Introduction to Climate Policy and Regulation in Context of AI Applications

Vaibhav Chugh
Pallavi Das

5 August 2026

© Council on Energy, Environment and Water, 2026

Lecture flow

(15 mins) Part 1: Foundational understanding with the Scope and Implications

- why policy exists, what it targets, the equity fight, the recurring tradeoffs

(10 mins) Part 2: Landscape, What actors,developments and frameworks from across different global contexts.

- Major Global pacts, Multilateral Institutions and mechanisms for financing and execution
- Domestic policies in India around Climate

(25 mins) Part 3: Implications for innovation, deployment, and decision-making

- Understand how climate policy and regulation shape innovation, deployment, and decision-making in both theory and practice, particularly in the AI and climate space.

Part 1: Mental Model for Climate Policy



Why policy exists?

To protect global public good
A stable climate is a global public good. No one can be shut out of it, and one country's benefit doesn't reduce anyone else's.



Price negative externality
A polluter's emissions harm others — heat, crop loss, but that cost falls on third parties, not the polluter, so it never enters their decision.

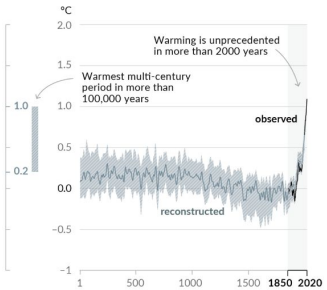


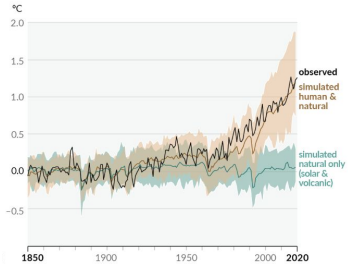
Collective-action / free-rider problem
everyone benefits when others act — so everyone waits for someone else to act.



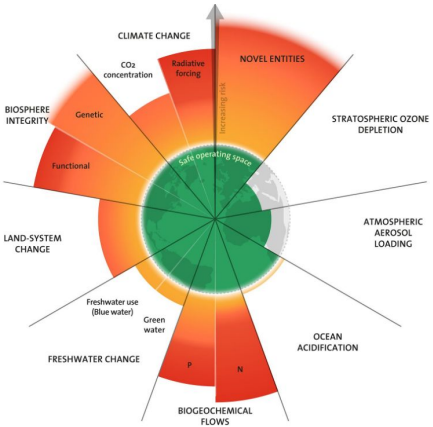
Why do we need climate policy?

Observed Changes

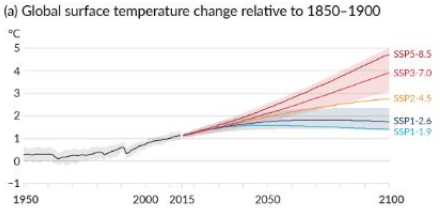


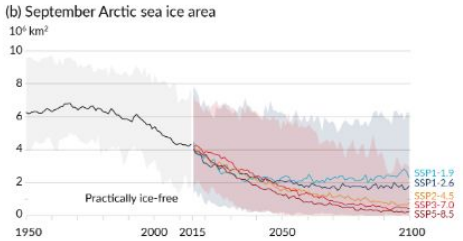


Safe Limits



Projected Changes





https://www.stockholmresilience.org/research/planetary-boundaries.html; https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/

What does policy target for climate change?

Mitigation

Reduce/prevent emissions

Use renewables, efficiency, carbon pricing

Historically the Global North's default frame

Adaptation

Manage unavoidable impacts such as temperature rise, extreme events

Nature-based-solution; parametric insurance; heat action plan

Existential for the exposed South

Means of implementation and policy instruments

Finance

Technology

Capacity Building

Policy Instruments

- **Market/price-based** — carbon taxes, emissions trading systems (ETS/cap-and-trade), removing fossil-fuel subsidies.
- **Regulatory (command-and-control)** — emission standards, efficiency norms, bans/phase-outs (e.g. coal, ICE vehicles), building codes.
- **Fiscal/financial** — subsidies, tax credits, green bonds, blended finance, feed-in tariffs, production-linked incentives.
- **Information/voluntary** — labelling, disclosure (e.g. climate risk reporting), corporate net-zero pledges.

Policy Trade-offs

- **Cost vs. Speed**
  - Carbon tax = fast action, regressive, politically fragile.
  - Standards = slower, durable, embedded in timelines

- **Scope vs. Depth**
  - Paris has all countries, weak enforcement.
  - Local mandates (cities, firms) have teeth, but don't scale

- ***Why it matters****:* Different countries choose different instruments because their starting conditions differ. No single "right" policy

Governance of policies

International
Treaties and norms set direction

National
Governments make binding commitments and design domestic policy

Subnational
States, provinces and cities often move faster and implement on the ground

Non-State Actors
Corporates, investors, civil society shape and implement policy efforts

A finite carbon budget — and who gets to use it

~876 $GtCO_2$

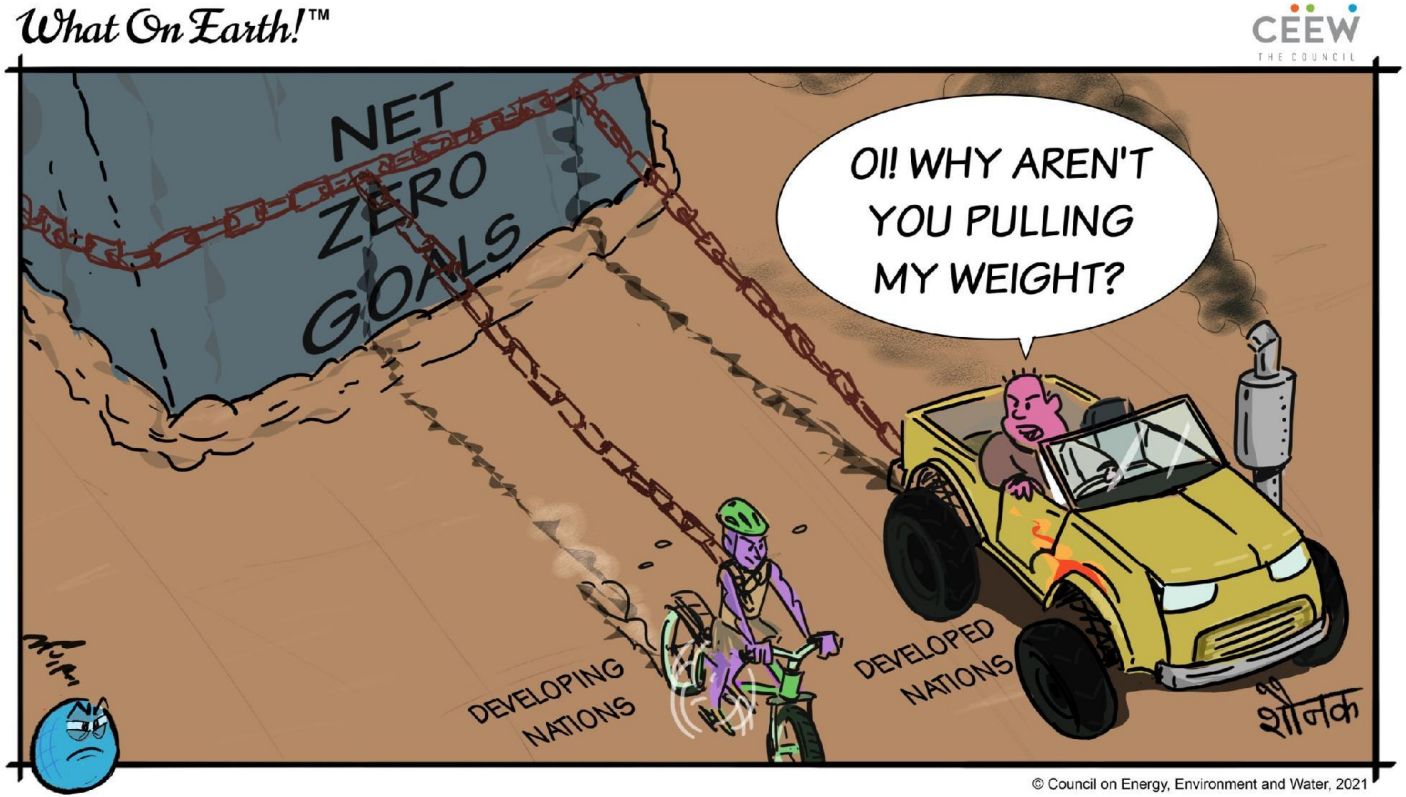
Finite carbon budget to remain below 2° C

Who gets to use this remaining space is THE distributive fight of climate policy

- **CBDR-RC:** Common But Differentiated Responsibilities & Respective Capabilities (UNFCCC, 1992): shared problem, differentiated duties by responsibility + capacity
- **Historical responsibility:** Cumulative emissions matter. The US, EU and developing countries have very high historic emissions
- **Carbon space:** split the budget by current emissions? per-capita? cumulative? Each answer implies radically different obligations

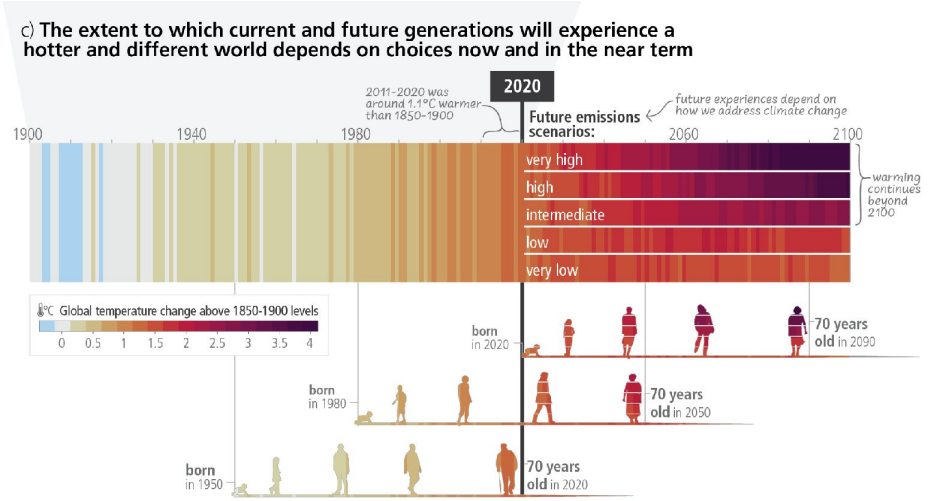
https://www.pik-potsdam.de/en/institute/departments/climate-economics-and-policy/carbon-clock/remaining-carbon-budget Accessed on 24 Jul 2026

Part 2: The Climate Policy Landscape



The Intergovernmental Panel on Climate change – keeping science ahead

- **It assesses, it doesn't research**. The IPCC conducts no new science of its own — it reviews and synthesises thousands of published studies to establish what is known, how confident we are, and where uncertainties lie.
- **Governments and scientists together**. Reports are written by hundreds of volunteer expert authors, then the Summary for Policymakers is approved line-by-line by all member governments.
- **Three working groups.** WG1 (the physical science), WG2 (impacts, adaptation, vulnerability), WG3 (mitigation). Assessment cycles run roughly every 6–7 years; the latest is the Seventh Assessment Report (AR7) underway



https://www.ipcc.ch/report/ar6/syr/downloads/figures/summary-for-policymakers/IPCC_AR6_SYR_SPM_Figure1.png

The UNFCCC exists to break that standoff by getting everyone to commit together, in public, repeatedly

Job of the **UNFCCC** (and Paris underneath it):

- It gets nearly **every country to the table together,** so no one is cutting alone while others coast
- It makes **commitments visible and public** (NDCs), so there is policy certainty for industry and accountability by countries
- It convenes **repeated meetings** (the annual COPs, mid year Bonn) to sustain dialogue, break dead lock in negotiations, bring accountability to commitments



Negotiation Grouping

- **African Group** – One of the UN's five regional negotiating groups, with 54 member states
- **Arab Group** – formally the League of Arab States, a regional organisation formed in 1945
- **BASIC (Brazil, South Africa, India and China)** – a coalition of four major emerging economies
- **European Union (EU)** – the 28 member states of the EU, with negotiations led by DG-Clima
- **Group of 77+China (G77+China)** – a large alliance of 134 developing nations
- **Least Developed Countries (LDCs)** – a group of the world's poorest nations, which evolves as economies change
- **Like-Minded Developing Countries (LMDCs)** – a group of developing countries, representing 3.5 bn people, with a strong focus on ensuring rich countries bear most responsibility for tackling climate change
- **Alliance of Small Island States (AOSIS)** – a group of 44 small islands and low-lying coastal states
- **Umbrella Group** – a cross-continent group of countries, many of which have been considered less-than-enthusiastic about climate change in the recent past, although in many cases these stances are now shifting

Source: https://www.carbonbrief.org/interactive-the-negotiating-alliances-at-the-paris-climate-conference/

International architecture — the regime shift (from coercion to peer pressure)

1992 UNFCCC

The frame. Sets principles — including CBDR-RC — but no binding targets

1997 KYOTO

Top-down: binding targets on Annex-I (rich) countries only.

2015 PARIS

Bottom-up: every country sets its own NDC, reviewed via a ratchet + 5-yearly global stocktake.

How the pact binds a country: India's NDC

- 47% emissions-intensity cut by 2035 compared to 2005
- 60% non-fossil electric capacity by 2035 compared to 2005
- 3.5 to 4.0 billion tonnes of CO_2 equivalent through forest and tree cover by 2035 (from 2005 levels)

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2245209&reg=3&lang=1

National level — different economies, different toolkits

The teaching point: there's no single "climate policy." What a country uses depends on its wealth, its emissions profile, and its politics.

The high-income "regulate + price" model

- **EU** — regulate + price: Green Deal / Fit for 55, EU ETS (cap-and-trade), CBAM (border tax), EU Taxonomy

The wider Global South — same logic, different pressures

- **China** — state-directed scale: national ETS (2021, power sector), dual-carbon goals (peak <2030, neutral <2060), and manufacturing dominance in solar, batteries, and EVs
- **India** — subsidy + mission-mode (deploy first, price later): National Solar Mission, PM-KUSUM (solar pumps), PM Surya Ghar (rooftop, 2024), FAME (EVs), LT-LEDS.
- **Brazil** — the *land-use* model: its mitigation story is dominated by Amazon deforestation, not energy; policy success = enforcement of forest protection. Also runs one of the world's oldest biofuel programmes (ethanol).
- **South Africa** — the *just-transition* model: heavily coal-dependent, and the test case for the **Just Energy Transition Partnership (JET-P)** — a multi-donor deal to fund a managed coal phase-out without collapsing coal-region livelihoods.
- **Indonesia / Vietnam** — also **JET-P** recipients: coal-heavy, fast-growing, negotiating externally financed transitions — showing climate finance and national policy meeting in practice.
- **Kenya / Ethiopia** — the *leapfrog / renewables-rich* model: already near-fully renewable grids (geothermal, hydro), so their frontier is access and resilience, not decarbonising a dirty grid.

Cross-cutting regimes, the frontier that crosses borders

What CBAM Does ?

The EU's carbon border adjustment taxes the embedded carbon in imports — steel, aluminium, cement, fertiliser, electricity, hydrogen.

Definitive (paying) phase began 1 January 2026; €100/tonne penalty for shortfalls. Expansion to ~180 downstream goods proposed for 2028.

Purpose: stop "carbon leakage" — firms relocating to dodge the EU ETS price.

Debate it triggers…

For Indian & other Global South exporters, CBAM is a direct cost at the EU border.

Seen from Delhi:

"green protectionism" — a rich bloc exporting its climate costs onto poorer producers.

Seen from Brussels:

a legitimate anti-leakage measure that levels the playing field.

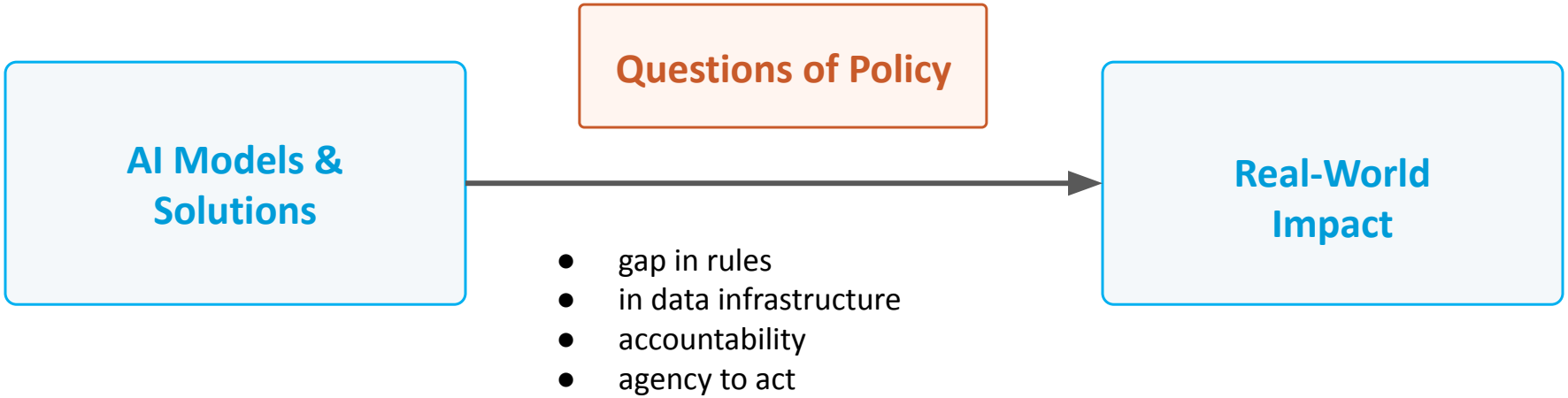
The best single example of a Western instrument reshaping Global South choices — and now WTO-contested.

Part 3: Implications for Innovation, Deployment & Decision-Making

What role policy has to play in Climate and AI?

Policy is Structured Formal, Slow but accountable, Works best when problems are well-understood and stable

AI is Exploratory, Iterative and flexible, ambiguous or evolving challenges



The same AI pitched to cut emissions is itself an unregulated emitter

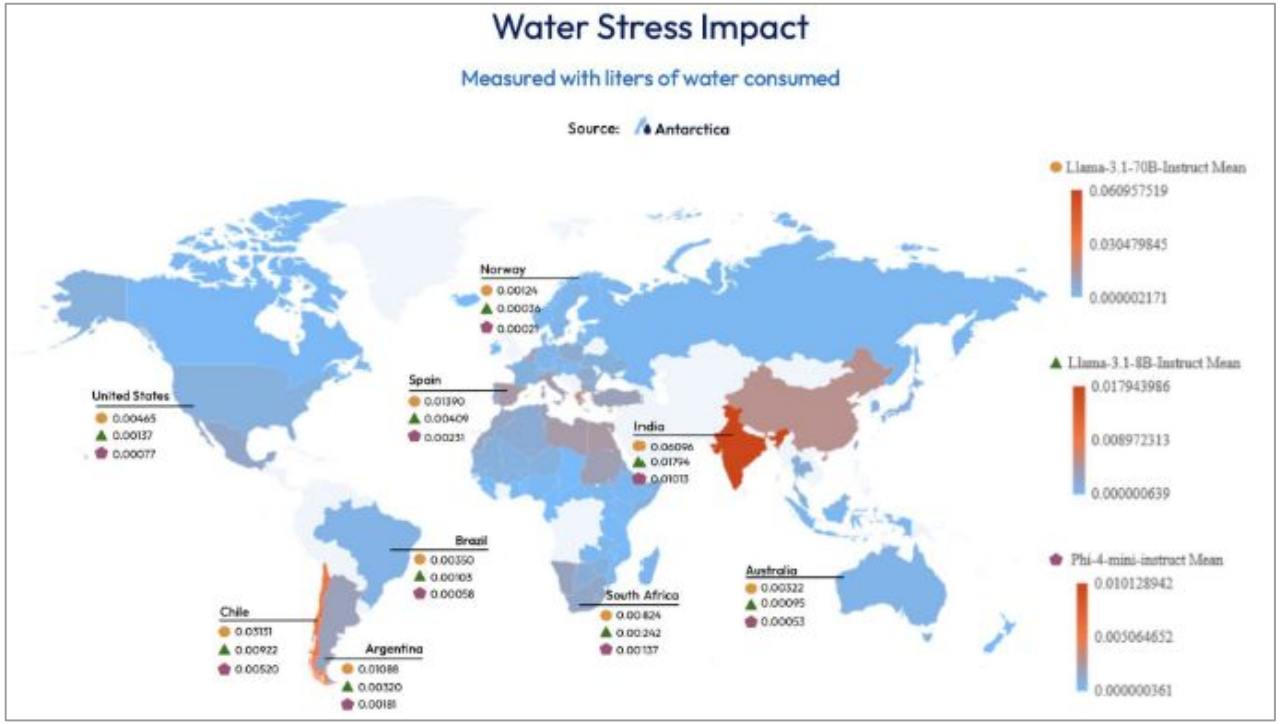
AI in power, transport and food could cut 3.2–5.4 billion tonnes CO_2e per year by 2035 — these reductions would outweigh the emissions from data centres and AI themselves. (Nature*)

Policy shapes AI	**Singapore's data-centre moratorium** - imposed a **moratorium on new data centres in 2019** over energy/water strain - Approval is now via a **competitive allocation** : requires best-in-class efficiency and at least 50% green energy. - ~30% reduction in IT-equipment energy, shifting the **efficiency burden from the building to the compute layer**
AI Shapes policy delivery	**India's CCTS to be launched in Oct'26** - CCTS, India's first compliance carbon market, goes live with first certificate trading **covering ~490 obligated entities,** among the world's largest ETS - Credits are issued against **verified emissions per unit of output** so the entire market lives or dies on MRV quality. - **Digital MRV** (satellite + AI) replaces slow manual audits with continuous, tamper-proof data, with providers reporting up to **70% lower verification costs**

*https://www.nature.com/articles/s44168-025-00252-3

The same AI query, sent to a data centre in India, produces a water stress impact 21 times higher than in Germany.

**Mathieu François , CEO of Antartica (LinkedIn POST)*

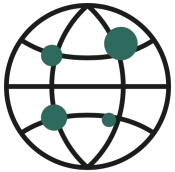


A medium-length query on **LLaMA 8B** consumes **8.96 mL of water in India.** The same query consumes **3.26 mL in Germany.** The volumetric gap is less than 3x.

But when weighted by regional scarcity, the **water stress impact is 17.1 in India and 0.8 in Germany.** The gap becomes 21x.

Scale this to the **70B model** on a long prompt and **India's stress impact reaches 122, compared to 5.7 in Germany.**

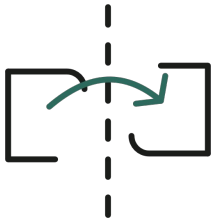
AI's carbon costs — three governance gaps



The geographic distribution of AI's carbon costs



Mandatory carbon accounting for AI training and inference



Cross-border accountability when AI built in one country harms people in another

What differentiates AI in Climate Action

Building AI solutions in Climate Action needs a broader approach

	Conventional	Climate Action
Purpose and Motivation	• Business Outcomes: Profit, market Share, IP	• Solving community-based or systemic issues: AI is part of larger problem
Who defines the problem Stakeholder/Engagement	• Business managers - Operational , Data Teams	• Affected communities, Public agencies, NGOs: Participatory workshops, Questionnaires, Co-Design sessions
Social and Ethical concerns	• More reactive and compliance oriented (Eg: GDPR)	• Built from start, Equity and Inclusion, Privacy protection
Data and Methods	• Structured , clean data, Internal data	• Often incomplete, messy data , Require interpretation, Cleaning, Contextual understanding
Accountability and Success	• KPIs: Sales, Cost reduction, User engagement	• Social benefits, Systemic changes, Accountability to communities involved

Escaping the “PILOT” trap .

DATA / PROBLEM

what gets built

Data needs to be **representative of the population** and **local** conditions

PULSE OXYMETER

Ground-truthed data, **owned by the institutions accountable** for acting

BRAZIL INPE (PRODES/DETER)

Identifying the **correct Problem statement** at the onset

DENGUE OUTBREAK MODEL / MP

DEPLOYMENT

whether it reaches anyone

Integrate AI into **existing workflows** along with mechanisms for action

DERNA FLOOD

Accuracy and benchmarks need to be established in **actual deployment** scenarios not Labs

GOOGLE DR SCREENING - THAILAND

DECISION MAKING

whether it can be relied on

For critical systems, Designing **Human in the loop** system

PERVS AI - BRAZIL

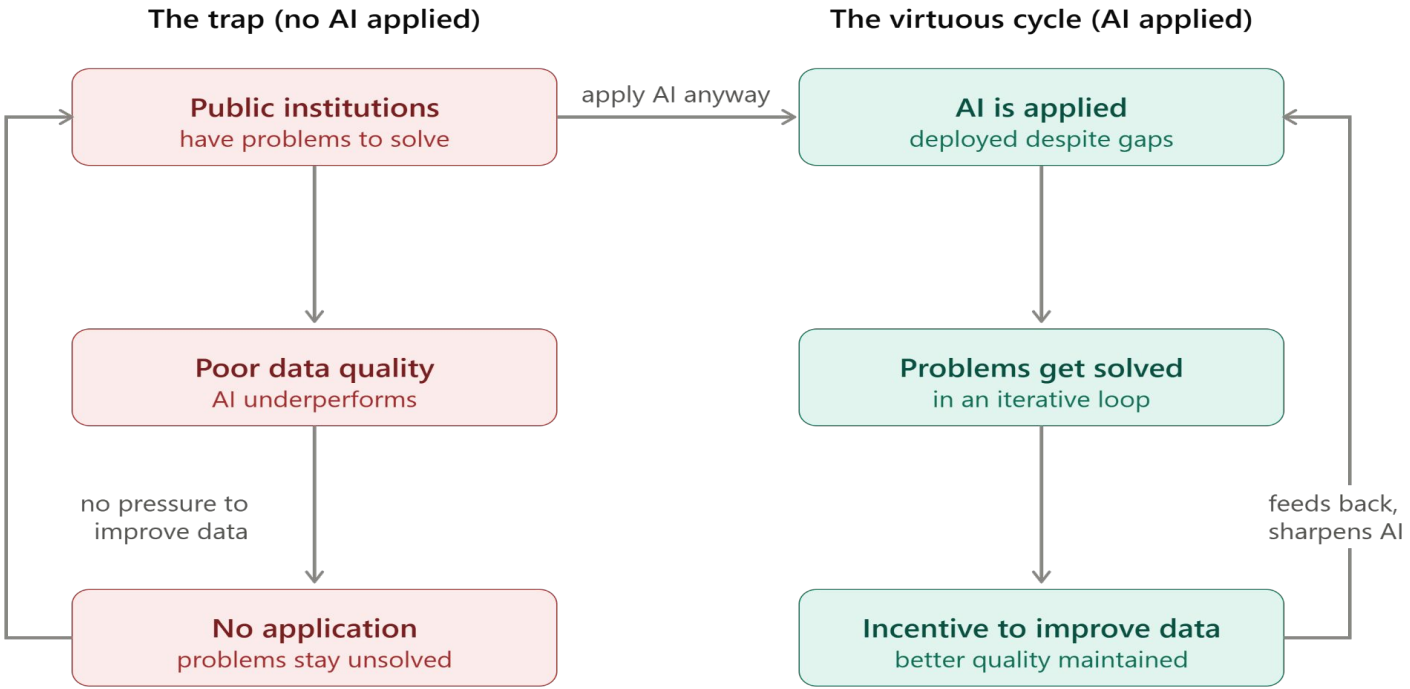
Institutional ownership beyond pilots and for iterative improvements that enable incremental operationalisation

AHMEDABAD HEAT ACTION PLAN

Communicating model uncertainty is as important as outcome

LIGHTNING - ORISSA CASE STUDY

Data quality isn't a precondition you wait for — it's an output of use



Case 1: Representative Data

The device everyone trusted was blind to darker skin

24,504

patients studied

JAMA Internal Medicine, 2022



Context

- During **COVID-19, pulse oximeters** became the frontline tool for monitoring blood oxygen.
- **Trusted** without question

Complication

- Calibrated on predominantly light-skinned populations and had been since 1989
- darker-skinned patients, it systematically overestimated oxygen levels making them appear fine when they were not

Actions

- FDA acknowledges problem, calls for updated standards.
- Devices remain unchanged. too embedded in clinical workflows.

The climate parallel: AI trained on Global North data misreads local ground truth the same way. Representative, sovereign data isn't a nice-to-have — it's the precondition for trust.

Case 2: Institutional ownership of data and outcomes

Brazil's INPE (PRODES / DETER) — sovereign, ground-truthed data owned by the accountable institution

~95%

accuracy of INPE's official PRODES record



Context

- Foreign deforestation projections led the World Bank to freeze loans to Brazil, amid constant disputes between agencies over the real numbers.

Action

- government tasked its own space institute with producing authoritative deforestation figures from actual satellite observations.
- PRODES delivers the official annual rate (~95% accuracy); issues daily alerts that, the enforcement agency, acts on — planning inspections and operations against illegal clearing

Result

- Under renewed enforcement from 2023, Brazil's share of global tropical primary forest loss fell from 43% to 30%, its lowest since 2015

The climate parallel: Risk intelligence is only actionable when the accountable institution owns the ground truth

CEEW THE COUNCIL

Case 4: A forecast isn't an outcome — someone still has to act on it

Derna, Libya · Storm Daniel · September 2023

4,300+

deaths, despite a 3-day warning



Context

- A meteorological warning was issued three days before Storm Daniel, and the forecast proved accurate.
- A dam failure had been flagged in a 2022 study — but no operational system was tracking or acting on that risk

Action

- As the dams collapsed in the early hours, officials had issued conflicting orders and a curfew that kept residents indoors, in the water's path.
- The ageing dams were known to be in disrepair, yet went unmaintained.

Result

- Over 4,300 people died. The forecast was right — the system around it failed: no clear mandate, no evacuation, no last-mile response.

The climate parallel: An accurate AI early-warning system is worthless without the machinery to act — clear mandates, evacuation logistics, communication that reaches people. The forecast is the easy part.

Case 5: Accuracy of AI models in actual deployment scenarios

Diabetic retinopathy screening · Thailand, 2019

>90%

lab accuracy · Beede et al., 2020



Context

- Thailand aimed to **screen 60% of its 4.5 million diabetic patients** — but had only **~200 retinal specialists, a 1:3,000 ratio.**
- A deep-learning tool promised specialist-level **reads in minutes instead of weeks**.

Complication

- In clinics, dim lighting caused the system to **reject 21% of the images** captured.
- **Slow internet** made uploads unreliable, 60 to 90 seconds per image.Nurses, who ran the screening, were never part of the design.

Consequence

- Nurses c**ould screen only ~10 patients per two-hour session; the tool created bottlenecks,** not speed.
- It made their **work harder** — they called it worse than what it replaced.

The climate parallel: A model that shines on benchmarks can fail in the field: Real extreme events and early warnings test climate AI in the real conditions and workflows where it must run.

Case 6: Institutional ownership for long-term impact

Ahmedabad Heat Action Plan · institutionalising an effective measure

46.8°C

Ahmedabad's May 2010 peak



Context

- May 2010's heatwave caused 1,344 excess deaths in Ahmedabad in a single month — a 43% jump over normal.
- Remote sensing shows low-income informal settlements run 5–10°C hotter than wealthier areas in the same city.

Action

- Correlating daily mortality with temperature revealed a trigger threshold (~41°C) where deaths spiked — the basis for a tiered alert system.
- Ahmedabad built South Asia's first Heat Action Plan: early warnings, a Cool Roofs programme, and mandatory rest hours for outdoor workers.

Result

- A follow-up study found the plan averts ~1,190 deaths a year (2,380 across the 2014–2015 summers studied).
- NDMA adopted it as the national template — heat action plans now run across 20+ Indian states.

The climate parallel: Long-term impact comes when an accountable institution owns and funds the tool beyond the pilot. Ahmedabad's plan endured because NDMA made it national policy

CEEW
THE COUNCIL

Case 7: Reliance on AI models that need a human in the loop

PrevisIA (Imazon, with the Pará State Prosecutor's Office)

~70%

accuracy for Human oversight



Context

- **Forest loss in Amazon, enforcement is slow** to react, as conventional monitoring only shows damage already done
- PrevisIA uses **AI on satellite imagery to forecast** where illegal deforestation will strike in advance, mainly by **detecting the roads cut first**, since 95% of clearing occurs within 5.5 km of a road

Action

- **risk alerts** through a public dashboard, but no fines and embargoes
- The State Office reviews the forecasts and builds action protocols: **AI predicts → humans decide → action follows**

Result

- It flagged roughly 5,000 km² of forest at high risk for 2024
- With ~30% false positives, **accuracy alone isn't enough.** Human oversight and institutional ownership are what let AI strengthen governance, not replace it.

The climate parallel: For high-stakes climate decisions, accuracy is necessary but not sufficient. Keep an accountable human institution in the loop — AI should strengthen governance, not stand in for it.

Policy doesn't decide whether AI happens

It decides

- **which problems are worth solving,**
- **whose data exists to solve them with,** and
- **what you must prove before anyone acts on your output.**

Case for Discussion: ‘Lightning’ Strikes

The 'Alarm Fatigue' & False Positives Problem (The Modelling Challenge)

For every 1°C increase in global ambient temperature, projections estimate 12% surge in lightning strikes

India is currently seeing increase in lightning frequency, concentrated in some states

- 96% of all lightning-related fatalities in India occur exclusively in rural areas
- 77% of those killed are daily-wage agricultural farmers working in open paddy fields or cattle grazers out in open pastures
- The peak window for fatal strikes occurs between 2:00 PM and 7:00 PM

- Lightning nowcasting deals with fast-moving, highly volatile convective clouds. Because these atmospheric systems change rapidly, predictive models frequently generate 'false positives'—forecasting a severe localized strike that ultimately dissipates or shifts course.
- If your AI model triggers three consecutive sirens in a village and no lightning occurs, the community will experience severe 'alarm fatigue' and actively ignore the fourth, potentially fatal warning.

As data scientists, how do you tune your model's sensitivity threshold? Is it better to tolerate false alarms to maximize safety, or risk missing a strike to preserve community trust?

Real Use case Example: Case Study of MP Vector Borne Disease Outbreak Predictions

Objective: **a forecasting model that looks weekly cases 1 to 8 weeks ahead for 51 districts of Madhya Pradesh, datasets of cases from 2017 to 2025.**

2 Questions for any District

STAGE 1 — Will there be cases in coming weeks?

Yes / No alert

▼

STAGE 2 — If yes, how large?

Grade the week into a severity band

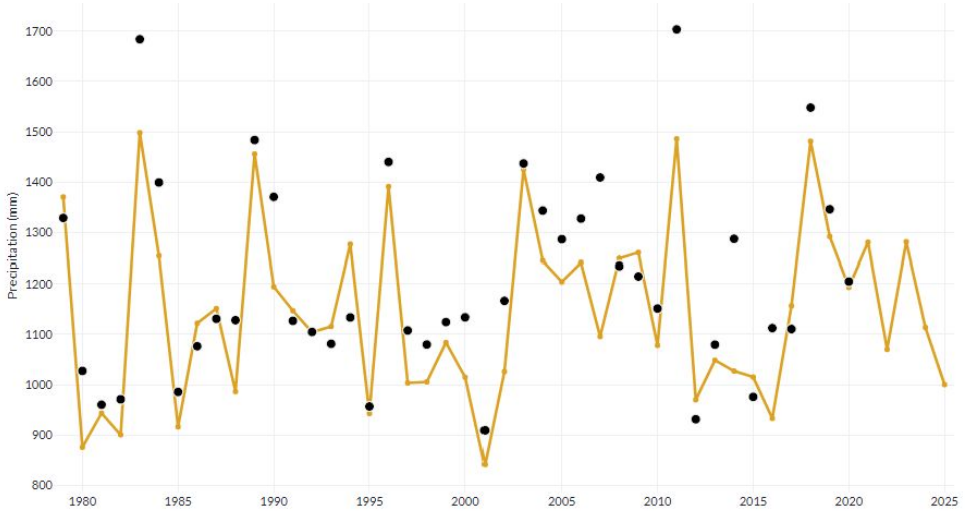
0	1–5	5–20	20+
None	Low	Medium	High

Climate datasets: Something that covers covers every district and week

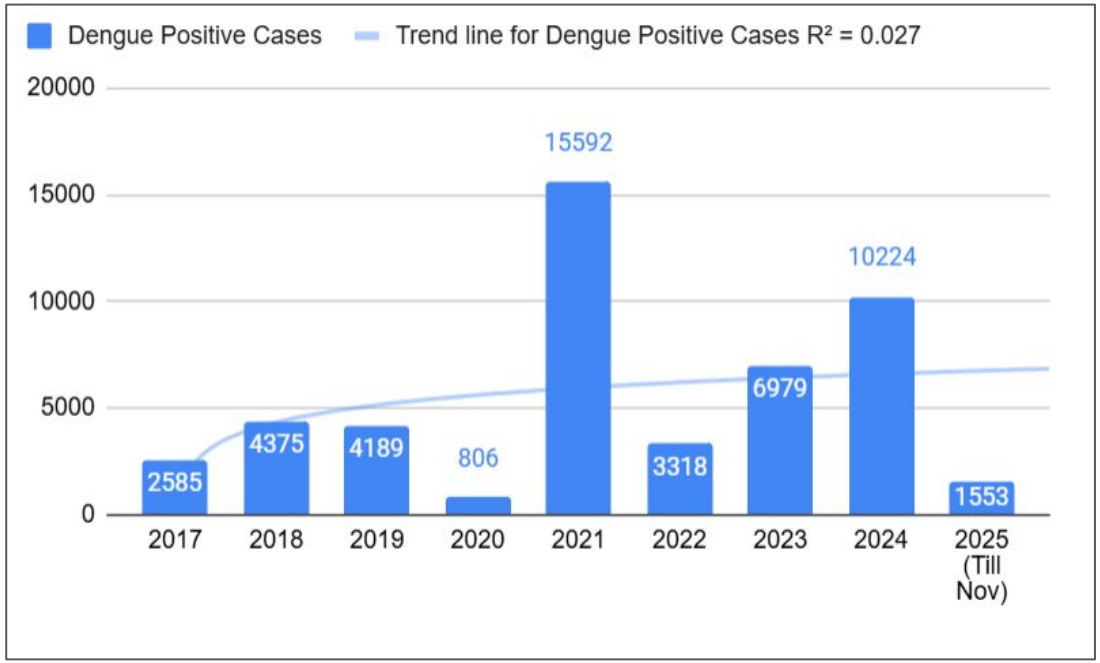
Station data is considered to be the ground truth

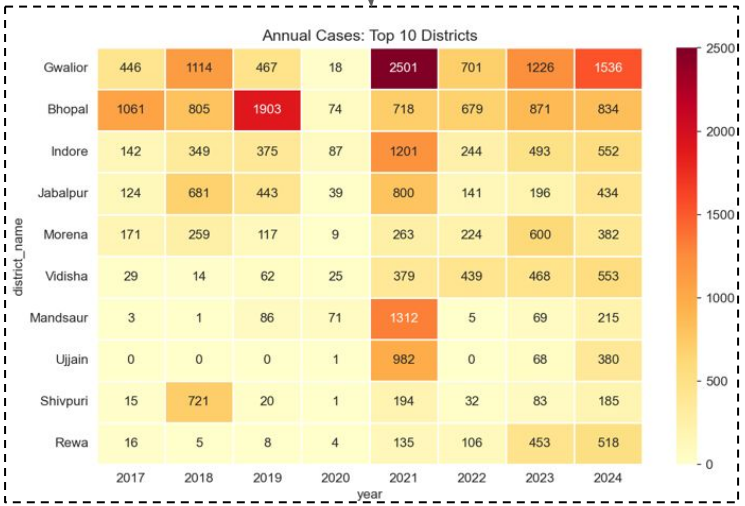


- For ERA5 and IMD Gridded with station data
- Comparing temperature and humidity in both against station daily data, these are almost the same
- Rainfall is better correlated for IMD

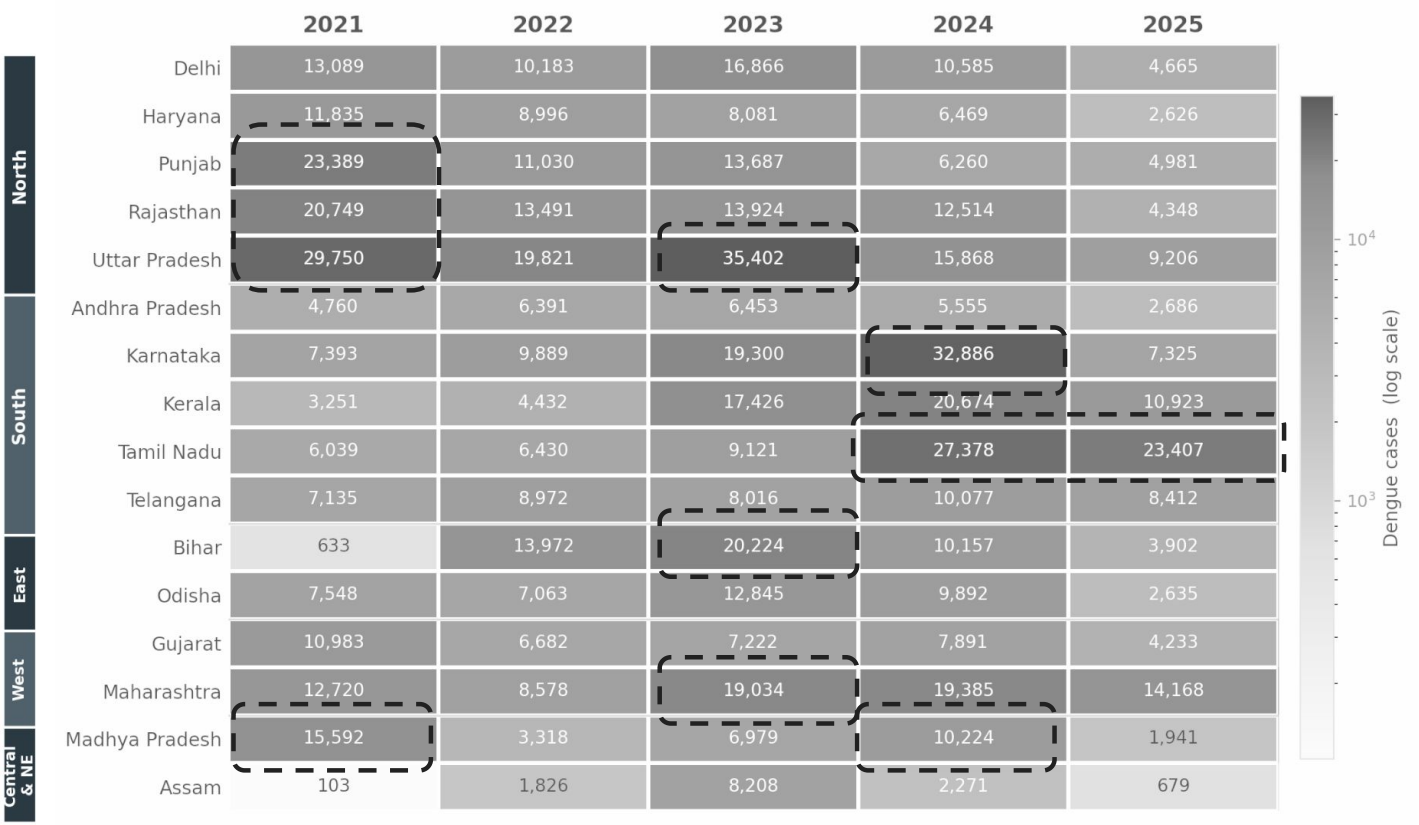


Exploratory data analysis from MP: Spikes in regions and years, which needed further investigation





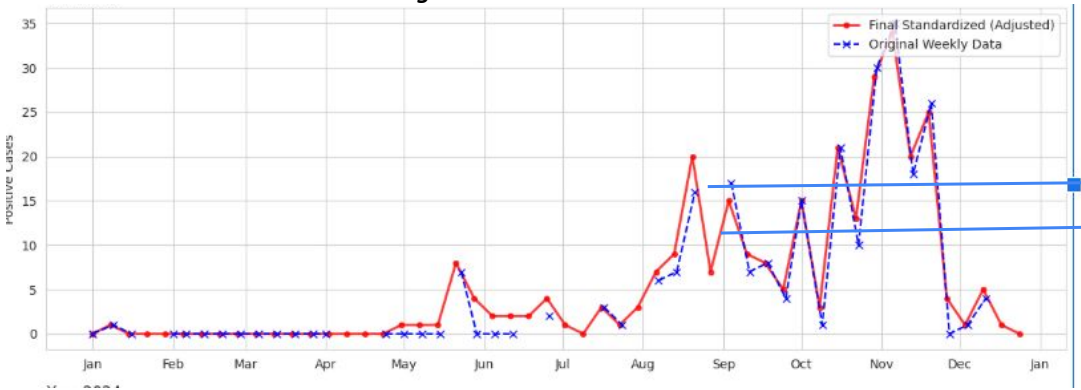
Across India, Different regions see Spikes in different years



- This means that the spikes depends on something which is influenced closely in space like entomology, serotypes
- It contradicts that spikes are purely due to post covid pandemic or migration

In Practical (On Ground): Data is incomplete and under-reported case data and needs corrections before modelling

Missing Case Data



Adjusted Cases using Monthly totals

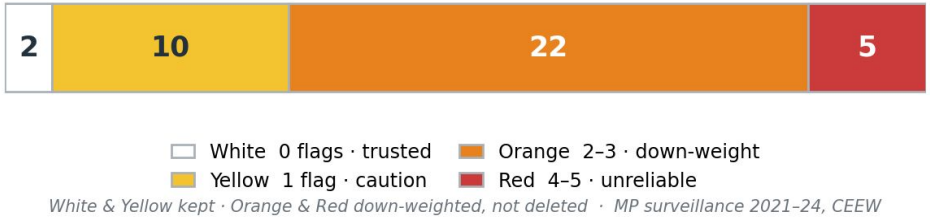
Reported Weekly Cases

Under-reporting Case Data

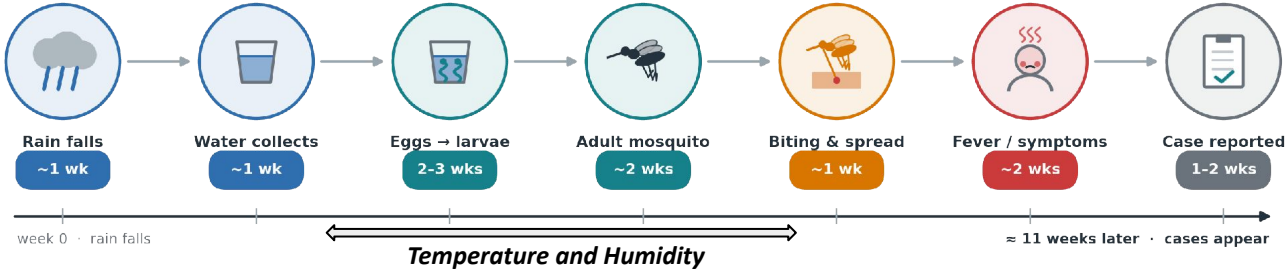
Every district-year is screened on 4 quality checks:

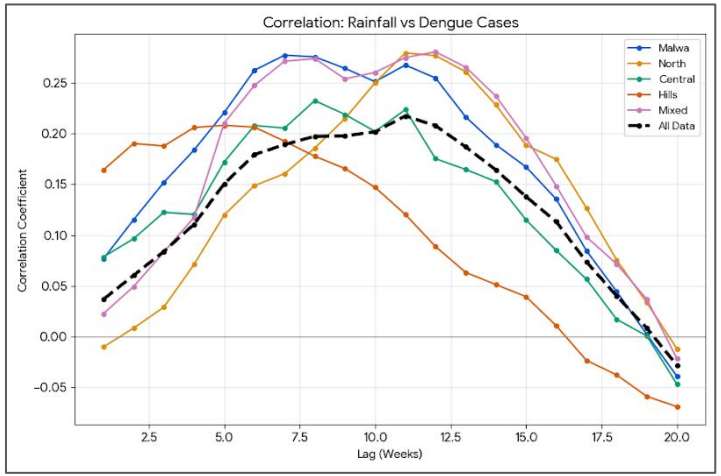
1. Zero / implausibly low sample counts
2. Missing weeks (no reporting at all)
3. **Sudden unexplained drops and spikes**
4. Internal inconsistencies (e.g. positives > samples)

39 Active districts: And District weeks



Lagged variables in Machine learning: Climate parameters leads cases by about 8–12 weeks and differs for each region's signature





Regions	Peak Rainfall Lags
Malwa Plateau: (e.g., Indore, Ujjain)	**8 weeks**
North/Gird: (e.g., Gwalior, Morena)	**11 weeks**
Central/Narmada: (e.g., Jabalpur, Hoshangabad)	**8 weeks**
Hills/Satpura: (e.g., Balaghat, Chhindwara)	**9 weeks**
Mixed/Bundelkhand: (e.g., Sagar, Rewa)	**12 weeks**

Machine Learning Model: a two-stage model, Stage 1 flags case-weeks (recall-tuned); Stage 2 grades how many

WHAT GOES IN — 56 features (M6)

Rainfall & precipitation · 15
optimal-lag rain, wet & heavy days; 4/8-wk sums, extremes

Temperature & humidity · 11
3-block temp & humidity; 12-wk max/min; 25–32° C band wks

Case history & baselines · 12
incidence lags 9–12; 7 district baselines; SARIMA anchor

Lab-testing & surveillance · 8
log samples & positivity, lags 5–8; rolling 4-wk averages

Vegetation & interactions · 10
NDVI habitat; rain×temp; mosquito-suitability; COVID flag

Seasonal calendar cues removed — spikes earned from climate, not the date.

What the case counts look like · 5,616 district-weeks, 2021–24

Count	Band	Share of weeks
0 cases	Quiet week	3,243 · 57.7%
1–5	Sporadic	1,139 · 20.3%
6–20	Elevated	795 · 14.2%
>20	Outbreak	439 · 7.8%

STAGE 1 — Will there be cases?

- Classifier scans case trend + climate and outputs a Yes / No alert for each district-week

→ **Alert threshold set at 0.35, not 0.5 — fires on weak signals so few real outbreaks slip through**

STAGE 2 — How many cases?

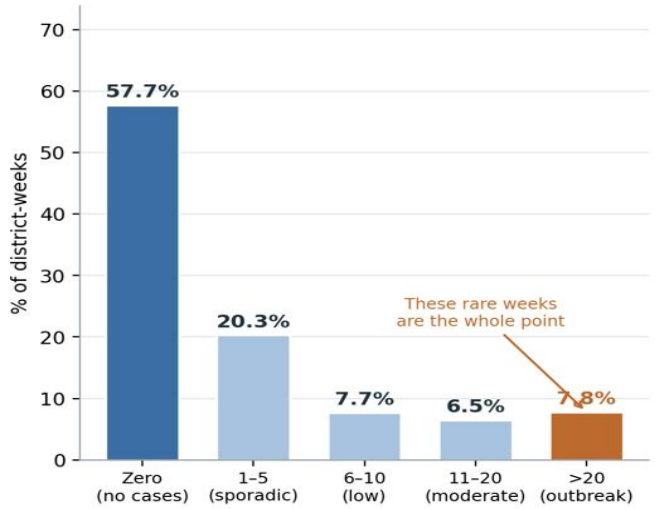
- Only runs if Stage 1 says Yes — a 3-model ensemble estimates the case count
- Tweedie loss: built for zero-heavy, skewed counts; predictions never go negative
  - Asymmetric Absolute mean Error

15 CEEW THE COUNCIL

Inclusion of two additions which makes accuracy better

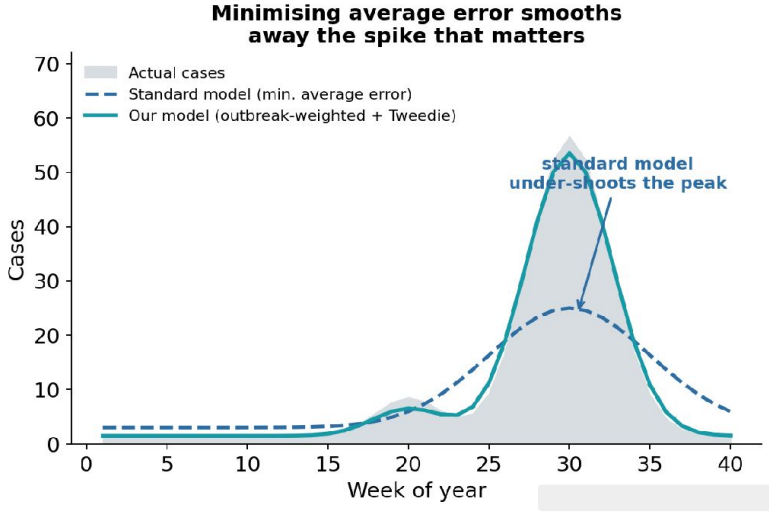
The Tweedie Loss Distribution:

- A model that treats all weeks equally, simply learn the dominant signal: "predict quiet."
- We use a Tweedie loss, designed specifically for zero-heavy, right-skewed counts, so predictions stay non-negative and the loss is concentrated on the dangerous tail



Asymmetric Mean Absolute Error (AsymMAE):

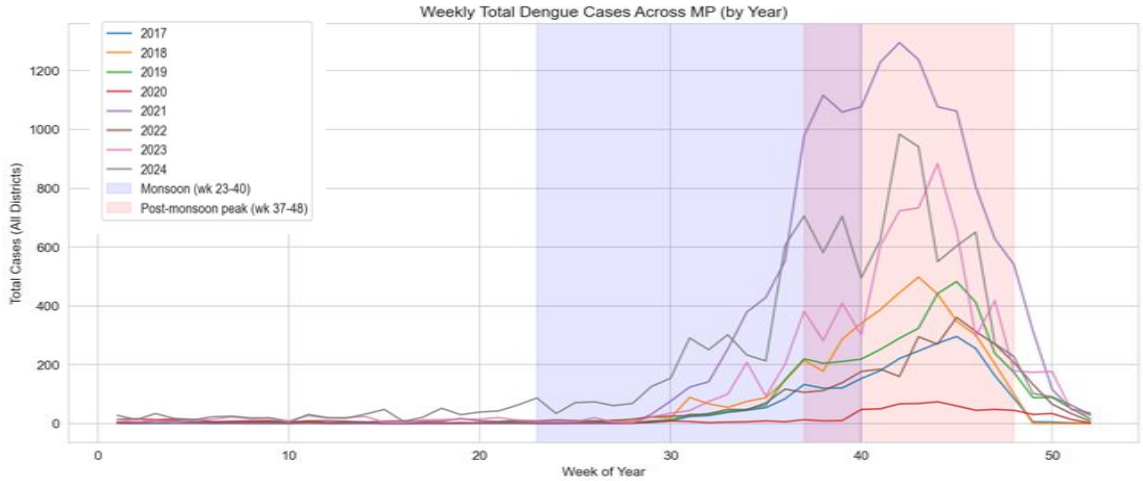
- Predicting "the average" looks correct most of the time, though we know this is not.
- An asymmetric loss function that penalises missing an outbreak far more heavily than over-predicting on a quiet week, the model is forced to chase the spike, not the mean.
- Outbreak weeks are further up-weighted 2–6× during training, so the model gets extra practice on the rare events that actually trigger health alerts.

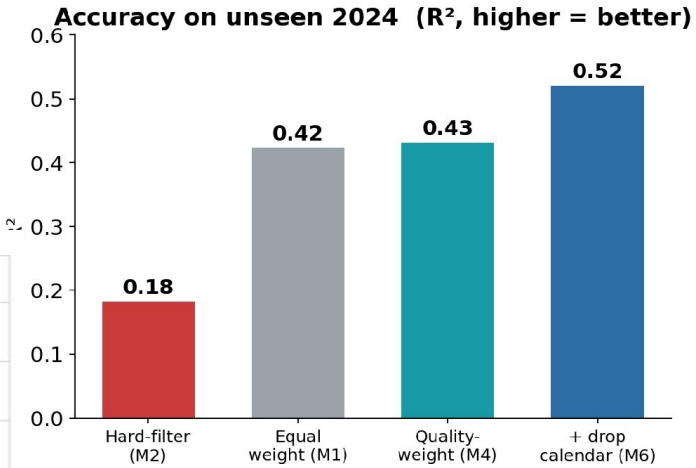


Dengue is seasonal, but what if there is an early/late monsoon

The season's shape is fixed; only its size changes — so we tested whether the model just memorises the date

- Season **shape is fixed**: ~**78% of cases in Sep–Nov**, October peak — every year. Only **size** (≈8×) and **duration** (≈8 vs ≈4 months) change.
- That makes the calendar a tempting shortcut — "predict a September spike because it's September".
- So we **removed all calendar features** (week-of-year, monsoon flags) and re-ran:
  - **Quality-weighted model → improved** (R^2 0.43 → **0.52**): learns the climate "why".
  - **Un-weighted model → got worse** (0.32 → 0.26): it needed the crutch.
- This helps the model to identify the real predictors for dengue cases i.e. climate changes.





M1 — Baseline: full data, no quality filtering, highest recall (0.85).
M2 — White-only: trained on clean data, removes bias, weakest
M4 — Quality-weighted: up-weights clean data, strong and low-complexity
M6 — Best: Quality-weighted, calendar removed, learns climate

Experiments: Six-model comparison (held-out 2024, clean districts)

What counts as an outbreak

- Outbreak week = cases above 5 cases in a district for a week
- M6 is scored against this threshold on the held-out 2024 data.

Model	Features	Season	MAE	RMSE	R^2	F1	Precision	Recall
M1: Keeping all data (no filtering)	60	Yes	4.28	9.85	0.42	0.85	0.85	0.85
M2: Filtering out under-reported weeks	25	Yes	4.85	11.73	0.18	0.80	0.80	0.81
M3: Per district lags included	69	Yes	4.38	10.69	0.32	0.81	0.89	0.74
M4: Weighted under-reported weeks	60	Yes	4.07	9.77	0.43	0.84	0.88	0.81
M5: Per district lags included + week# removed	65	No	4.64	11.16	0.26	0.84	0.87	0.81
M6: M4+M5	**56**	**No**	**4.03**	**8.97**	**0.52**	**0.85**	**0.87**	**0.84**

What our next steps are on the project

- **Increase Granularity:** Moving from district level to sub district level (ward/PHC) with data and GeoSpatial

- **Serotype switches:** Which strain is circulating (DEN-1/2/3/4). A shift, e.g. a DEN-2 rise can flip a quiet year into an outbreak.

- **Mosquito surveillance:** On-the-ground entomology, larval surveys and adult vector density (where and how many).

- **Pilot deployment in MP and Tamil Nadu:** Run the system in two states to test it across different climates and reporting systems.

What are the takeaways for the builders…

1. **Policy doesn't gate whether you build — it decides what's worth building**
2. **Your own compute is now a policy object**
3. **Representative, sovereign data is a precondition for trust, not a nice-to-have**
4. **Lab accuracy is not field accuracy**
5. **A forecast is not an outcome; institutional ownership is the deployment strategy**
6. **Optimise for the trust budget, not just the loss function**

Thank You