



How Algorithms Create and Prevent Fake News

Exploring the Impacts of
Social Media, Deepfakes, GPT-3,
and More

—

Noah Giansiracusa

Apress®

HOW ALGORITHMS CREATE AND PREVENT FAKE NEWS

EXPLORING THE IMPACTS OF SOCIAL
MEDIA, DEEPPAKES, GPT-3, AND MORE

Noah Giansiracusa

Apress®

How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More

Noah Giansiracusa
Acton, MA, USA

ISBN-13 (pbk): 978-1-4842-7154-4

ISBN-13 (electronic): 978-1-4842-7155-1

<https://doi.org/10.1007/978-1-4842-7155-1>

Copyright © 2021 by Noah Giansiracusa

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Shiva Ramachandran
Development Editor: James Markham
Coordinating Editors: Nancy Chen and Mark Powers

Cover designed by eStudioCalamar

Distributed to the book trade worldwide by Springer Science+Business Media New York, 1 New York Plaza, New York, NY 100043. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at <http://www.apress.com/bulk-sales>.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/9781484271544. For more detailed information, please visit <http://www.apress.com/source-code>.

Printed on acid-free paper

*Dedicated to my wife Emily and our parents:
Bob, Dorothy, Andy, and Carole.*

Contents

About the Author	vii
Acknowledgments	ix
Introduction	xi
Chapter 1: Perils of Pageview	1
Chapter 2: Crafted by Computer	17
Chapter 3: Deepfake Deception	41
Chapter 4: Autoplay the Autocrats	67
Chapter 5: Prevarication and the Polygraph	99
Chapter 6: Gravitating to Google	119
Chapter 7: Avarice of Advertising	151
Chapter 8: Social Spread	175
Chapter 9: Tools for Truth	217
Index	231

About the Author



Noah Giansiracusa received a PhD in mathematics from Brown University and is an Assistant Professor of Mathematics and Data Science at Bentley University, a business school near Boston. He previously taught at UC Berkeley, University of Georgia, and Swarthmore College. He has received national grants and spoken at international conferences for his research in mathematics, and he has been quoted several times in *Forbes* as an expert on artificial intelligence. He has dozens of publications in math and data science and has taught courses ranging from a first-year seminar on quantitative literacy to graduate machine learning. Most recently, he created an interdisciplinary seminar on truth and lies in data and algorithms that was part of the impetus for this book.

Acknowledgments

Thanks to Jordan Ellenberg, Cathy O’Neil, and Francis Su for inspiring me to write a book and helping me learn about the publishing industry. Thanks to Karen Hao, Will Douglas Heaven, Davey Alba, Jack Nicas, Kevin Roose, James Vincent, Issie Lapowsky, Carole Cadwalladr, Julia Angwin, Deepa Seetharaman, Jeff Horwitz, Sheera Frenkel, and so many other technology journalists for doing the hard work that this book relies so heavily upon. Thanks to Gerald Seidler, Jim Morrow, Henry Cohn, Dan Abramovich, Angela Gibney, Bernd Sturmfels, and my other teachers and mentors who helped me become a math professor. Thanks to Charlie Hadlock, Rick Oches, Lucy Kimball, and others at Bentley University for believing in me as a data scientist and providing me with the opportunities that led to this book. Thanks to Steffen Marcus for encouraging me to think and write about the broader context of math and technology. Thanks to Tom Taulli for putting me in contact with Apress, and to my team at Apress—Shiva Ramachandran, Matthew Moodie, Nancy Chen, Rita Fernando, and Mark Powers—for all their work shaping this book and helping get it across the finish line. Thanks to my parents, Bob and Dorothy, for homeschooling me and passing along a passion for lifelong learning. Thanks to my brother, Jeffrey, for paving the way forward for us both. And thanks to my wife, Emily, and our daughter, Claire, for more than words can convey.

Introduction

You might have heard rumors that the newsfeed algorithm at Facebook and the video recommendation algorithm at YouTube are spreading fake news, or that artificial intelligence (AI) can now rapidly generate convincing articles and make videos of people doing and saying things they never did, or that machine learning algorithms will save us from fake news by automatically detecting it and labeling assertions as true or false. But what do these claims even mean, and what should you believe? The main goal of this book is to help readers of all backgrounds—no knowledge of math, statistics, computers, algorithms, or journalism required—understand what’s really going on by collecting all the investigations, research, and stories about fake news and algorithms in one place and explaining it in a simple way while weaving it together into a coherent narrative. Another goal is to teach you about the publicly available tools that can help you do your own part in the fight against fake news.

“If we are not serious about facts and what’s true and what’s not, if we can’t discriminate between serious arguments and propaganda, then we have problems.” Barack Obama said this on November 17, 2016, just nine days after Donald Trump was elected to be his successor in the White House. Since then, there has been an increasing awareness of the scope and impact of “fake news,” a catchall label for *misinformation* (false information that is spread regardless of intent to mislead) and *disinformation* (deliberately false or misleading information). There has also been an increasing awareness of the role played by data-driven algorithms in the creation, dissemination, and detection/moderation of fake news. But the story of fake news and algorithms has been difficult for most of us to follow. It has unfolded in a wide range of academic publications, journalistic investigations, corporate announcements, and governmental hearings, and it involves many sophisticated technological concepts that sound mysterious. I strongly believe that the barriers to entering this important discussion are not nearly as high as they might seem, and this book is my attempt to lower them even further.

Chapter 1 sets the stage by exploring the economics of blogging and online newspapers, with an emphasis on the dynamics that have led to a proliferation of low-quality journalism. Data, in the form of clicks and pageviews, has transformed the news industry, and you’ll see how fake news peddlers have taken advantage of this. Chapter 2 looks at a new development in our ongoing battle to understand what’s real and what’s not: fake journalists with untraceable lifelike profile photos synthesized by AI, and entire articles written by AI

with the click of a button. You'll learn about the technology behind these advances (GPT-3 and deepfake GANs, with a gentle overview of machine learning along the way) and the impact they're having on journalism. Chapter 3 continues this line of investigation by turning to deepfake video editing—explaining how it works, what it can do, and the role it has played in politics. Chapter 4 is all about YouTube and its recommendation algorithm that automatically selects videos for you to watch. A history of this algorithm is provided, including brief discursions into deep learning and reinforcement learning, and empirical investigations into the way it works in practice are explored. This frames a discussion of fake news and conspiratorial content on YouTube, especially in the context of Brazil's 2018 election and the 2016 and 2020 US elections.

After several chapters on how AI can create and spread fake news, Chapter 5 asks if AI can help fight it by determining whether someone in a video is lying. This is part of an algorithmic reinvention of the polygraph that is currently being trialed at airports and elsewhere. Chapter 6 takes a deep look at one of the world's most popular sources of information: Google. The company's efforts to elevate quality content over fake news and harmful material are detailed, as are the various failures that have occurred along the way and the challenges that remain. Chapter 7 shows how Google supports the fake news industry financially through ad revenue and how Facebook's algorithmically distributed ads have been a persistent source of fake news and racism. Chapter 8 takes a thorough look at how fake news spreads across social media and how algorithms have been used to detect and mitigate this spread. Finally, Chapter 9 collects and explains some publicly available AI-powered fact-checking tools that you can use to make sure what you're reading is trustworthy and truthful.

Perils of Pageview

The Data-Driven Economics of Online Journalism

The economics of the Internet created a twisted set of incentives that make traffic more important—and more profitable—than the truth.

—Ryan Holiday, *Trust Me, I'm Lying: Confessions of a Media Manipulator*

Much of what we know, or think we know, about what is happening in the world we learn by reading the news. But nowadays “the news” means something different than it did in generations past. What we read primarily today are articles on the internet—everything ranging from casual blog posts to meticulously researched stories on national and international news sites. The transition of journalism from print to screen does not inherently mean what we read is less truthful than it used to be. However, this technological transformation has enabled a less overt but nonetheless extraordinarily influential economic transformation: the *datafication* of the journalism industry. The pageviews and clicks we all sprinkle across the internet are, as I will discuss, the digital fertilizer feeding a burgeoning garden of misinformation

and fake news. By tracing the financial incentives involved in the contemporary news cycle, I hope in this chapter to convey the alarming extent that data, unseen to most of us yet created by our actions and activities, is fundamentally shaping what we read every day and threatening the bulwark of traditional journalistic standards.

Propagation of Stories

Let me start with a taxonomy of sorts. At the bottom of the internet media food chain, if you will, are small blogs and websites that cover very focused issues, interests, or regions; these can be single author or multi-author. The next tier up comprises the blogs of newspapers, magazines, and television stations. This is a confusing middle ground because many of these blogs share the name, URL, and logo of a recognizable news source yet the editorial standards are generally lower than those of the parent organization, and many of the contributors lack the journalistic training one might expect from the parent organization. Then at the top are the official news sites, which can be regional but tend to draw a large national or international readership. This hierarchy is not about quality—indeed, some very focused small blogs produce content of extremely high quality, while some big-name national news sites consistently publish articles of seriously questionable accuracy. The levels here are more about the size of both the audience and the organization and about the scope of the content.

Information flows both vertically and horizontally through this internet news hierarchy. When the *Washington Post* breaks a big story, it is only a matter of hours before the *New York Times* covers it as well, and vice versa, often simply by reporting what was reported in the other newspaper's article. This is *horizontal propagation*, and it happens because even though the second newspaper cannot claim credit for breaking the story, it does not want its readership to obtain this information directly from the competitor newspaper. *Vertical propagation* happens in two directions. A big story broken at the top will be covered and duplicated by smaller news organizations and blogs because, similar to horizontal propagation, this is an easy way of keeping readers without doing much work; this is a *downward* flow of information.

While there is an obvious redundancy, hence an overall systemic inefficiency, to both horizontal propagation and downward vertical propagation, the only real harm to the truth-seeking reader is that important details might be omitted and facts distorted as the story is passed from organization to organization—though sometimes a more specialized blog will provide a valuable service by delving deeper into a particular facet of the story than would be appropriate for the top-level organization. It can be quite illuminating to find a story that was broken by one newspaper and then compare its coverage across a range of other newspapers and blogs; this is an excellent

way to uncover the ideological inclinations of different organizations, since the same set of facts will be colored by the different viewpoints involved.

The remaining form of journalistic propagation is the *upward* vertical flow, where stories start at small blogs and sometimes end all the way up at national news sites. This is one of the key topics of this chapter because it is responsible for a staggering amount of the misreporting and outright fake news that we see, and it is driven almost entirely by data and the economics of modern media. Before exploring this specific topic, it helps to take a step back and look at the financial forces driving blogs and newspapers; throughout, I take a broad view of “blogging” to include essentially all forms of posting written content online.

Economics of Blogging

Ostensibly, the revenue for blogs comes from selling advertisements. There are a variety of pecuniary mechanisms for online advertisements, such as the advertising company affixing a banner atop the blog and paying based on *pageviews* (the number of users who visit the blog where the banner is displayed), and in some cases the advertiser pays an additional sum when a reader on the blog clicks the ad link and proceeds to actually purchase a product from the advertising company. But the most common format is *pay-per-impression* and *pay-per-click* advertising, in which the blog places an ad somewhere on its website and is paid based on *impressions* (the number of times the ad is seen by a reader on the blog) or *clicks* (the number of times the ad is clicked by a reader on the blog). The bottom line is that to maximize ad revenue, the blog needs to maximize traffic.

But why did I write “ostensibly” in the preceding paragraph? Well, there is somewhat of a Ponzi scheme dynamic at play here. Advertising revenue tends to be relatively low even for popular blogs, so the real ambition of most blogs, even if they don’t admit it, is to gain sufficient popularity and traffic that a larger organization will buy them out and incorporate the blog into its larger website in order to increase traffic—often so that the larger website can boost its odds of being bought by a yet larger organization.

For example, Nate Silver’s technical yet surprisingly popular blog on political polls was launched in 2008, brought into the *New York Times* in 2010, acquired by ESPN in 2013, then transferred to the sister property ABC News in 2018. Arianna Huffington’s groundbreaking general news blog the *Huffington Post* was founded in 2005 with a one million dollar investment and sold to AOL in 2011 for three hundred and fifty million—but, quite tellingly, at the time of this sale, its ad revenue was only thirty-one million dollars per year. This tenfold purchase price to annual revenue ratio is rather extreme and suggests that AOL was banking on continued long-term growth as well as other factors like the prestige of adding such a popular online newspaper and bringing

onboard the superstar Arianna herself. At the end of the day, whether a blog aims to be bought out or not, the path to success is web traffic.

Next, let me turn from the economics of blogs to that of the bloggers themselves. In the early days of blogging, bloggers tended to be paid either a flat rate with a required minimum number of daily posts, or they were paid per post; in the mid-2000s, depending on the establishment, this rate was often a dismal five or ten dollars per post. A paradigm shift occurred when *Gawker* left this per-post payment system and instead paid each blogger a monthly salary that was augmented by a bonus based on the number of pageviews recorded by the blogger's articles. This shift made sense from an ad revenue perspective, and it quickly rippled across the blogosphere and ushered in a new era in which pageviews became the fundamental currency of blogging.

Gawker took things even further when it installed a large board in its office showing a live tally of the pageview statistics for all its bloggers and their posts (other blogs soon turned to similar methods as well). This led to an intense pageview competition among the bloggers at the company, designed to stimulate productivity, and it signaled a strong emphasis on analytics in which bloggers could not help but keep score of which articles generated the most pageviews.

This blogger remuneration system is blatantly reductionist: the reader's opinion of a blog post is irrelevant. In fact, it does not even matter whether the reader actually reads the post—once the link to a post is clicked, the pageview is recorded, and that's all that counts. An unfortunate but largely predictable consequence has been the proliferation of clickbait: catchy, often trashy, headlines that encourage clicks rather than bespeaking quality journalism.¹ A lengthy, methodically researched and fact-checked article provides no more financial value than a piece of vapid tabloid trash. This oversimplifies the situation as many readers follow certain blogs precisely because they consistently post high-quality articles, but many readers also click whatever stories are catchiest when scrolling through social media or news aggregators, and in these latter settings the name and reputation of the blog/organization is often a secondary factor in the decision to click—it is the headline that matters most.

An additional, and significant, dynamic is that blog posts tend to have short-lived pageview-generating lifespans. Consequently, bloggers and blogs, in their constant journey for increased traffic, are under intense pressure to produce as many posts as possible, as rapidly as possible. A traditional print newspaper had to produce content that filled one print edition per day; a cable news network has to produce content that fills twenty-four hours a day, three

¹And insidious techniques for gaming the system have inevitably, and unsurprisingly, flourished, such as posting slide shows in which the reader needs to click each slide one at a time, thereby artificially inflating the pageview metric.

hundred and sixty-five days a year; a blog has limitless space and is rewarded for attempting to fill this infinitude. This encourages rushed, sloppy writing and journalistic shortcuts; bloggers simply don't have time to fact-check. In fact, posts that generate controversy tend to also generate pageviews. Even worse, outright fallacies in news articles often entice disgruntled readers to leave comments complaining and/or correcting the article, but commenting on blogs usually involves multiple clicks and data trails that are dollars (well, pennies) in the pockets of the blogger.

Putting these observations all together, we see the perfect storm of conditions assaulting the foundations of journalism. Blogs and bloggers are almost all financially strapped, earning far less revenue than an outsider might expect, and so are in desperate need of more pageviews—whether to earn ad revenue directly or to raise the prospect of a lucrative buyout. This drives them to produce articles far too quickly, leaving precious little time to fact-check and verify sources. Even if they had time to fact-check, the pageview statistics they obsess over show that there is no real financial incentive for being truthful, as misleading articles with salacious headlines often encourage more clicks than do works of authentic journalism.

And let me be abundantly clear about this: it is the data-driven impetus of the blogging industry, and the vast oversimplification and distortion of multidimensional journalistic value caused by reducing everything to a single, simple-minded, superficial metric—the pageview—that is most responsible for this dangerous state of affairs. That some pageview-driven blogs thrive on thoughtful, methodical, accurate writing is truly remarkable in this market that is saturated with perverse incentives pressuring writers to engage in the exact opposite of these noble qualities. Let us all be thankful for the good blogs and good writing when we see it, for it is certainly out there but it struggles to rise above the ubiquitous clickbait filth pervading the internet.

Having presented the data-driven financial structure of blogs and bloggers, and the pernicious pressures it leads to, it is time now to turn back to the earlier discussion of the taxonomy of the blogosphere and the propagation of stories through it.

Up from the Bottom

Renée DiResta, a researcher at the Stanford Internet Observatory, recently wrote² in the *Atlantic* that “Media and social media are no longer distinct; consequential narratives emerge from the bottom up, as well as the top down, and bounce back and forth among different channels.” Recall that the

²Renée DiResta, “The Right’s Disinformation Machine Is Getting Ready for Trump to Lose,” *Atlantic*, October 20, 2020: <https://www.theatlantic.com/ideas/archive/2020/10/the-rights-disinformation-machine-is-hedging-its-bets/616761/>.

propagation direction I haven't yet directly addressed, despite claiming it is the one most responsible for our current morass of media mendacity, is the upward flow where stories start in small, typically special interest and/or geographically local blogs, and manage to work their way up the food chain, sometimes ending all the way at the top on national news sites. The questions we must ask here are: how and why does this happen, and why does this lead to less truthful news? The answers, as I next discuss, all essentially follow from the pageview economics of blogging.

All blogs and sources of news, even the highly regarded ones at the top, are in constant search for new stories. There is a fundamental inequality at play that the supply of actual stories (meaning real events transpiring in the world that ought to be reported) is substantially smaller than the supply of stories produced by blogs and online newspaper—because, as I discussed above, the pressure to accumulate pageviews compels writers to fill the limitless bandwidth of the internet at an unhealthy rate. This creates a dangerous vacuum in which bloggers at all levels are under immense pressure to constantly find stories wherever they can, and oftentimes to create something out of nothing, to keep the wheels of the modern media machine turning.

Blogs at the lowest levels of the hierarchy are typically underfunded and understaffed and tend to rely upon the small, close-knit nature of the community they are part of—meaning they often publish material based on suggestions from members of the community and follow leads on social media without really questioning their veracity. In many ways, this is quite reasonable: a respected national news station upon hearing some scandalous gossip regarding the Biden administration needs to be damn sure it is accurate before reporting it to the public, whereas a blog about Great Pyrenees dogs and their crazy antics is less concerned with the possibility that its posts might be construed as fake news. Generally speaking, smaller and more specialized blogs have fewer resources to investigate leads and less incentive to do so regardless.

The problem starts to arise, however, when we look at the middle rung in the hierarchy. Here, the bloggers are still desperate for stories, and they simply don't have time to search for them in traditional journalistic ways, so the obvious shortcut is to scour lower-tier blogs. Exciting posts that exhibit the potential to generate pageviews from a larger audience are quickly scooped up and refashioned by the mid-range bloggers. But these bloggers lack the time and resources to trace the stories back to their origins and fact-check them carefully, so a safe hedge is to simply report that that such-and-such blog (the lower-tier one) is reporting that such-and-such happened. You can't be wrong: whether or not that original story is true, it is unquestionably true that the story was featured on the blog in question.

Next, with enough horizontal propagation, the distinction between the story and the meta-story becomes blurred as bloggers quote each other and race

to share in the pageviews generated by this scoop. In time, the popularity of this story itself can become the story—for virality is newsworthy, isn't it?—at which point it is safe for national newspapers to elevate matters to the highest rung with headlines about this story taking the internet by storm. We saw this frequently in the final years of Steve Jobs: rumors of unknown provenance swirled about the shadows of the internet, gaining traction in unpredictable ways, and upon reaching a critical mass ended up influencing the stock price of Apple and in this way became *real* news, so to speak.

The upward creep of blog posts through the hierarchy happens in more direct ways as well. A national survey found³ that nearly nine out of ten journalists use blogs to research their stories, so even those at the top look downward for information. Moreover, the best way for a blogger to gain serious traffic is to have their stories picked up—and linked to—by higher-level organizations, especially national news sites. So, mid-level bloggers often submit their posts to news aggregators that are monitored by mass media journalists, and they even directly contact journalists in the hopes of getting interest from them—because, after all, even these journalists are in the constant hunt for pageview-generating popular stories.

Ryan Holiday wrote a marvelous book on this phenomenon, *Trust Me, I'm Lying: Confessions of a Media Manipulator*, based on his experiences of deliberately encouraging and exploiting for commercial gain this blogospheric form of upward mobility. In it, he describes how he can “turn nothing into something by placing a story with a small blog that has very low standards, which then becomes the source for a story by a larger blog, and that, in turn, for a story by larger media outlets.” He says that he often sees “uniquely worded or selectively edited facts that paid editors inserted into Wikipedia show up later in major newspapers and blogs, with the exact same wording,” a clear sign of journalistic shortcuts and how they can be taken advantage of. He insightfully, and frighteningly, summarizes the societal consequences of this game that he played for years as follows: “The news, whether it's found online or in print, is just the content that successfully navigated the media's filters. [...] Since the news informs our understanding of what is occurring around us, these filters create a constructed reality.” And remember, this constructed reality Holiday refers to stems from data-driven pageview economics. Data in the 21st century is supposed to provide a powerful new unvarnished window of truth into our world, but we see in this discussion of internet journalism that, alarmingly, it also undergirds a perilous perversion of our basic perceptions of the world.

³“National Survey Finds Majority of Journalists Now Depend on Social Media for Story Research.” *Cision*, January 20, 2010: <https://www.prnewswire.com/news-releases/national-survey-finds-majority-of-journalists-now-depend-on-social-media-for-story-research-82154642.html>.

A recent study⁴ by Harvard researchers on a disinformation campaign concerning mail-in voter fraud in the 2020 election details specific examples of fake news stories that originated in lower-tier publications with minimal editorial standards then launched upward through the system, spreading horizontally as they did so. For instance, a *New York Post* article from August 2020 relied on uncorroborated information from a single anonymous source, supposedly a Democratic operative, who claimed to have engaged in all sorts of voter fraud for decades to benefit the Democrats. Shortly afterward, versions of this story were put out by the *Blaze*, *Breitbart*, *Daily Caller*, and the *Washington Examiner*, and it eventually reached *Fox News* where it was covered on Tucker Carlson's show and on *Fox & Friends*. The Harvard researchers even argue, though without too much quantitative evidence, that popular news outlets are more to blame for the viral spread of disinformation than the much-maligned social media—at least in the specific context of discrediting the results of the 2020 presidential election. I'll revisit this topic in Chapter 8.

This state of journalistic affairs in which grabbing the reader's attention with flashy headlines and salacious content is more important than quality, and fidelity to truth is a mere afterthought, might sound familiar to the historically minded individual. Indeed, the so-called “yellow press” of the late 19th century and first few years of the 20th century—when papers with eye-catching headlines and scant legitimate content were hustled on street corners—had many of the same ills of today's online media ecosystem. To understand how we can dig ourselves out of this mess, it helps to look back and see how it was done in the past.

Historical Context

Theodore Roosevelt bemoaned that the newspapers at the time of his presidency “habitually and continually and as a matter of business practice every form of mendacity known to man, from the suppression of the truth and the suggestion of the false to the lie direct.”⁵ Just prior to his presidency, in one of the most extreme instances, fake news helped launch the Spanish-American War. William Randolph Hearst knew that the war would be a huge boon to his newspaper sales, but when one of his correspondents in Havana informed him that there would not be a war, Hearst fatefully responded: “You furnish the pictures, I'll furnish the war.” Hearst then published in his *Morning*

⁴Yochai Benkler et al., “Mail-In Voter Fraud: Anatomy of a Disinformation Campaign,” *Berkman Klein Center at Harvard University*, October 1, 2020: <https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020>.

⁵Frances Fenton, “The Influence of Newspaper Presentations Upon the Growth of Crime and Other Anti-Social Activity,” *American Journal of Sociology* Vol. 16, No. 3 (Nov. 1910), 342–371: <https://www.jstor.org/stable/2763009>.

Journal fake drawings of Cuban officials strip-searching American women, and his lucrative war soon followed.⁶

The solution to this problem of untrustworthy newspapers was the *subscription model*, ushered in by the *New York Times* around the turn of the century in a deliberate effort to make journalism more reliable. It worked and became the industry norm throughout the 20th century. With the subscription model, readers who are misled or disappointed by the content unsubscribe and turn to a competitor paper, so there is a direct financial incentive for the publisher to maintain quality, truthful journalism. In short, customers were finally paying for reputation, not just headline.

The 21st century in some ways turned journalism back to the 19th century, because unlike the 20th-century subscription model in which readers commit to one or two news sources, now with social media and news aggregators the news organization becomes secondary to the headline for many readers.⁷ Browsing the top stories in Google News is not so different from standing on a 19th-century street corner hearing the newsboys shout out the latest headlines in an effort to entice you to take the bait. But the key differences between now and then are (1) the scale enabled by the internet—instead of a handful of newspapers competing for street corner sales, there are countless sites competing for clicks—and (2) the detailed pageview data, which essentially render the entire journalistic blogosphere a vast quantitative experiment in maximizing clicks above all else. In short, contemporary pageview-driven news is the regrettable 19th-century yellow press on digital steroids.

There are some signs of hope, however. Just as the *New York Times* ushered in the print subscription model at the turn of the 20th century, the *Wall Street Journal* ushered in the online subscription model (the *paywall*) at the turn of the 21st century, a move that has been followed by the *New York Times*, the *Washington Post*, and many other highly reputed news organizations—and with great success at righting many of the earlier period's wrongs, one might argue. Readers pay monthly fees to these organizations in order to access and support quality journalism.

⁶Jacob Soll, "The Long and Brutal History of Fake News," *Politico*, December 18, 2016: <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535>.

⁷A study found that when Americans encounter news on social media, the degree to which they trust it is determined more by who shared it than by who published it: "people who see an article from a trusted sharer, but one written by an unknown media source, have much more trust in the information than people who see the same article that appears to come from a reputable media source shared by a person they do not trust." See "'Who shared it?': How Americans decide what news to trust on social media," *American Press Institute*, March 20, 2017: <https://www.americanpressinstitute.org/publications/reports/survey-research/trust-social-media/>.

One major downside to the subscription model is that it creates a financial barrier to quality journalism, and consequently people with less economic means are prone to rely on less accurate news—and this can lead to dangerous socioeconomic tensions and schisms. Indeed, it is a scary thought that middle- and upper-class Americans can afford to read the *New Yorker*, the *Atlantic*, and the *Wall Street Journal*, while the lower classes are relegated to free online newspapers supported entirely by ad revenue and therefore driven by pageviews.

Moreover, the subscription model simply is not an option for all but the largest organizations. One of the most positive aspects of the 21st-century media landscape is that it is far more democratized and diverse than ever before. No longer must we rely on a select few gatekeepers to tell us what is happening in the world. Voices that have traditionally been kept out of the mainstream press are now being heard for the first time. But nobody is willing to subscribe to dozens of different newspapers; due to the not-insignificant cost of a subscription, people choose which paywalls they are willing to overcome very selectively. The result is that usually only organizations with a large reach and broad audience have a chance of being financially supported by paying subscribers. For the rest, ad revenue is the only financial model available.

Fortunately, even in the realm of freely available blogs, there are glimmers of light. For instance, in the early days of the COVID-19 pandemic, a lengthy, technical, and well-researched blog post⁸ ended up drawing over forty million reads and possibly played an important role in shifting the political discourse on how governments should respond to the pandemic. This article was the exact opposite of clickbait, and it shows that in the right context genuine substance is capable of drawing pageviews at astonishing numbers. Just as many environmentally or socially oriented consumers now choose where to shop based on the views and values of the companies they buy from, perhaps news consumers are ready to recognize pageviews as influential currency and spend them more meaningfully and thoughtfully.

Before you become too sanguine, however, I'd like to relate some specific tales of pageview journalism driving the spread of fake news and shaping our political reality.

Examples of Fake News Peddlers

Paris Wade and Ben Goldman were both twenty-six years old in 2016 when the website they ran together, LibertyWritersNews.com, accumulated tens of millions of pageviews in the span of six months; ninety-five percent of the

⁸Tomas Pueyo, "Coronavirus: Why You Must Act Now," *Medium*, March 10, 2020: <https://tomaspueyo.medium.com/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca>.

site's traffic came from the eight hundred thousand followers they acquired on Facebook during this period. At its peak, their monthly revenue reached upwards of forty thousand dollars. Prior to this venture, they were both unemployed restaurant workers.

Wade and Goldman would studiously follow the analytics of their “news” stories after posting them to see what brought in the most readers. Here's a typical headline for one of their posts: “THE TRUTH IS OUT! The Media Doesn't Want You To See What Hillary Did After Losing...” Wade explained to the *Washington Post*⁹ that “Nothing in this article is anti-media, but I've used this headline a thousand times. Violence and chaos and aggressive wording is what people are attracted to.” Goldman added: “Our audience does not trust the mainstream media. It's definitely easier to hook them with that.” Wade followed up: “There's not a ton of thought put into it. Other than it frames the story so it gets a click. We're the new yellow journalists. We're the people on the side of the street yelling that the world is about to end.”

Why were Wade and Goldman so open with a journalist from the left-leaning, mainstream media *Washington Post*? Because they didn't care. They didn't believe a word of what they wrote on their website, but they knew their readership was never going to see—let alone trust—an article in the *Washington Post*, so they were happy to brag about their business success and have a laugh about all the suckers they have been duping with unabashedly fake news. In 2018, it was uncovered that Wade and Goldman were also involved in the fake news scheme run out of Macedonia before the 2016 presidential election that has generated a lot of press coverage for the possibility that it helped tilt the balance of the election to Trump. At the time when this Macedonian connection was first reported, Wade was running for Nevada state assembly; he lost to the Democratic contender—fortunately so, I think we can all agree.

Christopher Blair, along with some friends, launched a fake right-wing news site on Facebook during the run-up to the 2016 presidential election. He was profiled in a tell-all story¹⁰ in the *Washington Post*. But Blair had even less to hide than Wade and Goldman, for Blair's site was openly satirical. Indeed, Blair was a liberal blogger, and his site started simply as a practical joke among friends to poke fun at the extremist ideas spreading among the far right and to reveal the gullibility of people who couldn't tell obvious fake news from

⁹Terrence McCoy, “For the ‘new yellow journalists,’ opportunity comes in clicks and bucks,” *Washington Post*, November 20, 2016: https://www.washingtonpost.com/national/for-the-new-yellow-journalists-opportunity-comes-in-clicks-and-bucks/2016/11/20/d58d036c-adbf-11e6-8b45-f8e493f06fcd_story.html.

¹⁰Eli Saslow, “‘Nothing on this page is real’: How lies become truth in online America,” *Washington Post*, November 17, 2018: https://www.washingtonpost.com/national/nothing-on-this-page-is-real-how-lies-become-truth-in-online-america/2018/11/17/edd44cc8-e85a-11e8-bbdb-72fdbf9d4fed_story.html.

reality. Blair invented far-right, and far-fetched, stories about “California instituting sharia, former president Bill Clinton becoming a serial killer, undocumented immigrants defacing Mount Rushmore, and former president Barack Obama dodging the Vietnam draft when he was nine.” While doing this, he realized that “The more extreme we become, the more people believe it.”

Even though Blair’s site was openly satirical—it included fourteen disclaimers, one of which directly stated that “Nothing on this page is real”—for a time it became the most popular page on Facebook among Trump-supporting conservatives over fifty-five. His stories, which reached an audience of up to six million monthly visitors, were often taken seriously and wound up on the same Macedonian fake news farm that Wade and Goldman were involved in—despite Blair’s supposed attempts to cast his followers and likers and sharers as ignoramuses and pawns. Part of the problem with Blair’s approach here, as you’ll see throughout this book and especially in Chapter 8, is that social media provides news articles with a life and trajectory of their own and frequently strips articles of their original context and intent.

For a while, Blair liked to let people share his articles and then call them out for spreading his fake news—he thought that publicly embarrassing people would lead them to think more critically about what they shared online—but the site’s popularity among true believers grew at a staggering rate nonetheless. On his personal Facebook page, he once wrote: “No matter how racist, how bigoted, how offensive, how obviously fake we get, people keep coming back. Where is the edge? Is there ever a point where people realize they’re being fed garbage and decide to return to reality?” Perhaps Blair was underestimating the intense gravitational pull of the pageview-driven blogosphere—or perhaps he was well aware of it and simply enjoyed profiting from it financially.

In November 2016, *NPR* tracked down¹¹ the author of one particular fake news story that went viral during the election, to try to understand where such things come from. The article’s headline was “FBI Agent Suspected In Hillary Email Leaks Found Dead In Apparent Murder-Suicide.” It was published in what appeared to be a local newspaper called the *Denver Guardian*, and despite being completely fabricated, it was shared on Facebook over half a million times. The website for this newspaper had the local weather but only one news story, this fake one. Some clever online detective work led to the identity of the individual behind this fake local newspaper, who turned out to be Jestin Coler, a forty-year-old registered Democrat and father of two.

¹¹Laura Sydell, “We Tracked Down a Fake-News Creator In The Suburbs. Here’s What We Learned.” *NPR*, November 23, 2016: <https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs>.

Coler claimed he entered the fake news business in 2013 with similar intentions as Christopher Blair: “The whole idea from the start was to build a site that could kind of infiltrate the echo chambers of the alt-right, publish blatantly false or fictional stories and then be able to publicly denounce those stories and point out the fact that they were fiction.” After realizing how easily and rapidly his stories were spreading, Coler decided to capitalize on this endeavor and ended up forming a fake news company that employed a couple dozen writers and spanned an undisclosed number of websites, including the one for the *Denver Guardian*—a site that, according to Coler, collected over one and a half million views in a ten-day period. In describing the fake FBI agent story, Coler said: “Everything about it was fictional: the town, the people, the sheriff, the FBI guy. And then ... our social media guys kind of go out and do a little dropping it throughout Trump groups and Trump forums and boy it spread like wildfire.” As it and other fake stories written by his company spread across the country, Coler was making around twenty thousand dollars per month from ad revenue.

One consequence of the shifting economic forces in journalism has been the decimation of regional newspapers. As I discuss next, Coler’s fake Denver-based newspaper was not an isolated invention: nefarious entities have found strategic ways to fill the journalistic vacuum left behind as authentic local newspapers have gone out of business.

Losing Reliable Local News

Twenty percent of local newspapers across America have shut down over the past decade, and many of the ones that remain have had to significantly cut their staff due to financial pressures. This sad development was largely precipitated by the shift from print to online newspapers: most regional papers cannot possibly get enough web traffic to support themselves financially with ad revenue, and paywalls don’t work much better because if a reader is to pay for an online subscription to a newspaper, then it is usually going to be a well-known national paper rather than a regional one. Unfortunately, the loss of local reporters and the increased financial constraints and time pressures on the ones that remain have exacerbated the flaws described earlier in the news hierarchy that allow fake news to propagate and proliferate.

The disappearance of local newspapers has also been taken advantage of more directly through deliberate subterfuge. At the end of 2019, the *Columbia Journalism Review* (CJR), expanding on stories first reported elsewhere,

uncovered¹² a network of nearly five hundred websites masquerading as local news organizations, each “distributing thousands of algorithmically generated articles and a smaller number of reported stories.” I’ll turn to more sophisticated forms of automated story generation, based on cutting-edge artificial intelligence, in the next chapter; the “algorithmic” methods of automation used here, in contrast, are quite simple—essentially just bulk applications of copy-and-paste.

Almost half of these fake local news websites were set up by a single company, Metric Media, in a single year, and they all trace back to Brian Timpone, a conservative businessman who attracted outrage in 2012 for his “pink slime journalism” company Journatic that used low-cost automated story generation and was shown to have faked quotes and plagiarized rampantly. *CJR* found that during a two-week period leading up to the publication of its study, over fifty thousand stories had been published in this network, but “only about a hundred titles had the bylines of human reporters. The rest cited automated services or press releases.”

The websites in this *CJR* study, with names like *East Michigan News*, *Hickory Sun*, and *Grand Canyon Times*, are designed to look like ordinary local news organizations. They largely comprise easily mass-produced stories on topics such as local real estate prices, but strategically interspersed in this filler are political pieces—for instance, quoting local Republican officials on national right-wing talking points. These sites contain little information on funding sources or political usage, even though some were revealed to have been funded by political candidates and lobbying campaigns. They are, in short, a sinister weaponization of the trust people place in local news.

Just one year after the *CJR* study was released, the *New York Times* published an in-depth investigation¹³ of this deceptive Timpone-led network based on interviews with dozens of current and former employees and thousands of internal emails spanning multiple years. It found that the network had grown to over a thousand websites—more than double the number for the largest authentic newspaper chain in the country—and now operates in all fifty US states. These fake local news sites publish “propaganda ordered up by dozens of conservative think tanks, political operatives, corporate executives and public-relations professionals.” The sites in the network eschew journalistic standards such as fairness and transparency but stop short of outright fake

¹²Priyanjana Bengani, “Hundreds of ‘pink slime’ local news outlets are distributing algorithmic stories and conservative talking points,” *Columbia Journalism Review*, December 18, 2019: https://www.cjr.org/tow_center_reports/hundreds-of-pink-slime-local-news-outlets-are-distributing-algorithmic-stories-conservative-talking-points.php.

¹³Davey Alba and Jack Nicas, “As Local News Dies, a Pay-for-Play Network Rises in Its Place,” *New York Times*, October 20, 2020: <https://www.nytimes.com/2020/10/18/technology/timpone-local-news-metric-media.html>.

news. The editors assign articles to freelance writers with “precise instructions on whom to interview and what to write” and typically pay from a few dollars to a few dozen dollars per article. And they continue to surround these handwritten articles with lots of easily automated content—for instance, by pasting in press releases published elsewhere or by stitching together a local weather forecast with generic fluff to give the impression of an article written by a regional meteorologist. Drawing on nostalgia for the halcyon days of local news, in some cases these fake local news setups even deliver print copies of their papers, unsolicited, to residents’ houses.

In a November 2020 interview¹⁴ with the *Atlantic*, just days after Joe Biden defeated Donald Trump in the presidential election, Barack Obama described how the media landscape has changed since he first ran with Biden on his ticket—and the consequences this has had for the American political landscape. He said that in late 2008, even a Republican-owned small-town newspaper editor would meet with him and write an editorial that presented him as a liberal Chicago lawyer but a decent guy with some good ideas, and the local TV coverage was also fair. He lamented that “you go into those communities today and the newspapers are gone. If Fox News isn’t on every television in every barbershop and VFW hall, then it might be a Sinclair-owned station, and the presuppositions that exist there, about who I am and what I believe, are so fundamentally different, have changed so much, that it’s difficult to break through.” He went on to bemoan how “Now you have a situation in which large swaths of the country genuinely believe that the Democratic Party is a front for a pedophile ring. This stuff takes root.”

The disappearance of genuine local news organizations—a significant loss in American media, triggered largely by the economics of the internet—has produced a vacuum that’s been filled in unscrupulous ways. This has created a more polarized nation and fanned the flames of fake news.

Summary

American newspapers in the late 19th century were sold each day on an individual basis and competed for sales by having the wildest headlines even if the actual content was exaggerated or fabricated. The subscription model took over and dominated throughout the 20th century; it brought fake news under control by providing a financial incentive for journalists to write accurate, well-researched stories because misleading content would cause customers to cancel their subscriptions and turn to competitor papers.

¹⁴Jeffrey Goldberg, “Why Obama Fears for Our Democracy,” *Atlantic*, November 16, 2020: <https://www.theatlantic.com/ideas/archive/2020/11/why-obama-fears-for-our-democracy/617087/>.