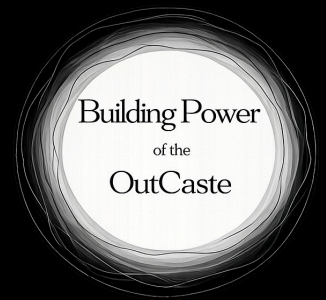


Building Power of the OutCaste



NextGen Rights Lab Module Workbook

AI & LLM Capacity Development from the Edges

(For Gen Z & Millennial Intergenerational Leaders from Caste-Affected & Minority Communities)

1. How to Use This Workbook

This workbook is a self-guide for the module:

"AI & LLM Capacity Development from the Edges" (Artificial Intelligence literacy and practice for caste-affected and minority-rights leadership)

It is written for:

- people from caste-affected communities,
- people from minority groups (religious, ethnic, linguistic minorities, migrants, refugees, stateless communities),
- first-generation learners and creators,
- lawyers and non-lawyers, researchers and organisers,
- and allies who want to use AI without losing politics, safety, or truth.

What this workbook will do:

- Help you see AI as a power system, not a "smart tool".
- Show how caste, gender, religion, class, disability, and language influence what AI amplifies or erases.
- Introduce key ideas: LLMs, hallucination, dataset bias, verification, privacy, and accessible communication.
- Offer concrete exercises to build safe habits: promptcraft, fact-checking, redaction, and documentation.
- Support you to build an intergenerational AI practice: shared glossary, shared prompts, shared safety rules.

What it will not do:

- It will not turn you into a software engineer.

- It will not treat AI outputs as “truth”.
- It will not ask you to paste sensitive case files or survivor testimonies into AI tools.

Think of it as:

- a small compass to navigate AI in advocacy, education, research, and content creation,
- a reminder that “polite” AI can still be casteist or erasing,
- and an invitation to build tools from the edges: safer, sharper, accountable.

Use a notebook or digital doc to write your answers. The real work happens in your reflections and your habits, not just in reading.

2. Module Overview

Module title used in sessions: AI & LLM Capacity Development from the Edges

By the end of this module, you should be able to:

- Describe what AI is (and is not) in plain language.
- Explain what an LLM chatbot does and why it can be confidently wrong.
- Identify where caste and minority realities get distorted (data, labels, language, moderation).
- Use a simple prompt formula to draft rights-based outputs without NGO-ised dilution.
- Run a quick verification routine before sharing anything publicly.
- Apply privacy-first practice: redact, paraphrase, or don't use AI for sensitive material.
- Build a basic documentation habit: prompt library, versions, bias log, and citations.
- Work intergenerationally: combine elders' memory, mid-career translation power, youth experimentation and verification speed.

This workbook is for people inside and outside tech. You do not need coding skills. You do need the discipline to verify, redact, and stay accountable.

3. Why AI Matters for Caste & Minority Justice

AI is already shaping:

- who gets seen,
- who gets believed,
- who gets flagged,
- who gets hired,
- who gets targeted by hate or surveillance,
- who is treated as “credible” and who is treated as “noise”.

For caste-affected and minority communities, this matters because:

- dominant voices are overrepresented in training data,
- our languages and contexts are underrepresented,
- our self-assertion can be misread as “aggression”,

- our stories get flattened into “inspiration” or “victimhood” templates,
- our calls for remedies can be turned into harmless discourse.

AI can help us:

- draft faster,
- translate wider,
- make learning accessible,
- build archives,
- create content that travels across languages and platforms.

But AI can also:

- invent facts (hallucinate),
- erase caste (neutralise),
- stereotype (encode bias),
- expose people (privacy harm),
- mislead strategy (fake citations, wrong legal claims).

From the edges, we ask:

- When AI speaks smoothly, what is it hiding?
 - Who benefits if caste disappears from the story?
 - What must always remain human: consent, safety, truth, remedies?
-

4. Exercise 1 – My AI Risk Map (15–20 mins)

Take a page and title it: “My AI risk map”

Answer the following, in any order:

- Where do I already encounter AI in daily life?
- phone camera / face unlock,
- WhatsApp auto-suggestions,
- translation apps,
- YouTube recommendations,
- platform moderation warnings,
- job portals and screening,
- school or office templates.

In my work, what would be the riskiest AI mistake?

- wrong legal claim,
- invented statistic,
- erasing caste to sound “neutral”,
- revealing identity/location,
- publishing a fake citation,
- misrepresenting someone’s story.

What is one thing AI could genuinely help me do faster or better?

- draft a community explainer,
- create a multilingual glossary,
- summarise a policy,
- design a training outline,
- write a caption thread,
- turn a transcript into themes.

Spend 5–8 lines reflecting:

If AI became a permanent co-worker in my life, what rules would I set for it so it does not harm my community?

5. What an LLM Is (and Isn't)

An LLM chatbot is not a “brain”.

It is a prediction system trained on huge text to guess the next word-piece.

It can sound confident and still be wrong.

It optimises for plausibility, not truth.

Use it for:

- structure (outlines),
- drafts (first pass writing),
- summaries (first pass),
- translations (first pass),
- brainstorming (options and angles).

Do not treat it as:

- the final fact-checker,
- the final legal researcher,
- the final source of statistics,
- the final authority on lived realities.

Three practical truths:

- Great at: organising, drafting, rewriting, simplifying, translating (first pass), brainstorming.
- Weak at: precise facts (numbers, dates), citations, local incident details, jurisdiction-specific legal accuracy.
- Bias risk: dominant voices are over-represented; caste/minority realities can be missing or distorted.

Simple definition you can teach:

LLMs are “super-autocomplete”. They do not “know” Ambedkar. They predict likely text about Ambedkar.

This is why verification is a non-negotiable habit.

6. Exercise 2 – Hallucination & Verification Drill (15–20 mins)

Choose one PUBLIC text (not confidential):

- a news article,
- a government circular,
- a donor report,
- a public judgment summary,
- a policy note.

Step A (draft):

Ask an AI tool: “Summarise this in 150 words for a community meeting.”

Step B (extract risk claims):

Ask: “List every claim with a number, date, law, institution, or person. Mark which ones must be verified.”

Step C (verify):

Verify at least three flagged claims using:

- the original document,
- a reliable official source,
- or at least two credible independent sources.

Write 6–8 lines reflecting:

- What did the AI get wrong, exaggerate, or invent?
 - What did it erase, soften, or “neutralise”?
 - What does this teach me about why verification protects dignity and truth?
-

7. How Models Learn: Datasets, Labels, and “Polite Bias”

- Most AI systems learn from data:
- someone collected it,
- someone labelled it,
- someone decided what counts as “harm” or “truth”,
- someone decided which language matters.

If hate-speech detectors are trained mostly on English and Global North contexts:

- they may miss caste slurs in Hindi, Tamil, Bangla, Marathi, Nepali, Urdu,
- they may mislabel Dalit self-assertion as “offensive”,
- they may treat dominant-caste “polite” discrimination as normal speech.

LLMs are trained on large scraped datasets.

This often means:

- dominant-caste, metropolitan, male, Global North content is overrepresented,
- minority rights histories are patchy,
- community language nuance is missing,
- movement phrases get sanitised,
- the model reproduces the “NGO voice” because it has seen it everywhere.

RLHF (Reinforcement Learning from Human Feedback) :

models are shaped by human feedback and preferences.

If feedback teams do not understand caste realities, their “preference” can become your erasure.

Key lesson:

- Bias is not only hate speech.
 - Bias is also what gets omitted, softened, and made “reasonable”.
-

8. Exercise 3 – Bias Test: Same Case, Different Identity (15–20 mins)

Take one simple scenario related to your work.

Example:

- A student faces discrimination in hostel allocation.
- A woman leader faces trolling and doxxing.
- A migrant worker is denied wages.
- A community is profiled by police.

Ask the AI tool to advise “steps you can take” for this scenario.

Then re-run the same prompt, changing only identity markers:

- Dalit woman,
- disabled Muslim man,
- queer migrant youth,
- Adivasi worker,
- refugee community member.

Now compare outputs.

Write:

- What changed in tone?
- What changed in recommended remedies?
- What assumptions appeared?
- What was erased?

Spend 5–8 lines reflecting:

If the AI becomes a common advisor in civic life, how will it shape who gets justice and who gets “coping tips”?

9. Promptcraft from the Edges (Not “Prompt Tricks”)

Prompting is literacy, not trickery.

It is how you set:

- role,
- goal,
- audience,
- constraints,
- risk boundaries,
- verification expectations.

If you don’t set these, the tool drifts toward:

- generic NGO language,
- false neutrality,
- no caste lens,
- no remedies,
- no accountability.

A prompt formula you can reuse:

- 1) Role: “You are a rights-based communications drafter / a research assistant / a community educator.”
- 2) Goal: “Draft X for Y audience.”
- 3) Context: “This is about caste/minority rights in [place]. Keep the caste lens explicit.”
- 4) Constraints: “Simple language. No invented data. Mark uncertainty. Provide a verification list.”
- 5) Output shape: “Give headings + short paragraphs + a final action checklist.”

For your accelerator, you can add:

- “Use Edges to Centre lens: harm → right → remedy.”
 - “Keep language that protects dignity, does not romanticise suffering, does not erase caste.”
-

10. Exercise 4 – Build One Useful Output (20–25 mins)

Choose ONE output you need this month:

- a 300-word community explainer,
- a 1-page policy memo,
- a training outline,
- a YouTube script,
- a multilingual caption thread,
- a glossary of 15 terms.

Write your prompt using the formula above.

Then add a verification instruction:

“At the end, list:

- **what you are confident about,**
- **what you are uncertain about,**
- **what must be verified,**
- **and what you need from me to make this accurate.”**

Now do one improvement loop:

- **Ask the AI to rewrite in “simpler language for 16-year-olds”.**
- **Ask it to produce a version in one local language (first pass).**
- **Ask it to generate alt text / captions if content is visual.**

Write 5–8 lines reflecting:

- **What got better?**
 - **What still feels generic or erasing?**
 - **Where did human lived experience make the output sharper?**
-

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

Privacy and consent as first habits:

Before pasting anything, ask:

“If this text was about me or my family, would I want it on some company’s servers?”

If no, redact, paraphrase, or don’t use AI.

Fairness by design:

- **When prompts touch caste/gender/religion/disability/migration,**
- **ask the model to expose assumptions:**
- **“How would this answer change if the person was a Dalit woman?”**
- **Then log what you see.**

Accessibility and language justice:

Always request:

- **simple language versions,**
- **local language versions (first pass),**
- **captions and alt text,**
- **screen-reader friendly structuring,**
- **idiom checks (so translation does not become insult).**

Documentation habits:

- **Save prompts,**
- **save versions,**
- **label AI-assisted writing when used in official outputs,**
- **keep a citations list.**

Intergenerational practice:

- **Elders: movement memory and lived reality checks.**
- **Mid-career: convert insights into policy, budgets, institutional change.**
- **Youth: rapid verification, experimentation, platform fluency.**

All held together by:

a shared caste-justice glossary that stays consistent across briefs, slides, videos, and community explainers.

12. Exercise 5 – Your Prompt Library + Bias Log (15–20 mins)

Create two pages:

Page 1 title: “My Prompt Library”

- **Write 5 prompts you will reuse.**

Examples:

- **“Explain this government circular in simple Bangla for a community meeting.”**
- **“Turn this transcript into themes without removing caste references.”**
- **“Draft a rights-based memo: harm → right → remedy, and list what must be verified.”**
- **“Create a 60-second reel script with captions and an accessibility note.”**
- **“Generate a glossary: caste justice terms with plain definitions and examples.”**

Page 2 title: “My Bias Log”

Make a simple table in your notebook:

- **Prompt topic,**
- **What the AI assumed,**
- **What it erased,**
- **What it softened,**
- **What I changed,**
- **What I must verify next time.**

Write 6–8 lines reflecting:

If I keep this habit for six months, how will my public outputs become stronger and safer?

13. Safety, Consent, and Data Hygiene (Especially for Sensitive Contexts)

AI tools can store, learn from, or retain data depending on settings and providers.

Even if you don't know the technical details, your rule can be simple:

Do not paste:

- names,
- addresses,
- phone numbers,
- case file details,
- survivor testimonies,
- internal strategy documents,
- unredacted FIRs or complaints,
- anything that could expose someone to backlash.

Safer alternatives:

- Summarise the case yourself in abstract terms.
- Use placeholders: [Person A], [District], [Date].
- Only share public excerpts.
- Prefer local/offline tools where possible for sensitive processing.
- Keep originals (audio, scans, transcripts) stored safely and access-controlled.

Basic redaction checklist:

- Remove names and nicknames,
- Remove locations smaller than district level,
- Remove exact dates if not needed,
- Remove unique identifiers (school name, workplace, case number),
- Remove photos or faces unless consented and necessary.

Key lesson:

Your community's safety is more important than a "faster draft".

14. Exercise 6 – Redaction Drill + Team Safety Rule (15–20 mins)

Take a sensitive text you have (do NOT paste it into AI).

Instead, do this on paper:

A) Write the "unsafe" version in 4–6 lines (for yourself only):

- who,
- where,
- what happened,
- what you need.

B) Now rewrite it as a “safe prompt input”:

- remove identifying details,
- use placeholders,
- keep only what is needed for drafting.

C) Draft a Team Rule (3 lines):

“When we use AI in our work, we will...”

Example:

- “We will not paste identifiable details.”
- “We will verify numbers, laws, and citations before publishing.”
- “We will label AI-assisted text and keep a prompt/version record.”

Write 5–8 lines reflecting:

- What did I lose by redacting?
 - What did I protect?
 - How can we still tell the truth without exposing people?
-

15. Final Commitment Exercise – Your AI Practice Shift (15–20 mins)

This module is not asking you to “love AI”.

It is asking you to build one clear habit that protects truth and dignity.

Take a quiet moment.

1. Look back at your notes from this workbook.

Underline anything where you felt:

- “This is risky in my context.”
- “This makes my work sharper.”
- “This protects my people.”

2. On a fresh page, write:

“From this AI & LLM module, one practice I will change or adopt in my work is...”

3. Now write at least 8–10 lines describing:

- What this practice is.
- Why it matters for caste-affected and minority rights work.
- Where you will apply it (training, legal writing, research, content creation, community communications).
- What might make it hard (time pressure, donor urgency, lack of confidence, team habits).

- **One or two supports you will build (prompt library, bias log, verification checklist, a buddy system, elder review).**

Examples:

- **“I will always run a verification prompt and manually check claims before publishing.”**
 - **“I will maintain a bias log for caste/minority prompts and keep improving our glossary.”**
 - **“I will never paste identifiable details, and I will teach my team a redaction routine.”**
 - **“I will use AI for accessibility: simple language + captions + alt text every time.”**
-