Generative AI, Deepfakes, & Disinformation: A Case for Fostering Civic-Minded AI Literacy

Cynthia Gerstl-Pepin

Abstract:

Social media and the internet have become significant sources for news and simultaneously effective venues through which to spread disinformation. The internet and social media play a key role in how citizens make decisions about voting. As generative AI increases the ability to easily replicate voices, images, and writing styles, it will be increasingly important to teach students and the public how to discern authentic and credible content from deepfakes and disinformation. This paper conceptualized a civic-minded AI literacy within the umbrella of media literacy education as a way to empower students to critically evaluate and navigate the increasing volumes of disinformation and fake news they encounter. It argues that media literacy standards and policies that incorporate AI literacy will be essential for preparing students to participate effectively as digital citizens. Civic-minded AI literacy as part of a robust media literacy education program across K-12 and higher education will be vital to support a healthy democracy.

Keywords: *Media Literacy Education, AI Literacy, Generative AI, Democracy and Education, Deepfakes, Digital Citizenship, Education Policy, Civic-Minded AI Literacy*

> *We are current and former employees at frontier AI companies, and we believe in the potential of AI technology to deliver unprecedented benefits to humanity. We also understand the serious risks posed by these technologies. These risks range from the further entrenchment of existing inequalities, to manipulation and misinformation, to the loss of control of autonomous AI systems potentially resulting in human extinction.* (Hilton et al., 2024)

The emergence of Generative Artificial Intelligence (GenAI) signaled alarm bells across the world as excitement about the new technology was tempered with concerns about what it might mean for society. Educators across K-12 and higher education grappled with potential issues such as AI generated plagiarism and how or whether to integrate the new technology into their teaching (Michel-Villarreal et al., 2023). Conversely, there are proponents who argue that GenAI holds much promise for personalized learning (Acar, 2024; Toner, 2024). But going beyond its potential utility for education, there are additional questions concerning its implications for democracy and existing inequities. The quote above from an open letter from AI experts alert us to some of the

more extreme dangers; highlighting the fact that confidentiality agreements (required by their employers) would restrict them from notifying the public about potential risks. It reflects emerging evidence that GenAI can be used to quickly spread and amplify disinformation and misinformation, influence politics, deepen political divides, reinforce social inequality, and undermine democracy (Bender et al., 2021; Kreps & Kriner, 2023; McQuade, 2024; Wack et al., 2025).

The pace at which digital technology has permeated society is breathtaking. According to the Pew Research Center (PRC) in 2000 approximately half of US adults used the internet compared to 96% who use it today (2024c). Approximately 91% of US adults now own smartphones with ownership being especially high 98% for 18-29 year olds (PRC, 2024d). The numbers are also striking for teens with approximately 95% having access to a smartphone (a 73% increase from 2014-15), 88% having access to a desktop/laptop, and 83% to a gaming console (PRC, 2025). These numbers show the extent to which digital technology has become embedded in our culture and the fabric of our democracy.

AIs, human created intelligent machines, have been around as early as the 1950s with the emergence of neural networks (Delipetrev et al., 2020; Ng et al., 2021). Contemporary AIs such as GenAI come out of deep learning–a type of machine learning that can "simulate the complex decision-making power of the human brain" (Holdsworth & Scapicchio, 2024). On a basic level, GenAI is a "machine-learning model that is trained to create new data, rather than making a prediction about a specific dataset" (Zewe, 2023). GenAI represents a shift in that it can quickly and autonomously generate images, text, video, and audio that appears to be authentic. Its ability to easily generate new realistic content is dangerous for democratic participation in that it can be used by human actors to sway political elections. For example, in the 2024 Presidential Primary, a deepfake of US President Joseph Biden's voice was used for robocalls discouraging New Hampshire residents from voting (FCC, 2024). Because the technology is so new, legislation and regulations in the US are not in place to ensure the democratic process is safeguarded.

Educating the public about GenAI is one way to counter the dangers of GenAI disinformation. In a democracy, public education has the potential to play a vital role in supporting an informed citizenry (Kahne & Bowyer, 2017; Gerstl-Pepin, 2007). Civic education (a non-partisan approach to teaching how democracy works and encouraging participation) has been seen as vital to a healthy and inclusive democracy in its potential to decrease polarization and increase broad participation (Fitzgerald et al., 2021). With the rise of GenAI's potential to influence political elections, media literacy education (a non-partisan understanding of how to assess media information credibility and accuracy) will also be vitally important. In particular, a subset of media literacy, digital literacy, which the National Association for Media Literacy (NAMLE) defines as the "knowledge, skills, and attitudes necessary to understand how digital tools interact with and impact society" (2024b, p. 5) has the potential to be helpful in countering the dangers of GenAI generated disinformation, propaganda, and misinformation (DiGiacomo et al., 2023; Kahne et al., 2016; Polizzi, 2020).

Finland has been at the forefront of teaching media literacy related to emerging technology (Salomaa & Palsa, 2019). The European Union has also enacted legislation citing the importance of media literacy but implementation across Europe is unclear (Sádaba & Salaverría, 2023). The United States and many other countries have not prioritized teaching about the dangers of these emerging technologies. Interest has been increasing in the US: Currently at least 21 states (less than half) have passed or considered bills related to media literacy education (Media Literacy Now, 2024b). However, any resulting curricular or pedagogical changes have not yet been fully assessed (DiGiacomo et al., 2023; Media Literacy Now, 2024a). Within the states that have enacted policy,

it is unclear the extent to which media literacy education is being implemented in schools and if it is being taught, how it is being taught, or whether it addresses GenAI. Research by DiGiacomo et al. (2023) indicates that media literacy policies across states are varied and inconsistent, often do not address equity issues, lack resources, training, and funding, and do not attend to "...the realities of what active citizenship looks like in today's digital age." (p. 348). Given that many states do not have legislation supporting media literacy and we lack substantive implementation information on the states that do (DiGiacomo et al., 2023), it is unclear the extent to which students in K-12 education in the U.S. are being taught media literacy. And, with the relative newness of AI, it is also unlikely widespread media literacy skills specific to AI are being taught consistently.

Considering AI threats to democracy and a lack of a coordinated effort nationally to teach media literacy, there is a growing need to educate students about GenAI and its potential role in spreading disinformation. Because the technology is so new, legislation and regulations in the US are not in place to ensure the democratic process is safeguarded. It is crucial for students and the wider public to understand how to critically evaluate information, especially in the digital age where sophisticated misleading content is easier to develop and disseminate at a rapid pace. Additionally, AI has also been shown to be biased and discriminatory across inequities such as race, gender, ability, and socioeconomic status (Bender et al., 2021; Curto et al., 2024; Ferrer et al., 2021; Lutz, 2019; Nadeem et al., 2020; Noble, 2018; Whittaker et al., 2019). AI can reflect societal divisions since it is trained on data society produces.

As GenAI tools like Claude, ChatGPT, Gemini, Llama, and Co-Pilot become more prevalent, there is a critical need for policies and guidelines that support a civic-minded AI literacy education. As GenAI's ability to replicate voices, images, and writing styles increases, it will be easier to create and spread deepfakes, fake news, and disinformation. For example, in 2025 AI generated voice and text messages impersonating US Secretary of State Marco Rubio were sent to a US governor and senator as well as at least three foreign ministers via the Signal messaging app (Shivaram, 2025). In another case, the BBC conducted a research study and found that news summaries generated by Microsoft's Co-Pilot, OpenAI's ChatGPT, Perplexity AI, and Google's Gemini AI chatbots "contained significant inaccuracies and distorted content from the BBC" (Elliott, 2025, para 7) raising concerns about AI spreading misinformation. With the rise of more complex and effective AI, citizens will need to be able to discern the credibility of digital information (Breakstone et al., 2019).

To address these issues, this article examines the potential for a nonpartisan civic-minded AI literacy education to offset the potential negative ramifications of GenAI for democracy. First, it conceptualizes how a civic-minded AI literacy can support participatory democracy. Then it examines how technological forms of disinformation, misinformation, and deepfakes threaten democracy. It then explores what a civic-minded AI literacy might encompass. The final section examines promising initiatives at both the K-12 and Higher Education levels and includes suggestions for future research and further development of civic-minded AI literacy.

Conceptualizing AI Literacy as Supporting Democracy

An informed citizenry can be seen as critical to a participatory democracy. But how does a citizenry become informed about politics and societal issues? Historically Americans were more civically engaged in their local communities via political parties, civic associations (e.g. clubrooms, committee tables), unions, churches, bowling alleys, card tables, and dinner tables (Putnam, 2020). National, international, and local news media outlets have also played a vital role

in sharing information on political and social issues. As the internet and digital forms of media have become more prominent in society, local news media have struggled to survive financially (Toff & Mathews, 2021). In-person forms of political participation have also shifted in favor of more digital forms such as social media and more global/national news outlets (Boulianne, 2020). The internet and social media now play a key role in how citizens make decisions about voting (Zhuravskaya et al., 2020).

Habermas (1991) conceptualized the public sphere as an ideal mediating space within democratic societies where citizens, the state (government), and the economy could come together as a public to engage in discourse about societal issues. The news media have long been an important source for political information and have the potential to serve a public sphere function (Garnham, 1994; Gerstl-Pepin, 2007). However, the public sphere is not immune to power inequities in society such as racism, poverty, ableism, and heteronormativity, which make it difficult for genuine dialogue to occur without inequities (Gerstl-Pepin, 2015). For example, Fraser (1994) offers the Civil Rights Movement as an example of how communication can be limited in the public sphere. It was not until violence against protesters was shown on television that citizens and politicians were persuaded to act (Bodroghkozy, 2012). So, while the news media space is limited, it is an important space for citizens to gain information about current issues.

Where citizens get their news has changed dramatically over the past decade. As of 2024, approximately 86% of U.S. adults sometimes or often consume news via a tablet, smartphone, or computer rather than traditional paper versions, with 57% saying they do it often (PRC, 2024a). In addition to traditional news media consumption happening electronically, an increasing number of adults (54%) are getting their news on social media (PRC, 2024b). These new arenas for news often lack the professional norms for fact checking and credibility of traditional, non-partisan news reporting. As a result, social media is becoming a central space for the spread of AI powered misinformation, disinformation and propaganda (Surjatmodjo et al., 2024).

At the same time, beyond their role as news sources, social media platforms lack guardrails concerning disinformation, propaganda, and fake news which make it increasingly important for digital news media consumers to identify disinformation and understand GenAI. In the US, Section 230 of the 1996 Communications Decency Act protects social media companies as forums for free speech, thus they cannot be held liable for content posted on their platforms (Moss, 2023; U.S. Department of Justice, 2020). In contrast, news organizations can be held liable for the propagation of misinformation. For example, Dominion Voting Systems was able to sue Fox News for defamation and reached a financial settlement for Fox's promotion of false conspiracy theories that Dominion rigged its voting equipment in the 2020 election in favor of Joe Biden (Bauder et al., 2023).

Social media platforms provide opportunity for communication across users via sharing, liking, or private messaging, and have the potential to be a democratizing source in society. For example, social media provided communication avenues for political movements in the 2011 Arab Spring (Frangonikolopoulos & Chapsos, 2012). However, invisible algorithms shape the content a viewer sees. This can limit the opportunity for users to see content that is different from their own political beliefs. Social media news feeds can serve as echo chambers in which the user's beliefs are reinforced (Cinelli et al., 2021). They can also serve as vehicles for spreading disinformation and limiting discussions and interactions across party lines (Cinelli et al., 2021).

Given these limitations, this paper conceptualizes these digital media forms as "thin" public spheres where communication and dialogue between citizens is restricted. They limit opportunity for discussion about issues with individuals across the political spectrum, social inequities

and power differentials can limit dialogue (Gerstl-Pepin, 2015). Within this context, a civic-minded AI literacy would acknowledge the ways in which power inequities may be embedded or operate in the digital world. So, while digital media plays an integral role in the US in terms of how citizens become informed about political issues, it is also a limited space in which power inequities may shape the news users select and receive.

How Disinformation, Deepfakes, & Misinformation Threaten Democracy

Propaganda is defined as "information, especially of a biased or misleading nature, used to promote or publicize a particular political cause or point of view" (Oxford English Dictionary, 2007). Disinformation and misinformation are key components of propaganda. Disinformation is when someone knowingly spreads false information while misinformation is when someone unknowingly spreads false information (McQuade, 2024). While propaganda has been around at least since the Ancient Greeks (Casey, 1944a), modern strategies emerged in World War 1 and World War 2. The US had an Office of Propaganda that sought to unify the country in support of war efforts using strategies such as the image of Rosie the Riveter to encourage women to work in traditionally male fields (National Archives, n.d.). The Nazis utilized the swastika as a symbol of racial purity to unify the Third Reich and support its genocide of over 6 million European Jews (O'Shaughnessy, 2009). The symbol was included on an array of items including clothing, flags, movies, and posters. In another example, the Nazis used the term "warmongers" to refer to the British Prime Minister Winston Churchill and US President Theodore Roosevelt even though it was Hitler who invaded Poland and set off WW2 (Casey, 1944b).

With the rise of the digital age, the potential for effective propaganda and citizen surveillance has greatly expanded to include an array of strategies including behavioral data collected on citizens through web tracking (data on individual internet searches), location information (e.g. where individuals are located when searching the internet), GPS tracking on smartphones or smartwatches, cross-device tracking (e.g. data across tablets, cell phones, watches), search engine optimization (e.g. digital strategies to increase internet traffic to a specific webpage or website), and digital ad buys based on behavioral data (Ghosh & Scott, 2018). In addition to the privacy issues it raises, this data collection allows for hyper-specific, targeted information campaigns. Strategies can now be coordinated and handled even more effectively and at a lower cost by GenAI.

The rise of digital news technology also coincided with shifts in the regulation of news media. The Fairness Doctrine policy of 1954 arose out of legislative concerns that broadcast networks could be used to support a specific political agenda and required that broadcast networks cover "contrasting views on issues of public importance" in news stories (National Archives, 2024). In 1987, the Federal Communications Commission (FCC) repealed the Fairness Doctrine over concerns that it limited free speech (Smith, 1999). This opened the door to partisan television and radio news coverage and ultimately to partisan news coverage such as the conservative leaning Fox News and the liberal-leaning MSNBC. Broadcast media was no longer legally obligated to provide "contrasting views" concerning issues of public interest in a democracy, leading to the further emergence in recent years of an array of partisan broadcast news outlets (Kawakami et al., 2020). Citizens have multiple ways to receive news and can select a partisan outlet if they so choose.

At the same time, the rise of the internet and digital media directly cut into the financial viability of paper-printed local and regional news sources (Hayes & Lawless, 2021). As digital online forms of traditional news media emerged, social media became a new source of news. As

noted earlier, about half of all adults reported getting some news from social media and of those, the most utilized platforms for news in 2024 were TikTok with 33%, YouTube with 32%, and Instagram with 20% (PRC, 2024b). In comparison to traditional news media sources, 'algorithmic personalization' shapes individual social media news feeds and has been shown to reinforce political polarization as it provides partisan content based on the personal data the platform collects (Chitra & Musco, 2020; Hobbs, 2020). Within this context news consumption has become both more polarized through the turn to partisan consumption by citizens, and less reliable due to the increasing popularity of nontraditional media that do not adhere to professional journalistic norms.

The 2016 US Presidential election signaled a significant shift in how social media and digital media could be used by actors external to the U.S. (foreign governments) and internally (citizens and organizations) to influence elections by spreading disinformation and increasing political polarization. For example, Russia's Internet Research Agency used trolls and bots on X (formerly Twitter) and YouTube to amplify polarizing viewpoints in support of Republican candidates (Badawy et al., 2018; Badawy et al., 2019; Golovchenko et al., 2020). Trolls are individuals who create "malicious accounts for the purpose of manipulation" (Badawy et al., 2018, p. 258) and bots (short for software robots) are automated software applications that can create content and interact "...with humans on social media, trying to emulate and possibly alter their behavior" (Ferrara et al., 2016, p. 96). During the 2016 US elections, social media platforms faced bipartisan criticism concerning bias, censorship, and misinformation as it became clear the platform's algorithms could function as "gatekeepers" of the news users would see online, fostering polarization (Calice et al., 2023).

In addition to Russian digital propaganda and disinformation strategies, the 2016 Presidential campaign signaled a shift towards a political propaganda strategy where facts reported by traditional news media were claimed to be false. The Trump presidential campaign used the disinformation strategy of presenting "alternative facts" and labeling unflattering accounts of the campaign as "fake news" to undermine and counter factual data presented by traditional news media sources (Barrera et al., 2020). "Alternative facts" is a disinformation strategy where false information is labeled true and facts are labeled as falsehoods. "Fake news" is a term used by politicians where traditional verified news stories are labeled as false as a way to undermine traditional news media reporting (Aïmeur, et al., 2023; Allcott & Gentzkow, 2017; Lee & Hosam, 2020). Additionally, due to algorithmic news curation which seeks to amplify sensational stories, these two strategies can be spread widely across partisan media platforms and social media (Calice et al., 2023). An additional challenge relates to confirmation bias, which is the tendency for individuals to believe alternative facts or fake news when it fits their own belief system (Ling, 2020). Social media and partisan news sources could then be used to spread disinformation and misinformation under the guise of free speech. This strategy also spread disinformation that mainstream media as a news source could not be trusted (Martens et al., 2018). "Fake news" as a strategy undermines factual information across the political spectrum and thus, democracy which relies on the public's trust in credible information and institutions, such as news media, to function.

The Rise of GenAI & Deepfakes

The introduction of GenAI has helped to supercharge the spread of propaganda and disinformation (Guo, 2024; Ryan-Mosley, 2023). GenAI is accessible and can be used to easily create more realistic forms of disinformation, such as deepfakes, which refer to counterfeit still images, audio recordings, or videos created with the assistance of AI (Ryan-Mosley, 2023; Chesney &

Citron, 2019; Painter, 2023). Deepfakes can be spread quickly through social media via resharing and algorithm amplification, where individuals can inadvertently spread disinformation that they think is true. Deepfakes are particularly concerning for social media. As noted earlier, due to Section 230, social media companies such as Facebook and X (formerly Twitter) are currently not liable for spreading disinformation and misinformation that may be posted by users or advertisers on their platforms (O'Donnell, 2021).

Deepfakes on social media could be used to weaken election integrity in the US and globally in multiple ways including suppressing voter turnout, misrepresenting candidates in multiple ways including their platforms, health, behavior, and background (Ranka et al., 2024). In response to the deepfake robocalls of Joe Biden in the 2024 New Hampshire primary (Bond, 2024; Swenson & Weissert, 2024), the FCC (2024) made fake AI-generated political robocalls illegal. Still, AI-generated deepfakes quickly spread on social media and can cause damage before they are recognized or halted. In another example, in the 2024 Presidential election, a Russian propaganda deepfake video smearing democratic Vice-Presidential candidate Tim Walz, made false claims that he was inappropriate with a student, and was viewed 5 million times on X (formerly Twitter) within 24 hours of its posting (Menn & Nakamura, 2024; Nazzaro, 2024). It is unclear what impact this propaganda may have had on the election outcome, but what is becoming clearer is that AI generated propaganda can be persuasive (Goldstein et al., 2024).

To call attention to the dangers of GenAI, an anonymous individual using the name NeaPaw created CounterCloud, an autonomous, uncensored disinformation system utilizing OpenAI's GenAI (the same technology utilized by ChatGPT) to highlight the potential dangers of GenAI (Knight, 2023; Thompson, 2023). NeaPaw claimed the system costs $400 per month to operate 24 hours a day, 7 days a week. NeaPaw (2023) posted a YouTube video explaining CounterCloud's purpose (to counter disinformation) by creating deepfake news articles and comments to respond to Russian disinformation about the US and the Biden Administration. CounterCloud's tactics included challenging the inaccurate facts of Russian fabricated disinformation articles, creating news articles that counter disinformation by fabricating fake historical events, fake journalists with names, bios, and photos, and fake article comments. The video states that all the disinformation produced was password protected and not released publicly (NeaPaw, 2023). The CounterCloud example has to be examined carefully, given it was produced by an anonymous source. Since the autonomous disinformation system is not open for public scrutiny, the story cannot be assumed to be factual despite *Wired* magazine's (Knight, 2023) and the Business Insider's (Thompson, 2023) coverage. Still, it suggests that knowledgeable AI developers could use open source GenAI technology to create effective disinformation systems and deepfakes at scale, potentially reaching a larger number of citizens far more quickly than human generated misinformation.

AI Tech Companies Response to Deepfake Concerns

Major technology companies have been accused of disregarding safety and disinformation. Frances Haugen, a whistleblower who used to work at Facebook, released internal company documents showing that Facebook prioritized profits over the health and safety of its users (Haugen, 2021). Although social media and AI are mostly unregulated in the US, companies feeling the pressure from congressional investigations (e.g. Disinformation Nation…, 2021; Examining Social Media Companies'..., 2019) and pressure from policymakers have made efforts in the past to address the issue of transparency and accountability by moderating content internally utilizing

human oversight and technology (Nonnecke & Carlton, 2022). These have not always been effective particularly with how quickly disinformation can spread online (Gongane et al., 2022). Additionally, there is concern that moderation of social media could be seen as censorship of free speech (Gorwa et al., 2020; West, 2018). Amid rising criticism and concern related to election integrity, in February of 2024, 27 tech companies including Meta, Google, OpenAI, TikTok, Anthropic, and X, among others signed the AI Elections Accord (2024) to protect the 2024 elections. The signers had no legal requirement to comply with what they signed. Instead, they committed to operating under 7 principal goals addressing deceptive AI election content related to 1) preventing deceptive content generation, 2) creating provenance markers to identify deceptive content, 3) detecting deceptive content, 4) quick "responsive protection" when deception is uncovered, 5) evaluation of deceptive incidents for further prevention, 6) collaborating to increase public awareness of deception via media literacy, and 7) building resilience among citizens against "Deceptive AI election Content" (AI Elections Accord, 2024, para 13).

In September of 2024 14 of the 27 companies provided updates as to steps they took to protect the 2024 elections. For example, Microsoft (2024) reported that they created a public awareness campaign reaching 350 million people globally and about 30,000 people in the US. Their 'Check. Recheck. Vote.' campaign is accessible online and is very basic. It defines the term deepfake, how it might be used, and offers an online quiz. The campaign has 5 strategies for citizens to use to combat deepfakes in elections: 1) critically consume information by checking and rechecking sources and ensure it comes from a trusted news source, 2) before sharing or commenting on sensational news check its accuracy, 3) if you suspect disinformation or deepfakes report it via reporting channels, 4) register to vote and double check what you need in order to vote in your state, 5) keep up on new technologies used in the spread of deepfakes and disinformation (Microsoft, 2024). The campaign also suggests that readers engage with the News Literacy Project (2024) to become more educated to stay abreast of changes and for further information on political disinformation. Overall, the program addresses key considerations related to a civic-minded AI literacy including utilizing critical thinking skills, focusing on differentiating credible information from disinformation, not taking sensational news stories at face value, actively participating in voting, and staying up on how AI evolves. Unfortunately, the reported reach of Microsoft's campaign in the U.S. suggests very little impact.

Eight of the companies that provided updates on their work (Adobe, ElevenLabs, LinkedIn, Meta, Nota, OpenAI, TikTok, and Truepic) cited their involvement in the work of the Coalition for Content Provenance and Authenticity (C2PA). C2PA is a collaborative with the goal of creating provenance markers in digital media to assist with the identification of GenAI created media and may help in the future identification of deepfakes. Google also reported working on its own provenance system and McAfee shared that it created its own deepfake detection system. Thirteen of the companies had not reported any actions they may have engaged in. While it is a good sign that 27 social media platforms made this commitment, the accord only covered the 2024 election and only 14 companies provided updates on steps they took. In 2025, the website was deleted, and the materials are no longer available. It is unclear whether the companies will agree to protect future elections, and it is also unclear if the commitment to securing the elections made a difference. Since these companies are opaque in terms of their internal workings it is unclear the extent to which they have publicized these tools or if they have been utilized. Tools such as C2PA and deepfake detectors do hold promise for helping identify deepfakes if they are effective and widely used.

The AI Tech Accord highlights awareness on the part of Tech companies that their platforms can support the spread of AI generated misinformation. However, the relative lack of promotion of Microsoft's 'Check, Recheck, Vote' program and the lack of public awareness or use of provenance markers, suggests these actions are not enough. And after the election, social media companies seemed to move even further away from the principles of the AI Tech Accord. In January 2025, Meta announced that to support free speech, it was ending third-party fact checking and instead instituting a community notes approach (similar to how X operates) where users are responsible for reporting concerns (Kaplan, 2025). This only underscores the critical importance of developing the ability of users to be able to distinguish fact from fiction.

This shift, along with other warning signs such as the AI Right to Warn letter (Hilton et al., 2024), and Facebook documents released by whistleblower Frances Haugen (2021) indicate that we cannot rely on profit-driven Tech companies to self-regulate. Guardrails in terms of legislation and/or regulation can keep companies in check, and media literacy education for citizens, our K-12 schools and higher education provide a way for citizens to build resilience against propaganda. The federal level has resisted regulating AI deepfakes, apart from the Take it Down Act (2025) which criminalizes the digital publication of nonconsensual intimate images including deepfakes. In this vacuum state legislatures have been active particularly in 2024 and 2025. As of July 2025, 26 states legally restrict the use of AI deepfakes in political campaigns, 24 of those restrict AI generated political advertising without a disclosure of AI use, while the other two states, Texas and Minnesota, ban the use of deepfake ads (National Conference of State Legislatures, 2025). In the absence of federal guidance, a hodgepodge of legislation is emerging across states. This leaves citizens in 24 states with no protection concerning the use of AI deepfakes in political campaigns and these laws have yet to be tested.

Despite efforts in some states, legislation has not caught up to technological innovations, and media literacy education is also uneven and uncertain. This makes the need for a civic-minded AI literacy especially crucial for supporting democracy.

Towards a Civic-Minded AI Literacy

> *Media literacy should not be limited to learning about tools and technologies, but should aim to equip citizens with the critical thinking skills required to exercise judgment, analyse [sic] complex realities and recognise[sic] the difference between opinion and fact. (Directive (EU) 2018/1808, 2018, § 59)*

Often missing from conversations about new emerging technology in the US is the critical role of media literacy education can play in supporting civic participation in our digital democracy. As the quote above shows, this has been a concern in the European Union (EU). With the rise of the digital age and the spread of disinformation and propaganda, media literacy is gaining the attention of policymakers and citizens in the US, but definitions of media literacy and its subset digital literacy are not uniform.

This paper utilizes the National Association for Media Literacy Education (NAMLE)—a national leader in media literacy education—definition of media literacy as a guide for AI literacy specifically. NAMLE (2024a) defined media as "...all electronic or digital means and print or artistic visuals used to transmit messages" and literacy as "…the ability to encode and decode symbols and to synthesize and analyze messages" (para. 3). Media literacy then is "the ability to access,

analyze, evaluate, create, and act using all forms of communication" (NAMLE, 2024a, para. 3). NAMLE's (2024b) conception of media literacy education includes four central components: 1) critical media literacy (an inquiry-based, critical thinking approach to acknowledging how power, systems, structures, representations, and ideologies may be embedded in media), 2) information literacy (finding, using, evaluating, analyzing, and reflecting on information), 3) digital literacy (understanding the interaction and impact of digital tools in society), and 4) news literacy (ascertaining the credibility of news sources including journalistic ethics and standards).

Media literacy education seeks to foster habits of inquiry and critical thinking skills that are vital to civic life. To support the democratic process media literacy must be nonpartisan. The goal is to promote critical thinking skills, not to tell students what to think.

Where Does AI Literacy Fit?

AI literacy emerged from the discipline of computer science, but it is often defined separately from media literacy education (see for example, Long & Magerko, 2020; Ng et al., 2021). AI literacy in this vein is more focused on defining AI, how it operates and how to identify it, create it, and utilize it (Laupichler et al., 2022). For example, the National Artificial Intelligence Advisory Committee's (2023) *Recommendations for Enhancing AI Literacy for the United States of America* do not mention media literacy education or connect it to civic responsibility although they do mention AI's role in disinformation and misinformation. Understanding how AI basically functions is especially important since the creators of AI do not always understand fully how it operates. However, adding the dimension of media literacy education expands AI literacy to support digital citizenship. This reconceptualizes AI literacy as an extension of digital literacy, a subset of media literacy education. A civic-minded AI literacy would then operate within the media literacy education umbrella (Kahne et al., 2016; Martens & Hobbs, 2015; Middaugh et al., 2022; NAMLE, 2023) and attend to the unique ways in which AI can be used to create and spread disinformation and misinformation.

This would extend the legacy of early propaganda education which arose out of an awareness of how the media can be manipulated to persuade its audience in one political direction or another (Hobbs & McGee, 2014). The rise of digital technology and the creation of social media and other digital forms of communication have opened new forms of communication and information sharing. These often segregate individuals, algorithmically pushing them into dialogue with mainly like-minded individuals. Algorithms may amplify false, sensational information because it attracts more attention, without journalistic standards for fact-checking. With such a large technological shift in how news is disseminated and how citizens make decisions in elections, a civic-minded AI literacy has the potential to offer a way to counter AI-generated disinformation, deepfakes, and fake news. Central to this approach would be critical thinking skills that allow individuals to discern facts and trustworthy information, and to be able to judge what is a fact versus an opinion. It should also focus on contributing to the public good.

Detecting deepfakes offers an example of a potential component of civic-minded AI literacy based on critical thinking. Deepfake detection can be challenging given how authentic they can seem, all the more as technology advances. There are at least three strategies for detecting AI generated deepfakes: 1) using AI detection technology, 2) adding provenance markers to digital content, and 3) using critical thinking strategies to assess content. All of these require human oversight to implement. As noted earlier some tech companies are developing AI detection technology

and working collaboratively to develop provenance markers such as C2PA (2024). Another strategy would educate people on how to detect politically motivated deepfake video, transcripts, and audio (Groh et al., 2024).

Appel & Prietzel (2022) proposed a theory of human-centered deepfake detection utilizing a reasoning model that posits that developing analytical skills (e.g. critical thinking) and an interest in politics can aid in identifying deepfakes. They suggest that the same cognitive processes individuals use to spot fake news stories could be used to spot deepfakes. Their theory's components include: 1) looking for audiovisual imperfections (technical flaws), 2) examining the context (e.g. is the source credible and verified by more than one reliable source), and 3) evaluating the content (e.g. does the message seem plausible) of the potential deepfake. For example, a video of a political candidate endorsing their opponent when both are close in the polls should cause uncertainty as to the authenticity of the video. Before sharing or believing the video, a viewer might search the internet for verification of the video by reputable news outlets. In another example, Kamali et al. (2024) offer five ways to visually detect deepfakes that would include looking carefully for: 1) anatomical implausibilities (e.g. missing fingers), 2) stylistic artifacts (e.g. inconsistencies in resolution & color), 3) functional implausibilities (e.g. detail rendering such as clothing looking distorted), 4) violations of physics (e.g. a shadow not reflecting the direction of the sun), and 5) sociocultural implausibilities (e.g. a politician in a situation saying something that is not in keeping with their political agenda). This is a very limited overview of the strategies identified by Kamali et al. (2024) but highlights potential critical visual analysis skills that could be developed in deepfake education.

Because it will be increasingly difficult to spot deepfakes as the technology improves, AI literacy that supports democracy cannot be static. While early deepfakes could be easily identified by a discerning eye, such as detecting missing fingers, the technology will make detection increasingly difficult. It is also unclear whether Tech companies can create or are willing to implement effective AI detectors that can move quickly enough to address concerning content before it spreads, particularly if moderation (fact-checking) is left up to users on the platforms. The passage of legislation concerning election deepfakes across 26 states suggests that more legislation from other states or the federal government may be forthcoming, but it is unclear whether the legislation would stop the quick spread and proliferation of deepfakes. Strategies for critically assessing deepfakes will be an important component of AI literacy.

Promising Initiatives in Support of AI Literacy

While the federal level has traditionally provided funding, guidelines, and regulations for K-12 and public higher education, the states provide most funding and policies and oversee the curriculum, so implementation of media literacy education is dependent primarily upon state legislatures and State Education Agencies (SEA). Private K-12 schools determine their own curriculum. While K-12 education is decentralized, higher education is even more decentralized. The most comprehensive way to support K-12 curriculum changes related to a civic-minded AI literacy would be for state policymakers and SEAs to create policies, funding, and professional development aimed at supporting their implementation. Because of the longstanding tradition of academic freedom for both public and private higher education institutions, curriculum is primarily the purview of the faculty, with the exception of state or regional accreditors, or professional accreditation requirements for certain fields. Therefore, curricular changes in higher education would have to come from faculty and/or administrative leadership.

Promising K-12 Initiatives

As mentioned earlier, as of 2024, 21 states have or are working on formal legislation related to media literacy (Media Literacy Now, 2024b). Two of the states – Delaware and New Jersey have passed legislation addressing media literacy state-wide and across all grade levels (Media Literacy Now, 2024b). California (AB 2876, 2024) was the first state to pass legislation aimed specifically at AI literacy across K-12. Scant research exists on how states are implementing these policies (DiGiacomo, 2023; Media Literacy Now, 2024a; NAMLE, 2024b). While media literacy is starting to gain political traction in some states, except for recent AI literacy education legislation in California (AB 2876, 2024), the legislation has not kept pace with how fast GenAI has developed.

New Jersey's S. 588 legislation was passed in 2023 and requires that information literacy be taught across K-12 schools. S. 588 defines information literacy as "...a set of skills that enables an individual to recognize when information is needed and to locate, evaluate, and use effectively the needed information" (2023, para. 1). As noted earlier, in the NAMLE definition, information literacy is one of the four components of media literacy (NAMLE, 2024b). S. 588 frames information literacy as encompassing "digital, visual, media, textual, and technological literacy" (para. 6). S. 588 requires that the Commissioner of Education work with the New Jersey Librarian to create an evaluation system to determine the effectiveness of district instructional programs in information literacy. It also mandates that teacher preparation programs and programs for alternate routes to teaching are required to include information literacy in their training programs. While New Jersey's legislation uses the term information literacy, the specific language of the bill includes key concepts of media literacy.

Delaware passed one of the most well-conceived bills, S.B. 195, The Digital Citizenship Act, which asserts that media literacy education in K-12 public education is vital to supporting democracy and citizenship (Delaware Department of Education, 2024; S.B. 195, 2022). S.B. 195 defines digital citizenship as "...a diverse set of skills related to current technology and social media including the norms of appropriate, responsible, and healthy behavior" (§ 4143) The Bill also defined media literacy as, "the ability to access, analyze, evaluate, create, and take action with all forms of communication, and encompasses the foundational skills of digital citizenship…" (§ 4143). S.B. 195 explicitly identifies curricular areas such as identifying credible sources and using multiple sources of information; acceptable use and purpose of social media platforms; the dangers of technology such as privacy issues, hacking, and online bullying; how to analyze, evaluate, access, and participate in digital forms of communication; persuasion techniques; identifying misinformation and bias, points of view that are excluded and included; and how media messages can shape society and culture. Delaware appears to have one of the more comprehensive and transparent media literacy requirements with explicit standards laid out across all grade levels and linked to the International Society for Technology in Education (ISTE) standards and the American Association of School Librarian (AASL) standards.

California passed S.B. 830 in 2018 requiring the California Department of Education (CDE) to provide web resources for state school districts on media literacy including related professional development for teachers. The bill defined media literacy as "...the ability to access, analyze, evaluate, and use media and encompasses the foundational skills that lead to digital citizenship" and digital citizenship as "...a diverse set of skills related to current technology and social media, including the norms of appropriate, responsible, and healthy behavior (§ 51206.4.a). In 2010, the CDE had adopted the Model School Library Standards for K-12 curricula which included

a focus on information literacy in terms of students accessing, evaluating, using, and integrating it into lifelong learning within school libraries (CDE, 2010). Then in 2023, California A.B. 873 recommended that the CDE expand media literacy curriculum into Science, Mathematics, History-Social Science, and English language arts/English language development across the K-12 education frameworks (Berman, 2023). In 2024, California then became the first state to pass legislation that expanded media literacy to include AI literacy with A.B. Bill 2876 (2024). AB 2876 (2024) utilizes earlier definitions for digital citizenship and media literacy from S.B. 830 and adds a definition of AI Literacy as "...the knowledge, skills, and attitudes associated with how artificial intelligence works, including its principles, concepts, and applications, as well as how to use artificial intelligence, including its limitations, implications, and ethical considerations" (§ 1a). The bill adds AI literacy education across the A.B. 873 media literacy curriculum frameworks (Berman, 2024). While the California legislation looks promising, the recommendations take effect the next time the curriculum is reviewed officially and as of now, the curriculum frameworks have not been updated (CDE, 2025)

While each of these state initiatives make inroads into ensuring all students in their states are being taught media literacy even using similar definitions in keeping with NAMLE, it is still unclear the extent to which these standards have been implemented in schools and districts across states. It is also unclear the extent to which professional development and district leadership has been provided to teachers. Research conducted by Baker et al. (2021) found that media literacy education was less likely to be supported centrally by school, district, or state leadership. They also found that teachers perceived more obstacles to teaching media literacy in schools with higher levels of students living in poverty or that were majority Black or Hispanic. Still, it is very promising that three states made such strong commitments to digital citizenship.

Promising Higher Education Initiatives

Higher education curricula are even more decentralized than K-12 education. If higher education institutions address AI literacy, it will be most likely at the initiative of higher education leaders, staff, and faculty in individual institutions or systems. Given the variety in higher education focus and organization, it is less clear what disciplines would be the home to media literacy training on GenAI. Digital Humanities, Journalism/Communications, Library Sciences, and Education are disciplines that may offer media literacy training as part of the curriculum. Library Sciences have standards related to higher education information literacy that cross all states (ACRL, 2016). Since digital media literacy education is inconsistent across K-12 public education nationally, higher education cannot assume that students have had media literacy education or AI literacy.

Higher education teaching and learning centers that support teaching innovation are also playing a role in supporting faculty as they think through how to integrate or how to handle GenAI in the classroom. For example, Barnard's Center for Engaged Pedagogy (CEP) and members of its Instructional Media and Academic Technology Services (IMATS) created a framework for AI Literacy to support faculty and students (Hibbert et al., 2024). The framework is composed of 4 components that each build on each other. First comes the goal to ensure students and faculty understand AI. Then they move to learning to use and apply AI. At the next level the goal is to learn to analyze and evaluate AI. At the final level with the appropriate training the goal is to create AI (Hibbert et al., 2024). As with definitions derived from computer science, this example does not connect to media literacy components such as information literacy or digital citizenship. Media

literacy can also be embedded in general education curricula. Students at Champlain College take "Navigating your Information Landscape", a required first year course centered on information and media literacy (Champlain College, 2025). Scant research exists on how higher education may seek to incorporate AI literacy as an outgrowth of the media literacy education umbrella into the curriculum broadly.

An emerging field in higher education, Public Interest Technology (PIT), is an interdisciplinary field that can serve as a vector for teaching AI literacy. The Public Interest Technology University Network (PIT-UN) connects over 60 higher education institutions across the US and globally (2025). It supports interdisciplinary learning, research, and curriculum that seeks to "advance the public interest and promote the public good in a digital age" (McGuinness & Schank, 2021, p. ix). This initiative is particularly promising because it is interdisciplinary and connects institutions across states, so it brings together scholars and teachers with the technological understanding of AI together with disciplines with expertise in media literacy education attentive to social equity and a civic-minded public good. The main challenge to AI literacy at both the K-12 and higher education institutions is for implementation across states, schools, programs, and classrooms so that it reaches students across economic, geographic and racial and ethnic boundaries.

Conclusion

Digital media play a key role in how citizens make decisions about voting and political participation. GenAI threatens democracy by making it easy to create and amplify disinformation and misinformation, making it much harder to make informed decisions about policies and voting. Political strategies designed to spread disinformation for political goals either coming from within the US or externally have eroded trust in media, institutions, facts, and ultimately democracy. Helping citizens and future citizens understand how to differentiate between fact and opinion and between credible and unsubstantiated claims is an important component of a civic-minded AI literacy.

AI literacy approaches are still under development across K-12 and higher education and vary by state and institution. A civic-minded AI literacy has the potential to empower students to critically evaluate and navigate the increasing volumes of disinformation and deepfakes they encounter. Central components for a civic minded AI literacy would operate under the umbrella of media literacy education and encourage critical thinking skills that are nonpartisan but acknowledge how power and inequity may be embedded in AI. It would be attentive to the ways GenAI can be biased and create, amplify, and spread misinformation, propaganda, disinformation, and deepfakes via invisible algorithmic manipulation. The intent here is by no means to assert that there is one way to educate students on AI literacy, rather to make a case for its importance and the importance of its connection to media literacy education. As an emerging field, more research is needed on understanding how AI literacy, media literacy, and digital citizenship is being conceptualized and implemented across K-12 and higher education globally.

AI literacy conceptualized as part of media literacy education would include learning about the potential dangers of GenAI for civic life. Given that technology companies are primarily driven by profit rather than supporting democracy and the public interest, US policymakers and educational leaders both at the federal and state level should encourage media literacy education that incorporates AI literacy and digital citizenship. Policymakers should also take up the issue of regulating AI as California has begun to do (with legislation such as S.B. 942) to ensure provenance markers and deepfake detection systems are available to users of digital media. More states

should consider mandating media literacy and AI literacy education for K-12 education and provide resources (e.g. professional development and funding) to support the implementation and effectiveness of media literacy and AI literacy efforts. Civic-minded AI literacy as part of a robust media literacy education program across K-12 and higher education is vital for a healthy democracy.

References

A.B. 873 2023-2024 Reg. Sess. State Assem. (Cal. 2023, Oct. 13). An act to amend Section 33548 of the Education Code, relating to pupil instruction. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB873

A.B. 2876. 2023-2024 Reg. Sess. State Assem. (Cal. 2024, Sept. 29). Pupil instruction: Media literacy: artificial intelligence: curriculum frameworks: instructional materials. An act to amend Section 33548 of the Education Code, relating to pupil instruction. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2876

Acar, O. (2024, February 19). With generative AI we can reimagine education – and the sky is the limit. World Economic Forum. https://www.weforum.org/agenda/2024/02/with-generative-ai-we-can-reimagine-education-and-the-sky-is-the-limit/

ACRL (2016, January 11). Framework for information literacy for higher education. https://www.ala.org/acrl/standards/ilframework

AI Elections Accord (2024, February). A tech accord to combat deceptive use of AI in 2024 elections. Munich Security Conference.
https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, *13*(1), 30.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-236.

Appel, M., & Prietzel, F. (2022). The detection of political deepfakes. *Journal of Computer-Mediated Communication*, *27*(4), zmac008.

Badawy, A., Addawood, A., Lerman, K., & Ferrara, E. (2019). Characterizing the 2016 Russian IRA influence campaign. *Social Network Analysis and Mining*, *9*, 1-11.

Badawy, A., Ferrara, E., & Lerman, K. (2018, August). Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)* (pp. 258-265). IEEE.

Baker, G., Faxon-Mills, S., Huguet, A., Pane, J. F., & Hamilton, L. S. (2021). *Approaches and obstacles to promoting media literacy education in US schools*. RAND.

Barrera, O., Guriev, S., Henry, E., & Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of public economics*, *182*, 104123.

Bauder, D., Chase, R., and Mulvihill, G. (2023, April 18). Fox, Dominion reach $787M settlement over election claims. AP. https://apnews.com/article/fox-news-dominion-lawsuit-trial-trump-2020-0ac71f75acfacc52ea80b3e747fb0afe

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Berman, M. (2023, October 14). California to teach media literacy in every grade. California State Assembly. https://a23.asmdc.org/press-releases/20231014-california-teach-media-literacy-every-grade

Berman, M. (2024, October 3). California to teach AI literacy in every grade. California State Assembly. https://a23.asmdc.org/press-releases/20241003-california-teach-ai-literacy-every-grade

Bond, S. (2024, May 23). A political consultant faces charges and fines for Biden deepfake robocalls. NPR. https://www.npr.org/2024/05/23/nx-s1-4977582/fcc-ai-deepfake-robocall-biden-new-hampshire-political-operative

Boulianne, S. (2020). Twenty years of digital media effects on civic and political participation. *Communication Research*, *47*(7), 947-966.

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., Saavedra, A. (2019). Students' civic online reasoning: A national portrait. Stanford Digital Repository. Retrieved https://purl.stanford.edu/gf151tb4868.

Bodroghkozy, A. (2012). *Equal time: Television and the civil rights movement*. University of Illinois Press.

C2PA (2024). About. An open technical standard providing publishers, creators, and consumers the ability to trace the origin of different types of media. The Coalition for Content Provenance and Authenticity. https://c2pa.org/

Calice, M. N., Bao, L., Freiling, I., Howell, E., Xenos, M. A., Yang, S., Broussard, D., Newman, T.P., Scheufele, D.A. (2023). Polarized platforms? How partisanship shapes perceptions of "algorithmic news bias." *New Media & Society*, *25*(11), 2833-2854.

California Department of Education (2010, September 16). Model School Library Standards. California Department of Education. https://www.cde.ca.gov/CI/cr/lb/schoollibstnds2017.asp

California Department of Education (2025, September 16). All Curriculum Frameworks. California Department of Education. Retrieved July 28, 2025, https://www.cde.ca.gov/ci/cr/cf/allfwks.asp

Casey, R.D. (1944a, July 1). What is propaganda: The story of propaganda. American Historical Association. https://www.historians.org/resource/the-story-of-propaganda/

Casey, R. D. (1944b, July 1). What is propaganda: What are the tools of propaganda? American Historical Association. https://www.historians.org/resource/what-are-the-tools-of-propaganda/

Champlain College (2025). Navigating Your Information Landscape. Champlain College. https://classlist.champlain.edu/show/course/number/Cor_103

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev., 107*, 1753.

Chitra, U., & Musco, C. (2020, January). Analyzing the impact of filter bubbles on social network polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (pp. 115-123).

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118.

Curto, G., Jojoa Acosta, M. F., Comim, F., & Garcia-Zapirain, B. (2024). Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. *AI & Society*, *39*(2), 617-632.

Delaware Department of Education. (2024). Media literacy & digital citizenship. Delaware media literacy standards. https://education.delaware.gov/educators/academic-support/standards-and-instruction/digital-de/instructional-resources/media-literacy/

Delipetrev, B., Tsinaraki, C., & Kostic, U. (2020). "Historical Evolution of Artificial Intelligence", EUR 30221EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-18940-4, doi:10.2760/801580, JRC120469 https://eprints.ugd.edu.mk/28050/1/2.%20jrc120469_historical_evolution_of_ai-v1.1.pdf

DiGiacomo, D. K., Hodgin, E., Kahne, J., Alkam, S., & Taylor, C. (2023). Assessing the state of media literacy policy in US K-12 schools. *Journal of Children and Media*, *17*(3), 336-352.

Directive (EU)Y2018/1808. *Directive (EU) 2018/1808 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL.* https://eur-lex.europa.eu/eli/dir/2018/1808/oj

Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation: House Energy and Commerce Subcommittee Communications and Technology, 117 Cong. (2021). https://www.congress.gov/event/117th-congress/house-event/111407

Elliott, O. (2025, February). Representation of BBC news content in AI assistants. BBC. Chrome. https://www.bbc.co.uk/aboutthebbc/documents/bbc-research-into-ai-assistants.pdf

Examining Social Media Companies' Efforts to Counter Online Terror Content and Misinformation. House Homeland Security Committee, 116 Cong. (2019). https://www.congress.gov/event/116th-congress/house-event/109710

Federal Communications Commission (2024, February 8). FCC Makes AI-Generated Voices in Robocalls Illegal. https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96-104.

Ferrer, X., Van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and discrimination in AI: a cross-disciplinary perspective. *IEEE Technology and Society Magazine*, *40*(2), 72-80.

Fitzgerald, J. C., Cohen, A. K., Maker Castro, E., & Pope, A. (2021). A systematic review of the last decade of civic education research in the United States. *Peabody Journal of Education*, *96*(3), 235-246.

Frangonikolopoulos, C. A., & Chapsos, I. (2012). Explaining the role and the impact of the social media in the Arab Spring. *Global Media Journal: Mediterranean Edition*, *7*(2).

Fraser, N. (1994). Rethinking the public sphere: A contribution to the critique of actually existing democracy. In C. Calhoun (Ed.), *Habermas and the public sphere* (pp. 109-142). MIT Press.

Garnham, N. (1994). The media and the public sphere. In C. Calhoun, Ed., *Habermas and the public sphere*. The MIT Press.

Gerstl-Pepin, C. I. (2007). Introduction to the special issue on the media, democracy, and the politics of education. *Peabody Journal of Education*, *82*(1), 1-9.

Gerstl-Pepin, C. (2015). Popular media portrayals of inequity and school reform in The Wire and Waiting for "Superman". *Peabody Journal of Education*, *90*(5), 691-710.

Ghosh, D., & Scott, B. (2018). Digital deceit: the technologies behind precision propaganda on the internet. New American Public Interest Technology & Harvard Kennedy School Shorenstein Center on Media, Politics, and Public Policy.

Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? *PNAS nexus*, *3*(2), 1-7.

Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A., & Tucker, J. A. (2020). Cross-platform state propaganda: Russian trolls on twitter and YouTube during the 2016 US Presidential Election. *The International Journal of Press/Politics*, *25*(3), 357-389.

Gongane, V. U., Munot, M. V., & Anuse, A. D. (2022). Detection and moderation of detrimental content on social media platforms: current status and future directions. *Social Network Analysis and Mining*, *12*(1), 129.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 2053951719897945.

Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. (2024). Human detection of political speech deepfakes across transcripts, audio, and video. *Nat Commun* 15, 7629 (2024). https://doi.org/10.1038/s41467-024-51998-z

Guo, Z. (2024, May). Online Disinformation and Generative Language Models: Motivations, Challenges, and Mitigations. In *Companion Proceedings of the ACM on Web Conference 2024* (pp. 1174-1177).

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of the bourgeois society*. The MIT Press.

Haugen, F. (2021, October 4). Statement of Frances Haugen. United States Senate Committee on Commerce, Science and Transportation Sub-Committee on Consumer Protection, Product Safety, and Data Security. https://www.commerce.senate.gov/services/files/FC8A558E-824E-4914-BEDB-3A7B1190BD49

Hayes, D. and Lawless, J. L. (2021). *The new hole: The demise of local journalism and political engagement*. New York, NY: Cambridge University Press.

Hibbert, M., Altman, E., Shippen, T., Wright, M. (2024, Jun 3). A framework for AI literacy. Educause Review. https://er.educause.edu/articles/2024/6/a-framework-for-ai-literacy

Hilton, J., Kokotajlo, D., Kumar, R., Nanda, N., Saunders, W., Wainwright, C., Ziegler, D., Anonymous, Anonymous, Anonymous, Anonymous, Anonymous, Anonymous. (2024). A Right to Warn about Advanced Artificial Intelligence. https://righttowarn.ai/

Hobbs, R. (2020). Propaganda in an Age of Algorithmic Personalization: Expanding Literacy Research and Practice. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.301 Available at: https://doi.org/10.1002/rrq.301

Hobbs, R., & McGee, S. (2014). Teaching about propaganda: An examination of the historical roots of media literacy. *Journal of Media Literacy Education*, *6*(2), 56-66.

Holdsworth, J., & Scapicchio, M. (2024, June 17). What is deep learning? IBM: Think. https://www.ibm.com/topics/deep-learning

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American educational research journal*, *54*(1), 3-34.

Kahne, J., Hodgin, E., & Eidman-Aadahl, E. (2016). Redesigning civic education for the digital age: Participatory politics and the pursuit of democratic engagement. *Theory & Research in Social Education*, *44*(1), 1-35.

Kamali, N., Nakamura, K., Chatzimparmpas, A., Hullman, J., & Groh, M. (2024). How to Distinguish AI-Generated Images from Authentic Photographs. *arXiv preprint arXiv:2406.08651*.

Kaplan, J. (2025, January 7). More speech and fewer mistakes. Meta. https://about.fb.com /news/2025/01/meta-more-speech-fewer-mistakes/#:~:text=Takeaways,system%20audits%20and%20additional%20signals.

Kawakami, A., Umarova, K., Huang, D., & Mustafaraj, E. (2020, July). The Fairness Doctrine lives on? Theorizing about the Algorithmic News Curation of Google's Top Stories. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (pp. 59-68).

Knight, W. (2023, August 29). It costs just $400 to build an AI disinformation machine. *Wired.* https://www.wired.com/story/400-dollars-to-build-an-ai-disinformation-machine/

Kreps, S., & Kriner, D. (2023). How AI threatens democracy. *Journal of Democracy*, *34*(4), 122-131.

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, *3*, 100101.

Lee, T., & Hosam, C. (2020, September). Fake news is real: The significance and sources of disbelief in mainstream media in Trump's America. In *Sociological Forum* (Vol. 35, pp. 996-1018).

Ling, R. (2020). Confirmation bias in the era of mobile news consumption: the social and psychological dimensions. *Digital Journalism*, *8*(5), 596-604.

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).

Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data. *Human Behavior and Emerging Technologies*, *1*(2), 141-148.

Martens, B., Aguiar, L., Gomez-Herrera, E., Müller-Langer, F. (2018). The digital transformation of news media and the rise of disinformation and fake news, JRC Digital Economy Working Paper, No. 2018-02, European Commission, Joint Research Centre (JRC), Seville

Martens, H., & Hobbs, R. (2015). How media literacy supports civic engagement in a digital age. *Atlantic Journal of Communication*, *23*(2), 120-137

McGuinness, T. D., & Schank, H. (2021). *Power to the public: The promise of public interest technology*. Princeton University Press.

McQuade, B. (2024). *Attack from within: How disinformation is sabotaging* America. Seven Stories Press.

Media Literacy Now (2024a, February). *Media Literacy Now Policy Report 2023: A State-by-State Status of Media Literacy Education Laws for K–12 Schools*. Media Literacy Now. https://medialiteracynow.org/wp-content/uploads/2024/02/MediaLiteracyNowPolicyReport2023 _publishedFeb2024b.pdf

Media Literacy Now (2024b, May). *Putting Media Literacy on the Public Agenda.* https://medialiteracynow.org/impact/current-policy/

Menn, J & Nakamura, D. (2024, October 22). U.S. officials say Russia smeared Tim Walz, might stoke post-vote violence. https://www.washingtonpost.com/national-security/2024/10/22/russia-election-disruption-tim-walz/. officials say Russia smeared Tim Walz, might stoke post-vote

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, *13*(9), 856.

Microsoft (2024, September 24). *Deepfakes and elections.Don't fall for deepfakes this election. Check. Recheck. Vote.* Microsoft. https://www.microsoft.com/en-us/corporate-responsibility/elections

Middaugh, E., Bell, S., & Kornbluh, M. (2022). Think before you share: building a civic media literacy framework for everyday contexts. *Information and Learning Sciences*, *123*(7/8), 421-444.

Moss, R. (2023, July 21). *The legal framework before the Supreme Court: What does it mean for consumers?* National Association of Attorneys General. https://www.naag.org/attorney-general-journal/the-future-of-section-230-what-does-it-mean-for-consumers/

Nadeem, A., Abedin, B., & Marjanovic, O. (2020). Gender bias in AI: A review of contributing factors and mitigating strategies.

National Archives (2024, September 5). Fairness doctrine. ttps://www.reaganlibrary.gov/archives/topic-guide/fairness-doctrine

National Archives (n.d.). Powers of persuasion. National Archives and Records Administration. https://www.archives.gov/exhibits/powers-of-persuasion

National Artificial Intelligence Advisory Committee (2023, November). Recommendations: Enhancing AI Literacy for the United States of America. https://ai.gov/wp-content/uploads/2023/12/Recommendations_Enhancing-Artificial-Intelligence-Literacy-for-the-United-States-of-America.pdf

National Association for Media Literacy Education (2023, May). Core principles of media literacy education. https://namle.org/wp-content/uploads/2023/05/NAMLE-Principles-Final-2.1.pdf

National Association for Media Literacy Education (2024a, September 4). Resources. https://namle.org/resources/media-literacy-defined/

National Association for Media Literacy Education (2024b). Snapshot 2024: The state of media literacy education in the U.S. https://namle.org/wp-content/uploads/2024/01/Snapshot-2024-State-of-Media-Literacy-FINAL.pdf

National Conference of State Legislatures. (2025, July 23). Artificial Intelligence (AI) in elections and campaigns. NCSL. https://www.ncsl.org/elections-and-campaigns/artificial-intelligence-ai-in-elections-and-campaigns?ref=51st.news

Nazzaro, M. (2024, October 23). Microsoft says Russia trying to smear Harris with deepfake video, AI. *The Hill*. https://thehill.com/policy/technology/4949612-russia-china-ai-influence-elections/

Neapaw (2023). CounterCloud—AI powered disinformation experiment. YouTube. https://www.youtube.com/watch?v=cwGdkrc9i2Y

News Literacy Project (2024, September 28). *You have the power to stop misinformation. Use it*. News Literacy Project. https://newslit.org/

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, *58*(1), 504-509.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY: New York University Press.

Nonnecke, B., & Carlton, C. (2022). EU and US legislation seek to open up digital platform data. *Science*, *375*(6581), 610-612.

O'Shaughnessy, N. (2009). Selling Hitler: propaganda and the Nazi brand. *Journal of Public Affairs: An International Journal*, *9*(1), 55-76.

Oxford English Dictionary. (2007). Propaganda. https://www.oed.com/dictionary/propaganda_n
Painter, R. W. (2023). Deepfake 2024: Will Citizens United and artificial intelligence together destroy representative democracy? *Journal of National Security Law & Policy*, *14*, 121.
Pew Research Center (2024a, September 17). News platform fact sheet. https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/
Pew Research Center (2024b, September 17). Social media and news fact sheet. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/
Pew Research Center (2024c, November 13). Internet, broadband fact sheet. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/
Pew Research Center (2024d, November 13). Mobile fact sheet. https://www.pewresearch.org/internet/fact-sheet/mobile/
Pew Research Center (2025, July 10). Teens and internet, device access fact sheet. https://www.pewresearch.org/internet/fact-sheet/teens-and-internet-device-access-fact-sheet/
Polizzi, G. (2020). Information literacy in the digital age: Why critical digital literacy matters for democracy. *Informed Societies: Why information literacy matters for citizenship, participation and democracy*, 1-23.
Public Interest Technology University Network (2025). The public interest technology university network. https://pit-un.org/about-us/
Putnam, R. (2020). *Bowling alone: The collapse and revival of American community, anniversary edition*. New York, NY: Simon & Schuster.
O'Donnell, N. (2021). Have we no decency? Section 230 and the liability of social media companies for deepfake videos. *U. Ill. L. Rev.*, 701.
Ranka, H., Surana, M., Kothari, N., Pariawala, V., Banerjee, P., Surve, A., ... & Mehta, S. (2024). Examining the implications of deepfakes for election integrity. *arXiv preprint arXiv:2406.14290*.
Ryan-Mosley, T. (2023, October 4). How generative AI is boosting the spread of disinformation and propaganda. *MIT Technology Review*. https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/
S. 588, 220th Leg. (N.J. 2022). An Act concerning public school instruction on information literacy and supplementing chapters 26 and 35 of Title 18A of the New Jersey Statutes. https://pub.njleg.gov/Bills/2022/S1000/588_I1.HTM
S.B. 195, 151st. Gen. Assem. (Del. 2022, August 29). An act to amend Title 14 of the Delaware code relating to media literacy. https://legis.delaware.gov/json/BillDetail/GenerateHtmlDocument?legislationId=78981&legislationTypeId=1&docTypeId=2&legislationName=SB195
S.B. 830, 2017-2018 Reg. Sess. State Assem. (Cal. 2018, Sept. 17). Pupil instruction: media literacy: resources. https://legiscan.com/CA/text/SB830/id/1766812
S.B. 942, 2023-2024 Reg. Sess. State Assem. (Cal. 2024, Sept. 19). California AI transparency act. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942
Sádaba, C., & Salaverría, R. (2023). Tackling disinformation with media literacy: analysis of trends in the European Union. *Revista latina de comunicación social*, (81), 17-32.
Salomaa, S. & Palsa, L. (2019). Media literacy in Finland: National media education policy. Retrieved https://www.unesco.org/creativity/en/creativity/policy-monitoring-platform/media-literacy-finland-national-media-education-policy

Shivaram, D. (2025, July 10). State Department investigating incident in which AI used to impersonate Marco Rubio. Morning Edition, NPR. https://www.npr.org/2025/07/10/nx-s1-5462844/state-department-investigating-incident-in-which-ai-used-to-impersonate-marco-rubio

Smith, C. R. (1999). The campaign to repeal the Fairness Doctrine. *Rhetoric & Public Affairs*, *2*(3), 481-505.

Surjatmodjo, D., Unde, A. A., Cangara, H., Sonni, A. F. (2024). Information pandemic: A critical review of disinformation spread on social media and its implications for state resilience. *Social Sciences*, *13*(8), 418. https://doi.org/10.3390/socsci13080418

Swenson, A. and Weissert, W. (2024, January 22). New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary. AP. https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-robocallf3469ceb6dd6130 79092287994663db5

Take It Down Act, 47 U.S.C. § 223 (2025). https://www.congress.gov/bill/119th-congress/senate-bill/146/text#:~:text=Text%20%2D%20S.,Congress.gov%20%7C%20Library%20of%20Congress

Thompson, P. (2023, September 3). A developer built an AI-powered 'propaganda' machine to using OpenAI tech to highlight the dangers of mass-produced AI disinformation. *Business Insider.* https://www.businessinsider.com/developer-creates-ai-disinformation-system-using-openai-2023-9

Toff, B., & Mathews, N. (2021). Is social media killing local news? An examination of engagement and ownership patterns in US Community news on Facebook. *Digital Journalism*, 1-20

Toner, (2024, January 10). Generative AI and Global Education. NAFSA: International Association of Educators. https://www.nafsa.org/ie-magazine/2024/1/10/generative-ai-and-global-education

U.S. Department of Justice (2020). Section 230 – Nurturing innovation or fostering unaccountability? Key takeaways and recommendations. https://www.justice.gov/ag/file/1072971/dl?inline=

Wack, M., Ehrett, C., Linvill, D., & Warren, P. (2025). Generative propaganda: Evidence of AI's impact from a state-backed disinformation campaign. *PNAS nexus*, *4*(4), https://doi.org/10.1093/pnasnexus/pgaf083

West, S. M. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, *20*(11), 4366-4383.

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., ... & West, S. M. (2019). Disability, bias, and AI. *AI Now Institute*, *8*.

Zewe, A. (2023, November 9). Explained: Generative AI. MIT News. https://news.mit.edu/2023/explained-generative-ai-1109

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, *12*(1), 415-438.