OCTOBER 2025

REPORT ON

Mapping LLM Tools for Public Discourse, Pluralism & Social Cohesion

AUTHORS
Matt DeVerna, David J. Grüning, Jen Hickey, Adnan Jaber, Julia Kamin, Brendan A. Miller, Rehan Mirza, Jiaxin Pei, Victoria Stanski

PLURALITY INSTITUTE
COOPERATE ACROSS DIFFERENCE

Council on Technology and Social Cohesion

Prosocial Design Network

ACKNOWLEDGEMENTS This report was made possible by Plurality Institute, The Council on Technology & Social Cohesion, Prosocial Design Network, with generous support from Google.org, Jigsaw, and The John Templeton Foundation. We are also grateful to all the participants of the February 2025 LLMs and Public Discourse Convening who contributed to this map.

INTRODUCTION: MOTIVATION, EVENT, PROCESS, DIMENSIONS

LLMs are powerful tools that can be leveraged to support and improve processes that were once dependent on either human cognitive work or less agile automation. In February 2025, Plurality Institute and Council on Technology and Social Cohesion gathered over seventy researchers and technologists to answer the question: how can we harness that enormous power to foster public discourse, pluralism, and social cohesion in digital spaces?

As part of that gathering, participants collectively mapped out existing and potential uses of LLMs to foster public discourse that could either be integrated into platforms - including mass social media, deliberative platforms, online groups and news comments sections - or function as independent online tools.

That collective work resulted in an initial **public dataset** of 70 LLM tools.

This Mapping Report is a companion to that dataset. It shares key insights and maps of the current landscape of LLMs for Public Discourse.

We offer these insights and maps not as the definitive word on LLM tools for Public Discourse; that would be impossible given this is a quickly evolving field. Instead we hope this map and report will begin the conversation of the existing and potential ways LLMs can be put in service to build a world where discourse, pluralism and social cohesion can thrive.

THE EVENT & BUILDING THE MAP

The event that inspired this report, hosted by Plurality Institute and Council on Technology and Social Cohesion, brought together researchers and technologists in Berkeley, California on February 27th and 28th to share and catalyze the creation of LLM tools to foster discourse.

The convening included an evening panel showcasing LLM projects, followed by a daylong workshop where participants mapped the landscape of potential tools, identified promising applications, worked in teams to develop new

projects and, finally, used approval voting to collectively award $50k to support those projects.

One of the primary goals of convening was to draw on the collective knowledge and creativity of the diverse attendees to produce a map of existing and potential uses of LLM tools that foster public discourse and pluralism. Our intention in building the map was to:

- Lay out where there are existing tools and work being done to build LLM tools for public discourse
- Reveal where there are opportunities to develop uses that can have a meaningful impact on public discourse
- Ultimately inspire innovative work and give direction on where our collective efforts can focus and have maximum impact in the future.

The map takes two forms: a dataset that itemizes each LLM tool; and graphical depictions to visualize the landscape of existing tools, provided in this report.

Both required that we first identify the relevant attributes for each LLM tool to be captured as columns in the dataset and as dimensions - the latitudes and longitudes, so to speak - of the maps. To select those key attributes and dimensions, participants joined a pre-event video call to surface the most salient elements of LLM tools they wanted represented. The full list of dimensions selected are discussed below.

Participants began to build the dataset before the event and continued on the day of the gathering during an interactive session in which participants filled out Cards that briefly described LLM tools that are "**In use"**, **"In development"** or **"Ideas"** - and added those cards to a map placed around the room. That map captured two initial dimensions: **Space** where the LLM tool would engage - e.g. on mass social media platforms or in deliberate platforms - and **Time** it would engage relative to users' actions - i.e. proactively, interactively or reactively.

With several exceptions, all the LLM tools "In Use" and "In Development" tools collected up to and during the convening can be found in the dataset. The exceptions are tools that we could not find adequate documentation on. Following the event, the authors of this report reached out to participants to help further refine the dataset which we present now with this report.

WHAT ARE LLMS?

Large Language Models (LLMs) are a specialized type of AI and Machine Learning, which all use algorithms to learn from data and make predictions or decisions. LLMs are unique in that they use vast amounts of unstructured data to understand and generate language. Other terms that are near synonyms for LLMs include GenAI, GPT (as in ChatGPT) and Foundation Models.

THE MAP ATTRIBUTES AND DIMENSIONS

Event participants identified over a dozen attributes of LLM tools that could serve as insightful dimensions for the map. Several - for example "Potential Impact" - were noted as particularly important to include, yet given the difficulty of objectively assessing them, we decided it was beyond the scope to include those dimensions in the maps and this report.

Below are descriptions of the Primary Dimensions that are included in the graphic maps, as well as the Secondary Dimensions that can be found as additional columns in our dataset.

PRIMARY DIMENSIONS

TIME

LLM tools can be called on to support users at various stages in their engagement with content and other users on a platform, including as they initially enter an online space, encounter and interact with content, respond to other users, etc. While there is no uniform timeline for user engagement, we broadly mapped LLM tools temporally, placing tools that play a role

- **Proactively**, before users engage with content and other users,
- **Interactively**, during engagement as, for example, they post, repost or respond to content
- **Reactively**, after engagement when, for example, they have posted something that is either inflammatory or praiseworthy, or a tense exchange has emerged

SPACE

Digital spaces are varied and some LLM tools may only make sense in the context of a particular online space. For the event we mapped LLM tools that can be integrated in the following digital spaces.

- **Deliberative platforms** (see Digital Platforms box for more detail on these platforms)
- **Comments sections** in news and article platforms
- **Mass social media** platforms
- **Online groups and communities** with volunteer moderators (often within mass social media platforms)
- **Miscellaneous** (i.e. any LLM tool outside of the spaces above)

Given the parallels between comments sections, online groups and mass social media platforms, we collapsed those three spaces into one for the report.

GOAL

While the uniting theme of all the LLM tools in our dataset is fostering discourse and pluralism, each tool took on a more specific goal toward that larger aim. To categorize those goals, we asked participants, post hoc, to briefly summarize what their LLM tools aimed to achieve, and then sorted those goals into eleven categories. We then noticed those categories each could be organized into one of the four broad "Civic Signals" that New_Public uses to classify the purposes of online civic spaces: Welcome, Connect, Learn and Act.

WELCOME	CONNECT	LEARN	ACT
Ensure safety Promote inclusion & diversity	Support healthy conversations Foster flourishing communities Bridge divides and reduce polarization	Facilitate deliberation Expand knowledge & awareness Reduce misinformation	Support consensus building or decision making Empower citizens and communities Support governance

FUNCTION

Likewise, LLM tools cover a range of core functions designed to enhance how people engage, understand, and share information within public spaces. For the purpose of this report, we categorize the functions as:

- **Content analysis and summarization** features that involve distilling large volumes of discourse, mapping opinions, and making sense of public input into coherent insights.
- **Moderation and safety** tools focus on keeping discussions civil and constructive through automated toxicity detection, real-time feedback, and moderation support.
- **Facilitation and collective intelligence** enable meaningful dialogue at scale through features like guided discussions or providing tools for scale/engagement that can inform decision making.
- **Personalized and inclusive participation** tools tailor the experience, adapt communication, and enable natural, inclusive dialogue for specific users or various contexts.
- **Information retrieval and fact-checking** capabilities surface reliable sources, verify claims, and provide core support for accurate informative discourse drawing on existing resources.

SECONDARY DIMENSIONS

- **Status:** We differentiate between LLM tools that have been fully developed and are actively "In Use," tools that teams are working on and are still "In Development," and tools that are "Ideas" that were generated by participants as part of the convening.
- **Integrated Tool, Self-Standing or Full-Stack Platform:** We likewise make a distinction between tools that are designed to be **integrated** into a digital platform (either by the platform itself or as a third party tool), tools that **operate independently** of other platforms, and tools that function as **full-stack** platforms themselves.
- **Agency:** We delineate how much a given LLM tool empowers individuals as opposed to potentially diminishing their agency (i.e. ability to freely choose and act).
- **Human Oversight:** Finally, we label LLM tools to distinguish the degree to which they operate under the supervision of a human or act autonomously.

MAIN MAPS

What goals and functions are LLM tools for discourse being developed and deployed? And are they engaging with individuals upstream, proactively fostering discourse - or at the point of engagement, or even reactively?

We discuss the landscape of LLM tools for public discourse in detail in the narrative section of the report. These maps give a birds-eye view of where current efforts are concentrated in using LLMs to foster discourse and social cohesion. The tools are categorized under the five main goal categories and labeled with their respective sub-goals.

For each digital space that we include in the map — **Social Media Platforms**, **Deliberative Platforms**, and **Miscellaneous** section - we chart when LLM tools in the dataset operate temporally (proactively, interactively or reactively), and toward what goal (Maps 1a, 1b, 1c) or with what function (Maps 2a, 2b, 2c).

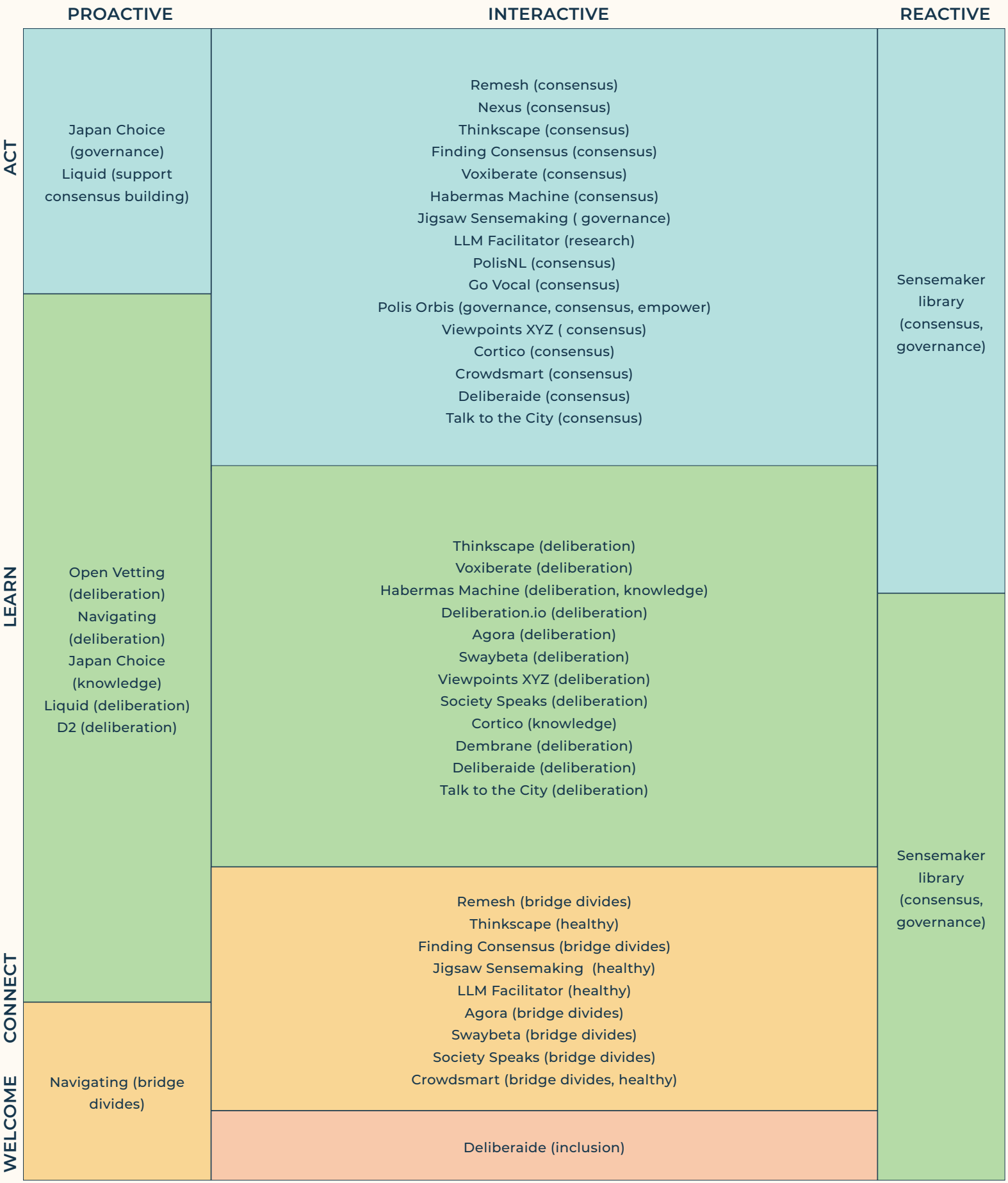
We use Mosaic charts to reveal where current efforts are concentrated, and conversely where fewer projects exist. For example, among projects using LLM tools to foster discourse on social media platforms, most engage users at the moment of interaction in order to build connection. More broadly, there is relatively less work directed at developing LLM tools that operate proactively.

GOALS + TIME

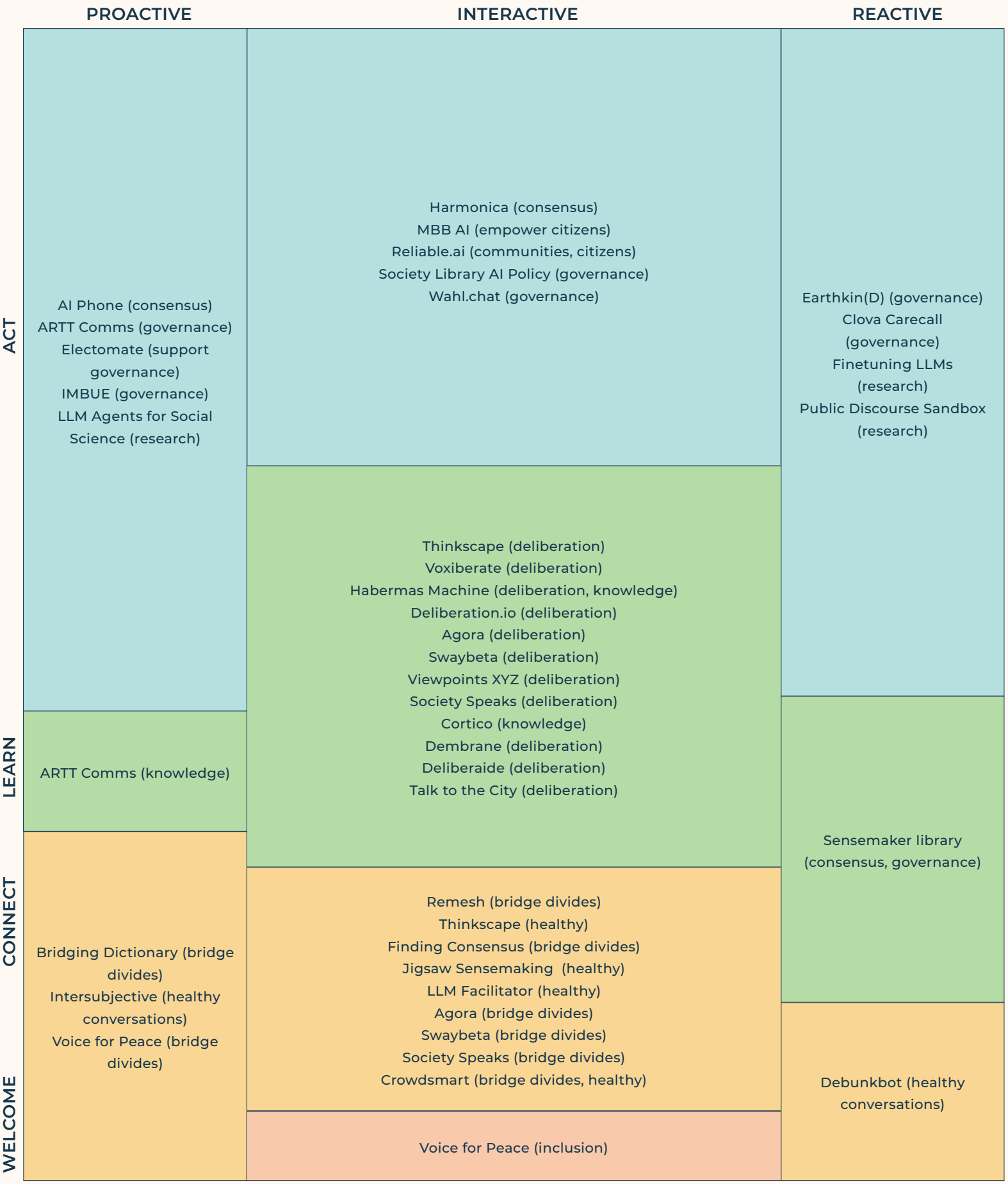
SOCIAL MEDIA PLATFORMS (MAP 1a)

	PROACTIVE	INTERACTIVE	REACTIVE
ACT			Filter Buddy (governance) Normsy (governance) Community Attributes (empower & governance) Perspective (empower) Filters (empower)
LEARN	Detoxigram (knowledge) AllStances (knowledge)	Muse of research (knowledge) BridgingBot (knowledge) NewsBridge (misinfo & knowledge) Muse of truth (misinfo) IsThisTrue (misinfo) Supernotes (misinfo) Factual (misinfo)	ARTT (misinfo) Harmful (misinfo)
CONNECT	Wapo (healthy) Sparkable (healthy & bridge) Detoxigram (healthy) AllStances (bridge)	ConvoWiz (healthy) IsThisTrue (healthy) Empath Assistant (healthy) MFM (healthy) CLR:SKY (healthy) OpenVetting (healthy) BridgingBot (bridge & healthy) NewsBridge (bridge)	Perspective (healthy) Filters (healthy) Harmful (healthy) Normsy (healthy & bridge) Meme (healthy & bridge) Counterspeech (healthy & bridge) TKI (bridge)
WELCOME	Wapo (inclusion) Detoxigram (safety)	Yahoo (safety) CLR:SKY (safety) MFM (safety)	Meme (safety) Counterspeech (safety) WeLivedIt (safety) Perspective (safety) CoPe (safety) Filters (safety) Harmful (safety)

DELIBERATIVE PLATFORMS (MAP 1b)

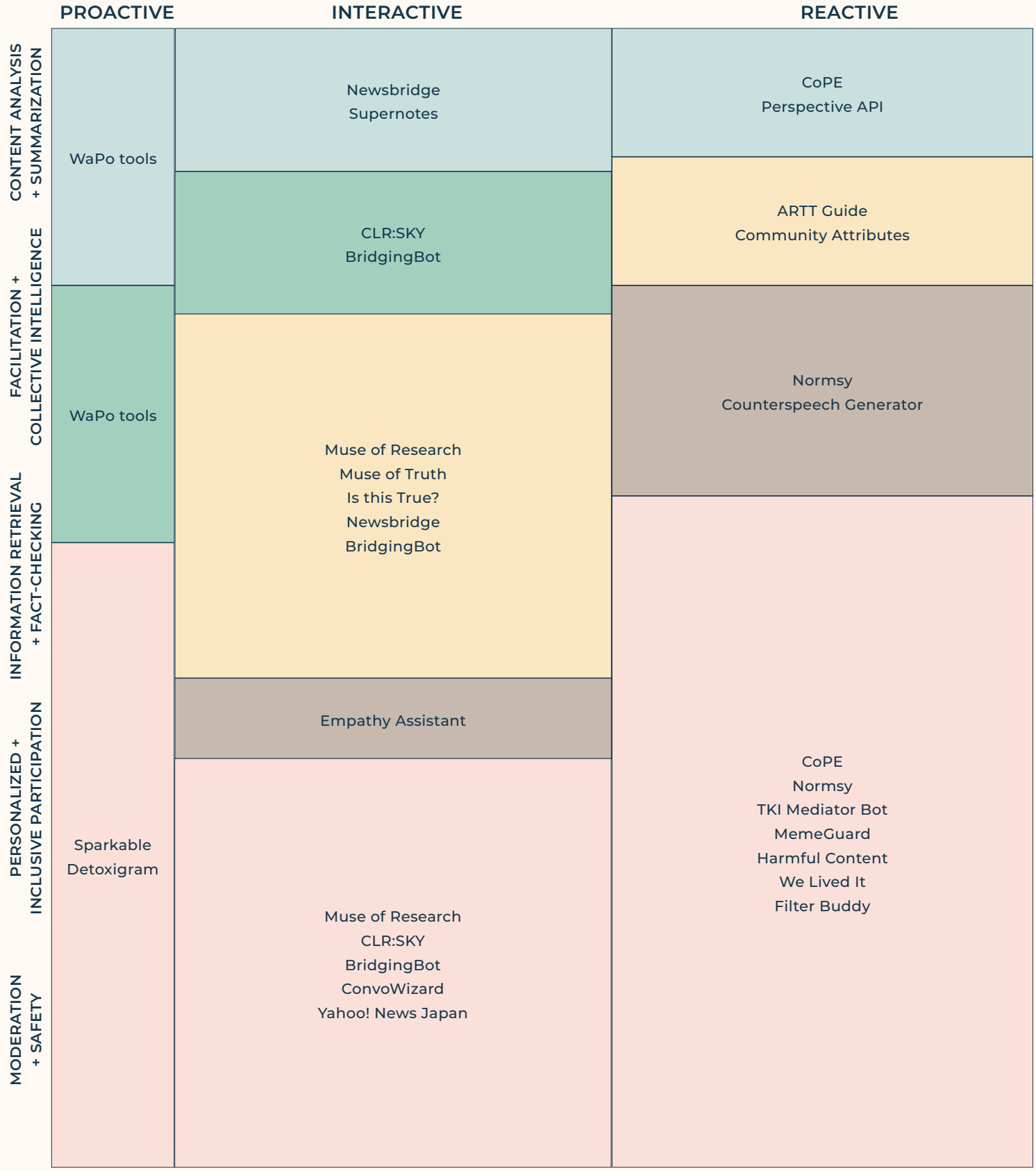


MISCELLANEOUS / OTHER DIGITAL SPACES (MAP 1c)

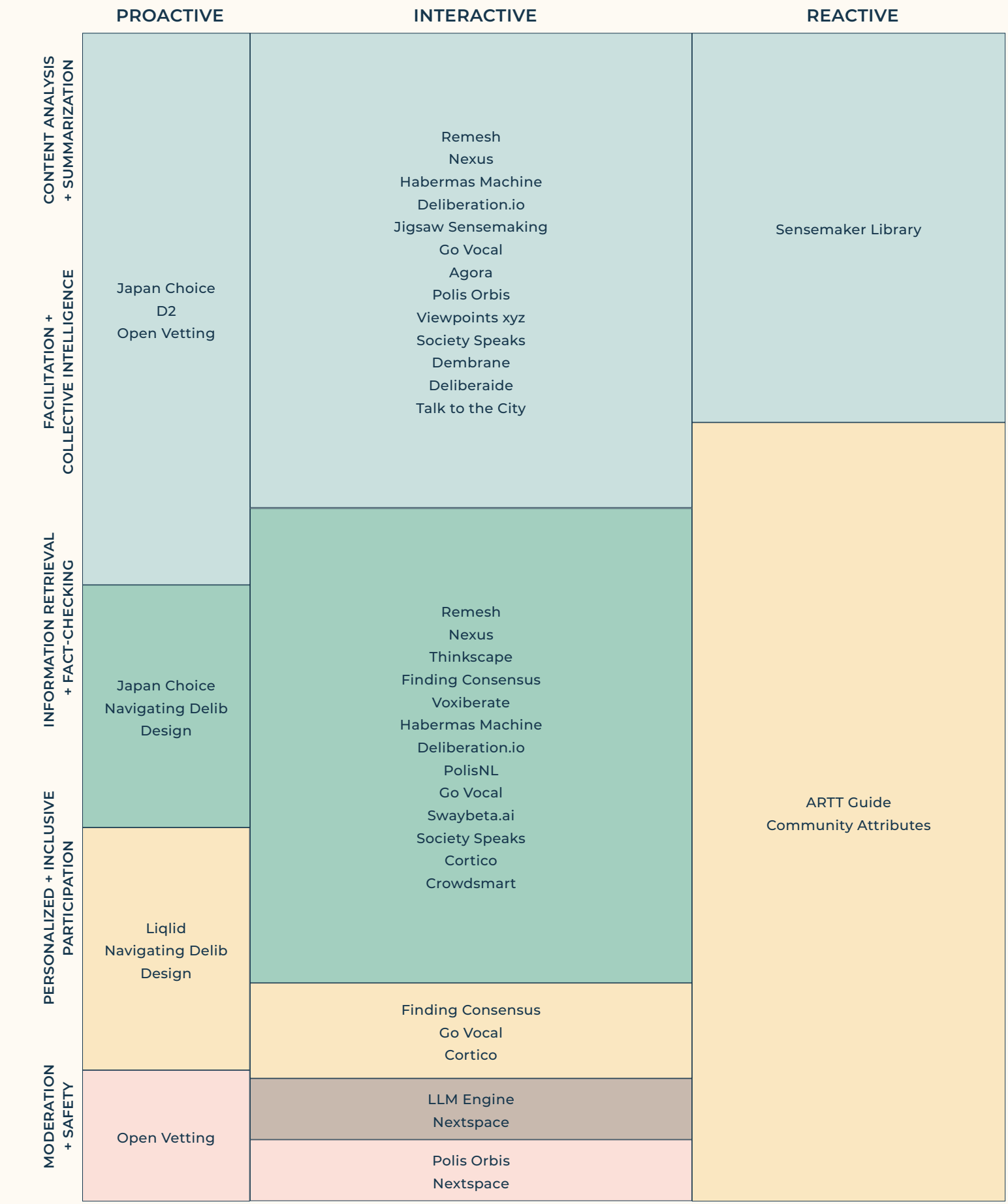


FUNCTION + TIME

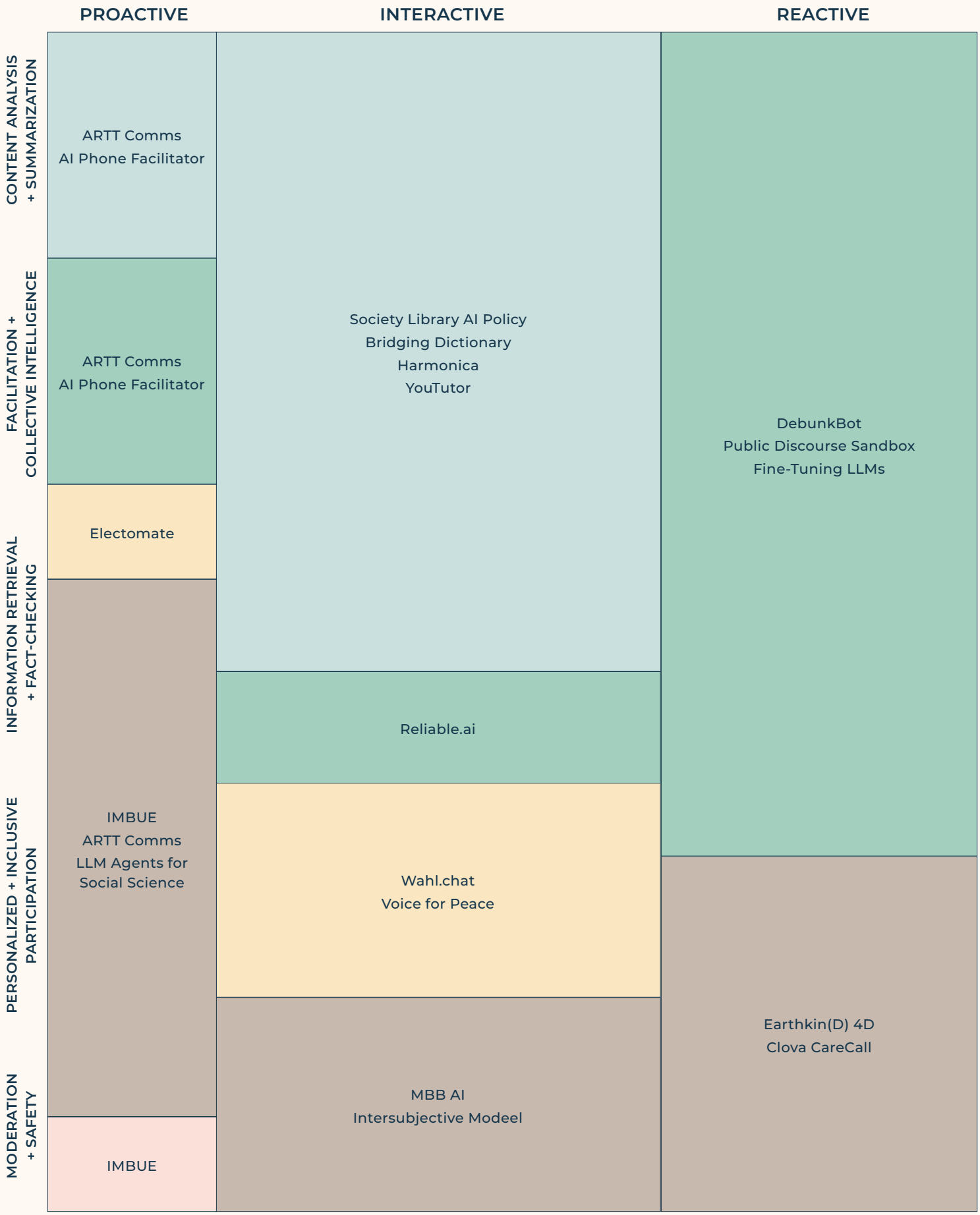
SOCIAL MEDIA SPACES (MAP 2a)



DELIBERATIVE PLATFORMS (MAP 2b)



MISCELLANEOUS / OTHER SPACES (MAP 2c)



LLM TOOLS FOR PUBLIC DISCOURSE ON SOCIAL MEDIA

Social media platforms - whether "mass" social media feeds, forums for online groups or comments sections of news publications - have been collectively described as the **"modern public square," spaces where people gather to express views, share information, and engage in public discourse**. Supporting these public squares - and in turn shaping their discourse - is a complex set of product design, content moderation, and broader governance decisions.

Large language models (LLMs) provide great opportunities for fostering discourse in these spaces, and as the maps above show, great energy is already being poured in from researchers, technologists, and civic organizations to build LLM solutions.

> "The most common goal across all 35 tools is to foster healthier conversations."

Participants in the convening identified 24 tools to promote discourse on social media platforms tthat are either currently in use (16 tools) or are in development (8 tools). Most of the tools identified are designed to operate on mass public square platforms (16 tools). Fewer were specifically designed to support volunteer moderators in online groups (5 tools) or for comments sections (2 tools). Of note - and as discussed below - only a handful of those tools are developed by and integrated into platforms by the social media companies themselves; instead, for the most part LLMs in this dataset operate as 3rd party tools that either intervene independently in those platforms (e.g. Normsy) or, more commonly, can be voluntarily adopted by users into their online experience (e.g. CLR:SKY or Muse of Truth).

VARIED GOALS AND FUNCTIONS WITH THREE POPULAR STRATEGIES

While most tools in the dataset are designed to achieve multiple prosocial goals, the most common goal across all 35 is to foster healthier conversations (18 tools). Many do so by interacting with users as they engage with others on the platform, for example giving users in-time feedback on whether their comments are likely to defuse or increase tension, but a fair number support healthier interactions merely by identifying and removing harmful posts and comments before users have a chance to see them.

While most tools in the dataset are designed to achieve multiple prosocial goals, the most common goal across all 35 is to foster healthier conversations (13 tools). Many do so by interacting with users as they engage with others on the platform, for example giving users in-time feedback on whether their comments are likely to defuse or increase tension (ConvoWizard), but a fair number support healthier interactions merely by identifying and removing harmful posts and comments before users have a chance to see them. Tools in that latter subset represent the second most common goal: ensuring the safety of users (8 tools); again, they achieve that by and large by removing user-generated content that might be harmful to others. A similar number of tools (7 tools) have the ultimate goal of bridging divides by using a number of strategies including increasing exposure to diverse views, countering divisive posts and specifically fostering healthier conversations for users across political divides. Two other common goals focus on improving our collective knowledge either by reducing misinformation (6 tools) or more generally expanding users' knowledge (4 tools). Other goals that tools aim to achieve are supporting governance (3 tools), empowering citizens and communities (2 tools), supporting consensus (1 tool), and promoting inclusion (1 tool).

The most common function among the tools is moderation and safety (14 tools). All five tools

supporting volunteer moderators included this function. Other important functions include information retrieval and fact-checking (7 tools), and content analysis and summarization (6 tools), with some offering multiple functions. A primary focus on "Moderation and Safety" reflects the importance of content moderation and product safety features in enabling healthy civic discourse. In turn, this suggests an emphasis on risk mitigation, underscoring a need to address negative, undesired content rather than enabling users to generate positive, desired content through facilitation, collective intelligence, and more inclusive participation.

THE ROLE OF HUMANS IN LLM-ENABLED TOOLS ON SOCIAL MEDIA

Given concerns about the risk of LLM hallucinations and their persuasive power, it will be important for many to know how much the tools being built keep humans in the loop or operate without oversight. Similarly, as healthy design should not be manipulative or coercive, we look at how much the LLM tools in the dataset support user agency.

As we looked over the 24 tools, we noticed that most used one of three strategies to foster discourse. One set of tools focuses on the detection and classification of healthy and harmful content, all toward improving the set of content that users are exposed to either by removing, ranking or alerting users to harmful and healthy content. Examples include Perspective AI, a widely adopted classifier of toxic content, and Detoxigram which alerts Telegram users to groups where there is high level of toxic content.

> "One set of tools focuses on the detection and classification of healthy and harmful content, all toward improving the set of content that users are exposed to either by removing, ranking or alerting users to harmful and healthy content."

Another set of tools give users context for the content they are seeing, either by summarizing content on a feed or by drawing in information from other sources. For example, Supernotes creates Community Notes to give readers context on X posts, while Muse of Truth is a bot X users can call on to fact check posts. A third set of tools generate content and interact with users to respond to, mediate and guide their contributions, sometimes in the moment, for example Thread with Caution which alerts users when the comment they are writing may raise tensions in a thread, or after the fact, like Normsy which inserts "counterspeech" when users post toxic content on social media platforms.

Out of the tools currently in use, five out of 16 tools are classified as requiring *full* human oversight, indicating that **these tools are designed to support, not replace, content moderation and other key features, including fact-checking and analysis.** Seven tools are marked as partial, meaning they automate some aspects of the process but still rely on human guidance. The minority of tools that do not require human involvement just summarize information, rather than attempting to moderate it. This reflects a strong emphasis on human judgment, empathy, and contextual understanding in content moderation and user discussions.

We likewise see a tendency among the tools to design with human agency in mind. Within the 16 identified tools, 15 tools are ranked as high to medium in agency, meaning the LLM empowers individuals as opposed to subverting their ability to freely choose and act. The users retain strong agency in shaping outcomes, interpreting content, and guiding interaction. Though this could be considered a strength, a potential concern is the 'burden of action' this places on the user. Each choice the user is prompted to make provides an 'activation cost', a small demand on the user's attention or effort which,

if accumulated, can detract from the overall experience or even become overwhelming for the user, leading to disengagement.

STAKEHOLDERS AND ROUTES OF DEPLOYMENT: TRADEOFFS

As noted above, most of the tools currently in use are designed to be adopted voluntarily by users either as third-party plugins or overlays, or as standalone platforms that require users to opt in alongside their use of mainstream social media. Of the few that are adopted by platforms, they either serve as content classifiers built to support moderation teams or in online news comments sections.

While we see existing tools for the most part operate independently, many could in principle be integrated directly into platform interfaces as native product features. While these different routes to adoption shape their potential impact, they also bring trade-offs. Third-party plugins, overlays, and standalone platforms can be developed independently of Big Tech platforms, which may be hesitant to adopt such solutions due to competing commercial priorities. These solutions also place a high activation energy on the user, who must be made aware of and then independently adopt the tool alongside their use of the main social media platform. This is likely to result in small usage relative to the mainstream usage of social media platforms. Moreover, more conscientious users already inclined towards healthier civic discourse are more likely to adopt such tools.

Tools oriented towards moderation and Trust & Safety teams, or product features that are directly integrated into social media platforms, will have more widespread reach and uptake. However, there is a much higher bar for adoption given that profit-oriented platforms, with exceptions, only integrate tools and other design elements that are commercially viable relative to existing product features and processes. Tools integrated into the platform as product features may be offered by but not "switched on by default", requiring users to navigate through settings menus to activate them. In such cases,the activation cost placed on the user is comparable to installing a third-party tool, though marginally lower.

"There is a much higher bar for adoption given that profit-oriented platforms, with exceptions, only integrate tools and other design elements that are commercially viable relative to existing product features and processes."

These tools and projects have emerged through the efforts of a broad range of stakeholders across sectors, including academic research groups (e.g., Cornell, University of Washington, LMU Munich), nonprofit organizations such as foundations and civic initiatives, and private companies spanning early-stage startups to established firms. While some tools are open-source and accessible via platforms like GitHub, others remain proprietary. As the space continues to grow, questions arise regarding how these stakeholders engage with each other through collaboration, competition, or knowledge transfer. Despite differing scopes and design priorities, shared goals and functional similarities may offer opportunities to align efforts in the future.

OPPORTUNITIES AND FUTURE DIRECTIONS

As we look for where there is further opportunity for LLMs to foster discourse on social media platforms, we observe that far more existing projects either engage users interactively, while they are consuming or posting content, or reactively, after they have engaged in potentially harmful behavior. Far fewer LLM tools in the dataset focus on proactively fostering healthy

engagement. We see this as an opportunity to explore ways LLMs can inform or engage users upstream - as they onboard onto a platform or before entering a conversation - to support healthier engagement.

In addition to identifying existing LLM tools that are either in use or being developed by teams, event participants generated new ideas - 24 of them - for where future LLM tools should go.

A number of ideas aim to support more proactive and social intelligent engagement, in line with existing user-oriented proactive tools. This includes a tool which encourages quieter users to participate, and one which helps users craft edgy or humorous comments that remain within community norms. Similarly, one aims to help users better understand the likely reception of their messages.

Another set of ideas centers on moderation support and group governance. Several concepts envision tools that reduce the burden on volunteer or platform-based moderators, providing guidance, real-time alerts, onboarding support, and crowd-sourced input.

A third area of development focuses on bridging divides and fostering mutual understanding, including tools that aim to mediate conflict or translate across political or cultural differences or explore how content exposure and idea generation can be designed to break filter bubbles or shift the tone of online dialogue.

Finally, some ideas explore sensemaking, such as a personalized digest tool designed to help usertdilms navigate complex online environments.

As the field evolves, emerging tools should continue to explore how LLMs can not only mitigate harm but also actively support prosocial engagement, inclusive participation, and healthier dialogue. Whether these innovations are scalable, adopted, or embedded within platform systems will ultimately depend not only on technical performance, but also on questions of governance, business incentives, and long-term resourcing.

SPOTLIGHT LLMS FOR PUBLIC DISCOURSE IN SOCIAL MEDIA: CLR:SKY

CLR:SKY is an LLM based-tool that anyone with a Bluesky account can currently use to get in-time feedback when navigating potentially heated conversations. It uses LLMs for three features as users are writing a comment: a “Toxicity Weather Report” that alerts users when their comments may cause offense; a GenAI Editor that suggests how a comment could be clearer and more empathetic; and a “Perspective Assistant” that will also suggest ways to acknowledge previous viewpoints in a thread. All toward nurturing more civil social media spaces “one conversation at a time.”

Like many tools we see in the Social Media landscape it requires users take action to use it, which makes it a tool that rates high in supporting user agency, but may also limit its impact. Are there ways for platforms to promote the use of these conversation assistants?

LLMS FOR PUBLIC DISCOURSE ON DELIBERATIVE PLATFORMS

Deliberative tech is designed to **foster understanding among large groups, exchange perspectives, and promote greater consensus on complex or contentious issues**. Unlike basic polling or social media comment systems, deliberative technologies are structured to promote equal participation, surface diverse perspectives, encourage reasoned exchange, and build consensus or shared understanding. For example, these tools may support processes like

citizen assemblies, participatory budgeting, or organizational consultations—enabling people to engage asynchronously or in real time, often at scale. At its core, deliberative tech seeks to enhance democratic legitimacy by fostering meaningful engagement between individuals, communities, and institutions.

Deliberative tech is trying to solve the problem of limited, unequal, and often polarized public participation in decision-making by creating digital spaces where people can engage in thoughtful and structured dialogue on complex issues. It aims to deepen deliberation, enabling communities to move beyond quick reactions or surface-level feedback toward more collaborative processes that reflect diverse perspectives and lead to better, more legitimate decisions. Through its design, it seeks to create more inclusive approaches especially within increasingly polarized environments. (See "Origin Story" text box for a deeper grounding in deliberative technology.)

Early deliberative tech largely focused on efforts to analyze, cluster and visualize opinions and positions of participants. The introduction of LLMs, while technically a moderate step in machine learning technology, has produced qualitatively new opportunities. LLM-based agents can now directly engage in conversations with human participants and play a variety of new roles, acting either as a support for a human facilitator, or potentially replacing the need for human facilitation in some functions. For example, an LLM might monitor time and keep people on agenda, fetch relevant online resources, provide summarization, or promote respectful speaking and active listening among participants. Or, a LLM might monitor participation to help encourage participants to speak more to promote inclusive engagement. LLMs can also enable sophisticated and more subtle textual analysis and comparison than was previously not possible, opening new opportunities to align ideas and find common ground.

LLM-based deliberative tech has the potential to both increase the quality and reach of deliberative events, but there are also significant risks. If the models on which it is based are inherently biased, these LLMs might reinforce existing power imbalances that would undermine their societal benefit. How can we develop trustworthy LLM-based interventions? How can we be assured of their fairness, and monitor and mitigate known problems like hallucination? What is the appropriate role of human oversight? Also, there is a risk that individuals lacking internet access or technological know-how are not able to participate in digital forums. How can we ensure this is not a barrier to participating in deliberation using digital tools? These are only some of the questions that need significantly more exploration to learn how to genuinely be inclusive.

Deliberative tech is for people and institutions who want to deliberate better together—to bridge gaps in understanding, build trust, and co-create solutions. Its value depends on design choices of the technology, as much as the process designed to foster the engagement. At this point, deliberative tech is not seen to wholly replace in-person interaction, rather compliment it.

DELIBERATIVE TOOL MAP: OBSERVATIONS

Within the workshop, **there were 28 platforms and tools identified in a range of stages** including 8 projects under development and 20 currently in use. In this report, we refer to platforms and tools interchangeably, but platforms are defined as stand-alone while tools can be integrated into platforms. Much of the technology is still evolving, so we don't make a specific distinction. The primary functions of the platforms in use include, with some offering multiple functions: facilitation and collective intelligence (12 tools), content analysis and summarization (13 tools), and information retrieval (5 tools).

Considering the 28 tools in use, their primary goals are: facilitates deliberation (15 tools), supports consensus building or decision making (16 tools), supports healthy conversations (6 tools), bridges divides and reduces polarization (6 tools), expands knowledge or awareness (4 tools), supports governance (4 tools), promotes inclusion or diversity (2 tools), and empowers

citizens and communities (1 tool). Based on these functions and goals, there are several observations about these tools.

WHAT MAKES THEM DELIBERATIVE?

Many of these deliberative platforms are built to facilitate productive conversation and enhance structured group thinking. Collectively, **these tools are primarily designed to improve the quality and inclusiveness of group dialogue, rather than just collect opinions or votes.** There is a focus on process-oriented deliberation—supporting participants to reflect, reason, and engage with diverse perspectives.

Depending on the context, many deliberative conversations are focused on building consensus across groups. These tools aim to foster consensus using bridging algorithms or inference. Consensus is often "soft" by highlighting shared values, common themes, or ranked ideas–not "hard" consensus (like voting thresholds or binding decisions). This softer approach allows for nuance, accommodates disagreement, and supports deliberation without coercion.

Given the need to manage complex ideas, there is a strong presence of content analysis and summarization functions because digital deliberation often generates large volumes of text and ideas. Tools that help synthesize, cluster, and visualize insights are crucial for distilling group thinking into usable outputs, especially for decision-makers or facilitators when engaging large groups. LLMs can accelerate the process of identifying consensus, which can expedite group thinking and agreement.

Iteration is an important part of deliberation. Some of these tools enable users to return multiple times, respond to others' input, and re-rank or reframe contributions. **This iterative design helps participants move beyond initial opinions and converge on shared insights through reflection and interaction**, which is important for trust-based consensus. Though, this can be challenging to have users engage multiple times because most individuals are used to surveys that are one-and-done.

WHAT IS THE ROLE OF HUMANS IN DELIBERATIVE TECH?

Out of the tools currently in use, 5 out of 28 tools are classified as requiring full human involvement, **indicating that these tools are designed to support—not replace—facilitators, moderators, or organizers.** This suggests a strong emphasis on human judgment, empathy, and contextual understanding in deliberative processes. 14 tools are marked as partial, meaning they automate some aspects of the process (e.g. summarization, ranking, content sorting) but still rely on human guidance. In particular, humans are needed to review language translations given current AI tools within LLMs don't have the necessary contextual information, especially for more sensitive topics within political deliberations.

> "This suggests a strong emphasis on human judgment, empathy, and contextual understanding in deliberative processes."

Most deliberative tech tools in the dataset prioritize human-centred design and consider human agency essential to deliberation. While automation is used to enhance efficiency and insight, very few tools aim to eliminate human facilitation altogether. With the data set created from the workshop, the current tools in use make deliberation more scalable with a focus on ethically grounded and context-sensitive approaches.

These technologies have been developed by a diverse set of stakeholders, including academic research initiatives (e.g. Stanford, MIT, Harvard, Carnegie Mellon), nonprofit organisations such as foundations and civic councils, and private companies ranging from startups to established

ORIGIN STORY: FOUNDATIONAL WISDOM, NEW TOOLS

Deliberation has deep roots in ancient traditions across Buddhist, Confucian, Islamic, Sub-Saharan, and Nordic societies, where it was used to settle disputes and inform governance. This practice evolved into early forms of democracy, notably in ancient Athens, where free male citizens gathered to debate public issues and make collective decisions. Emphasizing reasoned argument, listening, and the weighing of diverse viewpoints, Athenian democracy—though not inclusive—established the principle that legitimate governance requires public reasoning among equals.

Building on these ancient legacies, deliberative democracy evolved as a model that centers discussion and collective reasoning in political decision-making. Unlike aggregative methods—such as direct voting or interest-group bargaining—it asserts that legitimate outcomes arise from fair, inclusive, and reflective discussions, where participants justify their views, consider alternatives, and seek common ground or compromise. Modern democracies continue to strive for more inclusive deliberation, and technology, particularly LLMs, may play a crucial role in advancing this goal.
Civic tech began to emerge as a distinct field in the early 2000s, but its roots can be traced back further to the rise of open government in the 1970s and e-government initiatives and digital democracy movements in the 1990s. The field gained momentum alongside the open data movement to create digital tools, platforms, and systems designed to strengthen the relationship between people and government or to improve public life more broadly. This includes technologies that support government transparency, citizen services, participatory budgeting, civic data access, voting, public accountability, and community organizing. Over the past ten years, interest in deliberative democracy and civic tech has grown, and deliberative tech has emerged as an important subset of civic tech.

“While civic tech may facilitate basic engagement like reporting potholes, accessing public data, or submitting a comment, deliberative tech is explicitly designed to foster deeper understanding among the public, facilitate an exchange of perspectives, and promote greater consensus on complex or contentious issues.”

While civic tech may facilitate basic engagement like reporting potholes, accessing public data, or submitting a comment, deliberative tech is explicitly designed to foster deeper understanding among the public, facilitate an exchange of perspectives, and promote greater consensus on complex or contentious issues.

firms. Some tools are open source and available on platforms like GitHub, while others are proprietary, owned by businesses with specific market interests. While each tool is built with distinct functions and design priorities, many share overlapping goals around improving deliberation, civic engagement, and democratic decision-making. As the space becomes more crowded and competitive, **it is worth examining how organizations interact—whether through collaboration, competition, or knowledge-sharing—and what opportunities exist to align efforts and share resources** to reduce fragmentation in the field.

Within the 28 tools, at least 27 tools are ranked as high to medium in agency, meaning the LLM empowers individuals as opposed to subverting their ability to freely choose and act. It means users retain strong agency in shaping outcomes, interpreting content, and guiding interaction. It highlights a broader tension in deliberative tech: balancing efficiency and scalability with the need for user control, agency, and oversight to create legitimacy.

In addition, 27 of the 28 tools are interactive or proactive-interactive (not reactive), meaning users must dynamically engage with the tool in real time or through reciprocal inputs to generate content, in the form of comments, votes, or discussions. **The overwhelming majority of platforms aim to facilitate active, two-way engagement, a key feature for supporting meaningful deliberation.**

The analysis of integration types among the 28 deliberative platforms reveals a strong leaning toward standalone functionality. Seven platforms are described simply as "platforms," and four more as "self-standing," indicating that most tools are designed to operate independently without needing to integrate into existing systems. Only four tools are explicitly labelled as "integrated," suggesting a more modular or embedded use within broader institutional or civic frameworks. Overall, while most tools prioritize autonomy and ease of deployment, there may be an emerging trend toward integration-ready designs that may better support interoperability and long-term adoption within existing infrastructures.

SPOTLIGHT LLMS FOR PUBLIC DISCOURSE ON DELIBERATIVE PLATFORMS: ZKE

In Kenya, a group of young activists responded to toxic online polarization by creating zKE, a deliberative space that fosters constructive national engagement. Working in partnership with Siasa Place (a grassroots youth NGO) and The Situation Room (a popular national radio program), they sought to shift heated public policy debates sparked by Gen-Z protests away from divisive social media feeds and into inclusive civic dialogue. Over four months, they curated public policy conversations with thousands of young Kenyans, bridging analogue spaces—in-person youth assemblies (barazas) and live radio programming—with a digital deliberative ecosystem.

The zKE tech stack combined multiple tools to foster dialogue and consensus: a WhatsApp bot for real-time updates and contributions, an instance of Talk to the City for youth to share voice-note opinions, structured debates on pol.is to surface policy proposals with the highest consensus scores, and a Remesh meet-up to synthesize diverse perspectives. Large language models were used to analyze and summarize the wealth of input across these channels, turning unstructured contributions into insights that could inform collective policy positions. By integrating these platforms with youth assemblies and radio, zKE demonstrated how grassroots organizing and deliberative technology can be woven together to create a sustained, scalable forum to shape public policy and strengthen democratic participation.

FUTURE DEVELOPMENTS: WHERE IS IT GOING?

The future of LLM-based deliberative tech appears very promising, especially in helping deliberation to scale, but we are only in the opening chapter. **What currently exists is likely only a small fraction of the approaches and offerings we will see develop.** We must also be diligent in investigating biases and risks and work to mitigate them.

In particular, there is much to learn about LLM-based agents in deliberative contexts, for example, what can be safely delegated to them and where they should rather support a human facilitator or be overseen by one.

Deliberative tech itself is at an early stage, so the platforms people use and how they are extended and augmented are only partially known. This means how LLMs are integrated into deliberation experience will also grow and evolve over time.

We need to learn more about the various subfunctions under the facilitation and collective intelligence function - this function seems overly broad and ripe for deeper understanding and exploration. Along the way, we need to better understand what humans require to trust the use of LLMs within dialogues to enable greater application.

When we look at the various new ideas proposed in this space - 8 under development projects and 16 ideas for future tools - we see at least three emerging directions where there may be an identified need or opportunity:

- **Encouraging effective communication practices** like active listening and reframing
- **Promoting greater inclusivity and accommodation**, perhaps through interfaces customized on a participant-by-participant basis to support different languages, learning styles, and neurodiversity generally
- **Promoting the participant safety and engagement** that underlies active and meaningful participation

The potential for deliberative technologies to enable inclusive engagement and decision-making continues to grow. There are many questions to explore, especially around technology adoption by the public, policymakers, and broader institutions. As deliberative technologies evolve, **we must continue to build trust, ensure accessibility, and design systems that facilitate meaningful participation** for all stakeholders.

MISCELLANEOUS TOOLS

The “miscellaneous” tools in the dataset don’t neatly fit into our spaces for Social Media or Deliberative Platforms, but introduce other creative uses of LLMs to foster pluralism and civility in discourse.

About half of the tools that participants added to the miscellaneous dataset are LLM chatbots designed to converse with people via text or voice towards several ends. Three tools let people talk to or learn about non-human entities, such as a YouTube video or a political party. Two use voice and phone to reach out to people for conversations about peace-building or about their wellbeing. A similar tool acts as a survey taker. A couple of chatbots offer people practice conversations that, for example, build emotional skills. One in the dataset is built to decrease people’s beliefs in conspiracy theories. A final chatbot acts as a representative of a non-human entity — in this case “Nature” — that participates in deliberations.

Among the LLM projects in the miscellaneous dataset that are not chatbots, several use LLMs for research purposes either to simulate and study human interactions in complex dynamics, to study how humans interact with AI or to reveal insights into how groups use words differently. A couple of projects harness LLMs to support policymakers or non-governmental organizations in crafting legislation or more effective public messaging. One project is building a LLM tool to act as in-time “translators” to enhance understanding between people who speak the same language. Finally, three projects in the dataset are designed to improve LLMs by, training them with domain experts.

In looking at the goals of these "miscellaneous" LLMs, they are more heavily focused on supporting governance (6 tools), expanding knowledge and awareness (5 tools), supporting healthy conversations (4 tools), bridging divides and reducing polarization (3 tools), and empowering citizens and communities (3 tools). Several, however, also aim to support research and experimentation (2 tools), support consensus building or decision making (2 tools), foster flourishing communities (1 tool) and promote inclusion and diversity (1 tool).

The tools also cut across multiple functions, the most common being personalized and inclusive participation (6 tools), content analysis and summarization (6 tools), and facilitation and collective intelligence (5 tools), followed by information retrieval and fact-checking (3 tools) and moderation and safety (2 tools).

There is a broad range of human oversight. Certain tools require full oversight, especially those used in research or practitioner settings. Other tools which are more community-driven or self-standing (e.g. “Bridging Dictionary”) require lower oversight. Agency is generally high, as most tools are designed to support human decision-making, empowerment or participation. Some exceptions include tools that carry out the role of human mediators (e.g. “Voice for Peace”).

Given these tools are not associated with a particular social media or deliberative platform, they are self-standing tools, with integration existing mostly where personalization or communication assistance is central (e.g. “IMBUE”).

As with the other technology eco-systems, there is a broad range of actors involved in creating the tools, reflecting the diversity of use cases and approaches. This includes academia & research labs, such as Stanford University, Orbis (an EU-funded research project), and ABM simulations. Orbis also demonstrates the role a government can play either through funding or direct collaboration. Civil society and nonprofit contributions include Society Library and Voice for Peace.

SPOTLIGHT LLMS FOR PUBLIC DISCOURSE IN OTHER DIGITAL SPACES: WAHL.CHAT

Wahl.chat is a platform that lets German voters “chat” with political parties to understand their platforms or to more generally learn about and compare the policy positions of different parties. The LLM chat draws its answers from party platforms and is trained to give “neutral” answers.

By using a neutral tone, Walh.chat — and other LLM tools built on its model - have potential for exposing voters to more diverse views. More broadly, by virtue of being engaging, chatbots offer endless opportunities for drawing people in learning about civic topics that might otherwise feel dry or overwhelming.

WHERE IT’S GOING

These tools point to an emerging trend of AI as civic infrastructure to shape conversational, deliberative, and decision-making spaces. This is reflected in functions including personalization and facilitation. **Many of these systems move beyond simple content delivery to focus on personalization and facilitation, tailoring interactions to the needs of users while supporting more inclusive and productive exchanges.**

This trend spans both content and form: some tools emphasize textual mediation and consensus-finding, while others experiment with audio and video formats to foster greater presence, empathy, and accessibility in dialogue. Despite “miscellaneous” being a catch-all category, there are also common goals these tools are trying to achieve, whether it be bringing people together, enabling participation from underheard voices, or facilitating more effective

solutions such as policy making. Together, these tools point to a future where AI can be deployed as a public utility for discourse and governance.

At the same time, certain risks should be managed. Overreliance on AI could come at the cost of human agency. Tools that mimic human participation could be at risk of manipulation, erode trust, or falsely represent real-life figures if not carefully governed. **Whether these tools strengthen civil discourse will depend on design choices, governance frameworks, and the role of human oversight in their deployment.**

RISKS & CONCLUDING THOUGHTS

The landscape of LLM tools being built to foster social cohesion, which we share in this initial dataset and describe in this report, is still nascent but presents tremendous promise for how LLMs can massively fuel deliberative platforms, foster healthier conversations in all digital spaces and give individuals and groups the reins to expand their knowledge and include diverse perspectives.

And, yet, there are risks entailed with all LLMs and that we likewise need to be aware of as we continue to develop LLMs tools for good. While not comprehensive, we see several risks in particular:

LLM's inherent biases: LLMs have various types of biases and when we deploy LLMs at scale, whether on deliberation platforms or in other digital spaces, those biases might be translated with deleterious effects. In the case of deliberative platforms, a final deliberation outcome may absorb those biases if, for example, an LLM summarizes and aggregates public opinions and systematically overlooks unpopular opinions. An LLM specifically designed to build awareness may likewise sideline minority perspectives. Even an LLM built to facilitate healthy discussions could unwittingly alienate some users.

Oversight risk: Across all platforms, but in particular when LLM tools are used in deliberation and collective decision making, human beings may not be able to monitor and examine the processes that lead to any outcome.

Agency risk: As with all design interventions, when not carefully deployed LLM tools may diminish individuals' agency by "nudging" them in directions that are not in their best interest. More so with LLMs, which have been shown to be powerfully persuasive, even when individuals choose to adopt an LLM tool that tool may exercise an oversized persuasive power over its users.

Many of the projects we see in this report work explicitly to counter the risks above. But far more can be done to build in failsafes to reduce the risks of bias, lack of oversight and erosion to user agency.

Future collaboration: As we look to the continued development of LLMs for discourse, there are other questions we ask: As we design and build new deliberative and social media technologies to foster greater cooperation, will inclusivity be a core part of our approach? Will the key questions surrounding these tools be addressed transparently, or will they be shaped behind closed doors? To build trust and ensure these technologies benefit society, how can we create functional pathways that encourage shared dialogue and collaborative resources for these technologies? By moving away from a competitive, isolated approach, can we develop a structure that ensures deliberative tool design aligns with our shared goal of using new technologies to build stronger societies and governance systems that value public reasoning and participation?

We look forward to being in conversation with the growing body of researchers and technologists to answer those questions and continuing to harness the power of LLMs to foster discourse, pluralism and social cohesion.