Generative Artificial Intelligence, Semantic Entropy, and the Big Sort

Hal Berghel, University of Nevada, Las Vegas

While artificial intelligence (AI) presents several threats, the threat generative AI poses is immediate and existential.

According to the *Washington Post*, on Wednesday, 13 September 2023, "more than 20 of the most prominent AI builders, researchers, civil rights advocates and labor leaders huddled with senators about future regulation of the technology."[1] As an aside, we note that this is akin to asking corporate executives in the energy and transportation sectors how the world should protect itself against global warming. In any event, given the list of attendees and the structure of the meeting, the event appeared to be more of an invitation to gaslight than a fact-finding mission: these hearings were closed-door, the interviews were scripted, and there was no possibility of public input. This suggests disingenuous and vacuous political theatrics.

According to the *Washington Post* account, there was "unanimous agreement that the government needs to intervene to avert the potential pitfalls of the evolving technology." AIChat, or generative artificial intelligence (AI), was particularly worrisome for the Senate. While I share the concern, the high-tech executives whose corporations contributed to the problem are unlikely to advance much of a solution. What is more, in my view, a viable solution would not result from further regulation and new and more powerful federal agencies. Governments have a tradition of legislating technology badly.[2] I shall argue below that the solution lies in returning to the principles of a diversified well-rounded education.

GENERATIVE AI FORENSICS

The current litter of generative AI tools have reached the apex of automated bloviation. As H. L. Menken put it, bloviation produces "a sort of discourse that is ... puerile and wind-blown gibberish" suitable for yokels.[3] With generative AI, it is now possible to create syntactically well-formed content that, while it looks and feels like the product of mature thought to an unprepared mind,

EDITOR **HAL BERGHEL**
University of Nevada, Las Vegas; hlb@computer.org

it actually involves negligible cognitive investment. But bloviation is typically betrayed by signatures that provide hints that all is not as it appears.[4] So, if we are to deal with generative AI's automated bloviation, the fundamental question must be, How would one prepare oneself to recognize these signatures? What skills must a prepared mind have to detect AI-generated nonsense and falsehoods? We might think of such critical analysis as "generative AI forensics," a phrase that is likely to become commonplace in the near future. From a bureaucratic perspective, a related question would be, What can government do to help its citizens protect themselves from generative AI disinformation, deception, and the subversion of truthful communication? We shall argue that the answer to both questions is the same and that both are questions that the Senate should have addressed if it had any hope of addressing this pressing social issue.

THE SEMANTIC ENTROPY PROBLEM

We introduce the genesis of the problem with the following mini-thought experiment. Consider a hypothetical all-inclusive digital media library—an unfiltered, digital, and networked implementation of Vanavar Bush's concept of the memex.[5] We emphasize the unfiltered aspect of the library to draw it closer to the modern Internet. All archives filter information—formally, informally, or accidentally. (In a sense, the difference between the deep web and the surface web is the de facto filter: lack of interest in indexing.) So, in our hypothetical ideal digital media library, we assume that all media ever produced and recorded, irrespective of utility and correctness, is digitized, stored, and indexed on the Internet. One might draw a comparison between our digital media library

and the union of the surface, deep, and dark webs.

Next, further assume that we somehow have the ability to bifurcate the content of the library into veridical and nonveridical counterparts: news separated from fake news, legitimate imagery contrasted with deep fakes, history distinguished from mythology, facts from opinions, and so on.

For our thought experiment to be useful, we only need to concede that such bifurcation is possible, that we could, in principle, accurately associate each datum with one of the two branches of the bifurcation. (We leave aside how one might acquire the ability to do this.) We simply observe that the fact that we lack a correct differentiating algorithm at this moment does not undermine the claim that legitimate distinctions can be made between legitimate and bogus content: the absence of a dichotomy does not imply the absence of a distinction.

So, there we have it. All nonverifiable opinions, questionable beliefs, lies, fabrications, and the like are aggregated into the nonveridical branch of the dataset, and all verifiable statements, mathematical truths in their axiomatic contexts, scientifically confirmed predictions with their associated probabilities, verifiable observations, and so on form the veridical branch. The datasets that make up our bifurcated database accord with our common understanding of vetted versus nonvetted information, knowledge versus unknowledge, theorems and nontheorems, and even truth versus falsity, based on a commonsense understanding of epistemology.

Now we add an automated rendering engine (a large language model neural net will do) that extracts and repurposes information from the database. In terms of extracting and repurposing data, three situations suggest themselves depending upon the mix of veridical and nonveridical content extracted from the database. First, the output could be derived from properly

> We might think of such critical analysis as "generative AI forensics," a phrase that is likely to become commonplace in the near future.

vetted and verified data. Second, the output could consist of nonsense, misrepresentations, and falsehoods with no legitimate vetting or verification. Finally, and most likely, the output could be a mixture of both. This leads us to back to our fundamental question: How would one determine the veridicality of the output after the fact without being able to trace the content back into the source branches of the dataset?

If we can agree, as it seems we must, that the output can only be as veridical as the input, then we must ask how we can be expected to assess the output without, at a minimum, specific details concerning the sources and methods of the extraction. If the sources are fully disclosed, known to be produced by scientists, scholars, and legitimate investigative journalists, we might reasonably be inclined to accept the output as reliable. But just as surely, we must be far more questioning about output from anonymous, unreliable, partisan, bogus, and biased sources. But how do we know which sources to associate with individual output fragments?

From a societal point of view, this last step is the most critical one in evaluating and using the output. History has shown that spontaneously generated nonsense

and disinformation have a definite, measured, and predictable influence on people with low or negative cognitive inertia. So, we must take these distinctions seriously in the assessment of value of our hypothetical digital media library.[6, 7, 8, 9, 10] Failure to do so in the past has given rise to the current hyperpartisan climate present on social media, within echo clouds and chambers, and propagated through anonymous weaponized fake news sources that drive bogus content. Any serious review of social media, online services, and talk radio

nonveridical data in our renderings, together with the vicious cycle, ensures that entropy cannot decrease over time in our automated rendering environment. Observation suggests that contemporary communication has become more partisan, biased, and less reliable than in the recent past.[12] Be that as it may, the information unreliability must necessarily increase over time as more inconsistencies, falsehoods, misrepresentations, and the like are fed into, or back into, the system. On the surface, this appears to be a corollary to Claude

data. Linguistically, for any natural language, the number of syntactically well-formed sentences must necessarily exceed the number of those that are meaningful (whether true or false), which in turn must be greater than those that are veridical. The fundamental problem of generative AI is that the algorithms are insensitive to this reality. Paradoxically, this also makes generative AI algorithms an ideal object of study because their domain of operation interweaves computability theory, information theory, formal logic, linguistics, and philosophy. But, and this is the source of the problem we're addressing here, it's also an ideal instrument to unleash unprecedented amounts of disinformation and nonsense on an unprepared audience.

> History has shown that spontaneously generated nonsense and disinformation have a definite, measured, and predictable influence on people with low or negative cognitive inertia.

regarding polarizing issues like white supremacy, homophobia, antimulticulturalism, identity movements, ethnonationalism, junk science, antisemitism, racism, and sundry denialist agendas will reveal the effectiveness of such nonsense and disinformation. Add to these polarizing issues the psychological reinforcing effect of confirmation bias, cognitive dissonance, and the like, and it is easy to understand why the effect of "disinformedia" is hyperpartisanship. As the Cambridge Analytica experience demonstrated, online trolling for easily manipulated social lurkers is a thriving cottage industry.[11]

JUST HOW FAKE IS YOUR NEWS?

To return to our minithought experiment, since the hypothetical output will in turn be fed back into our digital media system, our rendering engine will be caught up in a vicious cycle of regurgitation that will produce ever-increasing output indeterminacy. This "blender effect" is analogous to the principle of entropy as it relates to the second law of thermodynamics. If we assume that the balance between input types remains mixed, the admixture of veridical and

Shannon's notion of information entropy applied to the reliability of the rendered output. But in reality, the entropy inheres in the admixture of the very bifurcated data themselves rather than in any communication mechanism. We might refer to this progressive deficiency in our digital library as "semantic entropy." It represents a higher-order problem than communication entropy, for the most critical "errors" are already present in the data and only compounded through subsequent communication.

Thus, the problem of identifying semantic entropy in a digital media database is far more consequential than that of error correction in a communication system. To use Shannon's terms, the problem doesn't arise from noisy channels but, rather, noisy data—the data are unreliable ab initio. To put an even finer point on this, nonveridical and "bogus" information creates data clumps that are permanently irreconcilable with the veridical data: one can never infer "bogus" directly from the veridical, and vice versa, as there is no logical connection between the veridical and the specious. What is worse, the potential upper bound on the amount of bogus data that can be produced is vastly greater than verifiable

In short, we have described an ideal environment that is analogous to the use of large language model generative AI on data mined from the Internet. The Internet is, in fact, a bifurcated database in our sense. The proportions of the bifurcations are, for all intents and purposes, unknowable. Further, the neural net generative AI model is analogous to our automated rendering engine. There is no way of avoiding the problem of semantic entropy as we have described it. It would appear that an attempt to reclassify the output of our hypothetical database as veridical or nonveridical after the fact is as likely as Maxwell's demon is to violate the second law of thermodynamics. This is precisely the reason that the legislation of generative AI is misguided and for my lack of confidence that anything of enduring value would result from it. The public would have been better served if the Senate had increased the budget of the National Science Foundation for the study of generative AI threat vectors. But that approach would certainly have produced partisan controversy[13] and, more likely, more innocuous political theater.

Generative AI is an ideal breeding ground for digitally generating gossip, specious religious doctrine, conspiracy theories, pseudoscience, sorcery,

witchcraft, deceptions, hoaxes, bogus legends, rumors, occultism, and humbug. It is the perfect postmodern complement to social media: the ideal Pavlovian pap portal.

BEYOND SCIENCE AND MATH EDUCATION

Once again, the problem with large language AI models based on nonauthenticated sources is that there is no way to verify the massive volume of output in any way that even distantly resembles scholarly peer review. We emphasize that this is a higher-order problem than Shannon addressed in his work on information entropy. To be sure, the fungibility of truth, correctness, and reliability of information has been our constant companion in the historical record. But the volume and velocity of generative AI output creates an existential crisis. So, given the reality that generative AI isn't going away, and that we have no way to deal with its effects at the moment, we need to identify the best defensive tactic for warding off the disinformation.

That tactic is education—not science, technology, engineering, and mathematics (STEM) education but enhanced general education. Disinformation favors cognitive misers and unprepared minds. While all formal education is helpful, some categories of education will be more helpful than others in detecting disinformation.[14] We note, first of all, that not all content will be equally amenable to disinformation. Most prominent in the disinformation-averse category will be noncontroversial content. Those themes that will be most likely weaponized will continue to be controversial themes that appeal to nonreality-based communities, delusionists, demagogues, dictators, cultists, zealots, fraudsters, cheaters, narcissists, sociopaths, and the like. So, those educational programs that are likely to be the most effective in identifying the mischief will be those that deal with such controversial themes as a matter of practice. Clearly, if one wants to spot disinformation in a discussion of human phylogeny, one needs domain knowledge in biology, whereas a critical analysis of astrology will require domain knowledge in astrophysics. But these domains are far less likely to be magnets for disinformation than topics that deal with historical revisionism, affirmative action, group alienation, climate change, wealth redistribution, religious retrenchment, and the like. In terms of domain knowledge, topics within non-STEM disciplines are far more likely to be targeted for disinformation. Students that routinely study these topics will be better prepared to deal with it.

> As the Cambridge Analytica experience demonstrated, online trolling for easily manipulated social lurkers is a thriving cottage industry.

In addition to domain knowledge, spotting disinformation requires proficient reasoning—historically circumscribed in philosophy, logic, and mathematics. But when it comes to disinformation, we especially require a third component: an understanding of the use and misuse of language. While we might subsume this knowledge under applied linguistics or information or communication theory, I prefer to label it "disinformatics."[15] But whatever we call it, the topics studied must include such things as the study of linguistic framing, the use of propaganda, biased messaging, perception management, confirmation bias, cognitive dissonance, pseudoscience, and, most critically, how computing and networking technology serves as an enabling technology for the above. With the exception of disinformatics, the ideal curriculum I am describing for detecting disinformation is what was referred to as essentially a diversified well-rounded education in the 20th century.

So, there we have it: a strong general education, paying special attention to the humanities and social sciences and that emphasizes reasoning ability, together with the inclusion of the study of disinformatics, provides the best defensive strategy for dealing with the problem of semantic entropy in digital information systems. These strengths are the critical ingredients of an effective educational environment that will help prepare students to detect and mitigate the type and variety of disinformation that will result from generative AI: fake news, false-flag messaging, historical revisionism, gaslighting, slander, astroturfing, post-truth reasoning, deep fakes, echo chamber toxins, denialisms, junk science, historical negationism, obscurantism, and sundry other maladies of the human predicament. What could the Senate do to help? Increase the support to non-STEM education just as it has for STEM education. If we want to overcome the effects of increased hyperpartisanship that will result from the social abuse of generative AI, and make any progress toward defragmenting society, that's how to do it.

It is worth mentioning in this regard that the drift in educational mission from traditional to STEM-focused was never motivated by pedagogy. It drew support from a form of technology capitalism that subscribed to the belief that education should be more valued for its job training than as a general public good. This was the hidden fulcrum upon which one of the great hoaxes perpetrated on the public in the recent history of higher education was built: the STEM crisis myth.[16] These two forces—the ascending influence of technology capitalism in higher education and the constant messaging of varieties of the STEM crisis myth—are largely responsible for distracting the public away from the historical commitment to traditional education as a public good. STEM education is not ideally positioned to explain why fact checking is

not effective with tribalists and members of nonreality-based communities and why facts are less relevant than tribal instinct in motivating extremists, hyperpartisans, and information warriors. The best strategy for governments is to extend support of education beyond STEM and STEAM and to go directly to STEALM.

DISINFORMATICS AND THE BIG SORT

This curriculum I have in mind would look familiar to the baby boomer generation: it's essentially a reinvigorated liberal curriculum like that practiced in public schools in the 20th century but with a more measured less ethnocentric

articles on the topic (search terms: alternative facts, posttruth politics, nonreality-based communities, and so on). There are even new words to describe the phenomena: *agnotology* (the study of disinformation), *agnoiology* (the study of ignorance), *aningmology* (the study of creating doubt), and c*ognitronics* (the study of perception distortion), all of which are ultimately related to the more established term *kakistocracy* (government by the least qualified). I challenge anyone to show how a background in calculus, engineering, and physics will be of use in identifying and mitigating this disinformation.

We must agree that the underlying problems are not new: formalized webs

epistemology. The resulting balkanization of democracies that results from this and related tribalization amounts to what Bill Bishop calls the "big sort."[21] While the precise causes of such a big sort have been the subject of intense scholarship for much of the past century without definitive results, there is little disagreement that a critical component is message framing that both reinforces ideological allegiances and biases and accelerates the divisiveness. Enter generative AI: the automated low-cost feeder technology for the big sort.

To be sure, the fungibility of truth, correctness, and reliability of information has been our constant companion in the historical record.

bias and an emphasis on disinformatics. It's suitability to the task appears obvious when one considers the nature of the disinformation onslaught: the willing acceptance and endorsement of tribal epistemology by wannabe social influencers.[17] The underlying belief systems consist of two premises: 1) influencers are entitled to their own facts, and 2) lying is a First Amendment right. I'll illustrate with a few of the more outrageous recent examples widely covered by commercial media: 1) Rudy Giuliani's infamous remark concerning the Mueller investigation that "truth isn't truth" (https://www.youtube.com/watch?v=Drc74nEZ-vY), 2) Kellyanne Conway's defense of Sean Spicer's false claim of the size of the crowd at President Trump's inauguration by characterizing it as an "alternative fact,"[18] and 3) the claimed equivalence of teaching evolution in the schools with "government abuse" by the recently elected speaker of the House of Representatives.[19] This tribal epistemology is so widespread at this point that there are even Wikipedia

of deception were championed by Plato in his *Republic* (for example, noble lies) and received considerable discussion by social critics like Aldous Huxley and George Orwell; investigative journalists, such as Ida B. Wells, I. F. Stone, and George Seldes; and scores of social scientists and historians over the years. But what is new is the magnitude of the problem enabled by technology. Social media, online messaging, microtargeting, commoditization of personal data, and now generative AI have increased the efficiency of disinformation delivery to the point where every partisan, authoritarian, despot, crook, and troll has embedded disinformation strategies into their business plans. One may comfortably identify those future targets of generative AI. In fact, in 2011, media personality Rush Limbaugh did the heavy lifting.[20] He circumscribed what he called the "four corners of deceit" (read: targets) for his partisan followers as government, academia, science, and media. Limbaugh sought to discredit those specific sources that are the most likely to oppose his own brand of tribal

My discomfort with the Senate hearing is twofold: 1) disinformation sources, by their very nature, do not yield to legislation because they are typically anonymous, geographically opaque, and weaponized, and 2) the hearing distracted public attention from the more viable strategies that could deal with the manageable dimension of the threats, namely, education. Alternative perspectives of these issues are provided by Bruce Schneier and Nathan Sanders[22] and historian Sophia Rosenfeld[23] as well as many of the references listed below.

The way that society will handle nearly foolproof audio, video, and textual deepfakery; disinformation; conspiracy theories; and sundry other forms of skullduggery and deceit is the time-proven method of a liberal (small "l") education augmented with a study of disinformatics. The existential crisis that will result from generative AI is epistemological. Any approach that is epistemologically agnostic will fail.

I would be remiss if I failed to draw attention to the 1945 Harvard report, *General Education in a Free Society*,[24] which delineates the issues and challenges of a general of liberal education much as I have described it. ▣

REFERENCES

1. C. Zakrzewsky, C. Lima, and D. DiMolfetta, "Tech leaders including Musk, Zuckerberg call for government action on AI," *The Washington Post*,

Sep. 13, 2023H. [Online]. Available: https://www.washingtonpost.com/technology/2023/09/13/senate-ai-hearing-musk-zuckerburg-schumer/

2. H. Berghel, "Legislating technology (Badly)," *Computer*, vol. 48, no. 10, pp. 72–78, Oct. 2015, doi: 10.1109/MC.2015.304. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7310956

3. H. L. Mencken, "Gamalielese again," in *H.L. Mencken on Politics: A Carnival of Buncombe*, M. Moos, Ed., Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1956, [Online]. Available: https://www.google.com/books/edition/On_Politics/m3rDQMFrmZMC?hl=en&gbpv=1&pg=PA46&printsec=frontcover

4. H. Berghel, "ChatGPT and AIChat epistemology," *Computer*, vol. 56, no. 5, pp. 130–137, May 2023, doi: 10.1109/MC.2023.3252379. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10109291

5. V. Bush, "As we may think," *The Atlantic Monthly*, pp. 101–108, Jul. 1945. [Online]. Available: https://cdn.theatlantic.com/media/archives/1945/07/176-1/132407932.pdf

6. K. Andersen, *Fantasyland: How America Went Haywire: A 500-Year History*, reprint. New York, NY, USA: Random House, 2018.

7. H. Berghel, "Lies, damn lies, and fake news," *Computer*, vol. 50, no. 2, pp. 80–85, Feb. 2017, doi: 10.1109/MC.2017.56. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7842838

8. D. Davies, "Fake news expert on how false stories spread and why people believe them," *NPR*, Dec. 14, 2016. [Online]. Available: https://www.npr.org/2016/12/14/505547295/fake-news-expert-on-how-false-stories-spread-and-why-people-believe-them

9. G. Kessler, "The fact checker's guide for detecting fake news," *The Washington Post*, Nov. 22, 2016. [Online]. Available: https://www.washingtonpost.com/news/fact-checker/wp/2016/11/22/the-fact-checkers-guide-for-detecting-fake-news/

10. G. Lakoff, H. Dean, and D. Hazen, *Don't Think of an Elephant! Know Your Values and Frame the Debate: The Essential Guide for Progressives*. New York, NY, USA: Chelsea, 2004.

11. C. Wylie, *Mindf*ck: Cambridge Analytica and the Plot to Break America*. New York, NY, USA: Random House, 2019.

12. Y. Benkler, R. Faris, and H. Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford, U.K.: Oxford Univ. Press, 2018.

13. J. Mervis, "Controversy over Truthy illustrates the power of social media to inform—And mislead," *Sci. Mag.*, Nov. 3, 2016. [Online]. Available: https://www.science.org/content/article/controversy-over-truthy-illustrates-power-social-media-inform-and-mislead

14. H. Rheingold and A. Weeks, "Crap detection 101: How to find what you need to know, and how to decide if it's true," in *Net Smart: How to Thrive Online*. Cambridge, MA, USA: MIT Press, 2012, pp. 76–109.

15. H. Berghel, "Disinformatics: The discipline behind grand deceptions," *Computer*, vol. 51, no. 1, pp. 89–93, Jan. 2018, doi: 10.1109/MC.2018.1151023. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8268033

16. H. Berghel, "STEM crazy," *Computer*, vol. 48, no. 9, pp. 75–80, Sep. 2015, doi: 10.1109/MC.2015.256. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7274416

17. D. Roberts, "Donald Trump and the rise of tribal epistemology," *Vox*, May 19, 2017. [Online]. Available: https://www.vox.com/policy-and-politics/2017/3/22/14762030/donald-trump-tribal-epistemology

18. A. Blake, "Kellyanne Conway says Donald Trump's team has 'alternative facts.' Which pretty much says it all," *The Washington Post*, Jan. 22, 2017. [Online]. Available: https://www.washingtonpost.com/news/the-fix/wp/2017/01/22/kellyanne-conway-says-donald-trumps-team-has-alternate-facts-which-pretty-much-says-it-all/

19. M. Walters, "Teaching evolution is 'Government Abuse': New house speaker mike Johnson praises creationist museum, claims it guides people to 'The Truth'," *MSN.COM*, Oct. 31, 2023. [Online]. Available: https://www.msn.com/en-us/news/us/teaching-evolution-is-government-abuse-new-house-speaker-mike-johnson-praises-creationist-museum-claims-it-guides-people-to-the-truth/ss-AA1j9CD2?ocid=winp1taskbar&cvid=4cca7acb3b234307fe861f69edc413e9&ei=12

20. R. Limbaugh, "The four corners of deceit: Prominent liberal social psychologist made it all up," *The Rush Limbaugh Show*, Apr. 29, 2013. [Online]. Available: https://www.rushlimbaugh.com/daily/2013/04/29/the_four_corners_of_deceit_prominent_liberal_social_psychologist_made_it_all_up/

21. B. Bishop and R. Cushing, *The Big Sort: Why the Clustering of Like-Minded American is Tearing us Apart*. New York, NY, USA: Mariner Books, 2009.

22. B. Schneier and N. Sanders, "The A.I. wars have three factions, and they all crave power, guest essay/opinion," *New York Times*, Sep. 28, 2023. [Online]. Available: https://www.nytimes.com/2023/09/28/opinion/ai-safety-ethics-effective.html

23. S. Rosenfeld, *Democracy and Truth: A Short History*. Philadelphia, PA, USA: Univ. of Pennsylvania Press, 2018.

24. *General Education in a Free Society: A Report of the Harvard Committee*, Cambridge, MA, USA: Harvard Univ. Press, 1950. [Online]. Available: https://archive.org/details/generaleducation032440mbp/page/n5/mode/2up

HAL BERGHEL is a professor of computer science at the University of Nevada, Las Vegas, Las Vegas, NV 89154 USA. Contact him at hlb@computer.org.