SECOND EDITION

Sovereignty-Aligned AI

for Environmental Negotiations

GAINFOREST · YOUTH NEGOTIATORS ACADEMY

MMXXVI

— A NOTE BEFORE WE BEGIN

This is a living document.

It follows the 2024 guide written for the Climate Youth Negotiator Programme, keeps its interviews and prompts, and updates them for tools that now act as well as answer. Navigating opportunities and recognising limitations to strengthen your advocacy.

—

WORKING DRAFT · 2026

Contents

— CHAPTER 1

Introduction

The UNFCCC negotiation space is dense with data, documents, histories, and constantly evolving terminologies. It is a place where precise language and a strong grasp of multifaceted issues can significantly influence outcomes. This is where technology such as large language models can provide an edge.

1.1 Why should we care?

For youth climate negotiators representing their nations at the United Nations Framework Convention on Climate Change, the stakes are high. They carry the concerns and aspirations of a younger generation, one that will inherit the consequences of today's decisions. Yet they often face difficulties in accessing vast volumes of climate data, understanding the nuances of policy language, or communicating their points effectively.

Language models can help with all three. They process, understand, and generate text based on very large collections of writing, which makes them useful for four things in particular.

Information analysis. Distilling large amounts of climate data and research into concise insights, so that negotiators can keep up with the latest findings.

Policy drafting assistance. Helping negotiators frame their points in language that resonates, adheres to convention, and still stands out.

Cross-cultural communication. Translating and contextualising information, which makes the negotiation environment more inclusive.

Simulation and training. Providing simulated negotiation scenarios so that young negotiators can practise before they step onto the global stage.

The technology itself is only half of the story. Using these tools well signals a broader shift, one where youth combine their commitment to climate action with the tools that shape how decisions get made.

1.2 How to use this guide

The first half explains what these systems are, what negotiators need, and how to ask a model for something useful. The second half is about agents, which read whole document sets and carry out a task in several steps. The last chapters cover who owns your data, how to run a model you control, and what it costs the planet. Every chapter carries prompts you can copy, and none of it needs a technical background.[1]

[1] The first edition was written in 2024 for the Climate Youth Negotiator Programme. The interviews in chapter 3 and the appendix are from that work.

— CHAPTER 2

What are large language models?

2.1 How we got here

In 1957 the psychologist Frank Rosenblatt began the perceptron programme at the Cornell Aeronautical Laboratory, and the US Navy demonstrated a perceptron machine the following year. A perceptron is a simplified mathematical model of a brain cell: it combines inputs, applies weights, and produces an output. It was an abstract model rather than a simulation of a brain, but modern neural networks descend in part from it.

Progress then stalled for decades, limited by the available data, the available hardware, and the available training methods. The field moved again when those three constraints loosened together. In 2012 Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton trained **AlexNet** on 1.2 million labelled images using graphics processing units, and reported a 17.0 per cent top five error rate on ImageNet.[1]

[1] The result came from the architecture, the data augmentation, and the dataset as much as from the hardware.

In 2017 Google researchers introduced the **Transformer**, which relies on attention rather than recurrence. Attention computes weights among the words in a passage according to context, so the model can relate distant parts of a text. Transformers train well in parallel, and they underpin almost every system in this guide.

OpenAI then applied the Transformer to generative pretraining. GPT and GPT-2 showed that a model trained only to predict the next word could pick up many capabilities without a separate trained model for each task. Researchers observed **scaling laws**, meaning that performance improved predictably as computation, data, and model size grew, provided model size and training data stayed in balance.

On 30 November 2022 OpenAI released ChatGPT as a free research preview, and the conversational interface put these capabilities in front of everyone.

Two more shifts followed. In September 2024 OpenAI announced o1, and in January 2025 DeepSeek released R1. Both spend extra computation at the moment of answering, generating longer working notes before they reply, which helps on problems that take several steps. And models began to be wired to tools and to loops, which is the subject of chapter 6.

2.2 The recipe

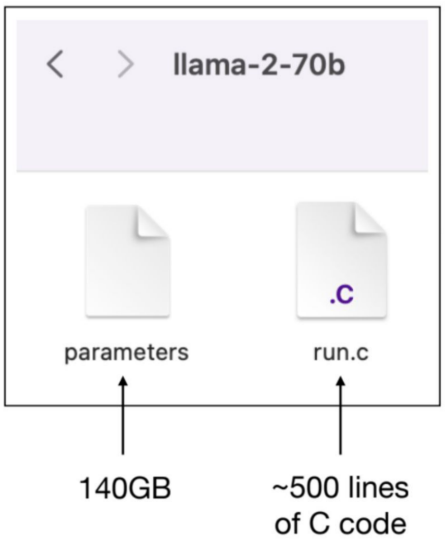
At its core a language model is a type of artificial intelligence designed to understand, generate, and work with human language. Think of it as a vast digital brain, trained on texts from books, articles, websites, and other sources, which it uses to generate human-like text based on the patterns it recognises.

ChatGPT, like other models in OpenAI's GPT series or Meta's Llama, is built in two stages: pretraining, then post-training. Post-training once meant a single round of fine-tuning. It now has two parts, and the second is why recent models can work through long problems.

Pretraining. The model is trained on a vast amount of text from the internet. It does not specifically know which documents were in its training set. The goal is to learn grammar, facts about the world, some reasoning ability, and above all how to predict the next word in a sentence. By doing this it absorbs a surprising amount of information, from trivial everyday facts to complex concepts.

Supervised fine-tuning. The first half of post-training. The model is trained further on a narrower dataset written and rated by human reviewers who follow published guidelines. Reviewers score possible outputs for a range of example inputs, and the model generalises from that feedback to a wide array of user requests. This is what teaches it to follow an instruction rather than simply carry on your sentence.

Reinforcement learning. The second half, and the newer one. Instead of copying a written answer, the model is rewarded for reaching a good result, so it learns which ways of working through a problem tend to pay off. This is what teaches a model to reason across a long problem and hold a chain of steps together before it replies.



On disk a model is two files: the parameters it learned, and a short program that runs them. Llama 2 70B is 140GB of parameters and about 500 lines of C.

2.3 Limitations

Researchers are still unravelling how these models work in detail. The general view is that pretraining embeds a wide array of world knowledge, and that post-training shapes how that knowledge gets used.

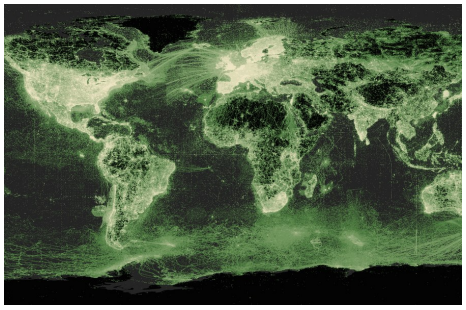
What remains true is that a language model predicts the next word in a sequence from statistical patterns. That method captures a great deal of human knowledge, and it also brings problems with it. Models reproduce the biases of the text they were trained on, and they produce confident statements that are false, which the field calls hallucination. A likely sentence is not a verified one.

2.4 Where the bias comes from

These models are trained on what the internet holds, and the internet does not hold the world evenly. This matters more for a climate negotiator than for almost any other reader.

Take the record of nature itself. The Global Biodiversity Information Facility is the largest open store of species observations on Earth, and roughly 83 per cent of its records come from North America and Europe. The remaining 17 per cent covers everywhere else, including the most biodiverse places on the planet.

The same imbalance runs through text. Far more has been written online in English, French, and Spanish than in Quechua, Twi, or Tok Pisin, so a model has seen far more of the first group. It will translate, summarise, and draft fluently for a delegation working in a well-represented language, and it will fail more often, and more confidently, for one that is not.



Every species observation in GBIF, plotted. The bright areas are where the records are, not where the life is. On raw counts, Belgium and Sweden come out more biodiverse than Brazil and Madagascar.

This is also where a great deal of hallucination comes from. Asked about a well-documented place, a model answers from a dense record. Asked about a thinly documented one, it produces the same confident sentences from far less material and fills the gaps with whatever the pattern suggests. The output does not look less certain. It only is.

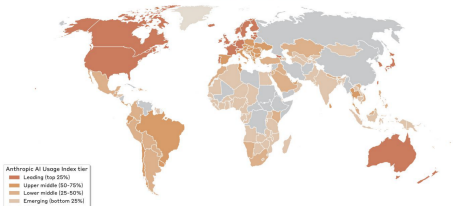
None of this is anyone's fault in a simple way. Collecting and publishing data takes infrastructure and money that many countries do not have, and some communities decline to hand over knowledge about their land and species, which is a legitimate position rather than a failure. The consequence arrives regardless: these tools work best for the places already best served by data, and worst for the places with the most at stake.

The same gap runs through who uses these tools at all. Anthropic's index of Claude use by country tracks national income closely, and twenty-five countries sit in a bottom tier registering almost no measured use.[2] Among them are Tonga, Samoa, Nauru, and Palau, which are parties to the same negotiation and among the most exposed to its outcome.

■ NOTE **What this means in the room**

Treat a model's fluency about your region as a measure of how much has been written about it, not of how much it knows. The less represented your country, your language, or your ecosystem, the harder you should check what it tells you.

[2] Anthropic, *Anthropic Economic Index*, 2025. Usage scales with GDP per working-age person at roughly a power of 0.69.



Claude use per working-age person, by country. Compare it with the map of species records above.

— CHAPTER 3

Stakeholder insights

Interviews with youth climate negotiators across several countries reveal a set of common challenges and a shared view of what would help.[1]

[1] Six negotiators from Liberia, Paraguay, Peru, Nigeria, Lebanon, and Indonesia. The full summary is in the appendix.

3.1 Key challenges and pain points

Language barriers. English is not a first language for most of the people in the room.

Technical language. The jargon of the UNFCCC takes years to absorb (Liberia, Paraguay).

Knowledge transfer. Technical knowledge and information are shared unevenly, and much of it is lost between cycles.

Historical knowledge. Awareness of past negotiations is hard to acquire (Lebanon).

Complex topic communication. Conversations with senior negotiators are difficult to hold under pressure (Peru, Nigeria), and standing has to be earned before the substance is heard.

Understanding party positions. Negotiation dynamics shift quickly and are hard to read.

Time constraints. There is never enough time to process the information that matters.

Expressing complex ideas. Fast and accurate expression in English is a persistent barrier.

3.2 Wished-for tools and resources

Language tools. More sophisticated language and grammar support built for UNFCCC texts rather than for general business writing.

Quick information retrieval. Platforms that scan long documents, pull out what matters, and confirm which text is the current one.

Customized training. Learning resources tailored to a negotiator's region, group, and level of experience.

The tools actually in use at the time of the interviews were Grammarly for writing and Google Drive for collaboration. One negotiator reported using no translation tools at all, on the grounds that the language of the negotiation does not survive translation.

— CHAPTER 4

LLMs for negotiations

Climate negotiations are multifaceted, touching on topics that range from carbon emissions and biodiversity to economic implications and sociopolitical dynamics. Navigating them requires three things at once.

Rapid access to information. Climate science evolves continuously, and negotiators need the latest data at their fingertips.

Effective drafting of agreements. Precision in language can decide whether an agreement succeeds or fails.

Multilingual communication. A global platform demands cross-cultural and multilingual engagement.

4.1 Where models help

Communication. Drafting and translating documents, so that information and agreements are communicated accurately between parties with different native languages.

Information gathering. Collecting and summarising large amounts of material on climate topics, including scientific research, policy documents, and historical agreements, so that negotiators stay informed.

Scenario work. Working through the possible consequences of different policy decisions, which helps a negotiator explore positions and outcomes before committing to one.[1]

Argument analysis. Summarising the arguments made by different parties during negotiations, which helps a negotiator see the key points and the counterarguments.

[1] This is sometimes called predictive modelling, which overstates it. Language models do not run climate or economic models. They restate and compare scenarios, and any number they produce needs checking against a real source.

4.2 Potential pitfalls

There are real benefits here, and there are caveats that matter in climate diplomacy.

Simplification of complex issues. These systems compress, and compression removes the qualifier that a party fought three sessions to insert.

Lack of emotional intelligence. A model does not read the emotional and political currents of a negotiation, and it does not know what was agreed in a corridor.

Lack of human expertise. Fluency in the register of climate policy is not the same as expertise in climate policy. Relying on a model for either can produce incomplete or inaccurate information.

Security and privacy. Discussions often involve sensitive material. Anything

pasted into a hosted system leaves your delegation, and the terms under which it is stored are rarely read.

Ethical considerations. Be transparent about automated assistance, keep responsibility for decisions with people, and watch for misuse.

Miscommunication and misinterpretation. A model can misread the intent behind a request and return text that sounds like your position without being it.

— CHAPTER 5

Using the models

5.1 Which model, and on whose terms

Before picking an assistant, know that there are two kinds, and the difference matters more than the brand.

Closed weights. ChatGPT, Claude, and Gemini run on company servers and are reached through an account. They are the strongest models and the quickest to start with. The free tiers handle drafting, translation, and summarising. Paid tiers cost around twenty dollars a month and add longer documents, file uploads, and the agent features in chapter 6. Everything you type goes to the company, and depending on your settings it can be retained, reviewed, or used to train the next model.

Open weights. Llama, Mistral, Qwen, Gemma, and DeepSeek publish their parameters, so the same model can be run by you, by your institution, or by a provider you choose rather than one you are given. Ollama puts one on a laptop in a few minutes. Polly, run with GainForest and the Youth Negotiators Academy at polly.ai4cop.org, is a hosted open weights model that keeps no record of what you send. Chapter 11 covers running your own.

A rule to start with. Use a hosted closed model for material that is already public, and an open weights model for anything that is not. Do not paste a position your delegation has not tabled into a service you do not control. Chapter 10 explains what is at stake in that choice, and it is worth reading before you open an account rather than after.

Whichever you choose, install the mobile app. Voice input is useful when you are walking between sessions, and dictating a rough thought in your own language often produces a better draft than typing a careful sentence in English.

5.2 Prompt engineering 101

The instruction you type is called a prompt. Writing one is the process of creating specific, well-defined instructions to help you solve a task, and the quality and specificity of what you write shapes the answer you get back.

A prompt can contain any of the following, and does not need all four.

Instruction. The task you want the model to perform.

Context. Background information that steers the model toward a better response.

Input data. The text, question, or document you want it to work on.

Output indicator. The type or format of the output you want.

5.3 What makes a prompt good

Six moves do most of the work, and they hold whichever model you are using.

Be clear, direct, and detailed. Say who you are, what you want, and what the output should look like. A model cannot infer the room you are standing in.

Show an example. One sample of the output you want teaches faster than a paragraph describing it.

Ask for quotes. Tell the model to quote the passage it is relying on. Grounding an answer in a retrieved document is the strongest guard there is against a confident invention.

Mark up the parts. Wrap each element in tags, `<context>`, `<task>`, `<text>`, so the model can tell your instruction apart from the document you pasted.

Let it think first. Ask for the reasoning before the answer. Models make fewer errors on anything with more than one step when they work through it in the open.

Give it a role. "You are a loss and damage researcher" produces different, and usually better, text than the same request with nobody speaking.

Then keep going. Refining a prompt is how the quality of the exchange improves, and the second attempt is almost always better than the first.

One habit is worth building from the start. Ask for sources, and ask the model to mark the parts it is unsure about. An answer you cannot check is an answer you cannot use in a plenary.

THE ANATOMY OF AN INTERVENTION

For the task negotiators do most, five parts make the difference: who is speaking and where, the one issue at hand, the evidence behind it, the framework it ties to, and what should be done.

---

[1]As a youth delegate from Malaysia at COP29, draft a two-minute intervention [2]addressing the loss and damage affecting children in Malaysia. Highlight [3]concerns such as rising sea levels threatening coastal communities, educational disruptions from climate impacts, and mental health challenges for children. [4]Offer specific recommendations for prioritising funding to protect children's rights and enhance their participation in climate discussions [5]within the Paris Agreement framework.

[1] Role and context [2] Focus topic [3] Evidence [4] Call to action [5] Technical content

---

Drop any one of the five and you can feel what goes missing. Without the role it drafts for nobody, without the evidence it produces sentiment, and without the last line it ties to no agreement anyone can act on.

5.4 Prompting for different use cases

1. **Data analysis** — FILE UPLOAD

   Given this spreadsheet, plot the average emissions increase over the last years, state your method, and list the assumptions you made about missing data.

2. **Translation** — ANY MODEL

   Translate this paragraph into Spanish, keeping the formal register used in decision text.

3. **Summarization** — WEB SEARCH ON

   Summarise the LDC Group's intervention at the Global Stocktake Technical Dialogue on 10 June 2023, and quote the passages you are drawing on.

4. **Information gathering to develop a position statement or prepare for a speech** — ANY MODEL

   You are the speech writer for a UN climate change negotiator. You are writing about the UK's capacity building efforts. I am the negotiator. Which five questions would you ask me to extract the information you need to draft a statement?

5. **Anticipate counter arguments and prepare rebuttals** — ANY MODEL

   Loss and damage is a key issue in the climate negotiations. You are a loss and damage researcher. Outline the key reasons why countries do not give importance to loss and damage. Produce a strong counter-argument against these points demonstrating the need for a loss and damage fund.

6. **Draft an opening statement or keynote speech** — ANY MODEL

   You are a speechwriter drafting a statement using the questions above. Produce a 300 word statement outlining the UK's climate capacity building efforts. The statement should mirror the writing style used by the UK Government.

7. **Decode the jargon** — ANY MODEL

   Explain what "means of implementation" means in this paragraph, in plain language, and give one example of how the phrase has been used in a previous decision.

8. **Find the document** — RETRIEVAL

   Which decision text carries the mandate for this work programme? Give me the document symbol and quote the paragraph it sits in.

9. **Translate for your own audience** — ANY MODEL

   Translate this intervention into Bahasa Malaysia for a community audience, keeping the meaning and dropping the UNFCCC jargon.

— CHAPTER 6

Agents

Everything up to this point works the same way. You type a request, you read an answer, you decide what to do with it. Four things have changed, and together they add a second way of working.

Context got long. A model can now hold an entire negotiating text, its previous revisions, and your delegation's briefing notes at the same time.

Retrieval became normal. A system can search a set of documents or the web before it answers, and quote what it found. That is what makes an answer checkable.

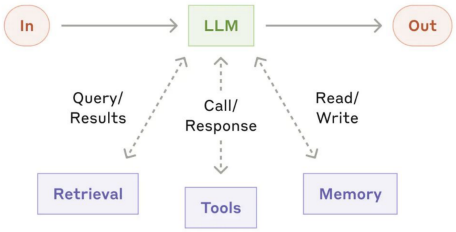
Models got tools. A model can be wired to a search index, a spreadsheet, a calendar, or a folder of files, so it can act on them rather than only describe them.

Systems got loops. Instead of answering once, a system can plan, take a step, look at the result, and try again until the task is done.

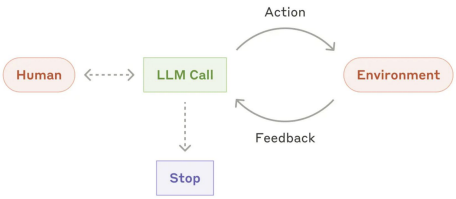
A model wired to tools, memory, and a repeat-until-done loop, running under permissions someone granted it, is what people mean by an agent. It has no judgment and no accountability. It has a task and a set of permissions.

■ RED LINE The line

An agent may prepare a position. It may not send, submit, or publish anything for a delegation without a named person authorising that act.



An augmented model. The same model, given something to search, tools to call, and somewhere to keep what it learned.



The loop. The model acts, reads what came back, and goes again until it stops. A person sets the task and the permissions, and is the only thing in the diagram that can be held responsible.

6.1 Three ways of working

Chat. You ask, it answers, you check. Best for drafting, translating, explaining, and rehearsal.

Retrieval. The system searches a defined set of documents first and answers with citations. Use it for anything factual about the negotiation: text history, party positions, previous decisions.[1]

Agentic. The system carries out a task in several steps using tools. Use it for repetitive work with a checkable result, such as tracking changes across revisions or assembling a daily brief.

[1] If a system tells you what a paragraph of decision text says and cannot show you the paragraph, treat it as a rumour.

6.2 Use cases

1. Text tracking FILES ATTACHED

   Compare revision 3 of this draft decision with revision 2. List every substantive change, quote both versions, and flag the changes that affect my delega-

tion's red lines.

2. Position mapping — RETRIEVAL

   From these submissions, build a table of party positions on Article 6.4 with a quote and a reference for each row. Mark any position you inferred.

3. History retrieval — RETRIEVAL

   Trace how the phrase "common but differentiated responsibilities" has been qualified in decision text since 2015, quoting each decision.

4. Overnight reading — FILES ATTACHED

   Here are the four texts published today. Tell me what moved, what stalled, and which three pages I must read before tomorrow.

5. Daily digest — SCHEDULED AGENT

   Every evening, collect the day's published texts from the tracks I follow, summarise what changed, and send me the list.

6. Bracket analysis — FILES ATTACHED

   For each bracketed option in this paragraph, explain who benefits, who resists, and what a fallback formulation could look like.

7. Rehearsal — CHAT

   Play a negotiator who opposes this text. Push back on my intervention, and do not be agreeable.

8. Back-translation — CHAT

   Translate my formulation into French, then translate it back into English independently, and tell me what drifted.

9. Handover — FILES ATTACHED

   From my notes across this session, write the handover document for whoever takes this file next cycle.

10. Data work — FILE ACCESS

    From this emissions spreadsheet, plot the trend for these countries since 2010, state your method, and list every assumption about missing data.

6.3 Working with an agent safely

An agent acting on a misreading produces consequences rather than a bad paragraph.

Check what you will repeat. Open one cited document and confirm the quote is there before you use it in the room.

Grant the smallest permission that works. A system that reads your files does not also need to send mail.

Treat every external document as untrusted. A submitted file can carry instructions aimed at the agent.

Decide who approves an AI-assisted text. With several tools in the chain, nobody else will.

— CHAPTER 7

Agents for emails

Email remains a primary mode of professional communication, and three things make it hard.

Volume management. Inboxes overflow, and prioritising them takes time you do not have.

Effective drafting. Clear, concise, actionable email takes skill.

Multilingual barriers. Global communication means working in languages that are not your own.

An assistant connected to your mailbox now reads the thread itself, so grant it read access and let it draft. Sending stays with you.

1. Meeting follow up

   Summarise the key points and commitments made during the recent negotiation meeting, attribute each to who made it, and draft a polite follow-up email to the attendees encouraging collaboration on the agreed actions.

2. Scheduling

   Generate a diplomatic and flexible email to propose multiple dates and times for the next round of meetings, expressed in the time zones of all participants.

3. Daily briefings and readouts

   Create a concise daily briefing email that highlights the progress made in the latest sessions, including new proposals, agreed points, and areas of contention requiring further discussion.

4. Smart prioritisation

   Develop an email to the negotiation team that outlines a strategy for prioritising agenda items based on their urgency, their potential impact on climate goals, and the feasibility of reaching agreement.

5. Triage

   Group my unread messages into: needs a decision from me, needs a reply, needs forwarding, needs nothing. Draft replies for the third group only and leave them unsent.

6. Register check

   Rewrite this email in formal diplomatic register without changing any substantive commitment, then list what you changed so I can check.

These systems default to a warm, agreeable voice that can concede things your original text did not. Ask what was changed, and read the list.

— CHAPTER 8

Agents for logistics

Organising schedules, creating checklists, and booking travel can become overwhelming during a session. Three problems come up repeatedly.

Time management. Prioritising tasks and allotting time to them is not straightforward.

Overlooking tasks. Essential items fall off the checklist.

Booking flights and hotels. Finding options that match your schedule and your budget takes hours.

This is the lowest risk work in the guide and the best place to start with an agent. The tasks repeat, the results are easy to check, and almost nothing here is confidential.

1. Bookings

   Identify the most cost-effective and convenient flights and hotels for a trip from [departure city] to [destination city] on [dates], with a preference for morning flights and accommodation close to the venue. Show me the trade-offs rather than one recommendation.

2. Creating checklists

   Generate a comprehensive checklist for a two-week international session that includes necessary documents, accreditation, clothing for variable weather, essential work equipment, and offline copies of key texts.

3. Plan day

   Look at my calendar and the session timetable, then propose a schedule that balances meetings, individual work blocks, exercise, and personal time, given these deadlines and commitments.

4. Progress tracking

   Create a template for tracking progress on the workstreams I follow, including milestones, deadlines, and a system for noting updates and next steps.

Logistics is also where an agent gets real permissions first, over your calendar, your mail, and your files. That is reasonable, and it is where the habit of granting a little more each time begins. The system that books your flights does not need to draft your positions.

— CHAPTER 9

Agents for other activities

Climate negotiations are not only figures and facts. They are interwoven with human emotions, cultural differences, and mental wellbeing.

Mental health strains. The realities of climate change and the pressure of negotiating take a toll on delegates.

Cultural misunderstandings. Misreadings that come from cultural difference lead to rifts and to slower dialogue.

This is the one part of the guide where a plain chat window beats an agent. These conversations want a model you talk to, not a system with permissions.

1. Cultural brief

   I am in a UN climate change negotiation with someone from Kenya. What should I be culturally aware of to advance my negotiation?

2. Emotion recognition

   I am feeling down right now, can you tell me something uplifting?

3. Debrief

   Here is what happened in that session and how I handled it. What would a more experienced negotiator have done differently?

Two cautions. A cultural brief produces a generalisation about a country, and you will meet a person, so treat what you get as orientation rather than prediction. And a model is a reasonable place to put a thought at one in the morning, but it is not a clinician, not a colleague, and not confidential. If the pressure is more than tiredness, speak to your head of delegation, your programme's support contacts, or a professional.

— CHAPTER 10

Sovereignty, and who owns what

These tools can equalise the multilateral space. They remove the language barrier that decides who speaks with confidence, they compress documents nobody has time to read, and they give a two-person delegation some of the reach of a fifty-person one. That is worth having. It also raises questions of integrity worth asking before you type.

Who owns the data. Text you paste into a hosted assistant leaves your delegation. Depending on the account and its settings, it can be retained, reviewed by staff, or used to train the next model. National positions, instructions from a capital, and anything told to you in confidence do not belong there.

Who owns the decision. A model produces text that reads like a position. If it drafts and nobody checks, the reasoning behind your party's position starts to live outside your delegation. Someone with a mandate has to own every text that leaves.

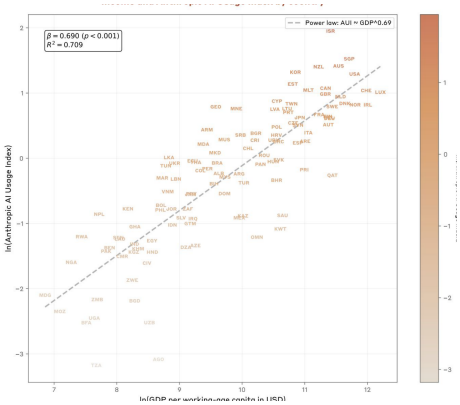
What you cannot see. OpenAI and Anthropic do not give back the reasoning traces. You get an answer, sometimes a summary of how it was reached, while the actual steps stay with the provider. You cannot audit the path, reproduce it, or show a colleague why the system said what it said.

Open weights and closed weights. A model's weights are the numbers it learned in training. ChatGPT, Claude, and Gemini keep theirs on company servers, and they are the strongest and easiest to use. Llama, Mistral, Qwen, Gemma, and DeepSeek publish theirs, so you can run the model on your own machine or a regional server. Open models handle translation, summarising, and drafting well, and trail on the hardest multi-step problems. Weights you hold cannot be repriced or withdrawn in the middle of a session.

Sovereignty-aligned AI means the delegation, region, or community sets the terms: which model is used, what data goes into it, who may inspect what happened, and the right to refuse. In the COP process the parties with the least capacity are usually the parties with the most at stake, which makes this practical rather than philosophical.

A working rule. Use the strongest hosted model for public material, background reading, drafting, and rehearsal. Use open weights, or a deployment governed by a contract your institution has read, for anything touching a position that is not yet public.

This guidebook exists to keep that trade-off visible. The chapters before it show what these systems do well. This one is about what you give up in exchange.



Claude use against income, by country. Adoption tracks GDP per working-age person closely, which is another way of saying these tools arrive last where they are needed most.

— CHAPTER 11

Deploying your own sovereign AI

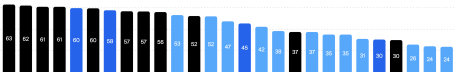
The previous chapter set out the trade-off. This one is about the strongest answer to it, which is running the model yourself.

Self-hosting means nothing leaves the building. There is no retention policy to trust, no terms of service that can change, and no account to be suspended in the middle of a session. It also builds the skill inside your institution rather than renting it.

Until recently this meant accepting a much weaker model. That is no longer the case.

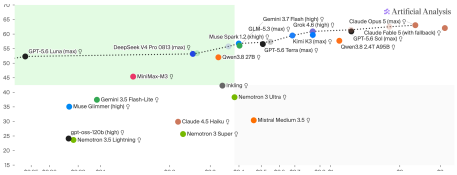
Kimi K3 and GLM-5.3 both score 60 on the Artificial Analysis Intelligence Index, three points behind Claude Opus 5 at 63, and Qwen3.8 follows at 58. Seventeen of the 27 models on the index have open weights, though the top of the table is still mostly proprietary.[1] A year ago the leading open model was thirteen points behind the leading proprietary one, so this gap has closed quickly and may close further.

Price is the second half of the argument. DeepSeek V4 Pro 0813 scores 53 at around a quarter of a dollar per task, inside that quadrant. Claude Opus 5 scores ten points higher at roughly ten times the price. That ratio is the practical reason a delegation can afford to run its own.



All 27 models on the Artificial Analysis Intelligence Index, 20 August 2026, in rank order. Black is proprietary, blue is open weights. The leading open model, Kimi K3, sits fifth at 60, three points off the top.

[1] The chart counts GLM-5.3 as proprietary. Z.ai publishes its weights under an MIT licence, so this book counts it as open, which is why the figure shows sixteen blue bars and the text says seventeen.



The same index against cost per task. The shaded area is the most attractive quadrant, high score and low cost, and the open models hold most of it.

11.1 What to run

MODEL	LAB	WEIGHTS	INDEX
Kimi K3 (max)	Moonshot	Open, use restricted	60
GLM-5.3 (max)	Z.ai	Open, MIT	60
Qwen3.8 2.4T A95B	Alibaba	Open, use restricted	58
DeepSeek V4 Pro 0813	DeepSeek	Open	53
Qwen3.8 27B	Alibaba	Open	52
MiniMax-M3	MiniMax	Open, use restricted	45
Nemotron 3 Super	NVIDIA	Open	26
gpt-oss-120b (high)	OpenAI	Open	24

State of play on 20 August 2026. Read the licence before you commit. Several of the strongest open models restrict commercial use, which for a delegation is usually fine and for a vendor building on top of one is not.

11.2 What it costs

An NVIDIA DGX Spark is a desktop machine with 128GB of unified memory, currently 4,699 dollars. Two of them connect directly to each other with no switch, giving 256GB for around 9,400 dollars. That pair runs a sparse model of the DeepSeek class at 55 to 60 tokens per second with a one million token context, which is enough to serve a working delegation.

A single unit runs a 27B to 35B model fast enough for sixty concurrent users, which covers translation, summarising, and drafting for a whole office. Three or four units at 384 to 512GB hold the largest open models without pruning them.

Set against a subscription for every delegate, the hardware pays for itself inside a cycle or two, and it keeps working after the funding ends.

11.3 How to set it up

Serve the model with vLLM or SGLang. Both present the same API as OpenAI, so any tool that talks to ChatGPT will talk to your own server after a change of address and key. For a first trial on a laptop, Ollama or llama.cpp are simpler. NVIDIA publishes setup playbooks for the Spark, and the community recipes for connecting two, three, and four units are public and reproducible.

If running your own is not realistic before the next session, the fallback is a hosted service with a zero data retention agreement in writing. Treat that as a stopgap rather than a destination.

11.4 One caveat worth knowing

The index above is an average of nine evaluations, and the open models do not trail evenly across them. The gap is widest on the hardest reasoning tasks and on hallucination, where the proprietary models are still clearly ahead. For a negotiator that second one decides everything. Point the model at your own documents, ask for citations, and open them.

NOTE **We are happy to help**

GainForest and the Youth Negotiators Academy advise delegations on this at no cost: sizing the hardware, choosing a model, getting it serving, and training the people who will use it. Write to team@gainforest.net.

— CHAPTER 12

The environmental footprint

A climate negotiator using a tool that burns energy and water should be able to say how much. The honest answer has two halves, and they point in different directions.

12.1 Your own use is small

Google measured a median text prompt to Gemini at 0.24 Wh of energy, 0.03 grams of CO2e, and 0.26 millilitres of water.[1] That is a few seconds of a laptop. A working day of heavy use is still less than a short car journey.

Polly reports this for every account, and the calculation is published rather than asserted.[2] It works from the model's active parameters:

- ENERGY PER TOKEN
  2 FLOP per active parameter → divide by chip throughput → multiply by chip power

  WATER AND CARBON
  1.1 litres per kWh for cooling → 480 grams of CO2e per kWh on an average grid

Reading your prompt costs about a quarter of what writing an answer costs, and re-reading a cached prompt costs a tenth, so the three are counted separately. On the sparse model Polly runs, which activates 13 billion parameters per token, a typical exchange comes to roughly 0.13 Wh and 0.14 millilitres of water. That lands just under Google's median.

Treat any of these numbers as good to a factor of two. They leave out the training of the model, the water used to generate the electricity, your own device, and the carbon embodied in the hardware.

12.2 The industry's use is not

The interesting figures are at the scale of the sector rather than the person. Data centre electricity demand is rising faster than grids are decarbonising, new capacity is being built in water-stressed regions, and the companies reporting the per-prompt numbers are the same ones whose total emissions have gone up since they started building these systems.

Both halves are true at once. Your own use is a rounding error, and the aggregate is a serious and growing claim on energy, water, and land. Saying the first without the second is the kind of accounting a negotiator would not accept from a party.

[1] Google, *Measuring the environmental impact of delivering AI at Google scale*, arXiv:2508.15734, August 2025. The figure includes idle capacity and data-centre overhead, which most published estimates leave out.

[2] The method and every constant are in `footprint.ts` in pi-village-core. A number a negotiator cannot interrogate is worse than no number at all.

12.3 What to do about it

Choose a smaller model when a smaller one will do. Translation and summarising do not need a frontier model, and a sparse model that activates a few billion parameters costs a fraction of a dense one.

Ask providers for their numbers, including the water. Ask which grid the request is served from. These are reasonable procurement questions and they are rarely asked.

Report your own figure when you use these tools in your work. A delegation that publishes the footprint of its own AI use has standing to ask others to do the same.

— CHAPTER 13

Conclusion and recommendations

13.1 Strengths and weaknesses

Strengths. High efficiency in processing and summarising the large volumes of text that climate policy runs on. The ability to generate reports, draft policy, and simulate negotiations in natural language, which saves time and resources. The ability to assist many people at once, from policymakers to researchers and activists.

Weaknesses. Potential bias in the models, which skews how information is processed unless it is monitored and corrected. Dependence on the quality and breadth of the training data, which limits how well a model handles nuanced climate issues. Difficulty interpreting complex legal and technical language accurately without human oversight.

Opportunities. Personalised communication strategies that engage different stakeholders in climate action. Integration with other tools for environmental data analysis, which supports better informed policy. Wider education and dissemination of climate policy information, making it accessible to non-experts.

Threats. Misinformation, where a model generates incorrect or misleading content from unreliable sources. Overreliance, which holds back the development of local expertise in climate policy. Ethical concerns about transparency and accountability in automated decision-making.

Agents add two more threats worth naming. Delegations with resources will run well-governed systems while smaller delegations use free tools with unclear terms, which widens the gap under the appearance of equal access. And with several people and several tools in a chain, responsibility for a text becomes hard to trace unless someone decides in advance where it sits.

13.2 Recommendations

Three things are worth building: tools for translating UNFCCC technical language, platforms for analysing historical negotiation data, and environments where young negotiators have the technology they need to contribute. Alongside those, six more.

Digital literacy. Training aimed specifically at navigating international policy databases and platforms.

Real-time support. Tools that provide clarification during negotiations rather than after them.

Virtual negotiation training. Simulated environments for practising negotiation scenarios.

Financial support platforms. Crowdfunding or micro-grant platforms addressing the financial barriers to attending negotiations, raised by negotiators from Liberia and Paraguay.

Cross-cultural communication training. Programmes addressing the misunderstandings identified by negotiators from Peru and Nigeria.

Mentorship programmes. Networks linking young negotiators with experienced ones, so that knowledge transfers between cycles.

For a negotiator, four habits cover most of it. Use these systems for preparation, language, and memory. Verify anything you will repeat aloud. Never let a system speak for your party. Keep doing enough of the research yourself that you could still do it without help.

For a delegation, write down which tools are permitted, what may be pasted into them, and who approves an AI-assisted text. One page is enough, and it is cheaper than an incident.

■

By harnessing the potential of AI, the youth of today are not just inheriting the future, but actively sculpting it; let the next chapter of our planet's story be written with the bold strokes of innovation and youthful leadership.

■

Questions and comments are welcome at team@gainforest.net.

— CHAPTER 14

Appendix: interview summary

Consolidated summary of interviews with climate youth negotiators, conducted to understand the challenges they face and how language models can assist their work.

Interviewees. Six negotiators, from Liberia, Paraguay, Peru, Nigeria, Lebanon, and Indonesia. Experience varied, with some new to the process.

14.1 Challenges and pain points

In the role. Language barriers where English is not a first language. The technical jargon of the UNFCCC (Liberia, Paraguay). Gaps in technical knowledge and in how information is shared. Lack of awareness of historical negotiations (Lebanon). Difficulty conversing with senior negotiators on complex topics (Peru, Nigeria). Difficulty adapting quickly to negotiation dynamics.

In specific situations. First COP experiences were daunting because of a lack of preparation (Paraguay, Peru). Public speaking in a high-stakes environment makes complex topics hard to articulate.

In gathering information. The UNFCCC website is the main source and is complex to navigate. There is insufficient time to process crucial information (Paraguay). Dissemination relies on internal communications and personal networks.

In communicating. Expressing complex ideas quickly and accurately in English is difficult. Levels of English proficiency vary and create barriers (Indonesia).

14.2 Tools

In use. Grammarly for writing assistance (Paraguay). Google Drive for document collaboration (Paraguay). No use of translation tools, on the grounds that the language is difficult to translate (Paraguay).

Wished for. More sophisticated language and grammar tools for UNFCCC texts (Paraguay, Liberia). Platforms for efficient document scanning and key information extraction (Lebanon, Nigeria). Tailored learning resources (Liberia).

14.3 Information processing and decision making

Staying updated. Through colleagues and shared resources, and through newsletters and online platforms (Liberia, Lebanon).

Approaching complex information. Teamwork and leveraging the expertise of others (Peru), and thorough research on unfamiliar topics.

Collaboration. Supportive communities among negotiators are actively fostered (Peru, Nigeria). Initiatives to improve language skills are underway (Indonesia).

14.4 Ideas and solutions

Desired improvements. Simplification of official documents, for inclusivity (Lebanon, Nigeria).

Repetitive tasks worth automating. Retrieval of historical negotiation detail (Lebanon).

Future vision. AI and digital tools are seen as potential aids in negotiations.

14.5 Opportunities for language models

Benefits. Translation and summarisation, which address language barriers and condense information.

Automation. Synthesis from extensive documents.

Reservations. Dependency on the technology could diminish critical research skills (Liberia, Paraguay).

— CHAPTER 15

AI glossary

Terms you will hear in the room, defined plainly.

15.1 Core terms

Artificial intelligence (AI)	Systems designed to perform tasks associated with human intelligence, such as prediction, classification, generation, or planning.
Machine learning	A method in which a system learns patterns from examples instead of receiving every rule explicitly.
Model	A learned mathematical system that maps inputs to outputs.
Large language model (LLM)	A model trained on large collections of text to predict and generate language.
Generative AI	AI that produces new text, images, audio, code, or other outputs.
Agentic system / AI agent	A model connected to tools, memory, instructions, and action loops so it can complete multiple steps.
Automation	Using a system to perform a task with limited direct human intervention.
Augmentation	Using AI to support rather than replace human work.
Human-in-the-loop	A person reviews, approves, or intervenes in an AI-assisted process.
Human-on-the-loop	A person supervises a system without reviewing every action.
Human-out-of-the-loop	A system acts without a person reviewing each decision or action.

15.2 Data and knowledge

Training data	Examples used to adjust a model's parameters during training.
Dataset	An organised collection of data used for training, testing, or analysis.
Data provenance	Information about where data came from and how it was collected or changed.
Data lineage	The history of how data moves through systems and transformations.
Structured data	Data organised into defined fields, rows, categories, or schemas.
Unstructured data	Material such as text, images, audio, or video without a fixed tabular structure.
Synthetic data	Artificially generated data designed to resemble real data.
Ground truth	A reference label or outcome treated as correct for evaluation.
Knowledge base	A curated collection of information that a system can retrieve.
Parametric knowledge	Information encoded in a model's learned parameters.
Factual knowledge	Information about facts or relationships that a model may have learned from its training data.
Retrieval-augmented generation (RAG)	A system retrieves documents or records before generating an answer.
Local knowledge	Knowledge grounded in the experience of a particular place or community.

Indigenous knowledge	Knowledge systems developed and maintained by Indigenous peoples and communities.
Data sovereignty	The principle that data is governed according to the laws, rights, and authority of the people or place it concerns.
Data minimisation	Collecting and retaining only the data necessary for a stated purpose.
Sensitive data	Data whose disclosure or misuse could cause harm.

15.3 Model inputs and outputs

Input / prompt	The instruction or material given to a model.
Output	The text, prediction, classification, image, or action produced by a system.
Inference	Running a trained model to produce an output.
Context window	The amount of text or other material a model can consider at one time.
Token	A unit of text or other data processed by a language model.
Embedding	A numerical representation of text, images, or other data used to compare relationships.
Classification	Assigning an item to one or more categories.
Prediction	Estimating a likely outcome from available data.
Forecast	A prediction about a future state or event.
Recommendation	An output that suggests a course of action.
Optimisation	Searching for the best result according to a defined objective and constraints.
Objective function	The formal quantity a system tries to maximise or minimise.
Proxy	An indirect measure used in place of a harder-to-measure goal.
Benchmark	A test used to compare model performance.
Evaluation	Measuring a model or system against defined criteria.
Uncertainty	The degree to which an output may be wrong or incomplete.
Confidence score	A numerical estimate associated with a prediction or classification.
Calibration	The relationship between predicted confidence and actual accuracy.
Explainability	The ability to provide an understandable account of how an output was produced.
Interpretability	The extent to which a system's internal behaviour can be understood.
Black box	A system whose internal operation is difficult to inspect.
Hallucination	A fluent but unsupported, false, or fabricated model output.
Bias	Systematic error or unequal performance associated with data, design, or use.
Fairness	Approaches for evaluating and addressing unequal treatment or outcomes.
Robustness	The ability to perform reliably under changed or imperfect conditions.
Adversarial attack	An input deliberately designed to cause a system to fail or behave unexpectedly.

15.4 Models and training

Neural network	A computational system made of connected layers that transform inputs into outputs.

Perceptron	An early mathematical model of a single brain cell, and a foundation for later neural-network research.
Transformer	A neural-network architecture that uses attention to process relationships among tokens.
Weights	The numbers a model learned during training. In practice, the weights are the model.
Open weights	Weights published so that anyone can run, inspect, or host the model themselves.
Closed weights	Weights kept on the provider's servers and reachable only through their service.
Pretraining	Initial training on a large body of data, usually by predicting missing or subsequent content.
Fine-tuning	Further training on a narrower dataset or task.
Post-training	Processes applied after pretraining to improve usefulness, behaviour, or alignment.
Reinforcement learning (RL)	Training through feedback that rewards some actions and penalises others.
Reinforcement learning from human feedback (RLHF)	Reinforcement learning using human preferences or evaluations.
Reasoning model	A model or system optimised to spend additional computation on multi-step problems.
Chain of thought	Intermediate reasoning text generated during problem solving.
Scaling laws	Observed relationships between model performance, model size, training data, and computing power.
Mixture of experts (MoE)	An architecture containing multiple specialist subnetworks, only some of which are used for each input.

15.5 Governance and responsibility

Alignment	Designing a system so its behaviour reflects specified goals or values.
AI safety	Research and practices aimed at reducing harmful or uncontrolled system behaviour.
AI governance	The rules, institutions, processes, and practices that shape AI development and use.
Responsible AI	Developing and using AI with attention to social, ethical, legal, and environmental impacts.
Value-sensitive design	Designing technology with explicit attention to human values and affected interests.
Sovereignty-aligned AI	AI designed and governed to preserve the agency of a defined community or place.
Fine-grained permissions	Restrictions on which tools, files, data, or actions a system can access.
Sandbox	An isolated environment in which a system can operate with limited access.
Human approval gate	A required human authorisation before an action proceeds.
Audit log	A record of system inputs, outputs, actions, and changes.
Model card	Documentation describing a model's intended use, limitations, and evaluation.
System card	Documentation covering a broader deployed system, including risks and mitigations.
Model drift	A change in performance as data, behaviour, or conditions change over time.
Versioning	Keeping track of changes to models, data, prompts, and software.
Rollback	Reverting a system or deployment to an earlier version.
Self-hosting	Running a model on hardware you or your institution control.
Zero data retention	A contractual promise that a provider stores nothing you send it.

— CHAPTER 16

References

1. Rosenblatt's perceptron and the 1958 demonstration, Cornell Chronicle
2. Krizhevsky, Sutskever, Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, NeurIPS 2012
3. Vaswani et al., *Attention Is All You Need*, Google Research 2017
4. Radford et al., *Language Models are Unsupervised Multitask Learners*, OpenAI 2019
5. Kaplan et al., *Scaling Laws for Neural Language Models*, OpenAI 2020
6. Hoffmann et al., *Training Compute-Optimal Large Language Models*, NeurIPS 2022
7. OpenAI, *Introducing ChatGPT*, 30 November 2022
8. OpenAI, *Learning to Reason with LLMs*, 12 September 2024
9. DeepSeek, *DeepSeek-R1*, 20 January 2025
10. METR, *Measuring AI Ability to Complete Long Tasks*, 2025
11. Anthropic, *Anthropic Economic Index*, 2025, usage by country and income tier
12. Global Biodiversity Information Facility, occurrence data and its geographic sampling bias, gbif.org and the GBIF data blog
13. Artificial Analysis, *Intelligence Index by Open Weights and Proprietary*, 20 August 2026
14. Artificial Analysis, *Intelligence Index vs. Cost per Intelligence Index Task*, 20 August 2026
15. NVIDIA, DGX Spark user guide and hardware overview
16. Google, *Measuring the environmental impact of delivering AI at Google scale*, arXiv:2508.15734, August 2025
17. Epoch AI, *How much energy does ChatGPT use?*, 2025
18. GainForest, `footprint.ts` in pi-village-core, the calculation behind Polly's per-account figures
19. GainForest, *Generative AI + Climate Guide 2024*, prepared for the Climate Youth Negotiator Programme ahead of COP29