altera quando as contingências de reforço são manipuladas (1.4). autorrelato em pesquisas consiste em comportamento verbal sob controle das contingências da coleta, e a amostragem arbitrária de estímulos pode enviesar o resultado das medidas (1.5). No Brasil, o quadro empírico limita-se a análises de conteúdo, relatórios de conjuntura e bases de agências checadoras, inexistindo registros comportamentais diretos de compartilhamento em rede aberta (1.6). Por fim, as mensurações de discurso de ódio variam radicalmente conforme os critérios lexicais e a plataforma de coleta adotados, não capturando a emissão de atos concretos contra os grupos-alvo (1.7).

Os capítulos subsequentes mobilizam diretamente este arcabouço: os Capítulos 2 e 3 operam sob o princípio de que toda taxa de falsidade é função do instrumento de medida; o Capítulo 5 adota as definições funcionais da seção 1.7; o Capítulo 6 fundamenta-se na evidência de que o compartilhamento representa uma fração ínfima da exposição; e o Capítulo 12 examina a discrepância entre o que o indivíduo relata verbalmente e o que ele efetivamente faz diante de contingências reais.

Referências

Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. 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), Article 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), Article 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), Article 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), Article 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), Article 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), Article 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

Os conceitos deste livro, um verbete de cada vez

Reforçamento, controle de estímulos, equivalência, prática baseada em evidências: verbetes técnicos escritos com clareza, com as fontes ao pé de cada um. Para estudar, revisar e citar.

80+ verbetes abertos

Acesse, de graça e sem cadastro

walden4.com.br/enciclopedia