

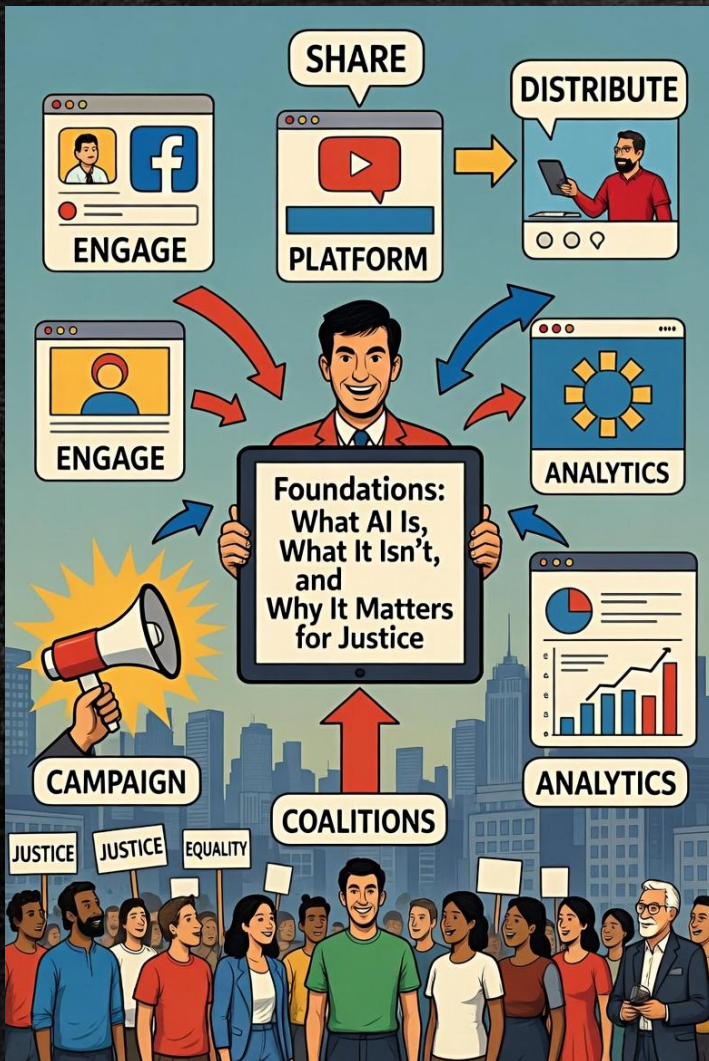
Edges to Centre: A Global Intergenerational Leadership Project for Marginalized Communities “Nothing About Us Without Us”

NextGen Rights Lab: Global South Intersectional Rights & Leadership Accelerator

Artificial Intelligence Literacy and Practice Track for Caste Affected
Intergenerational Leaders

Session – 1 (40 mins) - Foundations: What AI Is, What It Isn't, and Why It Matters
for Justice

1. Mapping the AI World in Plain Language (Not Tech Bro Mythology)



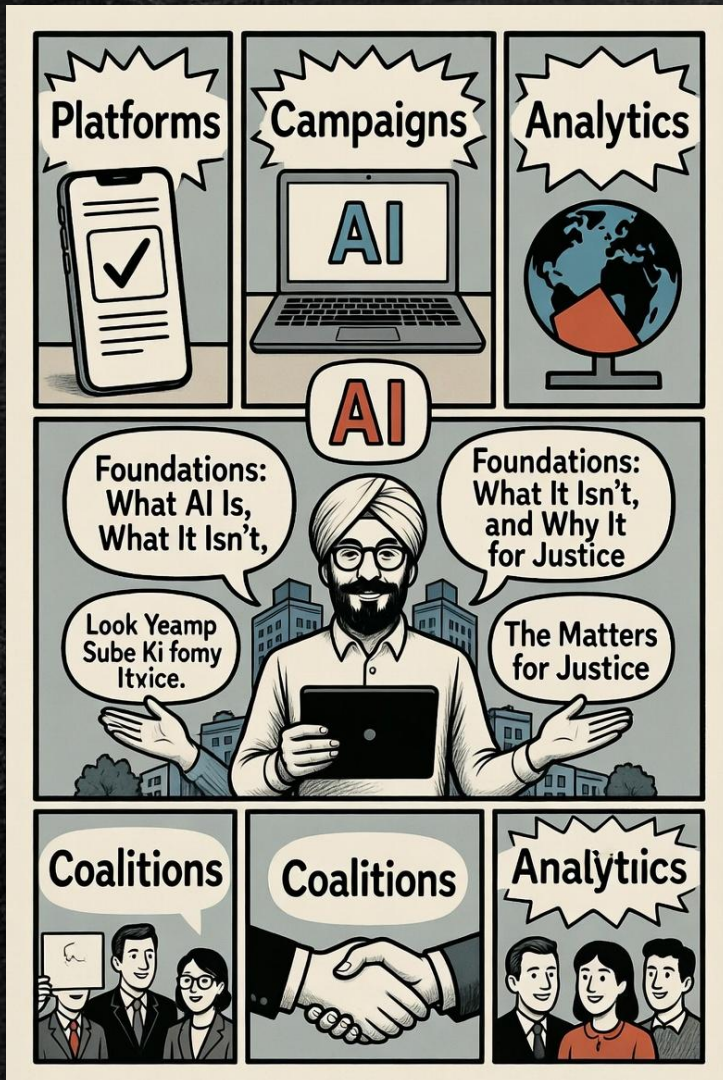
“AI” is actually a big umbrella term, and most of what you will touch in your work is a few specific types of systems, not some magical robot brain.

AI as the broad idea of machines doing tasks that we think need “intelligence”; machine learning as systems that learn patterns from data rather than following fixed if-then rules; and deep learning as a sub-type that uses many-layered neural networks to find complex patterns in images, text, and sound.

One example is when you unlock your phones with your face or fingerprint, that is a vision model; when you dictate a WhatsApp voice note and it turns to text, that is speech-to-text; when an app suggests auto-replies like “OK, thank you,” that is a simple language model. A large language models (LLMs) like ChatGPT or Claude are “super-autocomplete” systems that have been trained on huge amounts of text and learn to predict the next word, not to understand the world like a human being.

When you ask an LLM about Ambedkar, it is not ‘remembering’ Ambedkar the person; it is matching patterns from the text it has seen about Ambedkar.” This means LLMs can sound very confident and still be wrong, because they are optimising for plausibility, not truth, and that in caste-affected and minority contexts, wrong answers can be dangerous when they erase, stereotype, or misrepresent..

2. How Models Learn: Datasets, Bias, and Why Dalit/Bahujan Realities Get Distorted



Here we will focus on how models actually “learn” so you can see where power and bias get baked in. Most machine learning systems are trained on labelled examples (supervised learning), unlabelled patterns (unsupervised learning), or trial and error (reinforcement learning). In simple terms: in supervised learning, someone feeds the model pairs like “this is hate speech / this is not hate speech” thousands or millions of times; in deep learning, those examples update the model’s internal weights so it becomes better at predicting labels for new texts. The key point is: who collected that data, from where, in what language, and with what labels?

For example, if a hate-speech detector is trained mostly on English content from North America, it may miss caste slurs in Hindi, Tamil, or Bangla, or it may wrongly label Dalit self-assertion posts as “offensive” because annotators don’t understand context. For LLMs, they are trained on massive scraped datasets from the internet, books, and code, which means they reflect the over-representation of dominant-caste, Global North, male voices and the under-representation or misrepresentation of Dalit, Bahujan, Adivasi, Indigenous, Muslim, and queer communities.

Think about when you use AI in your work, assume the model is biased until proven otherwise; for any answer touching caste, religion, gender, or minority rights, you must check for stereotypes, erasures, and what is missing.

Have you heard of a “bias log”- a simple shared document where they note instances where the model got something wrong in a way that reflects social hierarchy, to build awareness over time

3. RLHF, Safety Layers, and Why the Model Sometimes Refuses You



We will unpack Reinforcement Learning from Human Feedback (RLHF) so you can understand why chatbots sometimes feel “safe but stubborn.” After the base model is trained on raw text, companies run a second phase where human annotators review possible model responses and rate them based on guidelines like helpfulness, harmlessness, and alignment with safety norms. These ratings are used to “reward” or “punish” certain outputs, pushing the model to prefer some patterns over others.

For example if annotators are told to penalise any output that looks like hate speech, the model will be more cautious when talking about caste slurs, but if annotators are themselves dominant-caste and not trained on local realities, they might penalise Dalit counter-speech while letting subtle casteist frames slip through as “normal opinion.”

When the model refuses to generate violent content or to identify private individuals, that is often RLHF at work, and as human-rights defenders, we want these guardrails, but we also have to watch for when those guardrails block legitimate documentation (for example, refusing to summarise news about pogroms because it contains graphic content).

Whenever the model refuses to help on a rights topic, do not immediately try to “jailbreak” it; instead, reframe the request in educational or analytic terms, such as “explain why caste-based violence is a human rights violation” rather than “describe in graphic detail this lynching,” and respect the boundary where requests could endanger real people.

4. Prompt Engineering as Everyday Literacy, Not a Secret Trick

This slide introduces prompt writing as a practical digital literacy skill that anyone working with AI tools should learn. Instead of treating AI prompts as a specialist or technical practice, the session encourages participants to approach prompting as a structured way of giving clear instructions and context.

You will learn a simple framework for writing stronger prompts:

- define the role,
- define the goal,
- set clear constraints,
- and ask the system to explain its reasoning and identify possible errors.

For example, instead of asking a broad question like “Explain manual scavenging,” participants are encouraged to provide context about audience, purpose, tone, region, and political framing. Adding instructions such as word limits, accessibility needs, or requests to avoid harmful language can also improve the quality and relevance of responses.

The session also emphasises that AI-generated outputs should always be verified. Participants are encouraged to ask follow-up questions that identify assumptions, possible inaccuracies, missing perspectives, or sources that require fact-checking. This is especially important when working on caste, labour, gender, religious discrimination, or human-rights issues where misinformation or oversimplification can cause harm.

You will also learn basic concepts such as tokens, context windows, and model memory in practical terms. Long conversations or large pasted documents can reduce accuracy, so shorter and more focused prompts often produce better results.

Finally, the slide introduces “temperature” settings as a way of controlling how predictable or creative an AI response will be. Lower settings are more useful for factual or structured tasks, while higher settings can support brainstorming, storytelling, or idea generation. Participants are encouraged to experiment carefully and understand these settings before relying on them in important work.

5. A Simple Taxonomy of AI Tools You Might Actually Use in Your Work

This slide introduces different types of AI tools that can support civic, advocacy, and community-based work in practical day-to-day situations. You will learn that large language model (LLM) chatbots, such as ChatGPT or similar systems, are only one category of AI tools and are mainly useful for drafting, summarising, translating, brainstorming, and organising information.

The session also introduces other useful AI systems. Vision-based tools can extract text from scanned or photographed documents such as FIRs, notices, or government circulars through optical character recognition (OCR). Speech-to-text systems can transcribe interviews, testimonies, meetings, or oral histories recorded on a phone. Translation tools can help create first-pass translations between languages such as English, Hindi, Bangla, Tamil, Nepali, and Urdu, although these translations should always be reviewed and corrected manually. Summarisation tools can help simplify long reports or PDFs, while retrieval-based AI systems can search through specific collections of policies, case law, or organisational documents.

You will also learn how to use these tools safely and responsibly. When working with sensitive material — including survivor testimonies, legal files, internal strategies, or community records — it is important to avoid uploading identifiable information into cloud-based AI systems whenever possible. Participants are encouraged to remove names, phone numbers, addresses, and other identifying details before using external tools and to prefer local or offline systems when handling high-risk information.

The slide includes practical examples of how these tools can support movement work. For example, you might scan a blurry government notice using an OCR tool and then ask an AI system to explain it in simpler community-friendly language. You might also record an elder's oral history, transcribe it using speech-to-text software, and then organise the transcript into themes with an AI assistant while keeping the original files stored securely and privately.

6. Cross-Cutting Principles: Privacy, Fairness, Accessibility, Documentation, Intergenerational Practice

This closing slide brings together the key principles that should guide every future use of AI tools in advocacy, organising, research, and digital work. It emphasises privacy and consent as everyday practices rather than technical afterthoughts. Before uploading text, images, testimonies, or documents into an AI system, participants are encouraged to ask whether they would feel comfortable if the same information about themselves or their families were stored on external company servers. If the answer is no, sensitive details should be removed, paraphrased, anonymised, or kept out of AI systems entirely.

The slide also introduces the idea of fairness by design. When working with prompts related to caste, religion, gender, disability, sexuality, or migration, participants are encouraged to actively question whose perspective is being centred and what assumptions the model may be making. One way to do this is by testing how responses change when different identities or social positions are introduced into the prompt.

Accessibility and language justice are presented as essential parts of responsible AI use. Participants are encouraged to ask AI systems for simpler language versions of complex ideas, multilingual explanations, alt text for images, or accessible descriptions that can support blind, low-vision, Deaf, or multilingual audiences.

The session also encourages participants to document and organise useful prompts, verified outputs, and shared terminology in collective resources that can support future community learning. AI-assisted writing or analysis should be clearly identified when used in reports, campaigns, or official outputs, and human review should remain part of all important decisions.

Finally, the slide frames AI use as an intergenerational responsibility. Elders can help ensure outputs remain connected to movement history and lived experience, mid-career professionals can connect AI-supported insights to policy or institutional change, and younger participants can contribute digital fluency, experimentation, and verification skills. Together, these practices support the development of shared, community-owned knowledge rooted in caste justice, accountability, and collective care.

Edges to Centre: A Global Intergenerational Leadership Project for Marginalized Communities “Nothing About Us Without Us”

NextGen Rights Lab: Global South Intersectional Rights & Leadership Accelerator

Artificial Intelligence Literacy and Practice Track for Caste Affected
Intergenerational Leaders

Session – 2 (40 mins) - Applied Prompt craft for Public Good: Education, Research,
Advocacy

1. A Discipline for “Asking AI” in Learning and Civic Work

This slide introduces a structured process for using AI tools in education, research, advocacy, and policy work. The goal is to develop consistent habits that make AI use more accurate, intentional, and accountable rather than relying on quick or unverified outputs.

Before beginning any task, participants are encouraged to clearly identify four things:

- the objective of the task,
- the intended audience,
- the relevant jurisdiction or legal context,
- and the level of risk involved.

For example, the objective might be understanding a law, drafting a policy outline, or simplifying technical information for community use. The audience could include students, local officials, journalists, donors, or grassroots communities, while the jurisdiction may relate to a specific country, region, or international legal framework. Risk assessment is also important, especially when dealing with sensitive cases, testimonies, or legal information.

Participants then learn how to turn this information into clearer and more useful prompts by specifying the role, context, audience, and desired format of the response. The slide also encourages participants to immediately verify AI-generated outputs through follow-up questions that ask the system to identify the laws, policies, judgments, assumptions, or uncertainties behind its answer.

You will also be encouraged to cross-check important claims using trusted websites, legal databases, reports, or independent sources before using AI-generated material in serious advocacy or professional work.

Throughout the session, AI is framed not as an authority or “oracle,” but as a fast and useful assistant whose work still requires human judgment, verification, and political awareness.

2. Using AI as a Study Partner Without Cheating Yourself

This slide explores how AI tools can support education and self-learning while maintaining academic integrity and critical thinking. Participants are encouraged to use AI systems to better understand difficult concepts, laws, judgments, and policy materials rather than relying on them to complete work automatically.

You will learn how AI can help summarise long readings, simplify complex language, generate examples, and explain ideas across multiple languages and levels of difficulty. For example, AI can be used to break down legal judgments into simpler points, explain concepts in accessible language, or create practice examples related to caste discrimination, labour rights, or minority protections.

The session also introduces ways to use AI for active learning. Participants can generate practice questions, flashcards, quizzes, or interactive exercises that test understanding and provide feedback in simpler language. This can support revision, self-study, and confidence-building, especially when working with technical legal or policy topics.

At the same time, the slide emphasises the importance of academic honesty and independent thinking. Participants are encouraged not to use AI systems to complete assignments, exams, or written work on their behalf. Instead, AI should be used as a support tool that helps strengthen analysis, improve structure, clarify arguments, and identify weak areas in drafts.

The overall goal is to use AI as a tutor, study partner, or learning assistant that supports your own thinking and skill development rather than replacing your voice, reasoning, or political understanding.

3. Research and Writing: Outlines, Literature Maps, Coding Frames, Summaries

This slide introduces practical ways to use AI tools during research, report writing, and campaign preparation. Participants will learn how AI can support the early stages of planning and organising ideas by helping structure reports, policy briefs, advocacy documents, or academic papers.

You will learn how to combine your own notes, research goals, and context into clear prompts that ask AI systems to suggest outlines, thematic sections, or possible areas for case studies and legal references. AI can also help identify important themes, keywords, debates, and source categories that may guide your literature review or research process.

The session emphasises that AI should support research rather than replace it. Participants are encouraged to locate and verify their own academic, legal, policy, and community sources rather than relying on fabricated citations or unverified references generated by AI systems. Once sources are collected, AI tools can then assist with summarising long documents, identifying patterns, or simplifying technical language while still requiring careful fact-checking.

The slide also introduces AI-supported qualitative analysis. Participants will learn how anonymised excerpts from interviews, testimonies, or focus groups can be used to generate possible thematic codes or categories for analysis. This can help organise large amounts of qualitative material and identify recurring experiences, themes, or forms of discrimination.

At the same time, the session reinforces strong privacy and data protection practices. Entire datasets containing identifiable testimonies or sensitive information should never be uploaded into remote AI systems. Instead, participants are encouraged to work with anonymised excerpts, remove identifying details, use local tools where possible, and maintain clear manual links between AI-generated analysis and the original source material.

4. Advocacy Prompt Patterns: Harm → Right → Remedy Across Languages and Formats

This slide explores how AI tools can support advocacy, public education, and campaign communication through a rights-based framework. Participants will learn how to structure prompts around a simple harm–right–remedy model that connects lived experiences to legal protections, political analysis, and demands for change.

You will learn how to begin with a short, anonymised description of a real-world issue and ask AI systems to help generate explainers, speeches, campaign messages, social media threads, or translated materials that frame the issue in terms of rights rather than charity or pity. AI can also help adapt the same message into multiple formats for different audiences, including local communities, media outlets, international organisations, or public meetings.

The session emphasises the importance of maintaining political clarity and movement language when using AI tools. Participants are encouraged to include key terms and definitions directly in their prompts so that concepts such as caste-based discrimination, labour exploitation, or minority rights are not diluted or replaced with softer or depoliticised language.

You will also learn how to critically review AI-generated outputs for harmful framing, stereotypes, or passive representations of affected communities. If the system produces language that portrays people only as victims or objects of sympathy, participants are encouraged to revise the output through follow-up prompts that restore agency, resistance, organising, and collective action to the narrative.

Throughout the session, AI is presented as a tool that can help strengthen advocacy communication while remaining accountable to community dignity, political accuracy, and rights-based storytelling.

5. Maths, Science, and Policy Calculations: Using AI as a Calculator With Context

This slide introduces practical ways AI can support calculations, estimates, and technical reasoning in advocacy, research, and policy work. Participants will learn how AI tools can assist with first-pass calculations, budgeting exercises, wage estimates, climate explanations, and other quantitative tasks that are often part of social justice and rights-based work.

You will learn how to ask AI systems to break down calculations step by step so that the reasoning behind the numbers remains clear and understandable. This can include estimating wage theft, comparing budgets, calculating programme costs, or understanding the financial implications of proposed policy changes.

The session also demonstrates how AI can help explain technical concepts in simpler language. Participants are encouraged to use AI to better understand economic, legal, climate, or policy-related terms that are often used in professional or institutional spaces. This can support stronger participation in discussions that may otherwise feel inaccessible or dominated by technical experts.

At the same time, the slide clearly distinguishes between rough estimation and expert verification. AI-generated calculations should be treated as starting points rather than final authority, especially when dealing with high-stakes issues such as legal penalties, public budgets, health data, or scientific claims. Participants are encouraged to verify important figures with trusted sources, official records, or qualified experts before using them in reports, campaigns, or advocacy work.

Overall, the session frames AI as a tool that can make technical knowledge more accessible while still requiring human judgment, verification, and political awareness.

6. Best-Practice Toolkit: Iterative Prompting, Personas, Checklists, and Bias Logs

This final slide introduces practical habits and tools that participants can begin using immediately in their everyday AI interactions. It encourages participants to treat AI prompting as an iterative process rather than expecting perfect results on the first attempt. Instead of trying to write the “perfect” prompt immediately, participants are encouraged to review the first response critically, identify where the system misunderstood the task, and refine the instructions step by step.

You will also learn how to use “persona prompts” to make AI responses more contextually grounded and useful. By specifying a role, audience, or political perspective within the prompt, participants can often generate outputs that are more relevant to legal education, community organising, advocacy, or public communication work.

The slide further introduces a simple self-check process for important AI tasks. Participants are encouraged to ask themselves whether they clearly defined the audience, jurisdiction, purpose, sources, and assumptions behind the request before relying on the output.

You will also learn how to ask AI systems to critique their own answers by identifying alternative viewpoints, possible weaknesses, or missing perspectives. This can help expose hidden assumptions, stereotypes, or incomplete reasoning within the model’s response.

Finally, the session encourages participants to maintain a shared “bias log” that documents harmful, inaccurate, or revealing AI outputs related to caste, religion, gender, migration, disability, or minority communities, alongside corrected versions and notes. Over time, this record can become both a teaching resource for future participants and a form of evidence when advocating for accountability, fairness, and better platform practices from AI companies and digital systems.

Edges to Centre: A Global Intergenerational Leadership Project for Marginalized Communities “Nothing About Us Without Us”

NextGen Rights Lab: Global South Intersectional Rights & Leadership Accelerator

Artificial Intelligence Literacy and Practice Track for Caste Affected
Intergenerational Leaders

Session – 3 (40 mins) - Work, Enterprise, and Collective Action: Using AI Without
Losing Rights

1. AI at Work: Concrete Tasks It Can Do for You (and Ones It Shouldn't)

This slide explores how AI tools can support the everyday responsibilities participants already manage in workplaces, collectives, community organisations, freelance work, or small enterprises. Rather than presenting AI as something abstract or highly technical, the session focuses on practical situations participants may already encounter.

You will learn how AI can assist with tasks such as researching organisations and competitors, identifying patterns in survey responses or workshop feedback, preparing job applications, improving CVs, organising notes, and drafting professional materials. For example, AI can help group open-ended responses into themes, compare organisations working on similar issues, or highlight relevant skills in a job application while still preserving the accuracy of your experience.

The session also emphasises that AI-generated outputs should always be treated as drafts that require review, editing, and human judgment. Participants are encouraged to maintain honesty and accuracy when using AI to rewrite or organise professional materials rather than exaggerating or fabricating qualifications, achievements, or experiences.

Strong attention is also given to privacy, confidentiality, and organisational safety. Participants are encouraged not to upload sensitive contracts, donor agreements, HR records, or identifiable workplace information into cloud-based AI systems. When working with sensitive material, key details should be anonymised, paraphrased, or masked wherever possible.

Throughout the slide, AI is framed as a support tool that can reduce repetitive labour, improve organisation, and strengthen communication while still requiring critical thinking, verification, and ethical responsibility.

2. Marketing and Content: Avoiding Stereotypes While Selling Real Products and Ideas

This slide explores how AI can support marketing, storytelling, and public communication without reinforcing harmful stereotypes related to caste, gender, religion, poverty, or marginalised communities. Participants will learn how to use AI tools to create product descriptions, social media posts, newsletters, and organisational messaging that centre dignity, skill, labour, and fairness rather than pity or charity narratives.

You will learn how to write prompts that clearly guide AI systems toward respectful and accurate language. For example, when describing products or services created by Dalit, Adivasi, minority, or women-led collectives, participants are encouraged to focus on craftsmanship, design, fair wages, sustainability, community ownership, and rights-based livelihoods instead of framing communities as helpless or dependent.

The session also emphasises the importance of reviewing AI-generated text carefully before publication. Participants are encouraged to remove language that exoticises communities, romanticises poverty, or exaggerates impact beyond what is truthful and accountable. AI outputs should reflect the real scale, values, and political commitments of the work rather than producing overly polished or misleading narratives.

Accessibility and language justice are also integrated into the communication process. Participants are encouraged to use AI tools to create simpler language versions, multilingual content, alt text for images, and communication that remains accessible to screen readers and diverse audiences.

Throughout the session, AI is framed as a communication support tool that can strengthen outreach and visibility while remaining grounded in honesty, dignity, accessibility, and community accountability.

3. CVs, STAR Stories, Mock Interviews: Practising for Biased Labour Markets

This slide introduces how AI can be used as a supportive training partner for people navigating biased and unequal job markets, including Dalit, Bahujan, minority, and queer candidates. The focus is on helping participants structure and communicate their real experiences more clearly, without changing or fabricating their truth.

You will learn how to use the STAR method (Situation, Task, Action, Result) to turn lived experiences into strong interview narratives. AI can help you shape these stories in a structured way, based only on the details you provide, while highlighting skills such as leadership, problem-solving, coordination, or advocacy.

The session also shows how AI can be used to prepare for interviews more effectively. Participants can generate role-specific practice questions, rehearse answers, and receive feedback on clarity, structure, and conciseness. This helps build confidence when applying for roles in research, advocacy, policy, or development sectors.

You will also learn how to use AI for bias-aware interview practice. This includes simulating difficult or inappropriate questions that candidates may face, and practising safe, professional, and assertive responses in advance. This prepares participants to navigate real-world workplace bias while maintaining dignity and self-protection.

Throughout the session, participants are reminded that AI should never be used to invent or exaggerate experience. Its purpose is to help structure and communicate real lived experience in a clear, professional format, while allowing individuals to retain their identity, context, and political integrity wherever possible.

4. CSOs and Movement Operations: Grants, Logframes, Memos, FAQs

This slide focuses on how AI can support the administrative and bureaucratic work of civil society organisations and movement spaces, while ensuring that all final decisions, language, and accountability remain with people.

You will learn how AI can help draft and refine common organisational documents such as concept notes, proposals, reports, and community-facing summaries. For example, AI can turn structured bullet points into a donor-friendly concept note that still clearly names caste discrimination, or simplify technical planning documents like logframes so they can be understood by community leaders and non-specialist audiences.

The session also introduces how AI can support programme design and evaluation thinking by suggesting possible indicators for measuring impact. This includes both quantitative and qualitative indicators that can help organisations reflect on whether their work is increasing awareness, access, or legal empowerment among affected communities.

You will also see how AI can assist with internal coordination tasks such as drafting board updates, donor FAQs, volunteer onboarding materials, and internal communication notes. Participants are encouraged to provide their own organisational language as input so that outputs remain aligned with their existing tone and values.

At the same time, the slide emphasises strong safeguards around confidentiality. Sensitive information such as donor negotiations, budgets, internal conflicts, or identifiable community data should be removed or anonymised before using any external AI system. All AI-generated drafts should be clearly marked and treated as works in progress that require careful human review before being shared or finalised.

5. Creator and Small Enterprise Workflows: From Product Listings to Policy Pages

This slide focuses on how AI can support practical, income-generating work for creators, small enterprises, and collectives from caste-affected and minority communities. The emphasis is on using AI to reduce repetitive work while strengthening clarity, communication, and reach.

You will learn how AI can help create multiple versions of product listings and creative descriptions tailored to different audiences. For example, the same product — such as a zine on Dalit feminist voices — can be described differently for global readers, Indian audiences engaged in caste discourse, or institutional buyers like university libraries, while keeping the core message consistent.

The session also shows how AI can support basic business infrastructure by drafting standard policy pages such as shipping, returns, privacy policies, or terms of service. Participants are reminded that these outputs must always be checked carefully and adapted to reflect their real practices, legal context, and organisational capacity.

You will also learn how AI can help create response templates for customer communication. This includes drafting polite, firm replies to objections or misunderstandings about anti-caste or rights-based messaging, helping reduce emotional labour while maintaining clarity about values and commitments.

For cross-border or multilingual work, AI can assist in translating product listings, FAQs, and outreach materials into multiple languages. However, the session emphasises that translations must always be reviewed and corrected by humans wherever possible, since cultural meaning, political context, and sensitive terminology can easily be distorted or lost.

6. Operations Copilots, Personal Planning, and Vendor Interrogation

This final slide introduces the idea of “operations copilots” — using AI to support everyday organisational, team, and personal coordination tasks in a structured and responsible way.

You will learn how AI can turn messy meeting notes or transcripts into clear, structured outputs such as summaries, decision lists, action points, deadlines, and open questions. These outputs can then be reviewed and integrated into project management tools and team workflows.

The session also shows how AI can help draft first versions of organisational systems and processes, such as standard operating procedures (SOPs). For example, participants can use AI to outline a step-by-step response plan for situations like reports of online caste harassment, including documentation, safety considerations, and referral pathways. These drafts are always meant to be reviewed, adapted, and formally approved by the organisation.

The slide also reinforces privacy-first practices for personal use. AI can be used for everyday planning such as travel, meals, or routines, but sensitive personal data such as ID numbers, full addresses, medical records, or unredacted legal documents should never be shared with external systems. Participants are encouraged to use placeholders and general descriptions when needed.

Finally, the session introduces a critical approach to working with AI vendors and consultants. Participants are encouraged to ask questions about data sources, accuracy across languages and demographics, audit processes, data deletion policies, and grievance mechanisms. This helps shift users from passive consumers of AI tools to informed negotiators who can demand transparency, accountability, and fairness from technology providers.

Edges to Centre: A Global Intergenerational Leadership Project for Marginalized Communities “Nothing About Us Without Us”

NextGen Rights Lab: Global South Intersectional Rights & Leadership Accelerator

Artificial Intelligence Literacy and Practice Track for Caste Affected
Intergenerational Leaders

Session – 4 (40 mins) - Safety, Limits, and the New Frontier: Multimodal,
Browsing, Analytics, Reasoning

1. Limits and Biases as Design Facts (Not Bugs to Wish Away)

This slide begins the final session by helping you understand that limitations and harms in AI systems are not random mistakes — they are a result of how these systems are designed, trained, and deployed.

You will learn that AI models are trained on data that over-represents dominant voices, languages, and perspectives. As a result, the lived experiences of Dalit, Bahujan, Adivasi, minority, and regional language communities may be missing, distorted, or simplified. When the system does not have enough reliable information, it may still produce confident-sounding answers that are incorrect or misleading (“hallucinations”).

The slide also introduces the idea of prompt injection in simple terms. This is a safety risk where external text (from websites, documents, or user inputs) can try to override instructions given to the AI system. For example, a malicious instruction hidden in a document may attempt to trick the system into ignoring safety rules or exposing information, which is why untrusted inputs must always be handled carefully.

You will also see examples of how harm can appear in AI outputs, such as oversimplifying caste discrimination, producing misleading translations, or making claims that erase context and structural violence.

The key practice taught in this session is to treat all AI outputs as suggestions, not truths. Participants are encouraged to be especially cautious with precise facts, numbers, or historical claims, and to avoid using AI for high-stakes judgments about individuals or communities, since hidden biases can influence outcomes in ways that are not immediately visible.

Throughout the slide, AI is framed as a tool that can assist thinking, but must always be checked, questioned, and grounded in lived reality and critical human judgment.

2. Reading Model Cards, Platform Policies, and Using Reporting Channels

This slide helps you understand how to read and interpret the documentation that comes with AI systems so you are not using these tools blindly.

You will learn about “model cards” — documents that describe who built a model, what kinds of data it was trained on (at a high level), what it is intended to be used for, and what limitations or known risks have been identified. These documents are a key way to understand what an AI system can and cannot reliably do.

You will also learn how to identify warning signs in model documentation or product pages. These include vague descriptions of training data, lack of transparency about testing in Global South contexts, or broad claims about “fairness” that are not supported by evidence or detail.

The session also explains platform content policies in clear, practical terms. These rules are not only about punishment or enforcement — they are also tools that users can use to challenge and report harmful AI outputs, especially when systems produce casteist, discriminatory, or misleading content.

You will learn a practical reporting approach for documenting harm. When an AI system produces harmful or biased output, it is important to record the full prompt, the full response, timestamps, and screenshots, along with any relevant policy references. This makes complaints stronger and harder to dismiss.

Finally, the slide encourages collective reporting and coordination. When multiple users report similar harms, platforms are more likely to respond seriously, making coordinated documentation and accountability efforts more effective than isolated complaints.

3. Multimodal Tools: Images, Audio, and the Ethics of Seeing and Being Seen

This slide introduces multimodal AI systems—tools that can work with images, audio, text, and sometimes video—and explains both their benefits and the ethical risks they raise.

You will learn that image-to-text tools can describe photos, charts, and scanned documents, which can support accessibility for blind or low-vision users and help extract information from reports, government documents, or visual data. Speech-to-text systems can transcribe interviews, meetings, or testimonies, while text-to-speech tools can convert written content into audio using different voices and accents, improving access for diverse audiences.

The slide also covers image generation tools that create visuals from text descriptions. While these can be useful for communication and education—for example, generating illustrations for workshop materials—they also introduce serious risks around consent, misinformation, and deepfakes.

You will learn that these tools should never be used to create realistic or misleading images of real people, especially survivors, opponents, or members of the public in false or harmful contexts. Participants are encouraged to critically evaluate any striking or sensational image they encounter and check whether it may be AI-generated rather than real.

The session also emphasises responsible use practices, including clearly labelling AI-generated images and educating communities about the existence of deepfakes. This helps build informed scepticism rather than panic, and supports safer, more transparent communication in digital spaces.

4. Advanced Data Analytics: Spreadsheets, Visualisations, and Protecting Community Data

This slide shows how AI-powered tools can support working with structured data such as spreadsheets, CSV files, and survey exports, when used carefully and responsibly.

You will learn a practical workflow using an example dataset of anonymised survey responses from Dalit youth about experiences of discrimination in educational institutions. This data may include fields such as age, gender, state, type of incident, and reported outcomes. AI tools can help organise and explore this data by answering structured questions and generating basic summaries or visualisations.

For example, you can ask AI to calculate how many incidents are reported per state and display the results as a bar chart, or to explore whether reporting patterns differ across gender categories. These outputs can help identify broad trends in large datasets and support initial analysis.

The session strongly emphasises data protection and anonymisation. Before using AI tools, participants must remove personal identifiers such as names, exact locations, phone numbers, and any free-text entries that could reveal identity. Where datasets are small or sensitive, information should be further aggregated, such as using age ranges instead of exact ages.

You will also learn the importance of transparency in analysis. Participants are encouraged to inspect how results are produced by checking intermediate steps, formulas, and transformations, rather than relying only on final outputs. It is important to verify whether any data has been dropped, misinterpreted, or transformed incorrectly during processing.

Finally, the slide reinforces that AI-generated patterns should not be over-interpreted. Correlations in data do not automatically indicate causation, and quantitative findings must always be interpreted alongside community knowledge, lived experience, and ethical judgment before being shared or used in public advocacy.

5. Safe, Purpose-Limited Web Browsing and Custom Assistants for Movements

This slide covers how AI systems that can browse the web or function as custom assistants can support rights-based research and organisational work, when used with clear limits and safeguards.

You will learn how to use web-enabled AI in a focused and purpose-driven way. Instead of allowing it to search broadly or rely on trending content, participants are encouraged to ask for specific documents or reports and request structured summaries. For example, AI can be used to locate a relevant UN Special Rapporteur report and extract recommendations that apply to caste-affected or marginalised communities.

The session emphasises verification as a non-negotiable step. Whenever AI provides summaries or cites sources, participants are expected to open at least one original source and cross-check whether the information is accurate, complete, and up to date.

You will also learn about custom AI assistants designed for organisational use. These systems can be trained or configured using internal documents such as reports, glossaries, policies, and guidance notes. This allows teams to create tools like a “caste justice FAQ assistant” that can respond using consistent, rights-based language and organisational knowledge.

At the same time, the slide highlights strict boundaries for safety and governance. Custom assistants should have clearly defined scope, limited access, and controlled update permissions. They must be prevented from answering outside their knowledge base, and sensitive or case-specific confidential information should never be uploaded into shared systems.

Overall, the session frames these tools as powerful but bounded systems that require careful design, human oversight, and strong data discipline to ensure they remain aligned with rights-based work.

6. Modern Reasoning, Deep Agents, and What Movements Should Demand

This final slide introduces the emerging language of “reasoning models,” “agents,” and “deep research” tools, and equips you to respond to them with clarity, caution, and informed expectations rather than awe or confusion.

You will learn that techniques such as chain-of-thought, self-critique, or tree-of-thought prompting are ways of asking AI systems to show their intermediate steps. This can be useful because it helps reveal assumptions, uncertainties, or errors that would otherwise remain hidden behind a single confident answer.

The session encourages you to explicitly request reasoning when working on legal, policy, data, or advocacy questions. Asking the system to explain its thinking step by step—and to highlight where it is guessing—helps you better assess the reliability of its output.

You will also be introduced to “agentic systems,” which are AI setups that can plan, use tools, browse the web, and carry out tasks with limited supervision. While these systems can improve efficiency, they also increase risk, since errors in interpretation or instructions can quickly lead to a chain of incorrect or harmful actions.

Finally, the slide shifts from individual use to collective responsibility and demands. It outlines what Dalit, Bahujan, Indigenous, and minority communities can and should expect from AI developers and regulators, including transparency about training data and biases, strong opt-out and deletion options, independent audits that include caste and Global South perspectives, accessible grievance mechanisms, and meaningful participation of affected communities in governance.

The session closes by reinforcing that you do not need to become an AI expert to engage with these systems. What matters is developing enough understanding to use them critically and strategically—so they remain tools for learning, organising, and survival, rather than systems that override rights, erase histories, or constrain political imagination.

Thank You for your attention.