Nevertheless, interest in AI lie detection has surged in the past few years. Some methods rely on video, others audio, and others text alone. They all suffer from a lack of transparency, exaggerated claims of accuracy, an unnervingly high rate of false positives, and bias that disproportionately impacts minority populations. This has not stopped them from being used for employment screening and fraud detection and occasionally even in the courtroom (despite the Frye standard), and from being trialed in airport security and other settings. Given all the flaws, overzealous commercialization, corporate secrecy, and embarrassing lack of even an attempt at scientific foundations, it does not appear that the polygraph—even when reinvented with AI—will ever be able to detect lies with the consistency needed to rein in fake news. Instead, the mythical ability to use fancy technology to peer into the mind and reliably detect deception is itself the fake news in this story.

If we can't use algorithmic lie detectors to unmask fake news and get to the truth in controversial matters, perhaps we should just do what hundreds of millions of people do every day: Google it. But be careful—there too the algorithms behind the scenes systematically distort our perception of reality, as you will see in the next chapter.

CHAPTER

6

Gravitating to Google

The Dangers of Letting an Algorithm Answer Our Questions

Search engines have come to play a central role in corralling and controlling the ever-growing sea of information that is available to us, and yet they are trusted more readily than they ought to be. They freely provide, it seems, a sorting of the wheat from the chaff, and answer our most profound and most trivial questions. They have become an object of faith.

—Alex Halavais, *Search Engine Society*

Billions of people turn to Google to find information, but there is no guarantee that what you find there is accurate. As awareness of fake news has risen in recent years, so has the pressure on Google to find ways of modifying its algorithms so that trustworthy content rises to the top. Fake news is not limited to Google's main web search platform—deceptive and harmful content also play a role on other Google products such as Google Maps, Google News, and Google Images, and it also shows up on Google's autocomplete

N. Giansiracusa, *How Algorithms Create and Prevent Fake News*,
https://doi.org/10.1007/978-1-4842-7155-1_6

tool that feeds into all these different products. In this chapter, I'll look at the role fake news plays in all these contexts and what Google has done about it over the years. In doing so, I'll take a somewhat more expansive view of the term "fake news" compared with previous chapters to include hateful racist stereotypes and bigoted misinformation.

Setting the Stage

On the morning of November 14, 2016, six days after the US presidential election in which Trump won the electoral college and Clinton won the popular vote, both by relatively wide margins, the top link in the "In the news" section of the Google search for "final election results" was an article asserting that Trump had won the popular vote by seven hundred thousand votes.[1] It was from a low-quality WordPress blog that cited Twitter posts as its source, yet somehow Google's algorithms propelled this fake news item to the very top. In response to this worrisome blunder, a Google spokesperson said:[2] "The goal of Search is to provide the most relevant and useful results for our users. We clearly didn't get it right, but we are continually working to improve our algorithms."

The next day, Sundar Pichai—just one year into his role as CEO of Google—was asked in an interview[3] with the BBC whether the virality of fake news might have influenced the outcome of the US election. Mark Zuckerberg had already dismissed this idea (naively and arrogantly, it appears in hindsight) as "pretty crazy," whereas Pichai was more circumspect: "I am not fully sure. Look, it is important to remember this was a very close election and so, just for me, so looking at it scientifically, one in a hundred voters voting one way or the other swings the election either way." Indeed, due to the electoral college, the election came down to just one hundred thousand votes. When asked specifically whether this tight margin means fake news could have potentially played a decisive role, Pichai said, after a pause: "Sure. You know, I think fake news as a whole could be an issue."

[1] Philip Bump, "Google's top news link for 'final election results' goes to a fake news site with false numbers," *Washington Post*, November 14, 2016: https://www.washingtonpost.com/news/the-fix/wp/2016/11/14/googles-top-news-link-for-final-election-results-goes-to-a-fake-news-site-with-false-numbers/.

[2] Richard Nieva, "Google admits it messed up with fake election story," *CNET*, November 14, 2016: https://www.cnet.com/news/google-fake-news-election-donald-trump-popular-vote/.

[3] Kamal Ahmed, "Google commits to £1bn UK investment plan," *BBC News*, November 15, 2016: https://www.bbc.com/news/business-37988095.

Less than a year later, Eric Schmidt, then the executive chairman of Alphabet, Google's parent company, publicly admitted[4] that Google had underestimated the potential dedication and impact of weaponized disinformation campaigns from adversarial foreign powers: "We did not understand the extent to which governments—essentially what the Russians did—would use hacking to control the information space. It was not something we anticipated strongly enough." He made that remark on August 30, 2017.

One month and two days later, on October 1, 2017, the worst mass shooting in modern US history took place in Las Vegas. Within hours, a fake news item was posted on the dubious website 4chan, in a "politically incorrect" channel associated with the alt-right, falsely accusing a liberal man as the shooter. Google's algorithm picked up on the popularity of this story, and soon the first result in a search for the name of this falsely accused man was the 4chan post—which was misleadingly presented as a "Top story" by Google. The response[5] from a Google spokesperson was unsurprisingly defensive and vague: "Unfortunately, early this morning we were briefly surfacing an inaccurate 4chan website in our search results for a small number of queries. [...] This should not have appeared for any queries, and we'll continue to make algorithmic improvements to prevent this from happening in the future." Google's wasn't the only algorithm misfiring here: Facebook's "Trending Topic" page for the Las Vegas shooting listed multiple fake news stories, including one by the Russian propaganda site *Sputnik*.[6] Schmidt's remark from a month earlier about Russian interference was oddly prescient—or frustratingly obvious, depending on your perspective.

One and a half months later, at an international security conference, Schmidt tried to explain[7] the challenge Google faces when dealing with fake news: "Let's say this group believes fact A, and this group believes fact B, and you passionately disagree with each other and you're all publishing and writing about it and so forth and so on. It's very difficult for us to understand

[4] Austin Carr, "Alphabet's Eric Schmidt On Fake News, Russia, And 'Information Warfare'," *Fast Company*, October 29, 2017: https://www.fastcompany.com/40488115/alphabets-eric-schmidt-on-fake-news-russia-and-information-warfare.

[5] Gerrit De Vynck, "Google Displayed Fake News in Wake of Las Vegas Shooting," *Bloomberg*, October 2, 2017: https://www.bloomberg.com/news/articles/2017-10-02/fake-news-fills-information-vacuum-in-wake-of-las-vegas-shooting.

[6] Kathleen Chaykowski, "Facebook And Google Still Have A 'Fake News' Problem, Las Vegas Shooting Reveals," *Forbes*, October 2, 2017: https://www.forbes.com/sites/kathleenchaykowski/2017/10/02/facebook-and-google-still-have-a-fake-news-problem-las-vegas-shooting-reveals/.

[7] Liam Tung, "Google Alphabet's Schmidt: Here's why we can't keep fake news out of search results," *ZDNET*, November 23, 2017: https://www.zdnet.com/article/google-alphabets-schmidt-heres-why-we-cant-keep-fake-news-out-of-search-results/.

truth. […] It's difficult for us to sort out which rank, A or B, is higher." He went on to explain that it is easier for Google to handle false information when there is a large consensus involved. A fair point in some respects, but it's hard to imagine how this applies to these past debacles—was there not a consensus in the 2016 election that Trump lost the popular vote, and in the Las Vegas shooting that an unsubstantiated rumor on an alt-right site was not actual news? What about just a few months later, in February 2018, when the top trending video on YouTube (which, as you recall from Chapter 4, is owned by Google and which has essentially taken over the video search portion of Google) was[8] an egregious conspiracy theory claiming that some survivors of the Parkland, Florida, high school shooting were actors?

If there really was a lack of "consensus" in these incidents, one has to wonder whether that was actually the cause of the problems with Google's algorithm as Schmidt suggested or whether he perhaps had it backward. Maybe the fact that Google's algorithm has propped up fake stories like these, thereby lending them both legitimacy and a vast platform, caused some of the erosion of truth that ultimately led to a lack of consensus on topics that should not have been controversial in the first place. In other words, did Google reflect a state of confusion, or did it cause one? In all likelihood, the answer is a combination of both. To start unravelling this complex issue, it helps to separate out the different services Google provides so that we can delve into the algorithmic dynamics underlying each one and explore the deceptive and hateful content that has surfaced on each one.

Throughout this chapter, I shall use the term "fake news" more broadly to include racist and bigoted content. I have largely resisted doing so in the book thus far because so much has been written on algorithmic bias already, so rather than overcrowding these chapters by retelling that tale, I prefer to encourage you to consult the excellent and rapidly developing literature on the matter. But when it comes to Google, which is such an intimate and immediate source of information for so many people, I cannot earnestly disentangle news-oriented disinformation from socially oriented disinformation of the kind found in racism, sexism, anti-Semitism, etc. For one thing, many fake news sites align with the white supremacist–leaning alt-right, so when Google feeds its users bigoted information, it is also priming them to fall for hardcore alt-right fake news material. And at a more philosophical level, one could argue that racist stereotypes are a form of fake news—they are in essence harmful disinformation that happens to focus on certain populations.

[8] Sara Salinas, "The top trending video on YouTube was a false conspiracy that a survivor of the Florida school shooting was an actor," *CNBC*, February 21, 2018: https://www.cnbc.com/2018/02/21/fake-news-item-on-parkland-shooting-become-top-youtube-video.html.

Google Maps

One of the most abhorrent examples of hateful disinformation on a Google platform occurred in May 2015, during President Obama's second term in office. It was reported[9] in the *Washington Post* that if one searched Google Maps for "N****r king" or "N***a house" (with the asterisks filled in), the map would locate and zoom in on the White House. This was not the result of algorithmic bias or some other subtle failure of AI, it was directly the result of racist users with malicious intent—or as some people call it, third-party trolling and vandalism. This, and other acts of vandalism, caused Google to suspend user-submitted edits to Google Maps at the time: "We are temporarily disabling editing on Map Maker starting today while we continue to work towards making the moderation system more robust."

An intriguing and thankfully less hateful act of vandalistic disinformation on Google Maps occurred[10] in February 2020 when an artist tricked the service into showing a nonexistent traffic jam in the center of Berlin. How did he pull this off? He simply piled a hundred borrowed and rented smartphones into a little red wagon that he slowly walked around the city while the phones' location services were enabled.

Most of the false information on Google Maps is not motivated by hate or artistry—it results from purely financial considerations, as I next discuss.

Fake Business Information

In June 2019, the *Wall Street Journal* reported[11] on the deluge of fake businesses listed on Google Maps. Experts estimated that around ten million business listings on Google Maps at any given moment are falsified and that hundreds of thousands of new ones appear each month. They claim that the "majority of listings for contractors, electricians, towing and car repair services and lawyers, among other business categories, aren't located at their pushpins on Google Maps." One motivation for someone to make

[9] Brian Fung, "If you search Google Maps for the N-word, it gives you the White House," *Washington Post*, May 19, 2015: https://www.washingtonpost.com/news/the-switch/wp/2015/05/19/if-you-search-google-maps-for-the-n-word-it-gives-you-the-white-house/.

[10] Rory Sullivan, "Artist uses 99 phones to trick Google into traffic jam alert," *CNN*, February 4, 2020: https://www.cnn.com/style/article/artist-google-traffic-jam-alert-trick-scli-intl/index.html.

[11] Rob Copeland and Katherine Bindley, "Millions of Business Listings on Google Maps Are Fake—and Google Profits," *Wall Street Journal*, June 20, 2019: https://www.wsj.com/articles/google-maps-littered-with-fake-business-listings-harming-consumers-and-competitors-11561042283.

fake listings is to give a misleading sense of the reach of one's business by exaggerating the number of locations and branch offices on Google Maps. Another motivation is to drown out the competition.

The owner of a cash-for-junk-cars business in the Chicago suburbs mostly relied on the Yellow Pages for advertising, but in 2018 he was contacted by a marketing firm that offered to broadcast his business on Google Maps—for a five-figure fee. He agreed, but then a few months later, the firm came back with a threat: if he doesn't start giving them half his revenue, then they will bury his Google Maps listing under hundreds of fictitious competitors. He refused, and sure enough they posted an avalanche of made-up competitors with locations near him so that it would be very difficult for customers to find his business amid all the fake noise. He drove around a few Chicago neighborhoods and searched on his phone for auto salvage yards as he went; he said that more than half the results that came up were fake. These fake listings pushed his business listing off the first page of Google Maps search results, and soon his number of incoming calls dropped by fifty percent.

Businesses do not pay anything to be listed on Google Maps, but before each one appears on the service, Google usually sends either a postcard or email or calls the business on the phone to provide a verification code that must be typed into Google Maps in order to have the listing approved. This precautionary measure is quite flimsy, and scammers have consistently been able to bypass it. In fact, doing so has become a business. The *Wall Street Journal* profiled a "listings merchant" who placed nearly four thousand fake listings on Google Maps each day from his basement in rural Pennsylvania.

This listings merchant claimed to have had a staff of eleven employees who ran a "mostly" legitimate service that helped clients improve their visibility on Google Maps. But he also claimed to have had a separate staff of twenty-five employees in the Philippines who used "unsanctioned methods to fill orders for fake listings" at a rate of ninety-nine dollars per fake listing. This fake listing service was "aimed at businesses that want to pepper Google Maps with faux locations to generate more customer calls." His employees gathered addresses from commercial real estate listings; to bypass Google's safeguards, they simply purchased phone numbers cheaply online and had Google's verification codes sent to these, then they routed these numbers to the clients once the Google Maps listings were approved. At the time of the *Wall Street Journal* article, however, this listings merchant said Google was investigating him, and tens of thousands of his listings had already been taken down.

Fake business is evidently big business on Google Maps: the site removed over three million false business listings in 2018. That figure comes from a company blog post[12] written by the director of Google Maps titled "How we fight fake business profiles on Google Maps." This blog post was published on June 20, 2019—the same exact date as the *Wall Street Journal* piece. It does not take a great stretch of the imagination to see this conspicuously timed blog post as a strategic effort to reduce the backlash that would surely follow the publication of the *Wall Street Journal* investigation. This post includes some other staggering figures, including that Google Maps has over two hundred million places and that "every month we connect people to businesses more than nine billion times, including more than one billion phone calls and three billion requests for directions."

The Google Maps blog post gives some examples of how people capitalize on fake business listings: "They do things like charge business owners for services that are actually free, defraud customers by posing as real businesses, and impersonate real businesses to secure leads and then sell them." (We know from the *Wall Street Journal* that there are more problems than just these.) The post also points out that as people find deceptive ways of gaming the system, Google is "continually working on new and better ways to fight these scams using a variety of ever-evolving manual and automated systems," but that as it does this the nefarious users find new deceptive methods and "the cycle continues."

These automated systems—algorithmic moderation, in other words—are closely held corporate secrets because revealing details about them would "actually help scammers find new ways to beat our systems." All the blog post really reveals is that (1) of the three million fake listings taken down in 2018, over ninety percent were removed by the internal systems before a user saw them, whereas the remaining ones were reported by users on the platform, and (2) more than one hundred fifty thousand accounts were disabled in 2018, a fifty percent increase over the previous year.

Perhaps the secretiveness of that blog post did not sit well with some, as just eight months later a new blog post[13] was published—still by the director of Google Maps, but with a different individual occupying this position—that, while still circumspect, went into more detail about the algorithmic moderation the site uses. This post said that Google Maps uses "automated detection systems, including machine learning models, that scan the millions of contributions we receive each day to detect and remove policy-violating content," and that for fake reviews specifically these machine learning models "watch out for specific words and phrases, examine patterns in the types of

[12] Ethan Russell, "How we fight fake business profiles on Google Maps," *Google blog*, June 20, 2019: https://www.blog.google/products/maps/how-we-fight-fake-business-profiles-google-maps/.

[13] Kevin Reece, "Google Maps 101: how contributed content makes a more helpful map," *Google blog*, February 19, 2020: https://www.blog.google/products/maps/google-maps-101-how-contributed-content-makes-maps-helpful/.

content an account has contributed in the past, and can detect suspicious review patterns." Still understandably vague, but I'll turn to the more general topic of machine learning for social media moderation in Chapter 8, so you'll hopefully get a sense of the methods Google Maps is alluding to here—as well as the challenges these methods face.

This second blog post goes on to explain that these automated systems are "not perfect," so Google also relies on "teams of trained operators and analysts who audit reviews, photos, business profiles and other types of content both individually and in bulk." The post also provides some interesting updated figures on content moderation: in 2019, Google Maps (1) removed more than seventy-five million policy-violating reviews and four million fake business profiles "thanks to refinements in our machine learning models and automated detection systems which are getting better at blocking policy-violating content and detecting anomalies for our operators to review"; (2) took down over half a million reviews and a quarter million business profiles that were reported by users; (3) removed ten million photos; (4) disabled almost half a million user accounts.

Google Images

In April 2016, an MBA student posted[14] on Twitter a disturbing discovery: doing a Google image search for "unprofessional hairstyles for work" returned almost entirely pictures of Black women, many with natural hair, while "professional hairstyles for work" returned almost entirely white women. Why was Google's image search algorithm so overtly racist? It's a complicated question, but the two main ingredients to the answer are that the algorithm naively absorbs information out of context and that it naively reflects overt racism permeating society.

Some of the images of Black women that came up on this particular search were from blog posts and Pinterest boards by Black women discussing racist attitudes about hair in the workplace. For instance, one top image was from a post criticizing a university's ban on dreadlocks and cornrows; the post illustrated the banned hairstyles by showing pictures of Black women with them and lamented how these hairstyles were deemed unprofessional by the university. The ban was clearly racist, whereas the post calling attention to it was the opposite, it was antiracist. The Google image search conflated these two contrasting aspects and stripped the hairstyle image of its context, simply associating the image with the word "unprofessional." In doing so, it turned an antiracist image into a racist one. In this inadvertent manner, racism on one college campus was algorithmically amplified and transformed into racist information that was broadcast on a massive scale: anyone innocently looking on Google for tips on how to look professional would be fed the horrendous suggestion that being Black is simply unprofessional.

There soon came to be a data feedback loop here. The MBA student's tweet went viral, which was largely a good thing because it helped raise awareness of Google's algorithmic racism. But this virality caused Google searches on hairstyles to point to this tweet itself and the many discussions about—all of which were calling attention to Google's racism by showing how Black women were labeled "unprofessional" while white women were labeled "professional." Once again, Google vacuumed up these images with these labels and stripped them of their important context, and in doing so the racist effect actually became stronger: a broader array of searches began turning up these offensive associations. In other words, Google image search emboldened and ossified the very same racism that this tweet was calling attention to.

Just a few months after this Google hairstyle fiasco, a trio of researchers in Brazil presented a detailed study on another manifestation of racism in Google's image search—one so abhorrent that the academic study was promptly covered prominently by the *Washington Post*.[15] The researchers collected the top fifty results from Google image searches for "beautiful woman" and "ugly woman," and they did this for searches based in dozens of different countries to see how the results vary by region. This yielded over two thousand images that were then fed into a commercial AI system for estimating the age, race, and gender of each person (supposedly with ninety percent accuracy). Here's what they found.

In almost every country the researchers analyzed, white women appeared more in the search results for "beautiful," and Black and Brown women appeared more in the results for "ugly"—even in Nigeria, Angola, and Brazil, where Black and Brown populations are predominate. In the United States, the results for "beautiful" were eighty percent white and mostly in the age range of nineteen to twenty-eight, whereas the results for "ugly" dropped to sixty percent white—and rose to thirty percent Black—and the ages mostly ranged from thirty to fifty, according to the AI estimates. This form of racism and ageism was not invented by Google's algorithm, it originates in society itself—but the algorithm picks up on it and then harmfully presents it to the world as an established fact. Thankfully, Google seems to have found ways of improving its algorithm in this regard as image searches for beauty now yield a much more diverse range of individuals.

[14] Leigh Alexander, "Do Google's 'unprofessional hair' results show it is racist?" *Guardian*, April 8, 2016: https://www.theguardian.com/technology/2016/apr/08/does-google-unprofessional-hair-results-prove-algorithms-racist-.

[15] Caitlin Dewey, "Study: Image results for the Google search 'ugly woman' are disproportionately black," *Washington Post*, August 10, 2016: https://www.washingtonpost.com/news/the-intersect/wp/2016/08/10/study-image-results-for-the-google-search-ugly-woman-are-disproportionately-black/.

Google Photos is a service introduced by Google in 2015 that allows users to store and share photos, and it uses machine learning to automatically recognize the content of the photos. It now has over one billion users who collectively upload more than one billion photos to the platform daily. But just one month after its initial launch, Google had to offer an official apology and declared itself "appalled and genuinely sorry" for a racist incident—an incident that Google's chief social architect responded[16] to on Twitter by writing "Holy fuck. [...] This is 100% not OK." What had happened? A Black software engineer and social activist revealed on Twitter that Google Photos repeatedly tagged pictures of himself and his girlfriend as "gorillas." Google said that as an immediate fix it would simply discontinue using the label "gorilla" in any capacity, and the company would work on a better longer-term solution.

Two and a half years later, *Wired* conducted a follow-up investigation[17] to see what Google had done to solve this heinous mislabeling problem. It turns out Google hadn't gotten very far from its original slapdash workaround: in 2018, "gorilla," "chimp," "chimpanzee," and "monkey" were simply disallowed tags on Google Photos. Indeed, *Wired* provided Google Photos with a database of forty thousand images well stocked with animals and found the platform performed impressively well at returning photos of whatever animals were requested—except for those named above: when those words were searched, Google Photos said no results were found. For all the hype in AI, the highly lauded team at Google Brain, the heralded breakthroughs provided by deep learning, it seems one of the most advanced technology companies on the planet couldn't figure out how to stop its algorithms from tagging Black people as gorillas other than by explicitly removing gorilla as a possible tag.

The problem here is that even the best AI algorithms today don't form abstract conceptualizations or common sense the way a human brain does—they just find patterns when hoovering up reams of data. (You might object that when I discussed deep learning earlier in this book, I did say that it is able to form abstract conceptualizations, but that's more in the sense of patterns within patterns rather than the kind of anthropomorphic conceptualizations us humans are used to.) If Google's algorithms are trained on real-world data that contains real-world racism, such as Black people being referred to as gorillas, then the algorithms will learn and reproduce this same form of racism.

Let me quickly recap the very public racist Google incidents discussed so far to emphasize the timeline. In May 2015, the *Washington Post* reported the Google Maps White House story that the office of the first Black president in the history of the United States was labeled with the most offensive racial slur in existence. One week later, Google launched Google Photos and within a month had to apologize for tagging images of Black people as gorillas, a story covered by the *Wall Street Journal*, among others. Less than a year later, in April 2016, the *Guardian* reported that Google image searches for unprofessional hairstyles mostly showed photos of Black women. Just a few months after that, in August 2016, the *Washington Post* covered a research investigation that showed Google image search results correlated beauty with race. Oh, I almost forgot: two months earlier, in June 2016, it was reported in many news outlets, including *BBC News*,[18] that doing a Google image search for "three black teenagers" returned mostly police mugshots, whereas searching for "three white teenagers" just showed smiling groups of wholesome-looking kids. These documented racist incidents are just a sample of the dangers inherent in letting Google's data-hungry machine learning algorithms sort and share the world's library of photographs.

Google Autocomplete

Google's autocomplete feature suggests popular searches for users after they type in one or more words to the search box on Google's homepage or to the address bar in Google's web browser Chrome. It is supposed to be a simple efficiency tool, like the autocomplete on your phone that helps you save time by suggesting word completions while you are texting. But Google searches are a powerful instrument that billions of people use as their initial source of information on just about every topic imaginable, so the consequences can be quite dire when Google's autocompletes send people in dangerous directions.

Suggesting Hate

In December 2016, it was reported[19] in the *Guardian* that Google's suggested autocompletes for the phrase "are Jews" included "evil," and for "are Muslims" they included "bad." Several other examples of offensive autocompletes were also found. In response,[20] a Google representative said the following:

[16] Loren Grush, "Google engineer apologizes after Photos app tags two black people as gorillas," *The Verge*, July 1, 2015: https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas.

[17] Tom Simonite, "When It Comes to Gorillas, Google Photos Remains Blind," *Wired*, January 11, 2018: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/.

[18] Rozina Sini, "'Three black teenagers' Google search sparks Twitter row," *BBC News*, June 9, 2016: https://www.bbc.com/news/world-us-canada-36487495.

[19] Carole Cadwalladr, "Google, democracy and the truth about internet search," *Guardian*, December 4, 2016: https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.

[20] Samuel Gibbs, "Google alters search autocomplete to remove 'are Jews evil' suggestion," *Guardian*, December 5, 2016: https://www.theguardian.com/technology/2016/dec/05/google-alters-search-autocomplete-remove-are-jews-evil-suggestion.

"Our search results are a reflection of the content across the web. This means that sometimes unpleasant portrayals of sensitive subject matter online can affect what search results appear for a given query. These results don't reflect Google's own opinions or beliefs." This refrain—that Google isn't racist, it's merely reflecting racism in society—has been a recurring defense throughout all these scandals. Google's response to this *Guardian* article went on to explain that its algorithmically generated autocomplete predictions "may be unexpected or unpleasant," but that "We do our best to prevent offensive terms, like porn and hate speech, from appearing."

On the official company blog, Google explains[21] that autocomplete is really providing "predictions" rather than "suggestions"—meaning it is using machine learning trained on the company's vast database of searches to estimate the words most likely to follow the words the user has typed so far.[22] In other words, Google is not trying to suggest what you should search for, it is just trying to figure out what is most probable that you will be searching for based on what you have typed so far. The Google blog explains that the autocomplete algorithm makes these statistical estimates based on what other users have searched for historically, what searches are currently trending, and also—if the user is logged in—your personal search history and geographic location.

Google is smart enough to moderate (both algorithmically and manually) the results of this machine learning prediction system. The company blog states that "Google removes predictions that are against our autocomplete policies, which bar: sexually explicit predictions that are not related to medical, scientific, or sex education topics; hateful predictions against groups and individuals on the basis of race, religion or several other demographics; violent predictions; dangerous and harmful activity in predictions." It says that a "guiding principle" here is to "not shock users with unexpected or unwanted predictions." In case you lost track, this means Google has said that its autocompletes may be "unexpected or unpleasant," but they aren't supposed to be "unexpected or unwanted." Confused yet? I know that I am.

A Google spokesperson said the company took action within hours of being notified of the offensive autocompletes uncovered by the *Guardian* article.

[21] Danny Sullivan, "How Google autocomplete works in Search," *Google blog*, April 20, 2020: `https://www.blog.google/products/search/how-google-autocomplete-works-search/`.

[22] This general idea of next-word prediction is somewhat similar to the GPT-3 system discussed in Chapter 2. However, GPT-3 was pre-trained on huge volumes of written text and then in real time considers only the prompt text. Google's autocomplete, on the other hand, uses pre-training data that is more focused on searches, and its real-time calculation uses not just the text typed into the prompt so far but also many other factors, as discussed shortly.

However, the *Guardian* found[23] that only some of the offensive examples listed in that article were removed; others remained. Evidently, Google's "guiding principle" is difficult to implement uniformly and incontrovertibly in practice. A little over a year later, in a February 2018 UK parliamentary hearing, Google's vice president of news admitted that "As much as I would like to believe our algorithms will be perfect, I don't believe they ever will be."

An investigation[24] was published in *Wired* just a few days after this UK hearing, finding that "almost a year after removing the 'are jews evil?' prompt, Google search still drags up a range of awful autocomplete suggestions for queries related to gender, race, religion, and Adolf Hitler." To avoid possibly misleading results, the searches for this *Wired* article were conducted in "incognito" mode, meaning Google's algorithm was only using general search history data rather than user-specific data. The top autocompletes for the prompt "Islamists are" were, in order of appearance, "not our friends," "coming," "evil," "nuts," "stupid," and "terrorists." The prompt "Hitler is" yielded several reasonable autocompletes as well as two cringeworthy ones: "my hero" and "god." The first autocomplete for "white supremacy is" was "good," whereas "black lives matter is" elicited the autocomplete "a hate group." Fortunately, at least, the top link for the search "black lives matter is a hate group" was to a Southern Poverty Law Center post explaining why BLM is not, in fact, a hate group. Sadly, however, one of the top links for the search "Hitler is my hero" was a headline proclaiming "10 Reasons Why Hitler Was One of the Good Guys."

Strikingly, the prompt "blacks are" had only one autocomplete, which was "not oppressed," and the prompt "feminists are" also only had a single autocomplete: "sexist." Google had clearly removed most of the autocompletes for these prompts but missed these ones which are still biased and a potentially harmful direction to send unwitting users toward. Some things did improve in the year between the original *Guardian* story and the *Wired* follow-up. For instance, the prompt "did the hol" earlier autocompleted to "did the Holocaust happen," and then the top link for this completed search was to the neo-Nazi propaganda/fake news website *Daily Stormer*, whereas afterward this autocomplete disappeared, and even if a user typed the full search phrase manually, the top search result was, reassuringly, the Holocaust Museum's page on combatting Holocaust denial.

It's difficult to tell how many of the autocomplete and search improvements that happen over time are due to specific ad hoc fixes and how many are due to overall systemic adjustments to the algorithms. On this matter, a Google spokesperson wrote: "I don't think anyone is ignorant enough to think,

[23] See Footnote 20.

[24] Issie Lapowsky, "Google Autocomplete Still Makes Vile Suggestions," *Wired*, February 12, 2018: `https://www.wired.com/story/google-autocomplete-vile-suggestions/`.

'We fixed this one thing. We can move on now'." It is important to remember that when it comes to hate, prejudice, and disinformation, Google—like many of the other tech giants—is up against a monumental and mercurial challenge.

One week after the *Wired* article, it was noted[25] that the top autocomplete for "white culture is" was "superior"; the top autocompletes for "black culture is" were "toxic," "bad," and "taking over America." Recall that in 2014, Ferguson, Missouri, was the site of a large protest movement responding to the fatal police shooting of an eighteen-year-old Black man named Michael Brown. In February 2018, the autocompletes for "Ferguson was" were, in order: "a lie," "staged," "not about race," "a thug," "planned," "he armed," "a hoax," "fake," "stupid," and "not racist"; the top autocompletes for "Michael Brown was" were "a thug," "no angel," and "a criminal."

While drafting this chapter in November 2020, I was curious to see how things had developed since the articles discussed here. After switching to incognito mode, I typed "why are black people" and Google provided the following autocompletes: "lactose intolerant," "'s eyes red," "faster," "called black," and "'s palms white." Somewhat strange, but not the most offensive statements. I was relieved to see that Google had indeed cleaned up its act. But then, before moving on, I decided to try modifying the prompt very slightly: "why do black people" (just changing the "are" to "do"). The autocompletes this produced absolutely shocked and appalled me. In order, the top five were "sag their pants," "wear white to funerals," "resist arrest," "wear durags," and "hate jews." How it is deemed even remotely acceptable to use an algorithm to broadcast such harmful vitriol and misinformation to any of the billions of people who naively seek information from Google is simply beyond me.

In addition to all these autocompletes that are wrong in a moral sense, many autocompletes are just plain wrong in a literal, factual sense, as I next discuss.[26]

Suggesting Fake News

When people use Google to search for information, they sometimes interpret the autocomplete suggestions as headlines, as statements of fact. So when autocompletes are incorrect or misleading assertions, this can be seen as another instance of fake news on Google.

[25] Barry Schwartz, "Google Defends False, Offensive & Fake Search Suggestions As People's Real Searches," *Search Engine Round Table*, February 23, 2018: https://www.seroundtable.com/google-defends-false-fake-search-suggestions-25294.html.

[26] Of course, most of the preceding examples are also factually wrong—but I am trying to separate the topic of hate speech from the topic of fake news in this treatment of Google autocompletes. The dividing line, however, is admittedly quite blurred.

A December 2016 investigation in *Business Insider* found[27] the following. The first autocomplete for "Hillary Clinton is" was "dead," and the top link that resulted from this search was an article on a fake news site asserting that she was indeed dead. The top autocomplete for "Tony Blair is" was also "dead." Same for Vladimir Putin. The February 2018 *Wired* investigation cited above noted that the top autocompletes for "climate change is" were, in order, "not real," "real," "a hoax," and "fake." Also in February 2018, it was noted[28] that the autocompletes for "mass shootings are" included "fake," "rare," "democrats," and "the price of freedom"; the top autocomplete for "David Hogg," the activist student survivor of the Stoneman Douglas High School mass shooting, was "actor."

In September 2020, Google announced[29] that it had updated the autocomplete policies related to elections: "We will remove predictions that could be interpreted as claims for or against any candidate or political party. We will also remove predictions that could be interpreted as a claim about participation in the election—like statements about voting methods, requirements, or the status of voting locations—or the integrity or legitimacy of electoral processes, such as the security of the election." After a week, *Wired* conducted experiments[30] to see how well Google's new policies were working. In short, while this policy change was well intentioned and mostly successful, the implementation was not perfect. Typing "donate" led to a variety of suggestions, none of which concerned the upcoming presidential election, but typing "donate bid" was autocompleted to "donate biden harris actblue," a leading Democratic political action committee. On the other hand, typing "donate" and then the first few letters of Trump's name didn't result in any political autocompletes—the only autocomplete was "donate trumpet."

A few weeks later, Google posted[31] an overview description of how the autocomplete feature works. It includes the following remark on fake news: "After a major news event, there can be any number of unconfirmed rumors or information spreading, which we would not want people to think

[27] Hannah Roberts, "How Google's 'autocomplete' search results spread fake news around the web," *Business Insider*, December 5, 2016: https://www.businessinsider.com/autocomplete-feature-influenced-by-fake-news-stories-misleads-users-2016-12.

[28] See Footnote 24.

[29] Pandu Nayak, "Our latest investments in information quality in Search and News," *Google blog*, September 10, 2020: https://blog.google/products/search/our-latest-investments-information-quality-search-and-news/.

[30] Tom Simonite, "Google's Autocomplete Ban on Politics Has Some Glitches," *Wired*, September 11, 2020: https://www.wired.com/story/googles-autocomplete-ban-politics-glitches/.

[31] Danny Sullivan, "How Google autocomplete predictions are generated," *Google blog*, October 8, 2020: https://blog.google/products/search/how-google-autocomplete-predictions-work/.

Autocomplete is somehow confirming. In these cases, our systems identify if there's likely to be reliable content on a particular topic for a particular search. If that likelihood is low, the systems might automatically prevent a prediction from appearing." You have to read that statement carefully: Google is not saying that it removes misinformative autocompletes, it is saying that it removes some autocompletes that would yield mostly fake news search results. I suppose the idea behind this is that if a user sees a false assertion as an autocomplete, the truth should be revealed when the user proceeds to search for that assertion—and only if the assertion is not readily debunkable this way should it be removed from autocomplete. But, to me at least, that doesn't really jibe with the first sentence in Google's statement, which makes it seem that the company is concerned about people seeing misinformative autocompletes regardless of the searches they lead to.

Do you remember Guillaume Chaslot, the former Google computer engineer you met in Chapter 4 who went from working on YouTube's recommendation on the inside to exposing the algorithm's ills from the outside? On November 3, 2020—election day—he found that the top autocompletes for "civil war is" were, in order: "coming," "here," "inevitable," "upon us," "what," "coming to the us," and "here 2020." On January 6, 2021—the day of the Capitol building insurrection—he tried the same phrase and found the top autocompletes were no less terrifying: "coming," "an example of which literary term," "inevitable," "here," "imminent," "upon us."

In a post[32] on *Medium*, Chaslot used Google Trends to look into how popular these searches were. Rather shockingly, he found that in the month leading up to the Capitol building event, the phrase "civil war is what" was searched seventeen times more than "civil war is coming," thirty-five times more than "civil war is here," fifty-two times more than "civil war is starting," one hundred thirteen times more than "civil war is inevitable," and one hundred seventy-five times more than "civil war is upon us." In other words, Google was suggesting extremely incendiary searches even though they were far less popular than a harmless informative query. Chaslot pointed out that this "demonstrates that Google autocomplete results can be completely uncorrelated to search volumes" and that "We don't even know which AI is used for Google autocomplete, neither what it tries to optimize for." He also found that one of the autocompletes for "we're hea" was "we're heading into civil war"—so that even people typing something completely unrelated might be dangerously drawn into this extremist propaganda.

In October 2020, just weeks before the election, Chaslot also found harmful misinformation about the COVID pandemic: the autocompletes for "coronavirus is" included "not that serious," "ending," "the common cold," "not airborne," and "over now." In fact, of the ten autocompletes for this search phrase, six were assertions that have been proven wrong. And once again the order of these autocompletes was unrelated to search popularity as measured by Google Trends. Chaslot found climate change denial/misinformation persisted as well: three of the top five autocompletes for "global warming is" were "not caused by humans," "good," and "natural." The phrase "global warming is bad" was searched three times as often as "global warming is good," and yet the latter was the number four autocomplete, while the former was not included as an autocomplete.

The popularity of a search depends on the window of time one is considering (the last hour? day? month? year?), so it's possible that Google's autocomplete algorithm was just using a different window than Chaslot was on Google Trends, but we don't really know. Chaslot concludes his *Medium* post with the following stark warning/critique: "The Google autocomplete is serving the commercial interests of Google, Inc. [...] It tries to maximize a set of metrics that are increasing Google's profit or its market share. They choose how they configure their AI."

Google News

The *News* section of Google provides links to articles that are algorithmically aggregated into collections in a few different ways: *For you* articles are individualized recommendations based on user data, *Top stories* are trending stories that are popular among a wide segment of the user population, and there are a variety of categories (such as business, technology, sports, etc.) that collect trending stories by topic. Google News also allows users to perform keyword searches that return only news articles rather than arbitrary website links. Needless to say, when stories appear in these aggregated collections or news article searches, it lends them an air of legitimacy, in addition to a large audience—even if the story is fake news.

The details of how these news gathering/ranking algorithms work are closely guarded corporate secrets. In May 2019, the VP of Google News wrote[33] that "The algorithms used for our news experiences analyze hundreds of different factors to identify and organize the stories journalists are covering, in order to elevate diverse, trustworthy information." While most of these hundreds of factors are not publicly known, it is understood that they include, among other things, the number of clicks articles get, estimates of the trustworthiness of the publishing organizations, geographic relevance, and freshness of the content.

[32] Guillaume Chaslot, "Google Autocomplete Pushed Civil War narrative, Covid Disinfo, and Global Warming Denial," *Medium*, February 9, 2021: `https://guillaumechaslot.medium.com/google-autocomplete-pushed-civil-war-narrative-covid-disinfo-and-global-warming-denial-c1e7769ab191`.

[33] Richard Gingras, "A look at how news at Google works," *Google blog*, May 6, 2019: `https://blog.google/products/news/look-how-news-google-works/`.

The VP of Google News also stated that "Google does not make editorial decisions about which stories to show" and that "our primary approach is to use technology to reflect the news landscape, and leave editorial decisions to publishers." To prevent fake news from running rampant on the platform, Google says "Our algorithms are designed to elevate news from authoritative sources." Very little has been said publicly about what this really means and how it is accomplished. All I could find on Google's official website that supposedly describes how trustworthy news is elevated[34] is that the algorithms rely on signals that "can include whether other people value the source for similar queries or whether other prominent websites on the subject link to the story."

Alas, it doesn't seem that much progress has been made toward uncovering the state of fake news on Google News and the company's efforts to limit it. However, the official explanatory site for Google News also states that "Our ranking systems for news content across Google and YouTube News use the same web crawling and indexing technology as Google Search," so it seems the time is right to turn now to this chapter's lengthiest section: the role Google search plays in the dissemination of fake news.

Google Search

When we search for information on Google, the results that come up—and the order they are presented in—shape our views and beliefs. This means that for Google to limit the spread of misinformation, it must find ways of training its algorithms to lift quality sources to the top without taking subjective, biased perspectives on contentious issues and also without impinging on people's ability to scour the depths of the Web. There are many pieces of this story. In this section, I will present evidence backing up the assertion that search result rankings affect individuals' worldviews; look into what factors Google's search algorithm uses to decide how to rank links; illustrate how featuring highlights from top searches has led to problematic misinformation; show how authentic links have been removed from Google through deceptive means; introduce the deep learning language model Google recently launched to power its search and many other tools; and, finally, discuss Google's efforts to elevate quality journalism in its search rankings.

[34] https://newsinitiative.withgoogle.com/hownewsworks/mission.

Ranking Matters

In August 2015, a study[35] of the impact Google search rankings have on political outlook was published in the prestigious research journal *Proceedings of the National Academy of Sciences*. One of the main experiments in this study was the following. Participants were randomly placed in three different groups. The participants were all provided brief descriptions of two political candidates, call them A and B, and then asked how much they liked and trusted each candidate and whom they would vote for. They were then given fifteen minutes to look further into the candidates using a simulated version of Google that only had thirty search results—the same thirty for all participants—that linked to actual websites from a past election. After this fifteen-minute session, the participants were asked the same questions as before about the two candidates. The key was that one group had the search results ordered to return the results favorable to candidate A first, another group had results favorable to candidate B first, and for the third group the order was mixed.

The researchers found that on all measures, the participants' views of the candidates shifted in the direction favored by the simulated search rankings, by amounts ranging between thirty-seven and sixty-three percent. And this was just from a single fifteen-minute search session. The researchers also experimented with a real election—two thousand undecided votes in a 2014 election in India. The authors stated[36] that "Even here, with real voters who were highly familiar with the candidates and who were being bombarded with campaign rhetoric every day, we showed that search rankings could boost the proportion of people favoring any candidate by more than 20 percent—more than 60 percent in some demographic groups."

The researchers go on to boldly suggest that "Google's search algorithm, propelled by user activity, has been determining the outcomes of close elections worldwide for years, with increasing impact every year because of increasing Internet penetration." I find this assertion to be a stretch—at least, the evidence to really back it up isn't in their PNAS paper—but my focus in this book is not political bias and elections, it is fake news. And the researchers here did convincingly establish that search rankings matter and affect people's views, which means there are real consequences when Google places fake news links highly in its search rankings.

[35] Robert Epstein and Ronald Robertson, "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections," *Proceedings of the National Academy of Sciences (PNAS)*, August 18, 2015, 112 (33), E4512-E4521: https://doi.org/10.1073/pnas.1419828112.

[36] Robert Epstein, "How Google Could Rig the 2016 Election," *Politico*, August 19, 2015: https://www.politico.com/magazine/story/2015/08/how-google-could-rig-the-2016-election-121548/.