Generative AI, Democracy, and Civic Engagement – Opportunities, Risks, and Implementation Strategies[1]

Executive Summary

- **GenAI for Policy Accessibility:** Generative AI, especially large language model (LLM) chatbots, can bridge the gap between complex government policies and public understanding. AI assistants can translate legalistic or technical policy documents into plain language explanations tailored to different audiences, improving transparency and inclusivity pmc.ncbi.nlm.nih.gov. Multilingual AI tools further enhance accessibility by delivering information in citizens' native languages, which is crucial in linguistically diverse nations and among marginalized groups. For example, the European Union already uses LLM-based translation to publish policy documents in all official languages, ensuring all member-state citizens receive the same information simultaneously botpenguin.com. In India, a voice-enabled AI assistant on the UMANG government services app supports eight regional

[1] This is Chapter 12 of *Co-Intelligence Applied*, an anthology co-created in February 2025 by OpenAI's Deep Research in cahoots with Robert Klitgaard of Claremont Graduate University. https://robertklitgaard.com.

Keywords: Generative AI, Democracy, Civic Engagement, Deliberative Democracy, AI Ethics, Public Participation, Government Chatbots, Multilingual AI, Astroturfing, AI Transparency, Digital Inclusion, AI Bias, Citizen Empowerment.

languages, enabling rural and less-literate citizens to ask questions in their own language and get instant answers about public programs aisight.fractal.ai. Such innovations allow millions to access voting information, tax guidelines, and services that were previously out of reach, fostering greater public participation.

- **Enhancing Civic Engagement:** GenAI has the potential to significantly broaden and deepen civic engagement. AI-generated discussion prompts and conversational agents can draw more citizens into public consultations by lowering barriers to participation. Advanced LLMs can digest thousands of public comments or forum posts and produce concise meeting summaries or "argument maps" that highlight key points and diverse viewpoints ai.objectives.institute. This assists both citizens and policymakers by structuring complex debates into understandable themes. Early experiments indicate that AI tools can scale up deliberative democracy – consulting larger populations faster and extracting nuanced insights – while surfacing minority opinions that might otherwise be overlooked ai.objectives.institute. By automating the labor-intensive task of analyzing qualitative feedback, AI frees human facilitators to focus on decision-making and ensures that every voice (not just the loudest) is accounted for in the dialogue.
- **Opportunities vs. Risks:** Generative AI can either strengthen democracy or undermine it, depending on implementation. On one hand, AI-driven platforms like chatbots have improved government responsiveness and accountability. In South Africa, the "GovChat" chatbot on WhatsApp and Facebook Messenger enabled 8 million citizens to apply for COVID-19 relief grants without standing in line, and processes up to 12,000 queries per minute govinsider.asia. It also lets users rate public services and immediately flags community issues (like water outages) to local officials, making governance more proactive

and transparent govinsider.asia . These successes show GenAI's promise to boost public trust by making governments more accessible and evidence-driven. On the other hand, serious concerns arise around AI accuracy, bias, and misuse. LLMs are prone to "hallucinations," producing incorrect or fabricated information with a confident tone canons.sog.unc.edu. If a government chatbot gives citizens wrong legal advice or misleading voting information, the consequences for public trust and rights can be severe. Bias in AI systems is another danger: studies found that some generative models systematically favor certain ideological or majority viewpoints phys.org, which could marginalize minority voices or skew public discourse. Moreover, malicious actors or even governments themselves could misuse GenAI to manipulate opinion – for instance, by flooding forums with AI-generated comments to create a false impression of consensus (a practice known as astroturfing)lawfaremedia.org. These risks highlight the double-edged nature of GenAI in the civic arena.

- **Ethical Safeguards and Recommendations:** Harnessing GenAI for democracy requires robust ethical guidelines and policy frameworks. To prevent misinformation, governments and developers must rigorously test AI systems and incorporate verification mechanisms (e.g. retrieval of official data to support answers pmc.ncbi.nlm.nih.gov) before deploying them for public use. Bias mitigation strategies are essential: this includes diversifying training data to represent all community segments, building multilingual models (as Nigeria did by developing a local LLM covering Hausa, Yoruba, Igbo, and more dig.watch), and conducting independent audits of AI outputs for fairness. Transparency and accountability should be ingrained – AI civic tools ought to clearly disclose that they are AI, explain how they arrive at answers, and allow human oversight or appeal for contentious cases. Institutions like the Westminster Foundation for Democracy advise that AI tools

be **"transparent, accountable, and designed to enhance rather than replace human decision-making,"** with strong oversight bodies and data transparency to maintain public trust babl.ai. Finally, legal and regulatory measures may be needed to guard against the malicious use of generative AI in politics (for example, requiring verification for public comment submissions or labeling AI-generated content in political ads). With careful implementation guided by ethics and equity, GenAI can be a powerful ally in strengthening democratic participation. This report provides a detailed analysis of these opportunities and challenges, with a focus on global applications and developing country contexts, and offers policy recommendations for responsible use of AI in civic processes.

Literature Review: AI in Civic Engagement and Governance

Growing Use of AI for Public Participation: The integration of artificial intelligence into public sector communication and citizen engagement is accelerating worldwide. Governments at various levels have begun experimenting with AI-powered chatbots, virtual assistants, and text analysis tools to improve interactions with citizens canons.sog.unc.edu. Unlike earlier-generation rule-based chatbots limited to scripted answers, new generative AI systems can produce flexible, natural-language responses, making them more user-friendly for the public canons.sog.unc.edu. Academic and policy research underscores the potential benefits of these tools. Yun et al. (2024) highlight that combining Generative AI with public administration has *"significantly enhanced...dynamic interaction between public authorities and citizens,"* enabling new ways to communicate policies and receive feedback pmc.ncbi.nlm.nih.gov. This has prompted interest in using LLMs to translate bureaucratic information into citizen-friendly dialogue. For instance, an open-government study proposed a

Retrieval-Augmented Generation system with LLMs to answer citizen queries about policies, achieving around 88% answer accuracy in trials and boosting public engagement with policy information pmc.ncbi.nlm.nih.gov. Such promising findings have spurred pilot projects in multiple countries.

AI Chatbots for Government Services: One of the most widespread applications of AI in governance so far is the deployment of chatbots on government websites and messaging platforms to handle inquiries about services, benefits, and regulations. A recent review of state government chatbots in the U.S. found they can optimize workloads and reduce wait times for information by handling simple questions automatically govtech.com. During the COVID-19 pandemic, many governments introduced AI assistants to disseminate information and triage requests. For example, India's MyGov introduced "Saathi," an AI virtual agent, to provide up-to-date answers about COVID protocols and support in multiple languages accenture.com aisight.fractal.ai. Similarly, numerous social security agencies in Latin America and Asia have adopted chatbots to guide users through processes like filing claims or obtaining permits, available 24/7 and at lower cost than call centers issa.int. Crucially, these systems are increasingly multilingual and conversational. The Indian government's UMANG platform integration mentioned above is a case in point: by enabling voice queries in local dialects, it addressed the long-standing problem of citizens being unable to access programs due to language or literacy barriers aisight.fractal.ai. Such voice-based AI services underscore how technology can extend governance reach into rural and underserved populations, a theme often noted in ICT4D (ICT for Development) literature. Early evidence suggests that when done right, citizens respond positively. The Government Digital Service (GDS) in the UK tested a generative AI chatbot on the official gov.uk portal and *"found that people liked the*

experience" of asking the AI for help – though they also noted the importance of **"ensuring accurate answers"** as a remaining technical challenge bi.team. This indicates a public openness to AI-assisted services, provided reliability is maintained.

AI in Policy Information and Transparency: Beyond service delivery, AI tools are being explored to make policy and legal information more accessible. Legal informatics research has begun to use LLMs to parse legislation and regulations, automatically summarizing key points or implications for laypersons padolsey.medium.com pmc.ncbi.nlm.nih.gov. The European Union, which operates in 24 official languages, has leveraged advanced machine translation (now often powered by transformer-based models) to ensure every citizen can read EU laws and notices in their mother tongue botpenguin.com. This multilingual approach is increasingly being recognized as essential in developing nations too, where a colonial or official language may not be understood by large segments of the population. Recent initiatives in Africa (e.g., Nigeria's first multilingual LLM) and South Asia are explicitly focusing on expanding AI's language support to include "low-resource" languages so that AI-driven government platforms are inclusive dig.watch. Think tanks emphasize that without such efforts, the AI revolution could bypass non-English speakers and worsen information inequality techletter.co. In summary, current literature and practice show a strong trend toward using GenAI to break down complex policy content and linguistic barriers, thereby inviting more citizens into the democratic dialogue. However, they also unanimously stress diligence in maintaining accuracy and equity in these AI systems – a theme explored in depth in later sections of this report.

AI for Public Deliberation: Another emergent area in both research and practice is using AI to facilitate public deliberation and consultation processes. Traditional public meetings, surveys, and comment periods often struggle to process input from large numbers of citizens or

to engage a representative sample of the community. AI offers a way to "scale up" deliberation by analyzing qualitative data (opinions, arguments, narratives) far more efficiently than human staff. For instance, civic tech organizations have developed tools like "Talk to the City," which employ frontier LLMs to **"analyze large datasets"** of citizen contributions, **"extract key arguments,"** and cluster similar opinions together for easier review ai.objectives.institute. These tools can automatically generate summaries and even visualization of the debate (e.g., mapping how many people voiced concern about climate impacts vs. economic costs in a town-hall discussion on development). Early case studies using data from platforms like Polis (an online deliberation tool) show that AI can indeed produce coherent reports capturing the diversity and consensus levels of public opinion. Government pilots in places like Finland and Taiwan have similarly used machine learning to help categorize thousands of citizen comments on draft laws into thematic reports that policymakers can digest. The literature suggests such AI-assisted sense-making could vastly improve how governments handle citizen input, making participatory policymaking more feasible even on a large scale apolitical.co publicinput.com.

Deliberative Democracy Experiments: Researchers are also examining whether AI can directly support structured deliberative democracy exercises – for example, citizen assemblies or moderated online forums. One interesting direction is using AI to generate balanced discussion prompts or to act as a "virtual facilitator." A study in *Proceedings of the National Academy of Sciences* even tested an AI assistant that could intervene in heated online debates by suggesting evidence-based information in real time, in hopes of steering discussions onto a more factual and less polarized track pnas.org. While results are preliminary, they point to AI's potential as a mediating tool in discourse. Additionally, projects like IBM's **Project Debater** have demonstrated that AI can ingest large volumes of arguments on a topic and produce persuasive

summaries for each side of an issue. In a governance context, such technology might be used to help a city council quickly understand the strongest pro and con arguments citizens have about a proposed policy, thereby ensuring all perspectives are considered.

Summary of Findings: In review, current applications of Generative AI in civic contexts fall into a few broad categories: **(1)** Information provision (Q&A chatbots, virtual assistants) for government services and policies; **(2)** Translation and localization (multilingual support) to reach diverse populations; **(3)** Data analysis and summarization tools to manage public feedback and support consultations; and **(4)** Experimental uses in fostering or guiding public discourse. There is a growing body of evidence – from academic evaluations to government pilot programs – that these AI tools can improve efficiency, reach, and inclusiveness of public engagement. At the same time, recurring concerns about accuracy, bias, and public trust emerge throughout the literature. These set the stage for a closer analysis of how GenAI can be optimally used for civic engagement (sections to follow on Policy Accessibility and Public Deliberation) and what ethical considerations must be addressed to ensure it strengthens rather than hinders democratic governance.

Analysis of Opportunities and Challenges

1. Policy Explanation and Accessibility

Plain-Language Policy Translation

Government policies and laws are often written in dense, technical language that many citizens find inaccessible. In both developed and developing countries, this complexity creates a knowledge barrier – people may not understand how a new tax law works, what their rights are under a health regulation, or how to apply for benefits. Generative AI can help overcome this by

translating official jargon into plain language in real time. LLM-powered chatbots can be trained on corpora of government documents and then fine-tuned to output simplified summaries or Q&A responses. For example, a policy briefing paper in China and the U.S. used an LLM-based system to answer citizen questions about local regulations; it employed a retrieval mechanism to pull the relevant policy text and then explained it in simpler terms with over 85–90 percent accuracy compared to expert answerspmc.ncbi.nlm.nih.gov. Such a system means a resident could ask, "What does the new education policy mean for school fees?" and get a clear, accurate answer in seconds, instead of wading through pages of legal text. The opportunity here is a dramatic increase in transparency: when people actually comprehend policies, they are more likely to engage with them or comply, and less likely to spread misconceptions.

In developing countries, the need for plain-language explanation is even more pronounced, as literacy levels and educational access vary widely. Take **voting procedures** as an example – many democracies struggle with voter education. An AI assistant can inform a first-time voter how to check their registration status, where their polling place is, and what identification to bring, all in an easy-to-follow manner. If the citizen asks a detailed question ("What if I moved since the last election?"), the AI can dynamically provide the relevant rule (e.g., address update process) rather than giving a generic answer. Personalization is key: these models can adjust explanations based on the user's context, asking follow-ups if needed to clarify (much like a human clerk would). This level of responsive, personalized guidance could especially benefit marginalized groups who might feel intimidated to approach government offices in person.

Multilingual AI Assistants

Multilingual support is a critical dimension of accessibility. Many countries – particularly in Africa, Asia, and parts of Europe – have multiple official or widely spoken languages. Often, minority language speakers have limited access to official information, which undermines equal participation in civic life. GenAI can significantly narrow this gap. Modern LLMs like GPT-4 and others are capable of understanding and generating text in dozens of languages. By deploying chatbots that converse in local languages or dialects, governments can engage citizens who were previously excluded by language barriers. The **European Union's** use of LLM-based machine translation to issue policies in 24 languages is a large-scale example that *"enables citizens across member states to access the same information simultaneously, fostering transparency and inclusivity."* botpenguin.com

On a national level, **South Africa** has 11 official languages – one could envision a single AI service that explains a new municipal bylaw in Zulu, Xhosa, Afrikaans, or any language the user selects.

We are already seeing movement in this direction. **Nigeria** recently launched its first multilingual LLM, aiming to strengthen representation of languages like Hausa, Yoruba, and Igbo in AI applications dig.watch. The goal is explicitly to make AI tools (including those used for government services) more inclusive in a country with over 500 languages. Similarly, the Government of India's collaboration with an AI firm to enhance the UMANG citizen service app with voice AI was driven by linguistic diversity – the system was built to handle at least 10 major Indian languages and many dialects aisight.fractal.ai. During its pilot, millions of users interacted with this Hindi-and-English chatbot, posing over 1 million queries within weeks, which demonstrates pent-up demand for information in local tongues aisight.fractal.ai. By asking questions aloud in Hindi (or other languages soon), rural users could learn about crop insurance

schemes or scholarship programs without needing an intermediary or struggling through English-only text. This multilingual outreach can empower ethnic minorities and indigenous communities, integrating them more fully into the democratic conversation.

However, challenges accompany these opportunities. Current mainstream LLMs still perform better in English than in many low-resource languages. For instance, one analysis found GPT-4 scored 85% on an English test but only ~62% when taking the same test in Telugu, a major Indian language techletter.co. Such discrepancies mean that out-of-the-box solutions might give flawed answers in some languages or ignore subtle cultural context. Developing nations may need to invest in or adopt locally trained models (as Nigeria is doing) to ensure high quality. Additionally, translation is not just linguistic but also cultural – the AI must be aware of local examples or analogies to truly make policy "plain" for the target audience. Ensuring that multilingual AI assistants are accurate and context-aware in all languages will require careful training, evaluation, and continuous improvement.

Personalized Citizen Assistance

Another promise of GenAI in this domain is providing **personalized, reliable answers** to citizens on demand. Rather than sifting through brochures or webpages, a person can ask an AI assistant a specific question: "What day is trash pickup in my neighborhood?" or "How do I file taxes if I'm self-employed?" The AI, if connected to official databases and documents, can generate a to-the-point response tailored to the query. This feels like a one-on-one conversation, much more engaging than reading an FAQ page. Governments are exploring this in various service areas. For example, some U.S. city websites now have chat windows where residents can inquire about anything from permit applications to library hours. The **Behavioural Insights**

Team (BIT) in the UK conducted an experiment with such chatbots on government sites and found that users appreciated the convenience and interactivity bi.team.

In developing countries, where government offices may be far away or backlogged, a digital assistant can be transformative. Consider a farmer in a remote area needing information on a new agricultural subsidy – an AI assistant accessible via a simple SMS or WhatsApp interface (text-only or voice) could instantly provide the details and even walk them through the application steps. Indeed, leveraging popular messaging apps is a smart strategy that some initiatives use to increase uptake. The **GovChat** platform in South Africa chose to integrate with WhatsApp and Facebook Messenger precisely because those were already widely used by citizens govinsider.asia. This meant no new app download or literacy in English was required – people could just message in their own language. As Eldrid Jordaan (GovChat's CEO) noted, *"we focused on messaging platforms that citizens already have... As we focused on access, we were able to make the voices of citizens a lot clearer to governments."* govinsider.asia. In addition to providing information, such platforms can gather feedback, effectively creating a two-way interaction. Citizens ask about services and also report their satisfaction or complaints, which the AI can compile for officials.

The personalization, however, must be balanced with **reliability**. An AI should ideally provide consistent answers that align with official policy. Unlike a human clerk who might have off days, software can ensure every user gets the same correct information – but only if the underlying system is properly updated and maintained. This calls for integrating AI assistants with real-time government data (for example, tying the chatbot to the latest voter registration database ensures it gives the correct registration status). The GenAI system proposed by Yun et al. (2024) addressed this via Retrieval-Augmented Generation, ensuring the answers were

grounded in actual policy text pmc.ncbi.nlm.nih.gov. Such designs significantly reduce errors compared to a free-form generative model.

Challenges & Mitigations:

The main challenges in policy Q&A and accessibility are **accuracy, currency of information, and trust.** If an AI assistant mis-answers a citizen's question about their legal obligations, the citizen could suffer harm (like missing a payment deadline or violating a rule unknowingly). Hallucination is a known risk: generative models might fabricate an answer that *sounds* plausible but is dead wrong canons.sog.unc.edu. To mitigate this, best practices include:

- Using **restricted knowledge bases**: e.g., the chatbot only draws from a vetted set of government documents, rather than the entire internet, to minimize random fabrications.
- Implementing **verification and human fallback**: Some government AI systems escalate to a human operator if confidence in an answer is low, or they provide the source (link or reference to the law) with the answer so users can verifypmc.ncbi.nlm.nih.gov.
- Continuous **testing and training**: Governments should audit the AI's responses periodically. This is especially important in multilingual setups – each language output should be evaluated by native speakers for correctness. The public also needs a clear way to report erroneous answers, prompting timely corrections.

There is also a need for **digital literacy** initiatives alongside deployment. Citizens should be informed that these AI assistants exist and encouraged to use them, but also made aware that they are a tool which can occasionally err. Building trust will be gradual: as people fact-check the AI's advice against real outcomes, confidence will grow. The BIT/Nesta experiment revealed a cautious optimism – people *would* use an AI for certain tasks, but for high-stakes advice (like

health or legal issues) they needed assurance of accuracy bi.team. Thus, starting with relatively low-stakes but high-volume queries (e.g., general service info) is a prudent way to prove the AI's value before expanding to more critical consultations.

In conclusion, GenAI offers a powerful toolkit to make policy information accessible: through plain-language translation, multilingual communication, and personalized assistance. When implemented with care to ensure reliability, these tools can demystify governance for ordinary people. In doing so, they help fulfill a core democratic principle – an informed citizenry. The case of GovChat and UMANG show real-world impacts like millions of new service users, which is encouraging for broader adoption. The next section will turn to how AI can not just inform citizens, but also listen to them and elevate their voices in the policy process.

2. Public Deliberation and Civic Engagement

AI-Generated Discussion Prompts and Facilitation

One opportunity for GenAI in civic engagement is to act as a facilitator or catalyst for discussion. Public forums – whether town hall meetings, community workshops, or online discussion boards – often benefit from well-crafted prompts that get people talking productively. Traditionally, moderators spend a lot of effort framing questions in neutral, accessible ways. AI language models, with their vast training on human dialogue, can assist by generating draft discussion prompts on specific issues. For instance, if a city council wants to consult residents on a new transit policy, an AI could propose questions like "*What concerns do you have about the new bus rapid transit plan?*" or "*How might the transit changes impact your daily commute?*", possibly even tailoring the wording for different neighborhoods using localized context. This can save time and also offer variations that a human moderator might not have thought of (reducing unconscious framing bias). AI could also translate these prompts into various languages to

include non-English-speaking residents in deliberations, aligning with the accessibility points discussed earlier.

Beyond prompts, AI could act as a **"virtual facilitator"** in online settings. One envisioned role is monitoring a live chat or forum for toxicity or off-topic threads and gently steering the conversation back on course. For example, an AI agent might intervene with a message: "Several people have brought up road safety. Let's explore that: what specific safety improvements do you suggest for the new transit routes?" This kind of nudge keeps the discussion focused and inclusive. While the technology for nuanced real-time facilitation is still developing, early trials (such as the PNAS study where an AI injected facts into divisive discussions) indicate that AIs can contribute to more informative exchanges pnas.org. However, careful design is needed to ensure the AI doesn't inadvertently carry its own bias in choosing when or how to intervene – a point we'll revisit in ethical considerations.

Summarizing Public Input and Meetings

Once citizen input is gathered, whether through a public hearing, an online survey, or social media engagement, the volume of data can be overwhelming. This is a classic "big data" problem but in qualitative form: hundreds or thousands of comments need to be read, understood, and synthesized to extract common themes and unique perspectives. Traditionally, government staff or consultants do this manually, which is time-consuming and may inadvertently cherry-pick or summarize in subjective ways. AI text summarization offers a scalable solution. Modern LLMs can summarize long transcripts or sets of comments into concise paragraphs, preserving the key points. Importantly, they can do **multi-document summarization** – meaning they can ingest a whole collection of responses and produce an overall summary.

For instance, imagine an online consultation on environmental policy that receives 5,000 written submissions. An AI could analyze all submissions and output: "**Summary:** A majority of citizens support stricter air pollution controls, citing health concerns and referencing recent asthma statistics. A significant minority (about 30%) worry about the economic impact on industries and jobs. Several unique suggestions emerged, such as incentivizing electric vehicles and improving public transit, as solutions. There is a common call for more public education on recycling and waste reduction." Such a summary provides decision-makers a snapshot of opinion distribution and key arguments. Tools like the **PublicInput GPT Comment Analysis** illustrate this capacity in practice – their system *"automatically identifies key themes"* in public comments and even tags sentiments (positive/negative) publicinput.com. By grouping comments by topic, the AI helps officials quickly see, for example, all comments related to "public transit" vs. "vehicle emissions" and gauge support levels for each publicinput.com.

Similarly, for in-person meetings, AI can transcribe the audio (speech-to-text, which is quite mature with AI), then generate minutes or summaries highlighting the discussion points and any decisions made. Some legislatures and city councils have started using AI to draft minutes of meetings, which are then lightly reviewed by humans – significantly speeding up the publishing of public records. Parliaments foresee using AI for *"real-time debate transcription [and] document summarization"*, as noted in WFD's guidelines babl.ai. In a community meeting context, this could translate to having a summary ready to share with attendees right at the end of a meeting, helping ensure a common understanding of outcomes.

The advantage for civic engagement is two-fold: (1) **Feedback to participants:** When people see their input summarized and acknowledged, they feel heard, which can increase trust. AI can even personalize feedback – e.g., thanking a citizen for a specific suggestion in follow-up

communication. (2) **Informing policymakers:** Officials get a digest of public opinion that is more digestible than raw data. It reduces the risk of ignoring silent majorities or fixating on a few loud voices, because the AI treats all inputs systematically. Indeed, AI can be instructed to surface *diverse viewpoints* – not just the majority sentiment. For example, it might report, "Most comments favored X, but a few articulated concerns Y and Z." This ensures minority opinions are explicitly presented rather than lost in the noise. Researchers developing *"Talk to the City"* emphasize designing AI summaries to **"capture diversity and nuance of opinion"**, not just the average view ai.objectives.institute.

A challenge here is ensuring the summary is accurate and not misleading. AI summarization is powerful, but if the model has biases or is not guided correctly, it might underrepresent certain feedback. One mitigation is the "argument mining" approach: extracting key points first (with AI) and then aggregating them. For instance, the AI could list distinct arguments it found (pro and con) and how frequently each was mentioned, rather than writing a narrative that could blur differences. This approach provides transparency – policymakers could see a list like: *"Argument A: mentioned in 45% of responses; Argument B: mentioned in 30%; Argument C (unique): mentioned in 5 responses,"* etc. Some AI tools indeed cluster similar comments so one can drill down into each cluster for more detailai.objectives.institute. This way, human analysts can validate that cluster summaries match the actual comments before relying on them. It combines efficiency with accountability.

AI in Structured Debates and Deliberative Democracy

Structured deliberative processes, such as citizens' assemblies, deliberative polls, or participatory budgeting forums, could also gain from GenAI support. These processes bring together diverse citizens to learn about an issue, deliberate, and often provide recommendations.

AI's role could be as an assistant to both organizers and participants. For example, during a multi-day citizens' jury on healthcare policy, an AI assistant could provide answers to factual questions that participants have ("what's the average cost of X treatment?", "how have other countries addressed Y issue?") by quickly retrieving information from a curated database. In this sense, the AI acts as a real-time research aide, ensuring the discussion is well-informed. It's important that such an aide is carefully controlled (using only vetted sources to avoid introducing misinformation).

AI can also help **moderators** by keeping track of the deliberation dynamics. It could alert the moderator if certain voices haven't contributed (maybe by analyzing the transcript and noticing that Person 5 spoke much less), allowing the moderator to invite that person to speak – indirectly, AI helps inclusivity. Or it might detect if the conversation has veered completely off topic and prompt a course correction suggestion. These uses are experimental but within reach given current AI capabilities in text analysis.

Another interesting application is **argument mapping**. In deliberation, especially on complex issues, it's useful to map out how different arguments relate – which ones support or counter others, where evidence was mentioned, etc. AI algorithms have been developed in research to perform *argument mining*, identifying premises and conclusions in text. Coupling that with generative abilities, AI could potentially produce an argument map diagram or a structured list of points for and against a proposal. IBM's Project Debater technology attempted something similar: after ingesting thousands of arguments from the public on a topic, it generated coherent narratives for both sides of the debate, including quoting some of the submitted points. Such output can then be presented to deliberation participants to ensure they

are aware of all facets of the issue. It's like giving the assembly a "brainstorm summary" that no single facilitator could have created in real time.

Encouraging Diverse Viewpoints: One of the promises of AI in deliberation is to help break echo chambers. Online, people often engage in homogeneous groups, but a well-designed AI platform could intentionally expose participants to counter-arguments or perspectives from different demographics. For example, an AI might say to a user, "I notice you emphasized economic concerns. Here is a perspective from someone who prioritized environmental concerns, to consider the other side." By serving as an impartial conveyor of viewpoints, AI can prompt users to reflect beyond their own experience, ideally leading to more nuanced opinions. This function must be handled carefully to avoid seeming intrusive or biased, but if done with transparency (perhaps framing it as "Here's what others are saying..."), it could enrich the deliberative experience.

Real-World Applications and Examples

Even though many of these ideas are in pilot or theoretical stages, some real-world instances shed light on AI's potential in civic engagement:

- **Taiwan's vTaiwan and Polis:** While not based on LLMs, Taiwan's vTaiwan platform used an AI-driven tool called Polis to facilitate large-scale discussions on tech policy. Participants shared opinions which the system clustered by agreement patterns, helping find consensus points in a debate that included thousands of citizens. Now, with GenAI, similar clustering can be accompanied by generative summaries in natural language.
- **"Talk to the City" prototype (global):** As discussed, this open-source project has been trialed to map public discourse on AI ethics. It successfully grouped opinions and provided summaries in multiple languages ai.objectives.institute, illustrating that such

tools can handle datasets from civic tech exercises and output human-readable reports. The fact that it's open-source means cities or civil society groups in developing countries could adapt it for their own consultations without huge costs. For example, a city in India or Kenya could input responses from a community survey into the tool and get an analysis in the local language.

- **PublicInput's GPT Analyzer (USA):** PublicInput, a civic tech company, integrated GPT to help local governments analyze open-ended feedback. A transportation department that received 3,000 comments on a road project used the tool to categorize sentiments and main topics in minutes rather than weekspublicinput.com. Officials could then easily see what the top concerns were (e.g., number of comments about pedestrian safety vs. cost). They reported that the AI categorization was a good first draft, which staff then reviewed – significantly cutting down manual labor while keeping humans in the loop for quality control.

These examples show a positive trajectory: AI helping to **scale up engagement** (more people consulted) and **scale down complexity** (making sense of many inputs). For developing countries, this could be game-changing. Often, resource constraints mean public consultation is either very limited or the input gathered never gets thoroughly analyzed. AI can provide the analytical muscle that was missing. Imagine a national government running a digital survey in multiple languages about its development plan, receiving 100,000 responses – with AI, even a small team could summarize that wealth of input and incorporate it into planning, thus legitimately claiming the policy was shaped by citizen voices.

Challenges and Considerations:

However, challenges mirror those in information provision, plus some unique to engagement:

- **Quality of AI summaries:** We must ensure AI doesn't introduce bias by under-reporting certain opinions. Bias can creep in if the model or prompt is not carefully calibrated. An oversight mechanism is needed. One approach is transparency: publish the raw data alongside the AI summary, enabling watchdog groups or interested citizens to verify the summary's fidelity.
- **Risk of homogenization:** There's a concern that AI, in summarizing, might "smooth out" the interesting edges of a debate. Minority or extreme viewpoints might get labeled as outliers and omitted, yet sometimes those perspectives contain important insights or at least need acknowledgment. It's essential that the parameters given to AI emphasize inclusion of divergent views. Otherwise, AI might inadvertently suppress minority voices – precisely what we want to avoid in democratic discourse.
- **Dependence on text input:** In many developing regions, literacy or internet access could limit who contributes written feedback. If AI tools only analyze text, there's a risk of excluding oral or in-person contributions. To counter this, governments could pair AI text analysis with initiatives to collect voice notes or conduct AI-assisted phone surveys (where speech is transcribed to text for analysis). This ensures even those who cannot type an essay can have their say, with the AI still able to process it.
- **Trust in AI outputs:** For recommendations coming out of a process (e.g., a citizens' assembly report drafted with AI help), participants and the broader public need to trust that the output genuinely reflects the deliberation. Building this trust might involve explaining the AI's role clearly: e.g., "This summary was prepared by an AI system that

reviewed all 10,000 comments; a team of facilitators then verified the content." Also, allowing citizens to see intermediate results (like clusters of opinions or draft summaries) and give feedback can create a collaborative feeling rather than a black-box scenario.

In sum, the use of GenAI in public deliberation is poised to **make engagement more efficient, inclusive, and insightful**, but it must be accompanied by transparency and care to truly strengthen democracy. When effectively used, AI can elevate public discourse by ensuring every comment is heard and by helping find common ground in polarized debates. The combination of human judgment and AI analytical power can lead to more responsive policymaking. As we embrace these opportunities, we must also squarely address the ethical and policy questions they raise – which is the focus of the next section.

3. Ethical and Policy Considerations

While GenAI opens exciting possibilities for civic engagement, it also presents significant ethical challenges and risks that must be managed to prevent inadvertent harm to democratic processes. This section examines these concerns in detail: from AI "hallucinations" and misinformation, to biases that may favor dominant groups, to the potential misuse of AI for undemocratic agendas. It also explores strategies to mitigate these risks, ensuring AI tools serve as a force for equity and trust rather than inequality and cynicism.

Accuracy, Hallucinations, and Misinformation

A foremost concern is that generative AI can produce information that is wrong or even entirely fabricated – yet delivered in a confident, authoritative tone. Unlike a search engine that only shows existing text, an LLM-based chatbot generates answers on the fly, which might include errors if the model's training data or prompt leads it astray. In a civic context, such **AI hallucinations** can be dangerous. A stark real-world example occurred with New York City's

experimental AI chatbot for public services. As reported, the chatbot gave users incorrect legal information – at one point asserting that it was legal for an employer to fire a worker for complaining about sexual harassment (which is absolutely not true)canons.sog.unc.edu. It even suggested a restaurant could serve cheese that had rat bites, a clear public health violation canons.sog.unc.edu. These kinds of mistakes aren't just harmless glitches; they could mislead citizens into illegal or harmful actions, or conversely, erode trust in public authorities once the errors come to light.

The risk of misinformation is amplified in domains where official information changes frequently (e.g., health guidelines during a pandemic, election procedures close to voting day) or where answers depend on specifics (e.g., eligibility for a benefit). If the AI's knowledge isn't up-to-date or it fills gaps with assumptions, people may act on outdated or false info. In developing countries, where citizens might have fewer alternative sources or lower media literacy, there's a vulnerability to believing whatever the government-provided chatbot says. That puts a special onus on accuracy.

Mitigation Strategies:

- **Retrieval-Augmented Generation (RAG):** This approach, used in some research prototypes, involves retrieving the relevant text from a trusted source (like the actual policy document or FAQ) and forcing the AI to base its answer on that pmc.ncbi.nlm.nih.gov. It greatly reduces hallucination since the AI isn't free-styling – it's quoting or summarizing an authoritative text. The system implemented by Yun et al. achieved high accuracy precisely by using RAG pmc.ncbi.nlm.nih.gov. Governments can implement this by maintaining databases of verified information (laws, guidelines, data) that the AI must use to answer questions.

- **Human oversight and verification:** Especially in critical applications, a human-in-the-loop can review AI outputs. For instance, if an AI assistant is answering questions about legal rights, the system could flag any low-confidence answers for a human expert to double-check before it reaches the user. Another model is to provide the AI's answer alongside the source material or a disclaimer, encouraging users to verify. The BIT/Nesta trial found people are concerned about trusting AI for important info bi.team; thus transparency about sources can help.
- **Regular updates and training:** The AI's knowledge should be synced with the latest official information. If a new law is passed or a procedure changes, the AI model (or its retrieval database) must be updated immediately. This may require organizational processes to ensure the tech team is in the loop on policy changes.
- **Limiting scope initially:** Many recommend deploying AI in narrow domains first (e.g., answering straightforward FAQs) and monitoring performance. Only after establishing reliability should the scope expand to more nuanced queries. This controlled rollout can catch issues early.

Despite best efforts, some misinformation may slip through, so it's vital to have a **redress mechanism**. Citizens should have an easy way to report an answer they believe is wrong ("report this answer" button), and there should be a process to correct the system and inform the user of the correction. This feedback loop not only fixes errors but also signals to users that the AI's output is not infallible, inviting a healthy skepticism that actually bolsters long-term trust once improvements are made.

Bias and Representation

AI models learn from data that often reflect societal biases – and they can inadvertently perpetuate or amplify these biases. In a civic context, this can manifest in several troubling ways:

- **Ideological Bias:** If an AI system is used to provide information or moderate content, it might lean towards certain political ideologies. A 2025 study found that ChatGPT's responses showed systematic deviations toward left-leaning perspectives in the U.S. context phys.org. On some topics, it would refuse to produce arguments from a right-leaning viewpoint, citing misinformation concerns, while readily producing left-leaning arguments phys.org. This kind of imbalance, if present in an AI deployed by government, could lead to accusations of partisanship or censorship. For example, an AI assistant might downplay conservative-leaning citizen concerns about a policy if its training data skewed liberal – or vice versa. Such bias could **"distort public discourse and exacerbate societal divides,"** as researchers warned phys.org. Clearly, an AI that is perceived as politically biased will undermine trust in any civic engagement effort.
- **Cultural & Racial Bias:** AI can also reflect majority cultural norms while misunderstanding or marginalizing minority group expressions. There is evidence that chatbots can pick up on the user's dialect or writing style and alter their responses – in one MIT study, chatbots showed less empathy in responses to users it assumed were Black, indicating a harmful bias in the model's interaction stylenews.mit.edu (from context, likely referencing the MIT News study on race and empathy). If a citizen from a marginalized community feels the AI "doesn't get" their issues or responds in a less respectful manner, it could alienate those who most need empowerment. Bias in language could be as subtle as not understanding slang or idioms used in a particular community, leading the AI to give nonsensical answers that discourage further participation.

- **Data Bias in Summarization:** When using AI to analyze public input, if the majority opinion is one way, a naive AI summarizer might ignore minority opinions altogether. For instance, if 80% of comments on a budget plan are from men in urban areas, and 20% from women in rural areas, the AI might focus on the majority's priorities unless specifically guided to also highlight differences. This can reinforce existing inequalities – the voices of historically underrepresented groups might once again be drowned out, now by an AI's synthesis. There's also risk of a **feedback loop**: if an AI assistant consistently provides answers or content aligned with majority viewpoints or the government's perspective, people with minority opinions might stop engaging (feeling it's futile), leading to less diverse input, which further biases the AI's knowledge base.

  Mitigation Strategies for Bias:

  Addressing bias requires intervention at multiple stages – data, model, and interface:

- **Diverse and Inclusive Training Data:** Ensuring the AI's training data (or fine-tuning data) includes a balanced representation of perspectives, dialects, and demographic groups can help it respond more equitably. For a policy chatbot, this might mean training it on questions posed by people of different ages, education levels, and communities, and including local idioms or examples. It also means scrubbing training data for harmful biases (e.g., if the data contains derogatory associations or stereotypes, those need to be corrected). Projects like developing local language models in Nigeria are vital – they create AI that understands Nigerian contexts, not just American or British ones dig.watch.
- **Algorithmic Fairness Techniques:** There is a growing field of AI research on fairness which can be applied. For example, classifiers or filters can be used to detect if the AI's

output or summaries disproportionately exclude certain keywords or topics that are proxies for minority issues. If an imbalance is found, the system can adjust (some use re-weighting of outputs, or ensuring at least one item from each cluster of opinions is included in summaries). In the context of ideological bias, one could explicitly prompt the model to provide a range of viewpoints or even use multiple models tuned differently and compare outputs. The phys.org article suggests *"transparency and regulatory safeguards to ensure alignment with democratic values"* phys.org, implying that external auditing of AI for bias should be standard.

- **Human Review Panels:** In critical deployments, have a diverse group of humans review the AI's responses and summaries periodically. For instance, a panel of community representatives could evaluate whether the answers given by a city's service chatbot are culturally appropriate and unbiased. Their feedback can then guide further fine-tuning. This participatory approach in AI oversight not only catches biases but also signals to the community that their perspectives are valued in shaping the tool.
- **Designing User Interfaces for Inclusion:** Bias can also be mitigated by how the AI system interacts with users. For example, letting users specify their language or preferences can avoid the AI making assumptions. In engagement platforms, showing multiple summaries or asking users "Did this summary capture your viewpoint?" can help validate that different angles are covered. If the AI is used to propose policy options, it could be required to present pros and cons for each option equally, to avoid slanting the outcome.

Finally, there's an ethical duty to **avoid over-reliance on AI** especially in areas like policing content or deciding which public inputs are relevant. Automating too much can

inadvertently encode systemic biases into decisions. The WFD guidelines for parliaments emphasize **"human oversight in AI decision-making, cautioning against over-reliance"** babl.ai. In practice, this means AI tools should inform and assist, but final judgments – especially those affecting rights or resource allocation – should involve human deliberation where biases can be discussed and checked transparently.

Manipulation, Astroturfing, and Agenda Capture

Perhaps the most dystopian scenario is the misuse of generative AI to undermine, rather than enhance, democratic participation. Malicious actors (be it hostile governments, political operatives, or lobby groups) could employ AI to flood civic spaces with false or manufactured input. This goes beyond the AI giving a wrong answer; it's about AI being used to distort the *process* of engagement.

One major threat is **astroturfing at scale**. Astroturfing means faking grassroots movements – making it appear as if a lot of citizens support or oppose something, when in reality it's orchestrated. With GenAI, creating hundreds of fake personas each with a unique, coherent comment on a public consultation is alarmingly feasible lawfaremedia.org. Back in 2016, propagandists had to use clumsier bots that repeated the same messages, which platforms started to detect and block lawfaremedia.org. But today's AI can generate endless variations of an opinion, complete with personal anecdotes or localized details, making fake submissions almost indistinguishable from real ones lawfaremedia.org. As Guggenberger and Salib (2023) noted, *"ChatGPT, or similar AI, will make effective astroturfing practically free and difficult to detect… They can mimic humans, producing an infinite supply of coherent, nuanced, and entirely unique content."* lawfaremedia.org. An AI-run fake Facebook account might behave normally for months (sharing recipes, chatting about sports) to build credibility, then start

voicing strong political opinions to sway otherslawfaremedia.org. This kind of infiltration can greatly mislead both the public and policymakers about what the "public" wants. A city council might receive 5,000 comments supporting a policy that were actually AI-generated at the behest of a special interest group, drowning out 500 genuine critical comments.

Another angle is **deepfakes and impersonation**. While this report focuses on text, generative AI can also produce synthetic voices or images. There have been incidents of deepfake audio of officials or fake "citizen" videos spreading misinformation. In the context of civic engagement, one could imagine a deepfake video of a community leader endorsing a particular government initiative being circulated to manufacture consent. Or AI-generated faces and profiles populating a virtual town hall to give the illusion of consensus. These are not just hypothetical – researchers and security experts are already warning how *"generative AI can flood the media, internet, and even personal correspondence, sowing confusion for voters and officials alike."* journalofdemocracy.org. Freedom House has also documented how some authoritarian regimes use swarms of bot accounts (sometimes human-curated, increasingly AI-driven) to drown out dissenting voices online and project an image of mass support for the regime carnegieendowment.org.

Impact on Trust: The misuse of AI in these ways could deeply erode trust in democratic institutions and civic processes. If people start suspecting that any online discourse might be bot-driven, they may disengage entirely or distrust legitimate outreach by government. Lawfare's analysis predicts that once **"bot pundits become indistinguishable from humans, [people] will start to question the identity of everyone online, especially those with whom they disagree."** lawfaremedia.org. Healthy debate relies on assuming others are genuine citizens voicing their views; if that assumption breaks, deliberation falters. Moreover, should a scandal arise where a

government or political actor is caught using AI sock-puppets to push an agenda, it can lead to public outrage and cynicism (e.g., "town hall meetings are just theater, they already rigged the feedback").

Mitigation Strategies:

This is a complex problem requiring technical, procedural, and legal responses:

- **Authentication of participation:** One defense is to implement better verification for civic participation platforms. For example, when submitting feedback on a government survey, perhaps one must log in with a digital ID or at least pass a CAPTCHA or other bot-detector. However, sophisticated bots can sometimes pass such tests, and requiring strong ID might deter real participants over privacy concerns. It's a balance – some identity verification (even an email or phone confirmation) raises the bar against mass fabrication.
- **AI detection tools:** Ironically, AI can help detect AI. There are algorithms being developed to identify AI-generated text by certain telltale patterns or using watermarks embedded by the text generators. Governments could employ these on incoming communications to flag which ones likely came from a bot. For instance, if a flood of comments arrives with suspiciously similar language distributions or metadata, they can be set aside for manual review. No detector is foolproof (and an arms race is ongoing as AI gets better), but it can catch unsophisticated attacks and reduce noise.
- **Transparency in public consultation:** Releasing the dataset of public comments (with personal info redacted) for anyone to analyze can crowdsource the detection of astroturfing. Journalists or civic hackers might notice patterns that a government missed. This transparency allows independent verification of the authenticity of engagement.

- **Regulations and norms:** At a higher level, new policies may be needed to penalize the use of generative AI for deceptive political influence. Already, jurisdictions are looking at requiring labels on deepfake videos or political bots. Election laws might need updating to treat mass AI-generated campaigning as a form of fraud. Of course, enforcement is tricky across borders (a foreign adversary won't heed domestic laws), but setting norms is still important. On the flip side, governments themselves should commit to not using AI in manipulative ways. There is a role for ethics: just because a ruling party could generate supportive comments doesn't mean it should – doing so would ultimately damage its legitimacy if exposed.
- **Public awareness:** Educating citizens that not everything online is as it seems is part of digital literacy. If people recognize the possibility of bots, they might be more critical thinkers (though too much skepticism can lead to nihilism, so again a balance). Initiatives like media literacy campaigns now may include segments on AI and how to spot unnatural activity.

We should note that these concerns are not limited to online; even in AI-assisted face-to-face processes, one must ensure the AI tools themselves are not biased to favor one outcome (for instance, an AI summarizer subtly framing an issue to sway the assembly). Thus, an ethical AI deployment will ideally be externally audited or open-sourced so that stakeholders can examine if there's any agenda baked into the code.

Ensuring Equity and Preventing Inequality Reinforcement

A recurring theme is the risk that AI could reinforce existing inequalities if not intentionally designed to counteract them. For example, wealthier or more tech-savvy citizens might benefit more from AI-enabled services, while others are left behind – a digital divide

issue. Or biases in AI could systematically disadvantage certain groups in subtle ways, as discussed. To prevent this, bias mitigation must be proactive.

Effective strategies include:

- **Co-creation with marginalized groups:** The people who are often left out of governance (rural villagers, slum dwellers, linguistic minorities, persons with disabilities) should be involved in the design and testing of these AI tools. GovChat's success in South Africa was partly attributed to its collaboration with government and focus on accessibility from the ground up govinsider.asia. By getting feedback from these communities early, developers can ensure the AI addresses their needs (e.g., adding a voice output for those who can't read at all, or simplifying language further for those with limited education). This helps avoid a scenario where an AI tool only serves the elite.
- **Localized AI development:** Relying solely on global AI models may perpetuate a Western-centric or urban-centric worldview. Encouraging local AI talent and localized models (like the Nigerian and Indian examples) builds tools better aligned with local contexts and values. It's a form of capacity building – developing countries need the expertise to adapt AI to their context, rather than just import solutions. International aid and partnerships can support this by providing resources and knowledge transfer for creating local language datasets, etc.
- **Continuous monitoring for disparate impact:** Once an AI system is live, governments should track metrics to see if certain groups are not using it or not benefiting. For instance, if an AI chatbot is mostly used by men and not women in a community, investigate why – maybe women have less access to phones or are less aware of it. Or if

the satisfaction ratings for the chatbot are lower in one region or ethnicity, treat that as a sign to improve the system for that group. This is analogous to how public agencies analyze service usage across demographics to identify gaps.

- **Adherence to Ethical Frameworks:** Global bodies like UNESCO and the OECD have released AI ethics guidelines emphasizing fairness, accountability, and **non-discrimination** oecd.org. Governments should incorporate these principles into their procurement and deployment of AI. For example, the OECD AI Principles call for AI that respects human rights and democratic values, which implies not using AI in ways that suppress dissent or violate privacy unjustly oecd.org. UNESCO's recommendation on AI ethics (2021) specifically warns against AI that exacerbates inequality and calls for inclusion as a key objective unesco.org. By embedding these into law or agency rules, there's a basis to hold AI systems accountable.
- **Independent Audits and Civil Society Oversight:** Just as budgets are audited, algorithms can be audited. Governments could mandate periodic algorithmic audits for public-facing AI systems to check for bias and accuracy. These could be done by independent experts or multidisciplinary panels (including ethicists, technologists, and citizen representatives). Public reports of these audits would enhance transparency and trust. Civil society organizations, like digital rights groups, should also have avenues to question and review government AI systems. Their vigilance can catch issues officials miss.

One positive aspect is that unlike human bureaucrats who might harbor biases unconsciously, AI systems' biases can potentially be identified and corrected if one has access to the system. It's easier to adjust a line of code or a training dataset than to change a lifetime of

human prejudices. So in some sense, AI offers a chance to *reduce* human bias in governance if we actively manage it. But that's contingent on the political will to do so.

Public Trust and Legitimacy

At the heart of all these considerations is the impact on public trust. Will GenAI in civic engagement strengthen citizens' faith in democratic institutions, or will it make them more skeptical? The outcome depends on how transparently and responsibly AI is implemented:

- If citizens find that AI-powered services reliably help them, save them time, and treat them fairly, they will likely view the government as more efficient and responsive, boosting trust. For example, GovChat's ability to rapidly connect people with services and allow them to rate those services created a sense of empowerment in South Africa – as Jordaan put it, *"We are able to see democracy come forth between the ballot boxes on a daily basis."* govinsider.asia. This daily responsiveness can build confidence that government cares about feedback beyond just elections.
- Conversely, any high-profile failures or abuses of AI will draw backlash. The New York City chatbot fiasco, once publicized, could make citizens wary ("if even NYC's AI gives bad info, can we trust ours?"). Therefore, managing expectations is crucial. An executive summary of a government report might explicitly state where AI is used and how it's supervised, to preempt fear of a "rogue AI" making decisions.

Public trust can be enhanced by **openness**. If a government rolls out an AI tool, explaining its purpose, how it works, what data it uses, and allowing people to opt out or give input, will make it seem like a public good rather than a sneaky surveillance or control tool. Especially in developing countries with histories of mistrust, transparency is key. It might be

wise to start with co-branded initiatives (e.g., with trusted NGOs or international orgs overseeing) to lend credibility.

Finally, capacity-building among public officials themselves is needed. They should understand AI's limits and not over-rely or over-promise. Policies should define accountability clearly: if an AI system misinforms, which office or person takes responsibility and remedies it? Clarity on this helps maintain accountability – citizens know there is still a human accountable, not a black box they can't challenge.

In summary, the ethical deployment of GenAI in civic contexts requires a concerted effort to mitigate harm: technical fixes, policy guidelines, oversight structures, and a culture of continuous improvement. Many organizations have begun sketching these solutions – for instance, the Westminster Foundation for Democracy's guidelines urging **"transparent, accountable"** AI with **"independent oversight bodies"** and **"clear policies on fairness and human rights"** babl.ai. These translate into concrete measures like algorithm registries, bias testing, and human-in-the-loop controls. Implementing such measures will be critical to ensure GenAI tools truly serve democracy. In the concluding section, we synthesize these insights into specific policy recommendations for governments and stakeholders looking to responsibly integrate AI into democratic processes.

Conclusion and Policy Recommendations

Generative AI holds significant promise for enhancing civic engagement and strengthening democratic governance. It can lower barriers to information, helping citizens understand policies and access services in familiar language. It can scale up public deliberation, making sense of thousands of voices and ensuring leaders hear the many rather than the few.

Importantly, it can do so in resource-constrained settings, offering developing countries a chance to "leapfrog" traditional hurdles in citizen outreach and government transparency pmc.ncbi.nlm.nih.gov techletter.co. However, this promise comes with profound responsibilities. If mismanaged, the same technologies could misinform citizens, amplify biases, or be weaponized to manipulate public opinion – outcomes that would erode trust in democracy.

The key to tipping the balance towards positive outcomes lies in **governance of the technology itself**. As this analysis has shown, technical solutions and policy frameworks must go hand in hand. Below, we offer specific policy recommendations for governments, international organizations, and civil society to ensure that GenAI's integration into civic processes is done ethically, inclusively, and effectively:

1. Establish Clear Ethical Guidelines and Standards: Governments should develop or adopt comprehensive guidelines for the use of AI in public sectors, akin to the WFD's "Guidelines for AI in Parliaments" babl.ai babl.ai. These should articulate principles of transparency, accountability, fairness, and human oversight. For example, standards could mandate that any AI system interacting with the public must disclose it is an AI and not a human, and provide a way to contact a human official. AI decisions or summaries that inform policy should be documented and reproducible. By setting these rules upfront, agencies have a clear framework to follow, and the public knows what to expect. International bodies (UN, OECD, etc.) can help by providing model guidelines and encouraging their adoption, ensuring consistency globally.

2. Invest in Multilingual and Inclusive AI Systems: To truly benefit developing nations and marginalized groups, targeted investment is needed in AI that supports local languages and contexts. Governments should fund the creation of open datasets and models for

their national and regional languages. Collaboration with universities and tech companies can accelerate this. The aim is to avoid an "English-first" AI paradigm; instead, prioritize tools like translation bots, voice assistants, and chatbots that operate in the languages of rural and underserved populations. For instance, expanding projects like India's UMANG voice assistant to cover all major Indian languages, or Africa-wide initiatives to build LLMs for Swahili, Yoruba, Amharic, etc., will make civic tech accessible to hundreds of millions more people aisight.fractal.ai dig.watch. Donor agencies and development banks could earmark grants for such AI localization efforts, seeing it as digital infrastructure for democracy.

3. Pilot and Evaluate AI Civic Tools in Controlled Settings: Before scaling up, governments should run pilot programs for AI chatbots or deliberation tools, ideally in partnership with academia or civic tech organizations that can rigorously evaluate them. Select a specific use-case (e.g., a city chatbot for waste collection queries, or an AI summary tool for one public consultation) and monitor its performance, user satisfaction, and any issues. Use surveys and focus groups to gauge if people felt better informed or heard. Importantly, involve a diverse user base in testing – including those from varying literacy levels and communities – to see if the tool works equitably. Publish the results (both successes and problems) to build knowledge. Gradual rollout with iterative improvement will help avoid major failures and allow the public to warm up to the technology. It also provides evidence to secure buy-in from stakeholders or legislators for broader implementation.

4. Implement Robust Fact-Checking and Update Mechanisms: Any AI system providing information to citizens should incorporate safeguards against misinformation. This includes using **retrieval-based responses** with citations to official sources whenever possible pmc.ncbi.nlm.nih.gov, and regularly updating its knowledge base. Agencies must assign

responsibility for feeding the AI new or amended policies (for example, when a new law is passed, the relevant ministry should promptly update the AI's documents). Additionally, integrate these systems with existing open data portals where feasible, so they pull the latest data. In high-stakes domains (legal, health, voting rights), consider a hybrid approach: AI drafts an answer and a trained human verifies it before it's delivered. While this may slow responses slightly, it's worth the accuracy for critical information. Over time, as the AI provenly learns, the human oversight can be dialed back.

5. Ensure Transparency and Public Awareness: Openness builds trust. Governments using AI for citizen-facing roles should be transparent about it. This means publicizing what tools are in use, what their capabilities and limits are, and what data they rely on. For example, if a council uses an AI to summarize public comments, it should announce that and perhaps provide a link to the raw comments and the AI's summary for comparison. When chatbots are launched, accompany them with FAQs clarifying that it's an AI, how it works, and how user data is handled. Moreover, educate the public on the purpose: emphasize that the goal is to serve them better (not to replace humans entirely or to snoop on them). Public-awareness campaigns or community workshops on "AI in our city hall" can demystify the technology. By engaging citizens in understanding the tool, governments can preempt fears and also get valuable feedback. In essence, treat AI deployments as a partnership with the community, not a top-down imposition.

6. Protect Against Misuse and Manipulation: Proactive measures are needed to secure democratic processes from AI-driven manipulation. Election commissions and public consultation processes should update their protocols to guard against fake AI-generated inputs. This could involve technical filters for bot detection on e-participation platforms, rate-limiting

input submissions, and requiring one person-one submission authentication where appropriate. In rulemaking public comment periods, agencies might use AI to flag clusters of near-duplicate submissions (a sign of automation) for separate review. Legally, governments should consider extending campaign finance or anti-propaganda laws to cover the use of AI – for instance, obligating disclosure if AI was used to generate large volumes of outreach or comments. Collaborate internationally to share intelligence on malicious AI usage (since fake campaigns often cross borders). On the flip side, use AI defensively: as mentioned, deploy AI tools to identify deepfakes and bot networks that aim to disrupt public opinion. It's an arms race that democratic institutions must be prepared for, ideally with help from the tech sector (which can provide detection tools) and civil society watchdogs.

7. Maintain Human Oversight and Accountability: No matter how advanced AI becomes, human officials must remain accountable for decisions. Policy recommendation: for any significant decision or output influenced by AI (e.g., an AI-drafted policy summary that will guide a vote), require a human sign-off. Create internal guidelines that delineate which tasks may be fully automated and which require review. By keeping humans in control of final outcomes, we ensure there's a clear point of accountability. If an AI gives a citizen a wrong answer and it causes harm, the agency running the AI should take responsibility and rectify it; it cannot shirk blame onto the algorithm. This should be communicated to citizens: that the AI is an aid, but government is still responsible for the services and info provided. Such assurances, backed by practice, will maintain public trust that democracy is about people serving people, with AI as a tool.

8. Foster Collaboration and Knowledge Sharing: The field of AI for civic engagement is new and evolving. Governments, especially in developing countries, should not have to

navigate it alone. International organizations, think tanks, and cross-country networks (like the Open Government Partnership) can facilitate the sharing of best practices, success stories, and cautionary tales. Developed nations deploying these tools should share open-source software or frameworks that others can adapt (with appropriate localization). Multilateral forums could establish an "AI and Democracy" working group to continually study and advise on emerging issues (much like election observation missions, but for AI usage). Additionally, involve multidisciplinary expertise – political scientists, ethicists, technologists, and importantly citizens – in evaluating and guiding AI implementations. This broad collaboration will help democracies collectively learn and improve the use of GenAI, while presenting a united front on values like equity, privacy, and freedom of expression in the AI era.

9. Empower Oversight Institutions: Parliaments, judicial bodies, or independent commissions (like human rights commissions or data protection authorities) should be empowered to oversee the use of AI in government. This might mean giving them the mandate and technical capacity to audit algorithms, review procurement of AI systems for bias or rights implications, and hear grievances from the public. For example, a "Public Digital Ombudsman" could be created to handle complaints about AI-powered public services. If a citizen feels an automated system treated them unfairly or didn't understand their input, they have recourse to an authority that can investigate and demand remedies. Empowering such oversight not only protects citizens but can provide systemic feedback to improve AI deployments over time.

10. Promote Digital Literacy and Inclusion: Finally, alongside introducing AI tools, invest in raising the digital literacy of citizens and civil servants. Training programs for public officials are needed so they understand how to use AI outputs critically (e.g., not taking an AI summary as gospel if it contradicts their knowledge). For citizens, incorporate into civic

education how to engage with digital platforms, how to verify information, and an understanding of what AI can and cannot do. This helps manage expectations and reduces the risk of manipulation. Inclusion efforts such as community tech hubs or mobile units can help rural or disadvantaged communities access these new digital channels. The promise of AI in democracy can only be realized if all segments of society can partake; otherwise, we risk creating a new elite of AI-empowered citizens versus others.

In conclusion, generative AI, like any powerful tool, can bolster democracy if wielded with care and foresight, or it can weaken it if left unchecked. The optimistic view – supported by early successes in places ranging from South Africa's GovChat to Taiwan's deliberative innovations – is that AI can make government more responsive, participatory, and transparent, thus renewing citizens' faith in democratic institutions. To achieve this globally, especially in developing countries, intentional actions must be taken to overcome resource gaps, language barriers, and existing inequalities. Responsible implementation, guided by ethical principles and continuous oversight, is imperative babl.ai phys.org. By following the recommendations above, policymakers and stakeholders can help ensure that the rise of Generative AI becomes a story of democratic revitalization – where technology is harnessed to amplify the voices of the people and to inform and improve the decisions that affect their lives, ultimately strengthening the social contract between governments and citizens.