AI for Public Policy

Climate Change AI Summer School, 2026

Carlos Rodriguez-Pardo, PhD

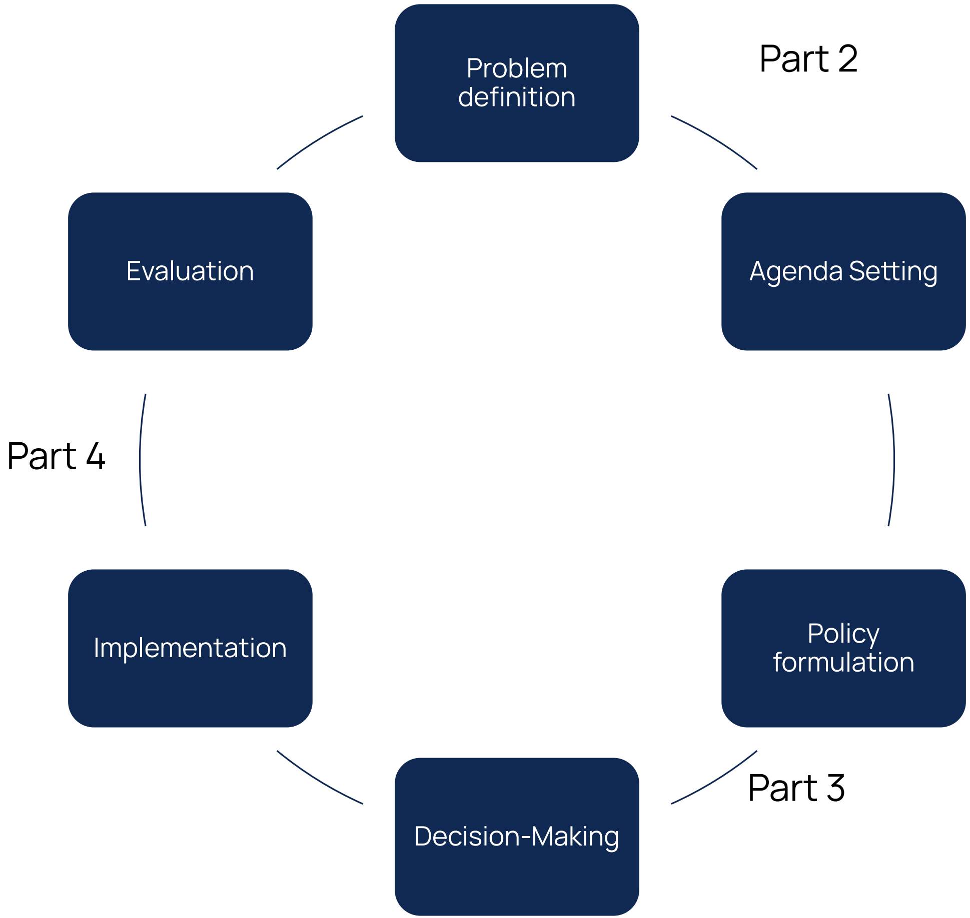
Politecnico di Milano / EIEE / CMCC

Agenda for today

	The question	The role of AI
Part 1: **Foundations**	What are the goals of climate policy analysis?	None yet, just an introduction
Part 2: **Measure**	What's happening, what's been tried, who is exposed?	NLP on policy corpora · computer vision & remote sensing · geospatial / Earth-system foundation models
Part 3: **Modelling futures**	What could happen under different policies and actors?	IAM emulators · neural PDE solvers · RL & multi-agent RL
Part 4: **Evaluation and deployment**	Did a policy work? Will it be accepted? Can it be trusted?	Causal ML · LLMs · agent-based models

The policy cycle

Where can AI help climate policy?



Which institutions are involved in policy decision-making?

International – UNFCCC, IPCC, EU

National & regional governments

Local governments

Sectoral actors: electricity providers, transport agencies

Private actors: Insurers, firms, banks, certifiers, etc

- Different actors with different problems and priorities
- Different timescales (decadal policies vs quarterly earnings)
- Different levers: Carbon taxes, renewable shares, electric vehicle % in public transport services
- Different spatial resolutions (continent-wise policies, city planning, building renovations…)

AI can help policymaking across these scales

What is climate policy analysis trying to achieve?

Estimate damages & risks

How bad, where, for whom, and when?

Compare options, evaluate tradeoffs

Mitigation vs. adaptation vs. accepting losses.

Design instruments

Taxes, standards, subsidies, plans, markets.

Evaluate what worked

Did our policies change anything, and under what conditions?

Decide under uncertainty

Robust choices despite deep unknowns.

Policy responses to climate change

Mitigation

Goal: prevent or reduce the GhG concentration in the atmosphere

Examples:
- Renewable standards
- Electric vehicle mandates
- Carbon capture and storage
- Carbon taxes
- Electrification

Adaptation

Goal: reduce harm from global warming that is inevitable

Examples:
- Flood defenses
- Building isolation planning
- Early warning systems
- Infrastructure

Loss and damage

Goal: Compensate what cannot be prevented or adapted to

Examples:
- Climate finance
- Attribution

Policy analysis needs to compare all three and understand their cost, tradeoffs and feasibility. For this, we need instruments like the social cost of carbon and damage estimates.

The policy instrument toolbox

Carbon pricing

Taxes & emissions trading

Regulation & standards

Bans, codes, efficiency rules

Subsidies & investment

Public R&D, renewables, renovations

Industrial policy

Green industrial strategy

Adaptation planning

Risk reduction & resilience

Climate finance

Who pays, who is compensated

Carbon markets & offsets

Credits, MRV, integrity

These policies may be competing for the same budgets

The social cost of carbon

The social cost of carbon is the economic damage caused by one extra emitted tonne of CO_2 expressed in today's money.

+1 tonne of CO_2 emitted today

↓

Causes additional global warming, over centuries

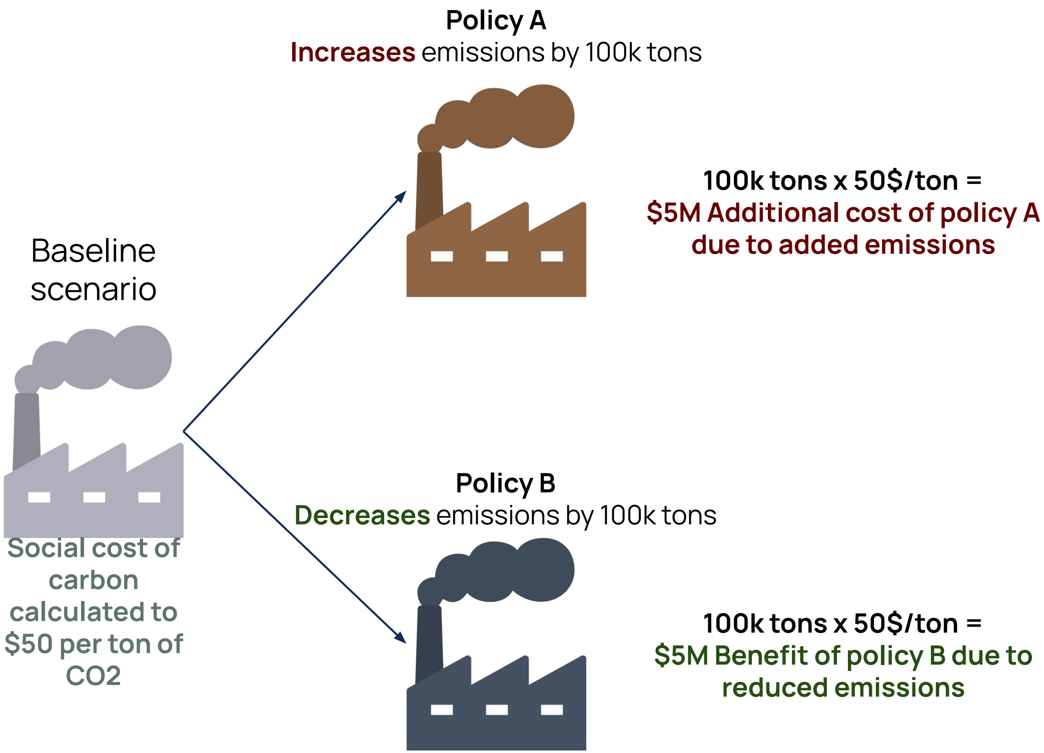
↓

Which causes future damages:
health, crops, capital…

↓

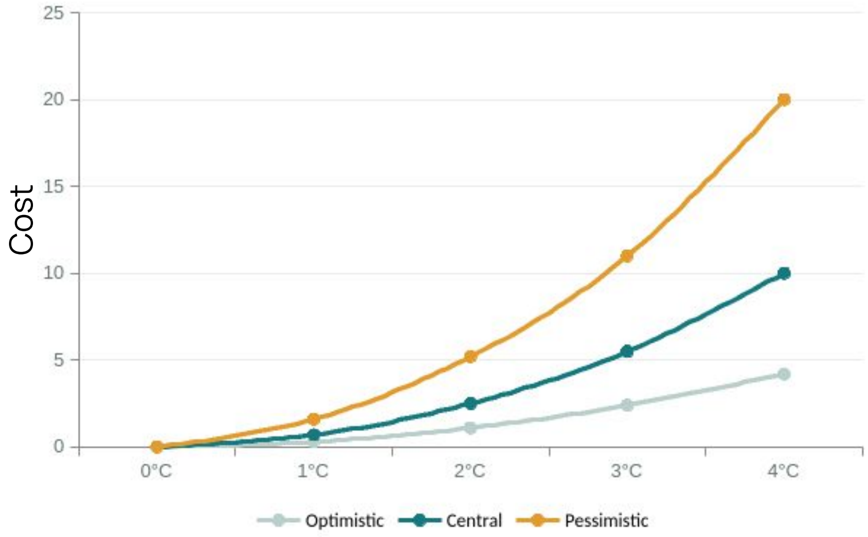
Discount those damages to present value:
$ / CO_2

All these arrows hide models, assumptions, and uncertainties



Damage functions

- Map global warming onto economic loss
- Estimated from historical data, extrapolated far beyond it
- These function are widely used in Integrated Assessment Models and other policy analysis instruments
- Functional form, tipping points and adaptation assumptions all contested
- Their uncertainty also have policy implications



Where does AI fit in all this?

Data

Processing and gathering insights from complex data sources:
Satellite images, policy documents, networks, sensors, economic datasets...

Models

Help in modelling the complex socioeconomic and climate systems:

- Foundation models
- Integrated Assessment Models
- Reinforcement Learning

Decisions

- Estimate possible effectiveness of different policies
- Understand if past policies were effective and under what conditions
- Optimise resource allocation

PART 2 — MEASURE

PART 3 — MODEL FUTURES

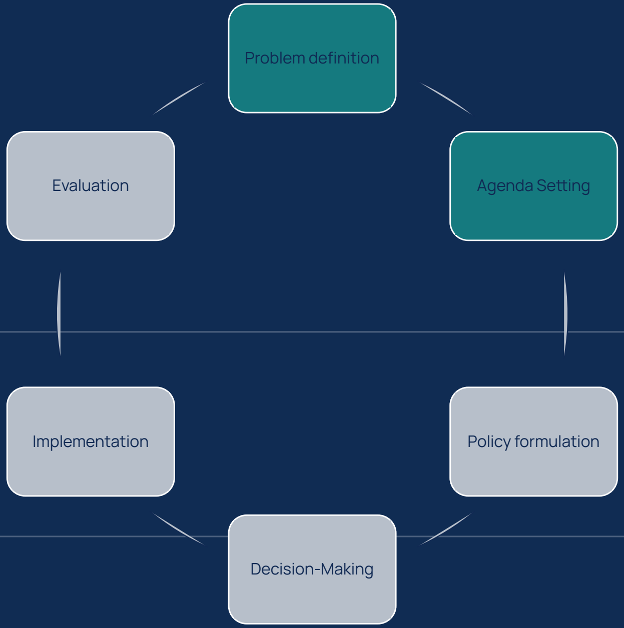
PART 4 — EVALUATE & DEPLOY

Useful references:

- Rolnick et al., Tackling Climate Change with Machine Learning
- Kaack et al. 2022, Aligning AI with climate change mitigation ·
- Ou et al. 2026, AI to support cross-disciplinary climate change research (Nature Climate Change)

Part 2: Measuring the policy world & climate impacts

- What policies exist?
- What's happening in the world?
- Who is exposed?



What do we need to know for effective policy design?

What policies currently exist?

- Policy databases
- Policy trackers

Where are the emissions happening?

- Inventories
- Estimates

Who is exposed? Who are the key actors?

Vulnerability mapping

What does the literature say?

Evidence maps, review papers,

Where are the evidence gaps?

Where are the uncertainties?
What evidence is missing?

How to know what has been tried and studied before?

Harmonisation of policy datasets

Input:
- Fragmented national policy documents
- Global, regional, sector-specific databases

Methods
- Natural language processing,
- Expert knowledge, rule-based learning
- Semi-supervised machine learning

Output:
- Comparable cross-country policy panels
- Who adaptod which policy and when
- Shared format, IPCC-aligned

Wu et al. 2024
https://www.nature.com/articles/s41597-024-03411-z

Mapping the scientific literature

Input:
- Hundreds of thousands of climate-policy papers
  - Impossible to read and analyse by any human or research group

Methods
- Natural language processing
- Text classification
- Topic mapping

Output:
- Evidence maps
- LIterature gaps, where evidence is thin

Callagan et al. 2025
https://www.nature.com/articles/s44168-024-00196-0

Measuring what's happening, at a policy-relevant resolution

High resolution and air quality measurement

- ML estimates of subnational urban GHG emissions
  - Yu et al. https://www.nature.com/articles/s41597-026-06691-9
- ML-driven classification of air quality stations
  - Renna et al (https://www.nature.com/articles/s41597-026-06797-0.pdf)

Reading and understanding policy documents

- NLP for the classification of political manifestos (Sanford et al. doi:10.1017/S0007123425100719)

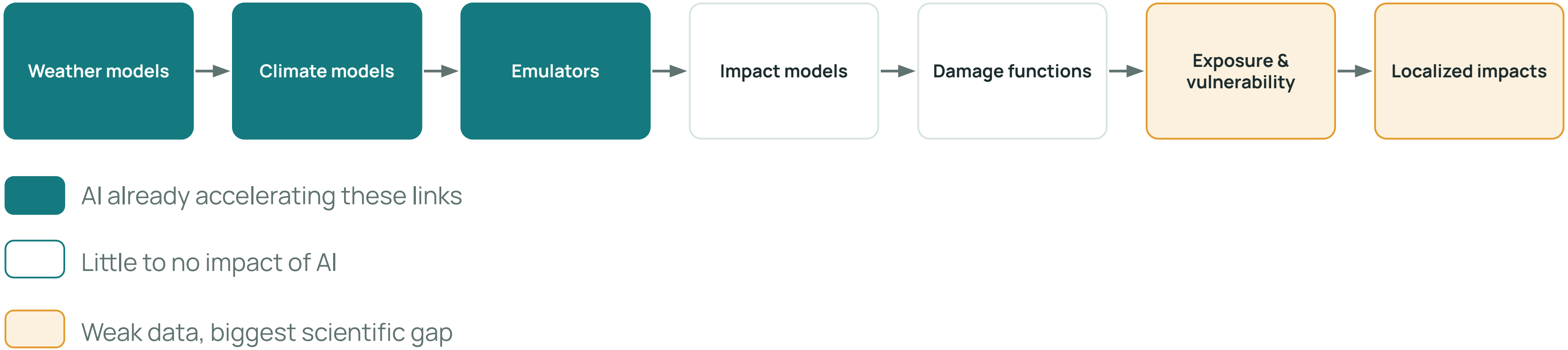
The integration challenge

- How to link policies, sectors, emissions, geographies, populations, outcomes?
- How can we turn disjoint datasets and data modalities into policy-relevant evidence?

References:

- Reichstein et al. https://www.nature.com/articles/s41467-025-57640-w
- Rodriguez-Pardo et al.: https://eartharxiv.org/repository/view/13019/

The impact modelling chain



AI for Weather and AI for Climate Science lectures at the CCAI26 Summer school will cover these topics deeply. In this lecture we focus on how climate impacts and how they enter policy decisions

Why resolution matters

Global grid cell

~100-1000km

- Is a country on track?
- What are the climate patterns in this area of the world?

Regions

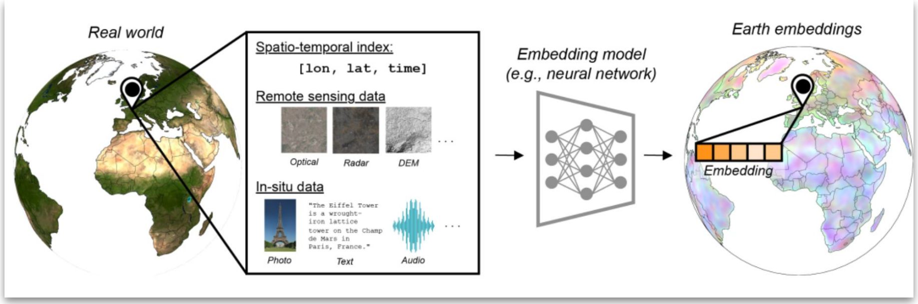
~10-100km

- Where should adaptation funds go to?
- Which cities are most vulnerable to climate extremes?
- What populations are more vulnerable?

City blocks

~10m-10km

- Which neighbourhoods need flood protection?
- Where are emissions coming from?
- Where should trees be planted?



Geospatial embeddings and earth-system foundation models can connect resolutions effectively: You learn representation once, then reuse it across several impact, reconstruction and exposure analysis tasks.

Examples include SatCLIP, GeoCLIP, Range, Clay, Tessera, AlphaEarth

Figure from: *Earth Embeddings: Towards AI-centric Representations of our Planet* (Klemmer et al. https://eartharxiv.org/repository/view/11083/)

Better data changes real decisions

Adaptation targeting

Who gets protected first and at what cost?

Climate finance & loss-and-damage

More accurate damage estimates can improve financial decisions

Ecosystems & nature

What are the climate impacts on biodiversity, where are they happening? Which species are most affected?

Insurance

Improvements in pricing, insurability, coverage, risk management

Infrastructure planning

What do we need to build and how? What are the chances of big floods, heatwaves, extreme snowfall?

Distributional assessment

Who will be most affected by climate change? What will the impacts be on equality, health?

Example: allocation of funds for adaptation

With coarse, low quality data:

- Funds may follow existing trends, biases, lobbying
- Funds may not be allocated to populations that need them most
- Taxpayer money is allocated suboptimally

With high resolution localised impact data

- Funds can follow exposure and vulnerability
- Funds can be allocated in a robust, defensible manner
- Decision makers better understand tradeoffs and limitations

Measurement is not neutral

False precision

- High resolution outputs may look exact, but many times they're not
- Many common datasets (eg ERA5) carry hidden assumptions, modelling decisions, and models
- Uncertainties compound in invisible ways

Data availability is not uniformly distributed

- Data coverage tends to be weaker exactly where vulnerability is higher
- Therefore, models tend to extrapolate and be less accurate in more vulnerable regions
- E.g.: Fairness and representation in satellite-based poverty maps: Evidence of urban-rural disparities and their impacts on downstream policy (Aiken et al https://arxiv.org/abs/2305.01783)

Documented does not mean effective

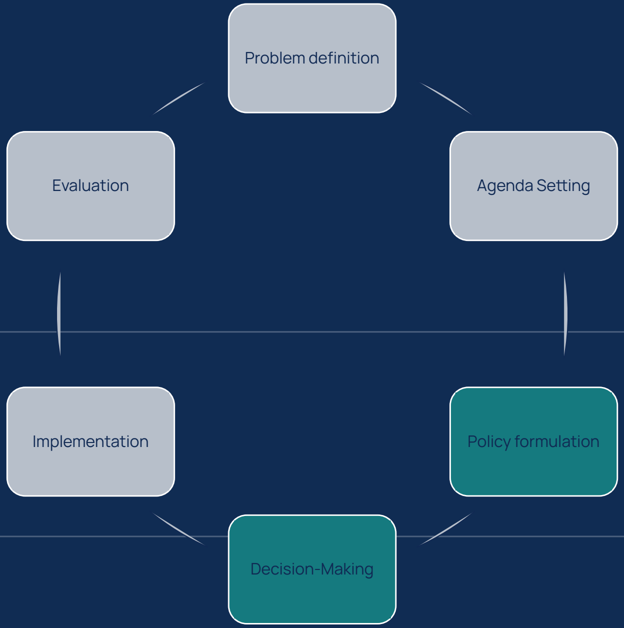
- Policy databases record what's written down, not necessarily what was effective

Hidden losses

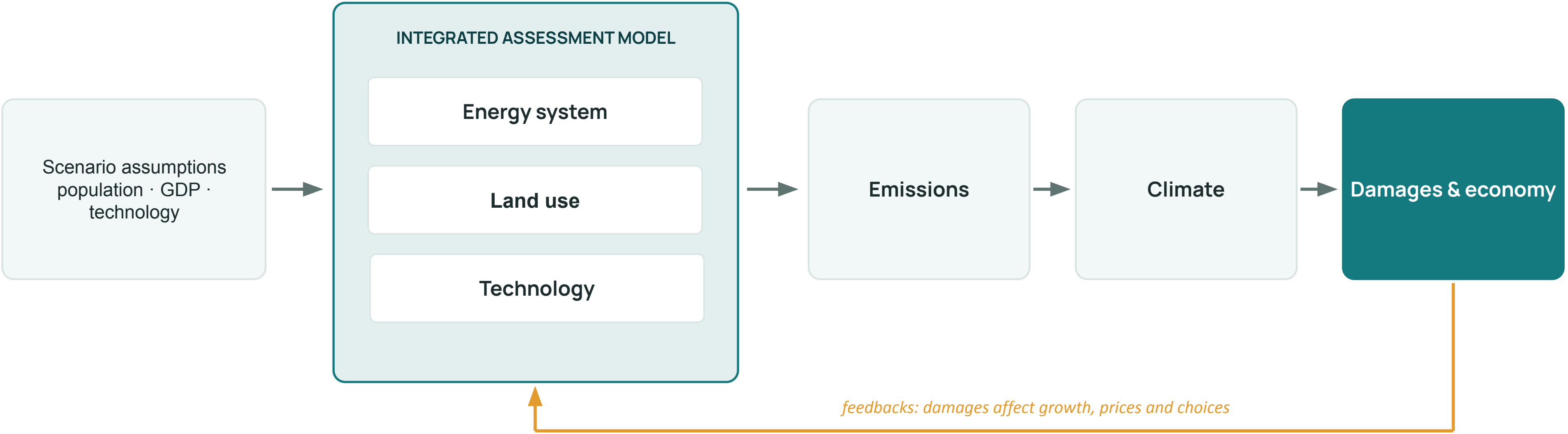
- Some variables are harder to measure than others. Damage functions can miss
  - Non-market damages
  - The distribution of damages
  - Cultural losses
  - Ecological damages

Part 3: Modelling futures & designing policy options

- How can we model the future climate and economy?
- How can we model the multiple actors involved in climate policy?
- How can we ensure our models are trustworthy?



Integrated assessment models (IAMs)



- IAMs model stylised worlds, they are useful to answer what-if questions, they are not doing forecasting
- IAMs come in different families:
  - Cost-benefit IAMs (DICE)
  - Process-detailed IAMs (REMIND, WITCH)

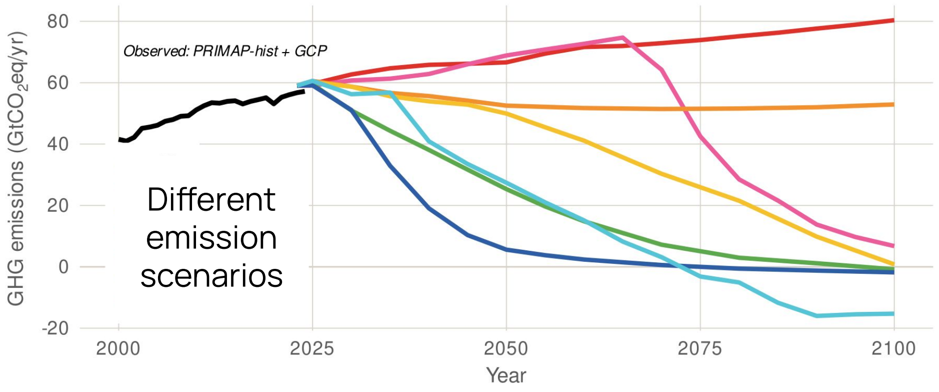
What IAMs inform

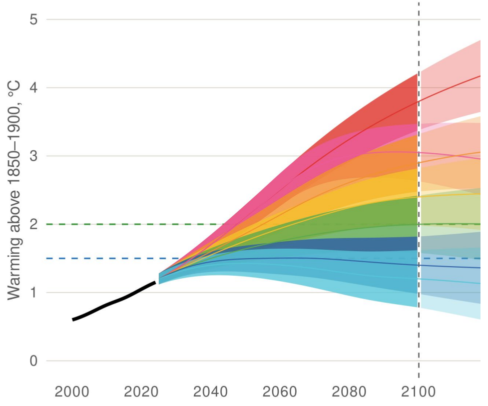
- Cost-effective pathways to temperature targets
- Technology portfolios: how much wind, CCS, nuclear, when, at what cost
- Carbon-price trajectories consistent with targets
- Regional burden-sharing and its (in)equity
- IPCC scenario databases

IAMC
Integrated Assessment Modeling Consortium
Founded 2007

ipcc
INTERGOVERNMENTAL PANEL ON climate change

Stylised emission pathways ($GtCO_2$ / yr)





AI for IAM emulation

- IAMs are useful but computationally really expensive
- Model ensembling to measure uncertainty can become intractable
- Machine learning emulators can help:
  - They learn to map from IAM inputs to IAM outputs
  - Once trained, they can answer instantly
- ML-IAM allows for interactive scenario exploration
  - Eg *Shin et al., ML-IAM v1.0 · Li et al. 2025, deep learning for global mitigation scenario variables* *(https://egusphere.copernicus.org/preprints/2026/egusphere-2025-5305/)*

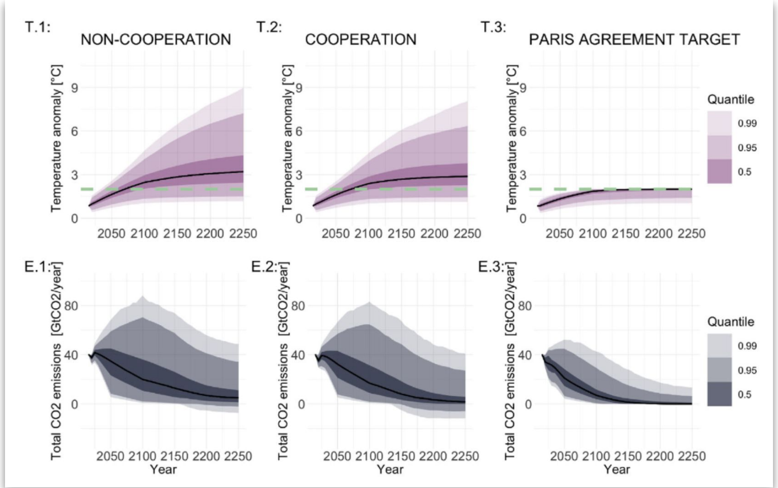
1 IAM scenario
hours to days of compute

10,000+ emulator evaluations in the same time

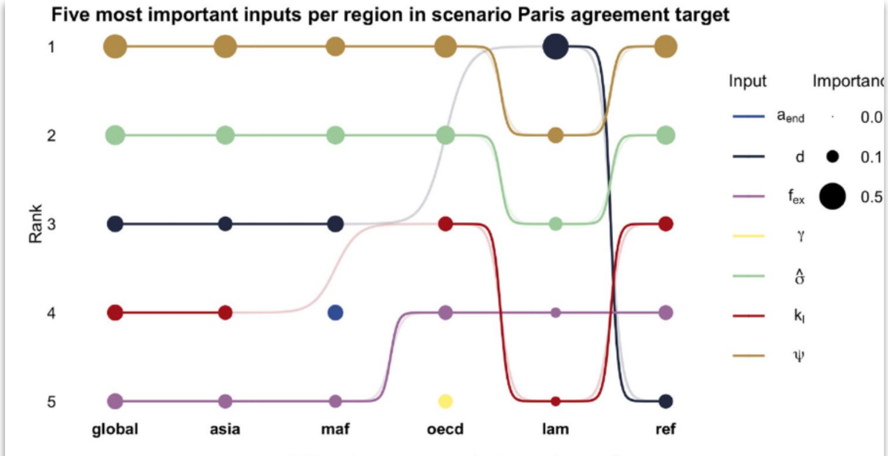
(illustrative)

Uncertainty Quantification, Sensitivity Analysis

Uncertainty quantification



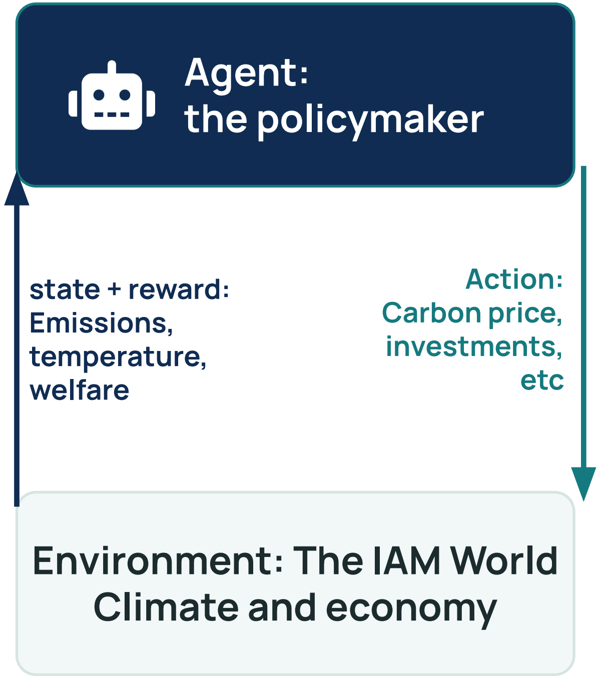
Sensitivity & black box explainability



(C) Paris agreement target scenario.

Figures from Chiani et al. *Global sensitivity analysis of integrated assessment models with multivariate outputs* https://onlinelibrary.wiley.com/doi/full/10.1111/risa.70002

Reinforcement Learning: policies as decision sequences



- Climate policy is sequential: act, observe, adjust — for decades
- RL searches over policy trajectories, not single settings
- Coupled with IAMs, RL **can reach corners of the solution space classical solvers never visit**
- **Useful output**: surprising, previously unexplored candidate strategies

Competing and cooperative agents

- Many policy actors, with different and ever-changing objectives and priorities (countries, firms, regulators)
- They cooperate, compete, free-ride, or ignore each other
- Multi-agent RL makes incentive misalignment and coordination failure visible and quantifiable

	B abates	B free-rides
A abates	Both pay, planet wins	A pays, B benefits
A free-rides	B pays, A benefits	**Nobody pays — everyone loses**

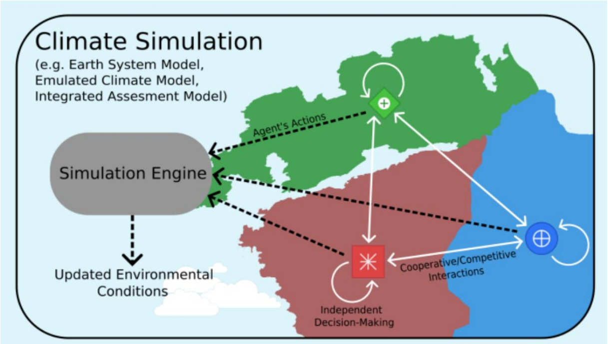


Figure from Multi-Agent Reinforcement Learning Simulation for Environmental Policy Synthesis (Rudd-Jones et al.)

New directions at the intersection of AI and IAM solvers

arXiv:2312.06527v1 [cs.AI] 11 Dec 2023

Can Reinforcement Learning support policy makers? A preliminary study with Integrated Assessment Models

Theodore Wolf
University College London | Carbon Re
theo@carbonre.com

Nantas Nardelli
Carbon Re
nanta@carbonre.com

John Shawe-Taylor
University College London
j.shawe-taylor@ucl.ac.uk

María Pérez Ortiz
University College London
maria.perez@ucl.ac.uk

Abstract

Governments around the world aspire to ground decision-making on evidence. Many of the foundations of policy making — e.g. sensing patterns that relate to societal needs, developing evidence-based programs, forecasting potential outcomes of policy changes, and monitoring effectiveness of policy programs — have the potential to benefit from the use of large-scale datasets or simulations together with intelligent algorithms. These could, if designed and deployed in a way that is well grounded on scientific evidence, enable a more comprehensive, faster, and rigorous approach to policy making. Integrated Assessment Models (IAM) is a broad umbrella covering scientific models that attempt to link main features of society and economy with the biosphere into one modelling framework. At present, these systems are probed by policy makers and advisory groups in a hypothesis-driven manner. In this paper, we empirically demonstrate that modern Reinforcement Learning can be used to probe IAMs and explore the space of solutions in a more principled manner. While the implication of our results are modest since the environment is simplistic, we believe that this is a stepping stone towards more ambitious use cases, which could allow for effective exploration of policies and understanding of their consequences and limitations.

1 Introduction

Climate is a high dimensional dynamical system with strong inter-dependent components and long time dependencies, all of which interact to produce highly non-linear responses and behavior. Climate is also highly conditioned on human behavior – another greatly complex system – such that is now necessary to reason about climate change from a socio-climatic perspective [Moore et al., 2022]. To make progress towards achieving some kind of solution in the face of extreme consequences, policymakers and advisory groups employ *Integrated Assessment Models* (IAMs), state of the art models for climate change that combine knowledge about human development (such as economical theories) together with planetary sciences such as ecology and geophysics [Parson and Fisher-Vanden, 1997]. Exploring and analysing the properties of the IAMs employed for large scale assessments [Pörtner et al., 2022] - to e.g. measure fidelity against the real world – is generally intractable from a computational perspective, which leads researchers to implement poor simplifying assumptions and decrease their effectiveness [Asefi-Najafabady et al., 2021]. Smaller IAM models aim to provide an alternative by employing fewer state variables and simpler sets of dynamics, making them amenable to mathematical probing and analysis [Kittel et al., 2017, Nitzbon et al., 2017]. The

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

RL for finding new solutions

Leveraging a Fully Differentiable Integrated Assessment Model for RL and Inference

Koen Ponse[*1] **Kai-Hendrik Cohrs**[*2*]
Phillip Wozny[*3,4] Andrew Robert Williams[5,6] Tianyu Zhang[5,6] Erman Acar[7]
Yoshua Bengio[5,6,8] Aske Plaat[1] Thomas M. Moerland[1]
Pierre Gentine[9,10] Gustau Camps-Valls[2]

[*]Equal contribution
[1]Leiden University, [2]Universitat de València, [3]Vrije Universiteit Amsterdam, [4] Tilburg University, [5]Mila - Quebec AI Institute, [6]DIRO, Université de Montréal, [7]University of Amsterdam, [8]CIFAR, [9]Columbia University, [10]LEAP NSF Science and Technology Center

Abstract

Integrated Assessment Models (IAMs) such as RICE have long provided a foundation for studying the coupled dynamics of the global economy and climate system. Traditionally, these models have been used in a forward-simulation mode, with parameters hand-calibrated and dynamics treated as fixed. In this work, we introduce RICE-N-JAX, a fully differentiable implementation of the multi-region RICE-N model in JAX. Beyond significantly accelerating the training of multi-agent reinforcement learning (MARL) agents, differentiability opens new research directions, including automated calibration to historical data, recovery of latent regional behavioral parameters (e.g., risk aversion and time preferences) and sensitivity analysis of economic and technological assumptions. Moreover, it allows us to treat policy design and international negotiation mechanisms as learnable parameters within a gradient-based optimization framework. We outline new research opportunities that arise when an IAM becomes a differentiable environment and discuss implications for climate–economics modeling, machine learning for climate policy and the fusion of data- and theory-driven approaches.

1 Introduction

The impacts of climate change are already evident: ecosystems are shifting and extreme weather events are intensifying, threatening livelihoods and economic stability and underscoring the urgent need for action [Pörtner et al., 2022].

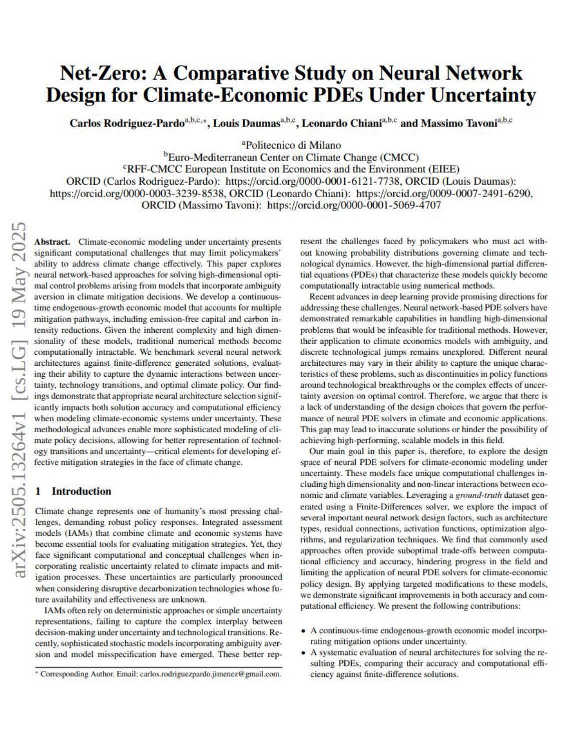
Climate change is a shared global challenge, yet mitigation entails trade-offs. Investments in low-carbon technologies and systemic transformation can constrain short-term growth and costs are unevenly distributed: wealthier nations can invest more readily, while developing countries must balance climate goals with basic development needs. This imbalance reinforces a collective-action dilemma, where self-interest threatens global progress [Gardiner, 2001, Erickson et al., 2015].

Integrated assessment models (IAMs) quantify these climate–economic trade-offs by linking CO_2 emissions, temperatures and growth dynamics. The pioneering Dynamic Integrated model of Climate and Economy (DICE) captures global interactions among population, technology, emissions and damages within a single economy [Nordhaus, 2007]. Its regional extension, the Regional Integrated model of Climate and Economy (RICE), disaggregates these processes across multiple re-

[*]Correspondence: k.ponse@liacs.leidenuniv.nl, kai.cohrs@uv.es

1st Workshop on Differentiable Systems and Scientific Machine Learning @ EurIPS 2025

Differentiable IAMs



arXiv:2505.13264v1 [cs.LG] 19 May 2025

Net-Zero: A Comparative Study on Neural Network Design for Climate-Economic PDEs Under Uncertainty

Carlos Rodriguez-Pardo[a,b,c,*], **Louis Daumas**[a,b,c], **Leonardo Chiani**[a,b,c] **and Massimo Tavoni**[a,b,c]

[a]Politecnico di Milano
[b]Euro-Mediterranean Center on Climate Change (CMCC)
[c]RFF-CMCC European Institute on Economics and the Environment (EIEE)
ORCID (Carlos Rodriguez-Pardo): https://orcid.org/0000-0001-6121-7738, ORCID (Louis Daumas): https://orcid.org/0000-0003-3239-8538, ORCID (Leonardo Chiani): https://orcid.org/0009-0007-2491-6290, ORCID (Massimo Tavoni): https://orcid.org/0000-0001-5069-4707

Abstract. Climate-economic modeling under uncertainty presents significant computational challenges that may limit policymakers' ability to address climate change effectively. This paper explores neural network-based approaches for solving high-dimensional optimal control problems arising from models that incorporate ambiguity aversion in climate mitigation decisions. We develop a continuous-time endogenous-growth economic model that accounts for multiple mitigation pathways, including emission-free capital and carbon intensity reductions. Given the inherent complexity and high dimensionality of these models, traditional numerical methods become computationally intractable. We benchmark several neural network architectures against finite-difference generated solutions, evaluating their ability to capture the dynamic interactions between uncertainty, technology transitions, and optimal climate policy. Our findings demonstrate that appropriate neural architecture selection significantly impacts both solution accuracy and computational efficiency when modeling climate-economic systems under uncertainty. These methodological advances enable more sophisticated modeling of climate policy decisions, allowing for better representation of technology transitions and uncertainty—critical elements for developing effective mitigation strategies in the face of climate change.

1 Introduction

Climate change represents one of humanity's most pressing challenges, demanding robust policy responses. Integrated assessment models (IAMs) that combine climate and economic systems have become essential tools for evaluating mitigation strategies. Yet, they face significant computational and conceptual challenges when incorporating realistic uncertainty related to climate impacts and mitigation processes. These uncertainties are particularly pronounced when considering disruptive decarbonization technologies whose future availability and effectiveness are unknown.

IAMs often rely on deterministic approaches or simple uncertainty representations, failing to capture the complex interplay between decision-making under uncertainty and technological transitions. Recently, sophisticated stochastic models incorporating ambiguity aversion and model misspecification have emerged. These better represent the challenges faced by policymakers who must act without knowing probability distributions governing climate and technological dynamics. However, the high-dimensional partial differential equations (PDEs) that characterize these models quickly become computationally intractable using numerical methods.

Recent advances in deep learning provide promising directions for addressing these challenges. Neural network-based PDE solvers have demonstrated remarkable capabilities in handling high-dimensional problems that would be infeasible for traditional methods. However, their application to climate economics models with ambiguity, and discrete technological jumps remains unexplored. Different neural architectures may vary in their ability to capture the unique characteristics of these problems, such as discontinuities in policy functions around technological breakthroughs or the complex effects of uncertainty aversion on optimal control. Therefore, we argue that there is a lack of understanding of the design choices that govern the performance of neural PDE solvers in climate and economic applications. This gap may lead to inaccurate solutions or hinder the possibility of achieving high-performing, scalable models in this field.

Our main goal in this paper is, therefore, to explore the design space of neural PDE solvers for climate-economic modeling under uncertainty. These models face unique computational challenges including high dimensionality and non-linear interactions between economic and climate variables. Leveraging a *ground-truth* dataset generated using a Finite-Differences solver, we explore the impact of several important neural network design factors, such as architecture types, residual connections, activation functions, optimization algorithms, and regularization techniques. We find that commonly used approaches often provide suboptimal trade-offs between computational efficiency and accuracy, hindering progress in the field and limiting the application of neural PDE solvers for climate-economic policy design. By applying targeted modifications to these models, we demonstrate significant improvements in both accuracy and computational efficiency. We present the following contributions:

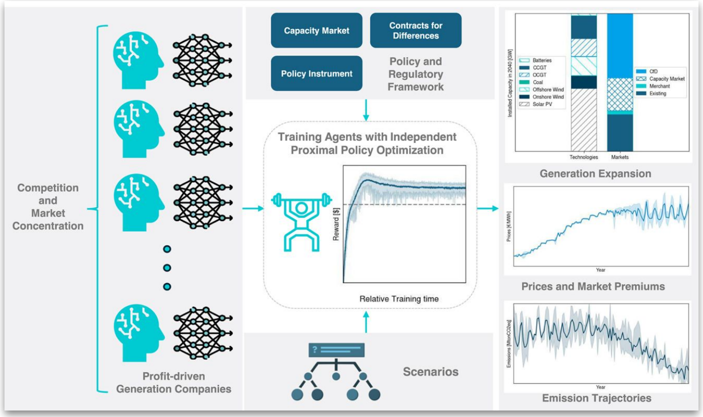
- A continuous-time endogenous-growth economic model incorporating mitigation options under uncertainty.
- A systematic evaluation of neural architectures for solving the resulting PDEs, comparing their accuracy and computational efficiency against finite-difference solutions.

* Corresponding Author. Email: carlos.rodriguezpardo.jimenez@gmail.com.

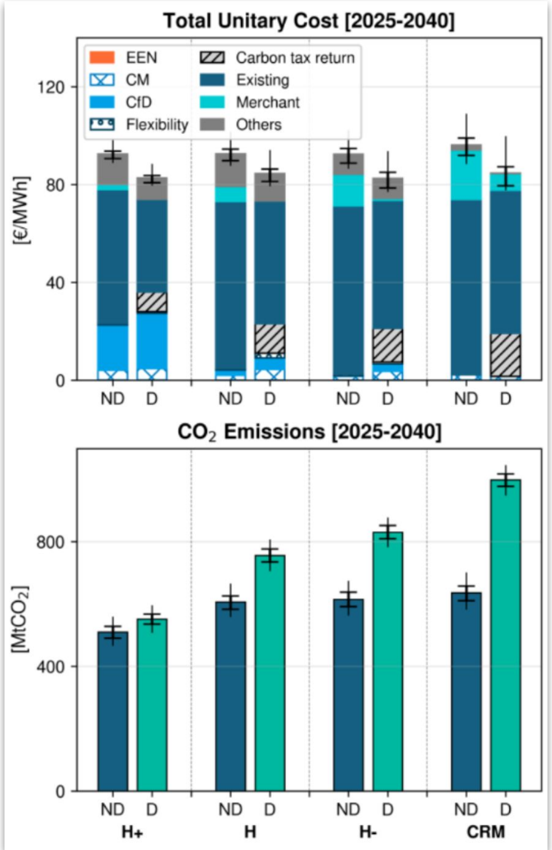
Neural PDE solvers for speed and uncertainty quantification

Multi-Agent RL for long term electricity market policy design

- **Question**: which market designs deliver deep decarbonization at acceptable prices?
- **Approach**: generators as learning agents, bidding and investing under alternative market rules
- Agents reveal strategic behaviour that equilibrium analysis misses
- Design choices visibly shift both price and decarbonization outcomes



González-Ruiz et. al. Assessing Long-Term Electricity Market Design for Ambitious Decarbonization Targets using Multi-Agent Reinforcement Learning



González-Ruiz et. al. EU-ETS under attack? The impact of carbon price suppression on the decarbonization of the power sector

What RL cannot decide

Values are not learnable

- RL optimises inside a formalised problem.
- What society should value is not something RL can or should optimise

A single RL reward cannot model complex, interconnected, multidimensional objectives

- Value functions in RL environments can flatten out plural, contesting goals.

Visibility bias

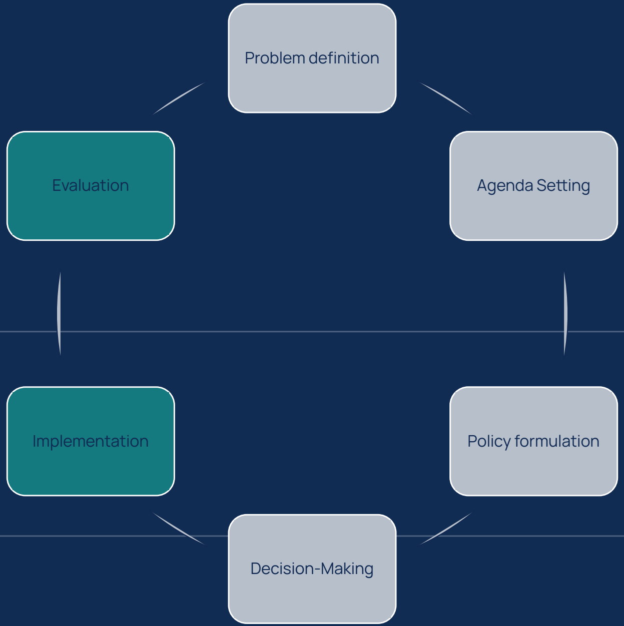
- State and action spaces only encode what is easily be measurable and actionable.
- Everything else gets ignored by design

Many constraints cannot be fully modelled

- Legal, cultural, societal norms, political restrictions, institutional limits are key for the feasibility of polices

Part 4: Evaluation and deployment

- How do we know a policy worked?
- Can we estimate if a policy will be politically popular or socially acceptable?



How do we know if a policy worked? - Causal inference

Randomized experiments

- The gold standard
- Randomly assign the policy, compare groups.
- Scientifically robust but hard to do in the real world.
- E.g.: how can you randomize carbon tax across countries?

Difference-in-differences

- Compare the change over time in policy-treated vs policy-untreated units.
- Very common in the literature

Synthetic control

- If you do not have a good comparison unit → Build a “twin” of a treated unit from a pool of untreated ones.
- E.g. Sweden’s carbon tax — transport emissions compared against a "synthetic Sweden" built from other OECD countries
- See Andersson, J.J. *Carbon Taxes and CO_2 Emissions: Sweden as a Case Study,*
- See Gomez-Mahecha *Colombian Carbon Tax: Evidence of Carbon Pricing in a Developing Context*

Instrumental variables

- Find a lever that moves the policy variable but has no other path to the outcome.
- E.g.: wind speed → more renewable generation → what happens to emissions? Weather is random; the answer is causal.

Causal ML can extend these tools: Finding high-dimensional confounders, heterogenous and non-linear effects, flexible functional forms, etc.

Prediction is not causal evaluation

Prediction

E.g.: “What will emissions be next year?”
This is great for planning and monitoring; but not that useful to understand whether a policy is effective

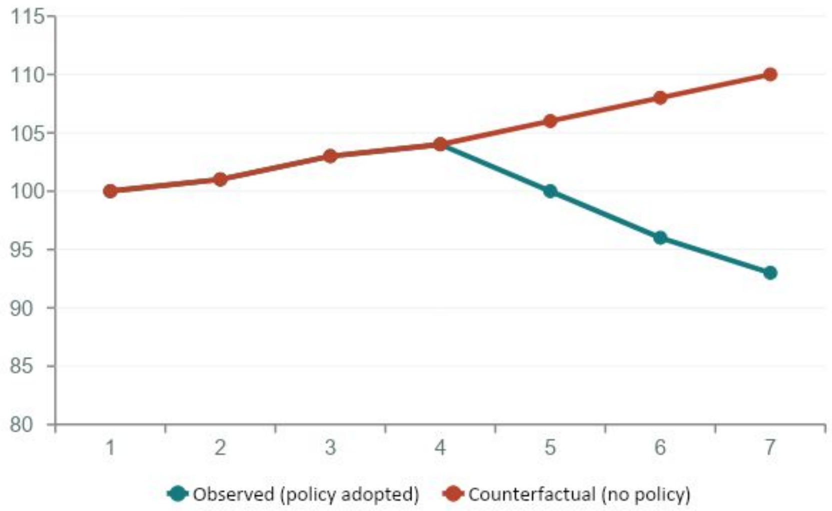
Causal evaluation

“What did this carbon tax change?”: needs a counterfactual,

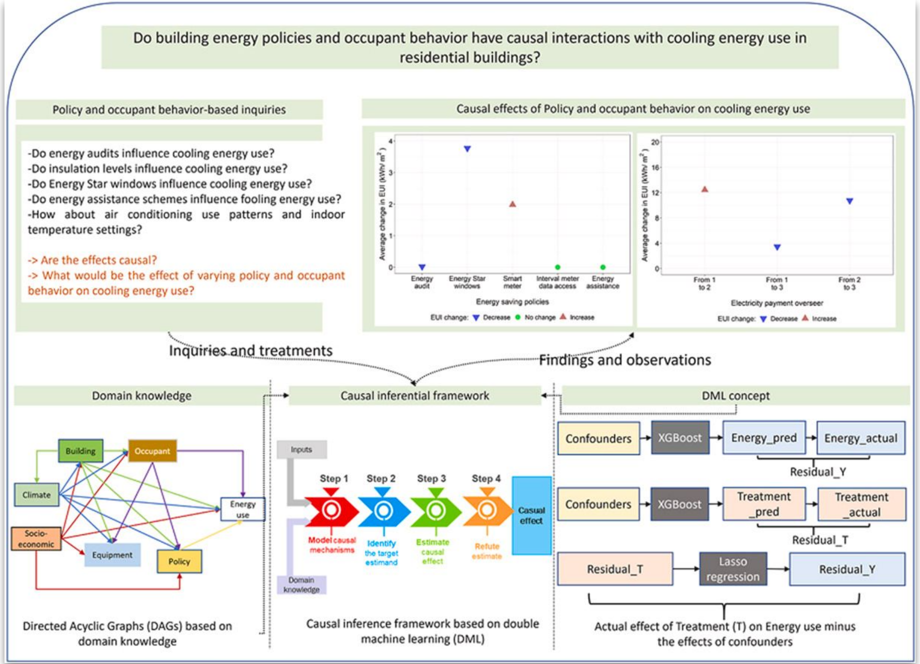
ML extends the causal inference toolkit to high-dimensional confounders and heterogeneous effects.

See Chernozhukov et al *Double/debiased machine learning for treatment and structural parameters*

The counterfactual, drawn (stylised)



Causal ML in climate policy analysis



From Nzivugira et. al *Causal effects of policy and occupant behavior on cooling energy*

OXFORD
JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

American Journal of Epidemiology, 2025, **194**, 185–194
https://doi.org/10.1093/aje/kwae171
Advance access publication date July 3, 2024
Practice of Epidemiology

A causal machine-learning framework for studying policy impact on air pollution: a case study in COVID-19 lockdowns

Claire Heffernan[1], Kirsten Koehler[2], Misti Levy Zamora[3], Colby Buehler[4], Drew R. Gentner[4], Roger D. Peng[5], Abhirup Datta[*,1]

[1]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, United States
[2]Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, United States
[3]Department of Public Health Sciences, School of Medicine, University of Connecticut, Farmington, CT 06030, United States
[4]Department of Chemical and Environmental Engineering, School of Engineering and Applied Science, Yale University, New Haven, CT 06520, United States
[5]Department of Statistics and Data Sciences, College of Natural Sciences, University of Texas, Austin, TX 78705, United States
[*]Corresponding author: Abhirup Datta, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Room E3640, Baltimore, MD 21205 (abhidatta@jhu.edu)

Abstract

When studying the impact of policy interventions or natural experiments on air pollution, such as new environmental policies or the opening or closing of an industrial facility, careful statistical analysis is needed to separate causal changes from other confounding factors. Using COVID-19 lockdowns as a case study, we present a comprehensive framework for estimating and validating causal changes from such perturbations. We propose using flexible machine learning–based comparative interrupted time series (CITS) models for estimating such a causal effect. We outline the assumptions required to identify causal effects, showing that many common methods rely on strong assumptions that are relaxed by machine learning models. For empirical validation, we also propose a simple diagnostic criterion, guarding against false effects in baseline years when there was no intervention. The framework is applied to study the impact of COVID-19 lockdowns on atmospheric nitrogen dioxide (NO_2) levels in the eastern United States. The machine learning approaches guard against false effects better than common methods and suggest decreases in NO_2 levels in 4 US cities (Boston, Massachusetts; New York, New York; Baltimore, Maryland; and Washington, DC) during the pandemic lockdowns. The study showcases the importance of our validation framework in selecting a suitable method and the utility of a machine learning–based CITS model for studying causal changes in air pollution time series.

This article is part of a Special Collection on Environmental Epidemiology.

Key words: machine learning; interrupted time series; causal inference; COVID-19; air pollution.

Introduction

In 2020, lockdowns and other restrictions were implemented worldwide due to the COVID-19 pandemic. This decreased traffic volume.[1] Traffic is a major source of certain pollutants, like nitrogen dioxide (NO_2).[2] Many studies have been published on the effects of COVID-19 lockdowns on ambient levels of air pollutants, including NO_2. Gkatzelis et al[3] reviewed over 200 such papers. A variety of statistical methods and analyses were used, including measurement of percent reduction[4-8] between the lockdown period and a reference period, comparison of concentrations in the same period over multiple years,[9,10] the difference-in-differences approach,[11,12] and regression models, including comparative interrupted time series (CITS)[13,14] or prediction of counterfactual concentrations.[15,16] Models of air pollution under different emission scenarios, including chemical transport models, were also used.[17-20]

These methods often suffer from some of the following limitations. First, meteorological conditions were not always considered in previous studies. COVID-19 lockdowns in the United States were generally imposed around March 2020, and many pollutant concentrations increase in winter and decrease in summer in parts of the Northern Hemisphere due to meteorological factors, including greater emissions from combustion-based heating sources during winter.[21,22] Not including weather conditions would confound the lockdown effect estimates.

Second, many researchers do not estimate a causal effect or rely on unstated and unrealistic assumptions. The COVID-19 lockdown constitutes an example of a "sharp" policy intervention with a potential impact on air pollution. Other examples of such policies or natural experiments include copper smelt strikes in the southwestern United States,[23,24] steel mill closures in the US state of Utah,[25] and emissions reductions during the 2008 Olympic Games in Beijing, China.[26] Conclusions about a decrease in air pollution have often been based on comparisons of average concentrations during and before the experiment. It is important to assess whether such common approaches provide a causal estimate.

Third, parametric forms such as linearity are often assumed, which, if inaccurate, would bias estimation despite the inclusion of meteorological data.

The contributions of this article are 2-fold. First, we propose a flexible machine learning–based CITS framework for studying

Received: May 14, 2023. **Accepted**: June 26, 2024

Two ways to price carbon

Carbon taxes

We fix the price, quantity adjusts

- Predictable costs for firms
- Uncertain total emissions

Cap & Trade (ETS)

We fix the quantity, price adjusts

- Certain cap on emissions
- Volatile prices
- Creates a market for permits

With both, markets create a ***measurement and verification problem:***

- Tonnes must be counted, verified and trusted, often including offsets from forests, farms and factories on the other side of the world.
- **This creates a verification problem that can be tackled using AI**

AI along the carbon-credit pipeline

Remote sensing

Forests, land use, agriculture, satellite image processing, methane plumes, industrial sites

→

Quantification

Biomass, fluxes estimation

→

Credit quality

Screening and rating credits

→

Integrity

Fraud detection, double-counting, market analysis

Limitations of AI in carbon markets

False precision in accounting

- A tonne "measured" by a ML model can carry compounding uncertainties: they cannot be treated as exact estimates

Gaming

- Measurable metrics can lead to gaming
- Carbon capture projects will optimise towards whatever the models reward

Tunnel vision

- Optimising the measurable CO2 tonne sides biodiversity, justice, governance co-benefits

Black-box verification

- Proprietary models cannot be audited
- Trust in carbon markets should not rest in private creditors

LLMs in the policy workflow

Policy datasets

Laws, reports, papers, drafts...

→

Retrieval and analysis

Find relevant documents, classify them, find patterns

→

LLM Interfacing

Chatbots to interact with policy databases.

Grounded anwers, with citations

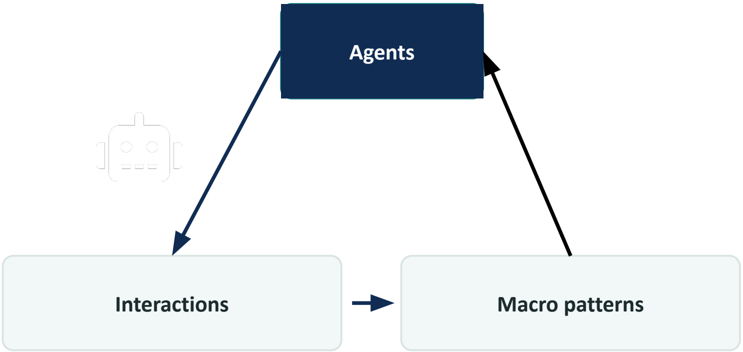
→

Human reviewer

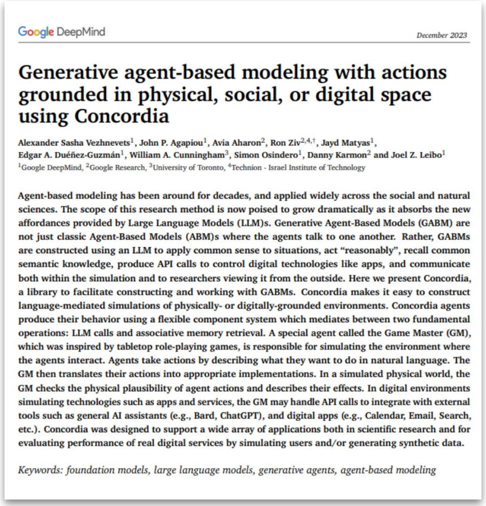
Human verification

- Policy document analysis
- Literature synthesis
- Q&A over policy corpora
- Analyses of large-scale text policy datasets

Agent-based models



- Heterogeneous actors: households, firms, lobbies
- Emergent macro dynamics from micro behaviour
- Policy acceptability explored before implementation, not discovered after
- ABMs can interact with predefined rules; but they can be modelled as LLM-interacting agents



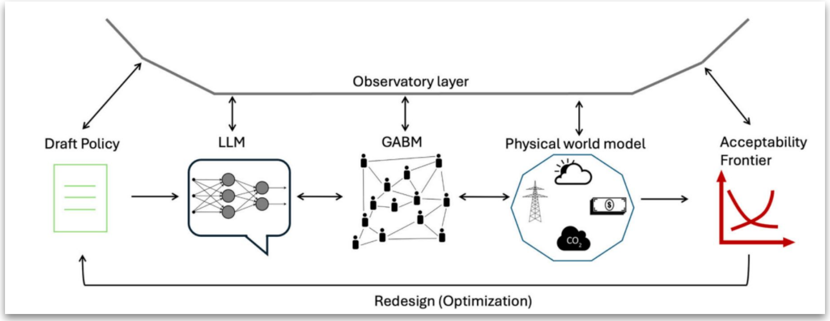
Google DeepMind — December 2023

Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia

Alexander Sasha Vezhnevets[1], John P. Agapiou[1], Avia Aharon[2], Ron Ziv[2,4,†], Jayd Matyas[1], Edgar A. Duéñez-Guzmán[1], William A. Cunningham[3], Simon Osindero[1], Danny Karmon[2] and Joel Z. Leibo[1]
[1]Google DeepMind, [2]Google Research, [3]University of Toronto, [4]Technion - Israel Institute of Technology

Agent-based modeling has been around for decades, and applied widely across the social and natural sciences. The scope of this research method is now poised to grow dramatically as it absorbs the new affordances provided by Large Language Models (LLM)s. Generative Agent-Based Models (GABM) are not just classic Agent-Based Models (ABM)s where the agents talk to one another. Rather, GABMs are constructed using an LLM to apply common sense to situations, act "reasonably", recall common semantic knowledge, produce API calls to control digital technologies like apps, and communicate both within the simulation and to researchers viewing it from the outside. Here we present Concordia, a library to facilitate constructing and working with GABMs. Concordia makes it easy to construct language-mediated simulations of physically- or digitally-grounded environments. Concordia agents produce their behavior using a flexible component system which mediates between two fundamental operations: LLM calls and associative memory retrieval. A special agent called the Game Master (GM), which was inspired by tabletop role-playing games, is responsible for simulating the environment where the agents interact. Agents take actions by describing what they want to do in natural language. The GM then translates their actions into appropriate implementations. In a simulated physical world, the GM checks the physical plausibility of agent actions and describes their effects. In digital environments simulating technologies such as apps and services, the GM may handle API calls to integrate with external tools such as general AI assistants (e.g., Bard, ChatGPT), and digital apps (e.g., Calendar, Email, Search, etc.). Concordia was designed to support a wide array of applications both in scientific research and for evaluating performance of real digital services by simulating users and/or generating synthetic data.

Keywords: foundation models, large language models, generative agents, agent-based modeling



From Manivannan, Spaiser et al. *Generative AI for climate governance and acceptability-constrained policy design*

LLMs & ABMs: Plausible, yet unvalidated

Hallucination

- LLMs tend to hallucinate information
- LLM chatbots are not usually capable of distinguishing whether they generated true or false information
- Policy work requires checkeable citations

Missing institutional context

- Models may not be aware of specific norms, precedents, mandates, red lines, evolving political contexts

Bias, confidentiality, overconfidence

- Models may be trained on biased datasets
- LLMs tend to sound certain of their outputs, even if they are not
- LLMs can leak sensitive data

Proprietary models

- Closed weights and shifting versions
- Reproducibility and sovereignty problems for public institutions
- Who owns the models? Who pays for them? Are private entities training models on critical public information?

AI's own environmental footprint

Compute-related

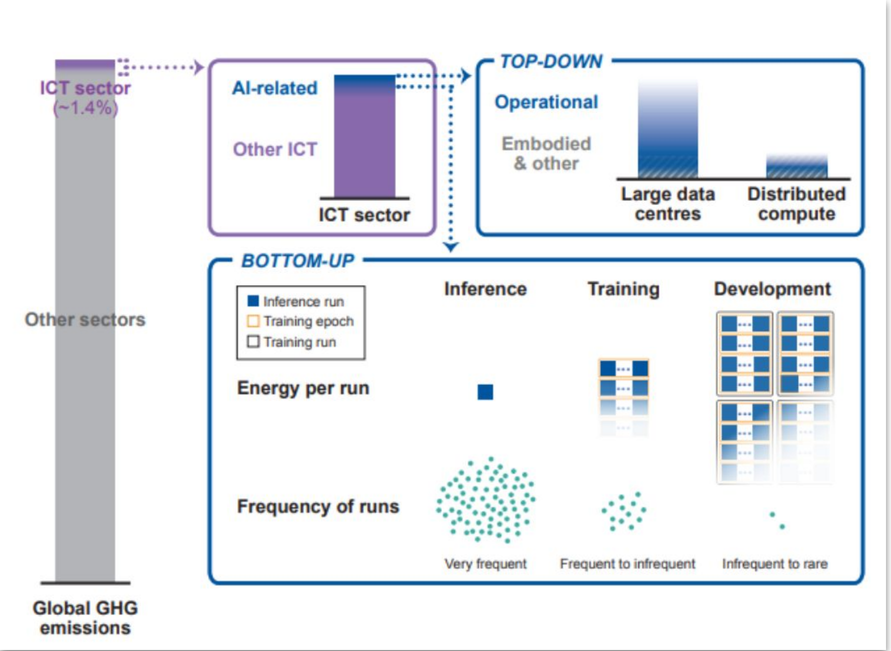
Training & inference energy, hardware, water

Effects of applying ML

Efficiency gains vs. rebound; who uses the tool, for what

System-level

Behaviour shifts, lock-ins, induced demand



From Kaack et al *Aligning artificial intelligence with climate change mitigation*

The deployment gap

Papers — thousands every year

↓

Pilots —a few survive contact with institutions

↓

Adopted tools:
very uncommon

Closing the gap:

- Benchmarks for policy usefulness, not just predictive accuracy
- Uncertainty well calculated and communicated in decision-relevant terms
- Embedded collaborations: co-design with the people and communities who will use it and could be affected by the technology
- Increase Institutional capacity and trust

See https://eartharxiv.org/repository/view/13019/

Open problems and takeaways

Open problems

1. Benchmarks for policy usefulness beyond accuracy
2. Robust policies under deep uncertainty, vs. a single 'optimal' one
3. Emulators & synthetic scenarios that keep interpretability and UQ
4. Safe, transparent LLM workflows for public institutions
5. Co-design methods with decision-makers and affected communities
6. Integrating hazards → impacts → social response → policy decisions
7. Localized damages: closing the resolution gap where data is thinnest

Integrated climate intelligence: A grand challenge

Every part of this lecture was one island: measurement, futures, uncertainty, agents, evaluation.

A key research agenda of the coming decade is to connect them: hazards, impacts, socioeconomic response and policy decisions in one loop; and to understand how AI itself affects climate action.

Stockholm Resilience Centre | Stockholm University

Research Edu

Stockholm Resilience Centre / Research / Synthesis reports / AI for a planet under pressure

AI for a planet under pressure

Editors: Victor Galaz & Maria Schewenius

Comment

https://doi.org/10.1038/s41558-026-02624-x

Artificial intelligence to support cross-disciplinary climate change research

Yang Ou, Carlos Rodriguez-Pardo, Alaa Al Khourdajie, Joeri Rogelj, Peijie Zhou, S. Karthik Mukkavilli, Massimo Tavoni, Leon Clarke, Yong Liu, Tengjiao Wang, Alice Oh & Haewon McJeon

Check for updates

Integrating knowledge across climate risks, societal responses and their interactions is a critical yet persistently challenging goal. We argue that advanced artificial intelligence frameworks, specifically foundation models, offer a new opportunity to unify these domains and support climate decision-making.

As human-caused climate change progresses and societies increasingly experience its harmful effects, research is providing insights into the physical, ecological, societal and economic consequences of climate change, and helping to identify potential responses[1]. Currently, such analysis often depends on complex multi-stage processes, where experts across physical climate modelling, climate impact assessment, and techno- and socio-economic analysis each contribute their part to a bigger picture. An example is the community-wide initiative producing global scenarios such as the IPCC's Representative Concentration Pathways and Shared Socioeconomic Pathways framework[2]. While this has enabled major advances, existing tools remain fragmented and resource-intensive, limiting timely and coherent policy support[2].

These long-standing limitations call for a renewed effort to pursue more agile and unified approaches enabled by recent advances in artificial intelligence (AI). First, fragmentation across climate physical modelling, impact assessment, and techno- and socio-economic analysis makes representing linked dynamics slow and disjointed. Second, while tools such as integrated assessment models (IAMs) and Earth system models (ESMs) are increasingly flexible within their own domains, inconsistent assumptions, diverging baselines and incompatible frameworks often complicate and delay cross-disciplines synthesis. Third, most workflows are not designed for rapid adaptation or interoperability across regions, sectors or decision types, which reduces responsiveness to new questions. Earlier initiatives such as the Platform for Regional Integrated Modeling and Analysis (PRIMA)[3] sought to bridge these divides, but were limited by rigid architectures and constrained computational capabilities. However, emerging AI tools allow us to revisit this integration challenge with new methods.

We argue that foundation models – that is, self-supervised systems trained on heterogeneous datasets and designed to generalize across a wide range of downstream applications – offer a promising path forward. By embedding a common logic and shared representation into a unified system architecture, such models would be capable of navigating diverse, policy-relevant tasks at accelerated speed. For example, advanced AI models like Aurora already demonstrate the ability to predict weather, air quality, ocean waves and cyclone tracks within a single unified system[4]. Unlike existing integrated tools, a foundation model embeds cross-domain knowledge into a shared, reusable architecture to enable faster iteration and more coherent decision support (see conceptual overview in Fig. 1). In this Comment, we outline the rationale for such a model, current modelling challenges, a technical framework for integration, and conclude with key opportunities and barriers to implementation.

Revisiting integration challenges in the age of AI

The challenge in building an integrated AI model for climate change analysis lies in how cross-domain structures could be translated into well-posed computational tasks. AI models typically require clearly defined inputs and outputs with consistent supervision, yet many climate-relevant applications involve exploratory or counterfactual reasoning that defies such rigid formulation, as they often lack the singular, verifiable outcomes required for model supervision. For instance, decision-makers often consider layered interventions that span fiscal policy, land-use regulation, adaptation efforts and innovation support. However, these are frequently distilled into stylized levers such as carbon pricing or technology support, limiting the ability to assess the potential efficacy of real-world policies. Notably, the literature rarely offers systematic evaluations of how specific instruments translate into intended outcomes[5], which in turn limits the historical data available to train predictive models of policy impact[6]. Translating these dynamic strategies into machine learning-friendly structures therefore raises core design questions: what is the model trying to predict? What constitutes 'ground truth' in a counterfactual world[7]? Should tasks simulate plausible transitions or push the boundaries and suggest innovative solutions? Should they quantify probabilistic outcomes, where likelihoods can be estimated, and compare strategies under deep uncertainty, where they cannot? These questions are amplified by sparse observations and the imperative for interpretability in high-consequence, societally embedded decisions.

AI is increasingly being leveraged across all three climate research domains, although often in isolation. In physical climate modelling, some efforts use machine learning to improve the resolution and speed of Earth system modelling[8], while others focus on enhancing the parametric representation of specific physical processes[9]. In the socio-economic domain, one recent proof-of-concept study[10] demonstrated how generative AI can replicate and expand mitigation scenarios from IAM data, while multi-agent reinforcement learning has been proposed to synthesize climate policy under uncertainty[11]. However, these applications are not typically designed to embed cross-discipline feedback. Physical emulators rarely account for human actions, while socio-economic models often use simplified climate inputs. A key next step, therefore, is to extend this paradigm

arXiv:2606.13704v1 [cs.CY] 9 Jun 2026

Position: AI Must Become Planet-Centered, Not Just Human-Centered

María Pérez-Ortiz [1]

Abstract

This position paper argues that contemporary AI paradigms are insufficient for supporting complex global goals and introduces Planet-Centered AI (PCAI) as a design philosophy and research agenda that reorients AI toward planetary-scale socio-ecological systems and their long-term trajectories. A planet-centered approach is grounded in systems thinking, treating Earth as an interconnected whole of which humans are part. We diagnose recurring limitations across AI frameworks—many of which remain human-centered—and show why these become especially consequential under current planetary conditions characterized by systemic risk, non-stationarity, and deep uncertainty. We then articulate how PCAI reshapes the AI lifecycle, from problem formulation and model design to evaluation and deployment, by emphasizing alignment with global agendas, developing system-aware AI foundations, trajectory-oriented evaluation, and monitorability. Finally, we advance a falsifiable claim: AI systems optimized without explicit consideration of systemic consequences are more likely to exacerbate systemic instability than to mitigate it.

1. Introduction

Over the past decade, the AI community has developed a range of paradigms to address the ethical, social, and technical risks of AI. Frameworks such as Human-Centered AI, Responsible AI, AI for Social Good/Sustainability, and AI safety have been essential in establishing that AI has far-reaching consequences, and that potential harms and broader impacts should inform algorithmic design and deployment.

Despite this progress, **this position paper argues that these paradigms are insufficient for enabling AI to meaningfully support societies in confronting complex challenges: ML must be reoriented toward a planet-centered paradigm that treats systemic risk, long-term impact, and global goals as first-order design objectives.**

As the world enters what is described as a polycrisis (Lawrence et al., 2024), risks arise not from isolated failures but from coupled systems whose interactions generate self-reinforcing and systemic dynamics. Consider the following example: rising ocean temperatures alter predator-prey dynamics, enabling jellyfish blooms that clog coastal power plant cooling intakes, triggering forced shutdowns and energy instability (Purcell et al., 2007). This is systemic risk, risks emerge from the coupling between systems. Similarly, climate change has been shown to aggravate more than half of known infectious diseases, through over a thousand distinct pathways linking physical, ecological, and social systems (Mora et al., 2022). In such settings, feedback loops, nonlinear interactions, and path dependence—where early interventions shape and constrain future outcomes—can amplify risks and lock societies into trajectories that are difficult to reverse (Delannoy et al., 2025; Steffen et al., 2018). Recent work shows that AI is increasingly entangled with polycrisis dynamics through material pathways (e.g., energy use, resource extraction, infrastructure lock-in) and informational pathways (e.g., shaping behavior, accelerating decision cycles, synchronizing systems), with the potential to intensify systemic instability (Creutzig et al., 2022).

It is precisely this systems perspective (Meadows, 2008)—focused on feedbacks, interactions, and trajectories—that remains absent from AI frameworks (Kondor et al., 2024). We argue that AI methods are poorly suited to supporting planetary challenges, which exhibit properties of "wicked problems" (Rittel & Webber, 1973): e.g. non-stationarity and feedback-driven dynamics. This mismatch matters because such conditions increasingly characterize high-stakes domains in which AI is deployed, such as climate governance, technology regulation, and public policy (Ilcic et al., 2025). Due to this misalignment, AI can generate systemic risks by interacting with or amplifying underlying system dynamics in unintended ways (Schön et al., 2025). Yet frameworks for anticipating and evaluating systemic effects remain limited, leaving concepts of systemic risk underdeveloped and inconsistently operationalized in AI governance (Carey, 2025; Stahl et al., 2023).

[1]Centre for AI, Department of Computer Science, University College London, London, UK. Correspondence to: María Pérez-Ortiz <maria.perez@ucl.ac.uk>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026.

The starter toolbox

DATA & TOOLS

- Climate Policy Radar
- IAMC scenario databases
- IAM emulators (e.g. ML-IAM) & open-source IAMs / energy models
- Population Dynamics Foundation Model
- Geospatial embeddings & TorchGeo
- Causal inference & econometrics packages

COMMUNITIES & VENUES

- Climate Change AI — you’re already here
- CCAI NeurIPS workshops
- IAMC — Responsible AI for IAMs working group
- EGU · AGU · SISC · WCERE: where climate, earth science and economics meet

Takeaway messages

AI can enter every stage of the policy cycle

- Measurement
- Futures
- Evaluation
- Workflows
- Verification

The constraint is rarely predictive accuracy

- Causal validity
- Values
- Trust
- Societal constraints
- Institutions
- Social acceptability

Integration is the frontier

- Across scales
- Across data types
- Across stages in the policy cycle
- Across institutions, communities
- Across the modelling chain

Start from policy-relevant questions and the problems of policymakers, not model capabilities or from what can be done with a particular dataset

Thank you for your attention!