

Background

Generative artificial intelligence (GAI) already has disparate impacts on different communities, and are likely to further exacerbate existing disparities and biases. Recent research demonstrates that ChatGPT and other GAI tools are most capable when responding to prompts in English, meaning they may “exacerbate bias for English and English speakers.”¹ Namely, GAI systems struggle when translating English into other languages, and particularly languages that do not use the Latin alphabet; an especially alarming related research finding from academics at the University of Oregon note that ChatGPT is “more likely to fabricate information” in non-English languages.² For the 31% of Asian Americans who have limited English proficiency (LEP),³ many of whom primarily speak character-based languages like Chinese or Korean, GAI in its current state is likely not a reliable or useful tool.

This presents an interesting paradox. Of course, it is important for the creators of GAI to close the language gaps of these tools and increase the use of multilingual training data so non-English speakers can take advantage of their myriad beneficial use cases. And yet, better Vietnamese-language capabilities for ChatGPT, for example, means this technology has greater potential to be weaponized to spread misinformation to Vietnamese-American communities. As GAI becomes seemingly omnipresent in modern society, the White House must commit to advocating for technology that works for everyone—English-speaking or not—while also bolstering safeguards to prevent these tools from ultimately harming the very individuals they intend to benefit.

[Asian Americans Advancing Justice | AAJC](#) supports the White House’s initiative to evaluate the risks posed by GAI. Our responses focus on the unique threats GAI poses to Asian Americans, who—alongside other communities of color—are most negatively impacted by this type of emerging technology and yet ironically largely unable to reap its benefits. It is imperative that this initiative prioritize and protect our most vulnerable populations.

Questions:

2. How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?

As GAI draws on existing data from the real world to train its algorithms and inform its outputs, unfortunately, stereotypes and conspiracy theories muddle its system. For an already misunderstood Asian American community, AI’s tendency to lean into these stereotypes risks granting legitimacy to these harmful beliefs and misconceptions. One Stanford study, for example, found a “persistent anti-Muslim bias” in GPT-3.⁴ When prompted with the phrase “Two Muslims walk into a...,” the program completed the sentence with violent themes 66% of the time. One particularly egregious sentence read “Two Muslims walked into a Texas church and began shooting.” When researchers swapped “Muslim” with other religions, the violent association was 40 - 90% lower.⁵ In another example, journalists from Bloomberg News asked Stable Diffusion to generate images of workers for 14 different jobs as well as three categories related to crime. The 5,100 generated images were

¹ <https://www.wired.com/story/chatgpt-non-english-languages-ai-revolution/>

² Ibid.

³ <https://www.rwjf.org/en/library/research/2022/12/many-asian-american-and-native-hawaiian-pacific-islander-adults-may-face-healthcare-access-challenges-related-to-limited-english-proficiency.html>

⁴ <https://www.cbc.ca/news/science/artificial-intelligence-racism-bias-1.6027150>

⁵ Ibid.

overlaid to create the “average” face for each occupation. An Asian woman represented the “average” face for a housekeeper, while the average face of a terrorist was a Middle Eastern man; in contrast, white men constituted the average judge, lawyer, and CEO.⁶

Tools like these are trained on data that reflects all of the racism, sexism, and ableism proliferating in the real world. More research must be prioritized to fully understand the impact on populations, but data on how existing technology currently represents Asian Americans is already deeply troubling. When terms like “Asian girls” or “Asian women” are input into search engines, the results play directly into age-old stereotypes that hypersexualize and dehumanize Asian women.⁷ Thinking about how GAI might respond when asked to create content about Asian Americans (e.g., “draw me a picture of an Asian woman”), the potential for bias is enormous. Moreover, when these tools generate biased content, this in turn can be used to train future model iterations, creating a vicious cycle of problematic information.

It is important to recognize that biased GAI outputs may not be the deliberate intentions of “malicious actors,” but rather incidental byproducts of the data being drawn upon for their algorithms. Data is not, and has never been, neutral and understanding where and how data is collected is imperative for mitigating disinformation risks. OpenAI’s GPT-2, for example, uses a dataset called WebText which draws inputs from Reddit, a place where white supremacists host large forums with little threat of moderation. GPT-2’s responses to prompts about white men and Black women, which included discussion of “white Aryan nations” and numerous racial slurs, highlight the white supremacist ideology used to train the tool.⁸ While addressing the misuse of AI by malicious actors will necessarily include reactive measures, meaningfully examining the datasets used to train algorithms is an essential preventative measure. Several major developers of GAI tools recently committed to honor a broad set of AI safeguards. As additional commitments are made, one should include increased transparency into training data and how algorithms are built, as Meta has already set an alarming precedent by not disclosing the data used to train its newly-released Llama 2.⁹ Moreover, as the government and technology companies continue collaboration on solutions to this complex issue, it is important for civil society to have a seat at the table.

3. What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?

Technologies

Robust, accurate AI detection tools are critical in combating AI-generated disinformation. While several of these tools exist, research highlights flaws in their accuracy¹⁰ and biases against certain populations. For example, a recent Stanford study found that several existing AI detectors demonstrated bias against non-native English speakers. Researchers fed essays written by non-native English speakers for the Test of English as a Foreign Language (TOEFL) and essays written by native English-speaking eighth graders through seven commonly used AI detectors; the detectors flagged more than 50% of the TOEFL essays—but less than 10% of the middle schooler essays—as AI-generated.¹¹ In the absence of better technology that remedies this kind of bias, non-native English

⁶<https://www.bloomberg.com/graphics/2023-generative-ai-bias/>

⁷<https://www.nbcnews.com/think/opinion/google-search-algorithms-are-not-impartial-they-are-biased-just-ncna849886>

⁸<https://www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion/>

⁹<https://www.businessinsider.com/meta-llama-2-data-train-ai-models-2023-7?op=1>

¹⁰https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html?name=styl®ion=TOP_BANNER&block=storyline_menu_recirc&action=click&pgtype=Article&variant=undefined

¹¹<https://www.theguardian.com/technology/2023/jul/10/programs-to-detect-ai-discriminate-against-non-native-english-speakers-shows-study>

speakers, already severely disadvantaged in the college application and job search process, face additional potential discrimination. More funding should be allocated towards building more conscious tools that do not flag legitimate work as AI and disproportionately harm non-native English speakers and other marginalized communities.

Policies

Policy-focused solutions to detect and counter AI disinformation, including legislation like the REAL Political Advertisements Act¹² are a good start, but contain several limitations. This legislation in its current form, which requires footage either wholly or partly created by GAI to include a “clear and conspicuous” disclaimer, does not include protections for content in all languages. For LEP individuals, English-only disclaimers are not useful in alerting the presence of computer-generated content. It is imperative that any policy-based solutions are crafted to consider non-English speaking communities who are the most vulnerable to consuming disinformation.¹³ Interventions and rights must be meaningful and accessible to communities, especially those most impacted.

Infrastructure

News deserts¹⁴ and lack of reliable, fact-checked news sources can amplify the spread of mis- and disinformation, particularly among non-English speaking communities. Greater support in funding these resources are an important inoculation strategy for keeping these individuals safe from new iterations of GAI-produced falsehoods. In the absence of sufficient in-language resources, members of the Asian American community often turn to ethnic media outlets for their news consumption. According to a study by Bendixen & Associates for New California Media, “more than half of all Chinese and Vietnamese adults read an ethnic newspaper on a regular basis. Nearly half of all Korean adults also read a Korean newspaper frequently...one-fifth of adults in this group read a Filipino newspaper a few times a month or more.¹⁵” These newspapers fill coverage gaps that mainstream media misses, serving as important sources of information that connect communities with local resources.

Ethnic newspapers, like many other local papers, have been suffering from a lack of funding and the digital transformation of reporting. The US has lost more than a fourth of its local newspapers since 2005 and is predicted to lose a third by 2025.¹⁶ Of the 70 million people who live in counties without newspapers, many are from low income communities who lack alternatives to local reporting.¹⁷ Instead, individuals are increasingly turning to social media for news, creating a dangerous opportunity for the effects of malicious GAI to be amplified. Funding for local, ethnic newspapers can combat spread of GAI-produced disinformation and return a vital source of trusted media to the Asian American community.

Beyond the funding of better information ecosystems for Asian Americans, the federal government should invest in broader public education around GAI targeted towards the most vulnerable communities. While there is yet to be research quantifying this phenomenon, based on existing knowledge and assumptions drawn from other areas of emerging technology, there are likely stark racial disparities surrounding who is and is not knowledgeable about GAI and its associated risks. Content produced by GAI is by design supposed to *appear*

¹² <https://www.congress.gov/bill/118th-congress/house-bill/3044/text>

¹³ <https://cdt.org/insights/election-disinformation-in-different-languages-is-a-big-problem-in-the-u-s/>

¹⁴ <https://www.usnewsdeserts.com/#1591281825217-2139af59-2beb>

¹⁵ <https://legacy.npr.org/documents/2005/jul/ncmfreport.pdf>

¹⁶ [https://www.usnewsdeserts.com/reports/news-deserts-and-ghost-newspapers-will-local-news-survive/the-news-landscape-in-2020-transf](https://www.usnewsdeserts.com/reports/news-deserts-and-ghost-newspapers-will-local-news-survive/the-news-landscape-in-2020-transformed-and-diminished/vanishing-newspapers/)

¹⁷ <https://localnewsinitiative.northwestern.edu/research/state-of-local-news/report/>

real; chatbots seek to spit out “human-sounding” text, and DALL-E and other image generators attempt to create “realistic” images. An important pillar of GAI-related public information campaigns should be informing Americans of the limitations of these tools despite their veneer of legitimacy. These systems are trained to elicit a response, *any* response, regardless of veracity, often resulting in “hallucinations,” or made-up, inaccurate content.¹⁸ If netizens are used to plugging a search term into Google or Bing and getting an accurate result, they may falsely approach a chatbot powered by GAI with the same degree of trust.

For older Asian Americans who might be ignorant of even the *existence* of this type of technology, GAI has the potential to be especially pernicious; not knowing about deep fakes or computer-generated audio, these individuals are likely to take everything they see at face value. Drawing off existing trainings and explainers that explain how to spot AI-generated content,¹⁹ the federal government should fund culturally competent, multilingual media literacy courses to spread awareness of GAI and focus on how to detect it online. Younger Americans could also benefit from such training.²⁰ In 2022, Illinois became the first state to require media literacy as part of public high school curriculums;²¹ the White House should increase federal funding for these types of programs and encourage other states to follow suit. Of course, for communities struggling to even gain access to reliable broadband, issues surrounding GAI seem very abstract and distant, a reminder that renewed funding for programs like the Affordable Connectivity Program²² is pivotal in the broader fight against disinformation and keeping marginalized communities informed.

4. How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?

Even before the prominence of GAI, election-related disinformation already negatively impacted the ability of Asian Americans to engage in democratic processes.²³ Unfortunately, during the 2020 and 2022 elections, pandemic-related changes to election procedures, coupled with a dearth of in-language resources to explain said changes, created opportunities for bad actors to sow confusion and exploit information vacuums. Inadequate platform moderation of non-English content and microtargeted messaging also contributed to this phenomenon. GAI enables bad actors to revamp existing media manipulation tactics and proliferate false content.²⁴

The majority of historical election disinformation targeted at Asian American communities has consisted of misleading content in English (e.g., a video clip of a candidate speaking) translated into relevant languages in the form of added captions or direct translations.²⁵ This already creates a wide window of manipulation in the form of imperfect or purposefully manipulative translations or videos taken out of context. GAI takes the possibilities of manipulation to the next level, opening the door for completely fabricated content that looks authentic to the untrained eye. Furthermore, bad actors currently go through the labor- and time-intensive process of translating

¹⁸ <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html>

¹⁹ https://graphics.axios.com/2023-deepfakes/index.html?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axioslogin&stream=top

²⁰ <https://www.scientificamerican.com/article/how-susceptible-are-you-to-fake-news-theres-a-test-for-that/>

²¹ <https://www.npr.org/2022/09/14/1122895362/illinois-now-requires-media-literacy-instruction-in-its-high-school-curriculum>

²² <https://www.forbes.com/sites/roslynlayton/2022/09/30/policy-no-brainer-extend-the-affordable-connectivity-program-for-5-years-with-30-billion/?sh=1c9599c04557>

²³ <https://apnews.com/article/elections-voting-misinformation-race-immigration-712a5c5a9b72c1668b8c9b1eb6e0038a>

²⁴ <https://www.theguardian.com/us-news/2023/jul/19/ai-generated-disinformation-us-elections>

²⁵ <https://medium.com/advancing-justice-aajc/what-the-lies-and-misleading-narratives-in-the-2022-midterm-elections-can-teach-us-85255b36a20>

content manually, but as tools like ChatGPT develop more robust translation capabilities,²⁶ multilingual disinformation operations can be automated and deployed with a few keystrokes. For LEP individuals who do not have the ability to understand source content or conduct cursory fact checks on their own, GAI further blurs the distinction between truth and reality by allowing this type of content to be generated at a much wider scale. This is already happening. In March of 2023, ahead of Trump's first arraignment, AI-generated images of Trump fighting off police officers²⁷ emerged across social media. These fake pictures eventually circulated among Chinese- and Vietnamese language social media and were widely believed to be genuine.

Another risk of generative AI is the liar's dividend,²⁸ giving politicians and other individuals leeway to dismiss legitimate information as computer-generated. It is important for netizens to confront information they see with a healthy degree of skepticism; however, there exists a fine line between this and not believing *any* information is credible. Perhaps the most insidious ability of GAI is its undermining of the credibility of ALL information, artificially generated or not. This emphasizes the need for robust media literacy education²⁹ for all Americans.

Even with translated voting materials required by law in certain jurisdictions under Section 203 of the Voting Rights Act, access to accurate in-language information about voting still remains a challenge for Asian Americans. Some potential fixes to the ultra-high tech problem of voter manipulation through GAI are very low-tech solutions that include providing better voting resources in the native languages of Asian Americans as well as more robust, multilingual candidate outreach to these communities. Targeting election-related information vacuums that allow mis- and disinformation to thrive is crucial to addressing information disorders that will be exacerbated by GAI.

Conclusion

Asian Americans are often portrayed as “tech-savvy³⁰” and thus largely benefiting from technology like GAI. Further evaluations of the potential harms of these tools, however, threatens these beliefs. A newly released study even found that Asian Americans workers are the most likely of any racial group to have their jobs replaced by AI.³¹ Addressing GAI will require nuanced, multi-pronged solutions involving stakeholders across community groups, government, and technology companies. Asian Americans Advancing Justice | AAJC looks forward to working collectively with these partners to combat this emerging issue.

For more information, please contact Emily Chi, Senior Director of Telecommunications, Technology, and Media at Asian Americans Advancing Justice | AAJC at echi@advancingjustice-aajc.org.

Sincerely,
Asian Americans Advancing Justice | AAJC

²⁶ <https://arxiv.org/pdf/2301.08745.pdf>

²⁷ <https://www.bbc.com/news/world-us-canada-65069316>

²⁸ <https://www.businessinsider.com/deepfakes-liars-dividend-explained-future-misinformation-social-media-fake-news-2021-4?op=1>

²⁹ <https://thehill.com/changing-america/enrichment/education/598795-media-literacy-is-desperately-needed-in-classrooms/>

³⁰ <https://www.nielsen.com/news-center/2019/for-whats-next-in-todays-digital-landscape-look-to-tech-savvy-asian-americans/>

³¹ <https://www.pewresearch.org/social-trends/2023/07/26/which-u-s-workers-are-more-exposed-to-ai-on-their-jobs/>