Distrust: Big data, data-torturing, and the assault on science
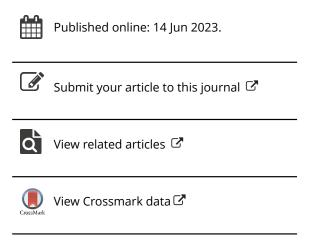
by Gary Smith, New York, NY, Oxford University Press, 2023, 323 pp., $29.98 (hardback), ISBN 9780192868459

Krishnendu Sarkar

Published online: 14 Jun 2023.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

BOOK REVIEW

Distrust: Big data, data-torturing, and the assault on science, by Gary Smith, New York, NY, Oxford University Press, 2023, 323 pp., $29.98 (hardback), ISBN 9780192868459

Distrust: Big Data, Data-Torturing, and the Assault on Science by Gary Smith leaves no doubt about society's growing distrust of science and it paves a way forward for restoring the credibility of the scientific community through pertinent anecdotes and arguments. Besides, it extends the horizon of Smith's previous book The AI Delusion by describing why human reasoning can be more trusted than the algorithms emanating from software black boxes that are more trained to understand correlation and not causality, let alone critical thinking. Furthermore, *Distrust* succinctly highlights the fear, uncertainty, and doubt that are impairing the trust in science with fake news, conspiracy theories, confirmation bias, and p-hacking, which primarily can be attributed to authors' publish-or-perish motivation for scientific publishing and scientific media's motivation for higher revenue in the post-truth world, which in the words of Mark Twain, Smith quotes "never let the truth get in the way of a good story" (p. 76). Arguably, the main message of *Distrust* to the readers is to think twice before they leap to embrace the overpromises of science and technology in their quest for universal and speedy solutions.

The book has 15 chapters categorized into five parts, namely, "Disinformation," "Data Torturing," "Data Mining," "The Real Promise" and "Peril of AI," and "The Crisis." Aside, a detailed introduction and extant Reference followed by a useful Index. Put together, it offers diverse cases ranging from COVID pandemic disinformation to bitcoin and demonstrates the danger of trusting ignorant and narrow algorithms like Zillow's house-flipping venture and impractical causal relationships of the stock price of Urban Tea with Donald Trump tweets. *Distrust* excavates such misrepresentations of the truth stemming from much-hyped big data and technology trends in pursuit of reposing faith in human expertise.

Chapter 1, "The Paranormal Is Normal," questions the claims made with extra sensory perception. For instance, reading another person's mind, Zener card guessing, identifying an unseen object, receiving something that has not occurred, and other supernatural or occult spiritualism, or fake realities like BBC's scripted and staged show *Ghostwatch*. Chapter 2, "Flying Saucers and Space Tourists," explains how pioneering products of science like the Internet and Social Media can later be used in certain adverse ways bringing discredit to scientists and science (p. 52). For example, the internet world's UFO saucer nets are basically Tully nests caused by whirlwinds. Chapter 3, "Elite Conspiracies," explains how ruthlessly a ruling class can use computer technology for cultivating conspiracies, as unstoppable viruses, to exercise control over their people. Further, how technologies like Pegasus smartphone spyware serve more the governors and not the governed. Furthermore, how animals and their robotic versions have also been a part of surveillance and harmful weaponry movements, which in a way does put science as a part of the problem? Chapter 4, "A Post-Fact World," deals mainly with motives of propaganda, the market of false news, and controlling the trend by controlling the narrative. In this regard, Smith refers to Islamic State of Iraq and Syria, as the first organization to weaponize Twitter (p. 89) and Russia's aggression in Ukraine in the garb of "peacekeeping" disinformation (p. 93). Chapter 5, "Squeezing Blood from Rocks," examines the formulated links to untoward incidents like COVID-19—its cure, and deaths. Actually what worked as the cure for COVID-19 was vaccines and masks and not Hydroxychloroquine, as was made out to be. There have been more than 200,000 publications on COVID-19 where the motivation was somehow to comply with the under-5% p-value statistical significance of empirical research for publication acceptance any which way, even negating the tests that did not work out or selectively deleting observations to reduce the p-value. Smith quotes Ronald Coarse: "If you torture data long enough, they will confess," and interestingly inserts a counter-quote from Stephen M. Stigler: "But confessions obtained under duress may not be admissible in the court of scientific opinion," (p. 111) echoing the essence of the chapter. Chapter 6,

"Most Medicines Disappoint," elaborates on the practical importance and efficacy of research results over the statistical significance and why many medically proven treatments flip/flop and fall short of expectations. As a result, many published papers in top-tier academic journals cannot be repeated with the same conclusions at industrial labs (p. 129). Moreover, about 50% of a workday is unreasonably spent interacting with computer screens instead of patents as per the data cited from the National Academy of Sciences. Chapter 7, "Provocative, but Wrong," brings back attention to p-hacking for fraudulently claiming fame, fortune, and funding with provocative findings that slip past even the editorial desks of top-rated journals. Smith points out that the highly recommended randomized controlled trials for empirical research sometimes can have false positives. Chapter 8, "Looking for Needles in Haystacks," highlights the fact that while algorithms are best at finding patterns and hypotheses from data, they still do not have an understanding of the sensitivity of statistical relationships (p. 157). On the contrary, humans have the ability to sense nonsense, a big reason why Smith subtly questions the blind trustworthiness of any cognitive computing systems with examples like Dr. Watson of IBM. Chapter 9, "Beat the Market," inquires how algorithms' insanity can be a human opportunity (p. 178) based on real instances of mindless data mining in the context of bitcoin trading and stock markets and predicting futures based on back-tested models. Chapter 10, "Too Much Data," argues that data mining is flawed, accentuated by the internet as it encourages people to data-mine the historical data and embrace false relationships. As a result, the number of coincidental patterns is growing exponentially and huge data systems like electronic health record systems cannot be trusted for accurate and relevant medical records and information about patients' illnesses. Chapter 11, "Overpromising and Underdelivering," leaves no doubt that computers are confined in Math World and never get to experience the real world (p. 207). Smith supports it with use cases about under-delivered promises ranging from Herbert Simon's in the late 1960s to Mark Zuckerberg's in the present times including other overpromises in areas such as medical image processing and large language models. Smith's moot point is that computers lack feelings, emotions, and problem sensitivity because they lack deductive and inductive reasoning for information ordering. Chapter 12, "Artificial Unintelligence," reiterates that algorithms are devoid of real-world sensitivities, which severely limits their intellectual capabilities to meaningfully deliver beyond-the-curve fit patterns. That is why, taking a cue from Warren Buffett, Smith warns not to trust AI algorithms that we do not understand (p. 227). Chapter 13, "Irreproducible Research," harps on the significance of reproducibility and replicability of research results for building public faith in the credibility of science, given the present situation where publication and citation counts are too easily manipulated resulting in bogus publications —the pizza papers, losada-line paper, Ivermectin as a COVID-19 cure are among the others exemplified. Chapter 14, "The Replication Crisis," draws attention to the publication bias that deviates from real science and real relationship for making useful predictions with new data. Smith cites that 90% of published medical research is flawed and unfit for replication with examples like the Zombie studies and Measles Mumps Rubella vaccine case that were later found with false positives. Chapter 15, "Restoring the Lustre of Science," concludes with recommendations that can not only restore trust in science but also help people to stay away from the pitfalls of disinformation, data torturing, and data mining through requisite literacy on fact-checking, recognizing statistical fallacies, and malicious bots.

In many ways, the book should empower 21st-century general readers and digital natives. More importantly, scientists and researchers would be encouraged to reflect and re-examine their past and present pursuits for societal, environmental, and economic development and to rightfully undertake course corrections in search of truth toward building a trustworthy future of science, by science, and for science.

Krishnendu Sarkar
NSHM Knowledge Campus, Kolkata, India
✉ krishnendu.sarkar@nshm.com 🆔 http://orcid.org/0000-0003-2151-0678

Check for updates