Artificial Intelligence Policy

Luis Videgaray

Sir Winston Scott Memorial Lecture

November 25, 2024

MIT
MANAGEMENT
SLOAN SCHOOL

2017: Mexican AI policy initiative at the UN

What is AI Policy?

- No formal/consensus definition available

- We can think of AI policy as **public policy intended to modulate the impact of AI in the world**
  - Maximize benefits
  - Minimize risks and costs

Who does AI Policy?

- National governments
  - Executive
  - Legislative
  - Judicial
- State and local governments
- Supra-national entities (the European Union)
- Multilateral organizations (WTO, WB, OECD, UNESCO, etc)

*In addition, non-governmental entities can have significant **influence** on AI policy*

What are the tools of AI policy?

1) Government Investment
   - R&D
   - Education
   - Infrastructure
   - Incentives (subsidies, tax breaks)
   - Governmental venture capital

2) Use of AI in government
   - AI for the delivery of public services
   - Decision making in government
   - Military use of AI

3) Laws and regulations

How to think of AI Policy in Barbados?

1) AI is, in many ways, a global equalizer

How to seize this moment?

2) The AI value chain is global

How to benefit from Barbados's position in it?

Four camps of debate on AI policy

"AI Ethics"

Privacy
Bias and fairness
Explainability
Manipulation, fakes

"AI Safety"

Alignment
Extinction level risk

Economic Impact of AI

Jobs
Inequality
Concentration
Stability

Geopolitics of AI

Strategic advantage
Security
Military

National AI Strategies

- The first national strategies were published in 2017
  - China, Canada, Japan, Finland and the UAE
- The US published its AI strategy in 2019;
- By now, more than 50 countries (or regions) have released AI strategies
- AI national strategies vary widely, each focusing on different aspects of AI policy.
  - Non-binding, mostly high-level documents
  - Establish priorities and goals
  - Describe intended activities

AI Principles

- Increasing calls for policy guardrails to ensure trustworthy and socially responsible AI.
- Most agree that policies should be defined through inclusive and interdisciplinary dialogue
- A multitude of “AI principles” documents have been published over the past five years
  - *Typically lists of* ***desired technical and socio-technical attributes*** *of AI systems*

Example: OECD/G20 AI Principles

1. Inclusive growth, sustainable development and well-being
2. Human-centered values: rule of law, human rights and democratic values
    i. Freedom
    ii. Dignity and autonomy
    iii. Privacy and data protection
    iv. Fairness; non-discrimination and equality
    v. Diversity
    vi. Social justice; labor rights
3. Transparency and explainability
4. Robustness, security and safety
5. Accountability

Key design choices in AI laws

- **Voluntary vs. binding rules** *(do we need AI laws?*)
- **Horizontal vs vertical rules** (*do we need one law or many laws?*)
- **New or existing regulators ?**
- **Regulating inputs or outcomes ?**

China: Provisions on Recommendation Algorithms

- Registry: online database for algorithms with social impact
    - Requires developers to submit training data and complete a security self-assessment
    - Filing becomes partially public upon successful registration
- User Rights vs recommendation algorithms include:
    - Turn off recommendations.
    - Delete personalization tags.
    - Receive an explanation for significant algorithmic impacts.
- Addresses some economic impacts of algorithms
    - Limits on algorithms creating exploitative work conditions for delivery workers
    - Provisions against anti-competitive practices and excessive price discrimination.
- AI providers should “uphold mainstream value orientations” and “actively transmit positive energy”
- *Subsequent laws focus on “deep synthesis” and generative AI tools*

European AI Act

Draft introduced in April 2021. Voted in Parliament March 2024.

- Cross-sectional (horizontal) law
- Emphasis on inputs (e.g. quality of training data)
- Four risk tiers: unacceptable, high, limited, and minimal
  - Unacceptable risks, such as social scoring by governments, are prohibited
  - High-risk AI systems be subject to rigorous regulations
- Pre-market self-assessments required for high-risk systems
- Creates new national AI regulators and a European AI Board and an AI Office
- Non-compliance could result in fines up to 6% of global revenue
- Sandboxes for innovation of SMEs
- *Recently added rules for "general-purpose AI", focusing on "systemic" risks and with exceptions for open-source models*

Brazil's Senate AI Draft Law

- Requires risk assessment before ***any*** AI system can enter the market
- Enumerates high-risk AI applications, including critical infrastructure, healthcare, autonomous vehicles, and ***credit scoring***.
  - Establishes a publicly accessible database for high-risk AI systems
  - Strict liability
- Individual rights to:
  - ***Information***
  - ***Explanation***
  - ***Challenge AI decisions***
  - ***Human intervention***
- Prohibits subliminal manipulation AI systems, social scoring by authorities, and certain biometric uses
- Sandboxes for innovation

US: California SB 1047 (failed)

Introduced by Senator Scott Wiener, aimed to regulate the development of advanced AI models

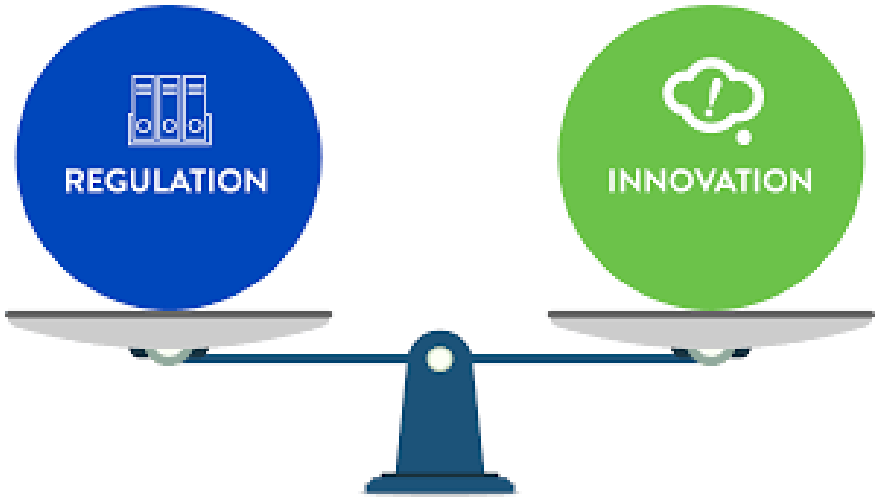
- Targeted large "frontier" AI models with computing power greater than 10^26 floating-point operations and training costs over $100 million.
- Required developers to implement safety measures like pre-deployment testing, cybersecurity safeguards, and post-deployment monitoring.
- Created whistleblower protections for employees of frontier AI laboratories.
- Established transparent pricing and prohibits price discrimination.

India’s shifting position

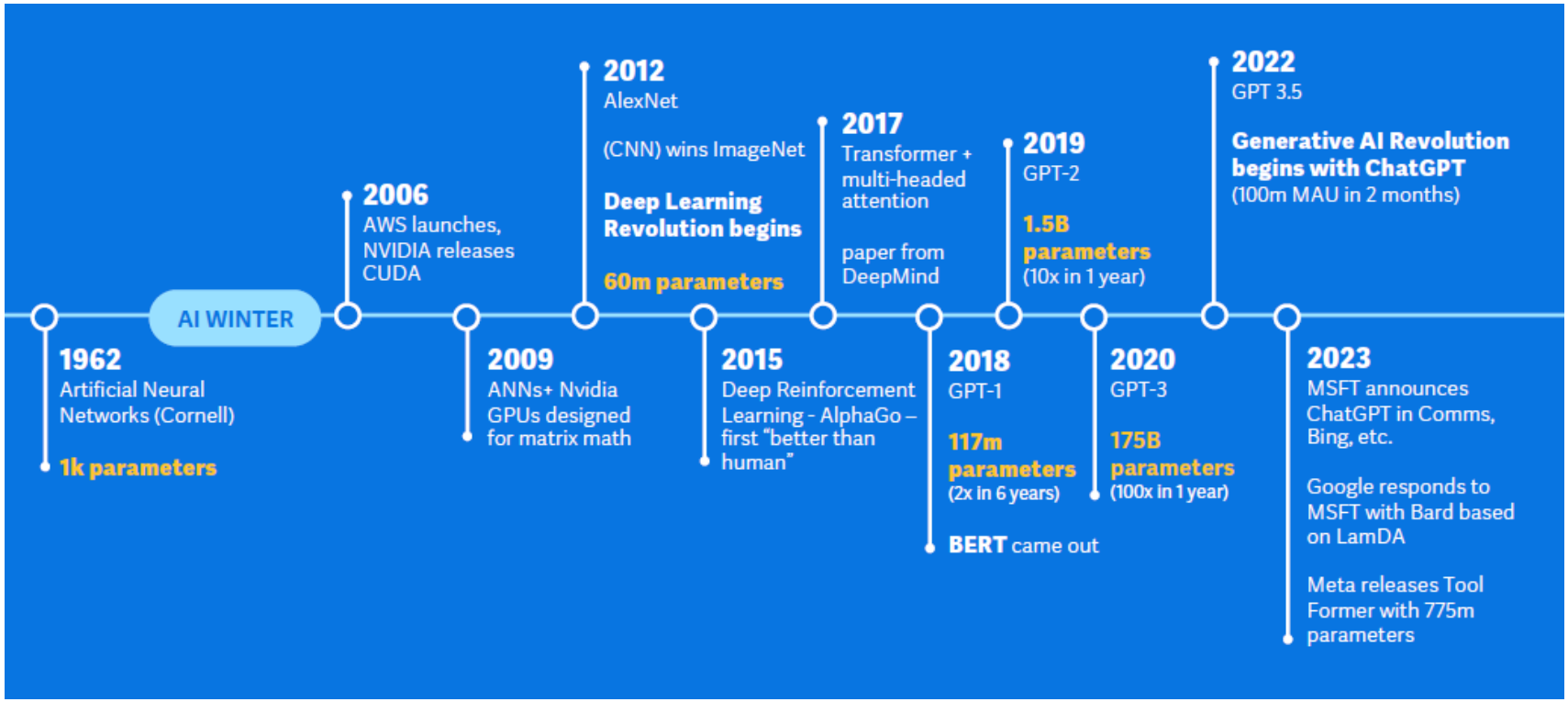
- In April 2023, it had informed Parliament that it was not eyeing any legislation to regulate AI.
- In March 2024, it required platforms to obtain explicit permission before implementing "unreliable" AI models or tools for Indian users.
- IT Secretary S Krishnan stated that the government will attempt to regulate AI, but not in a way that could stifle innovation.
    - He emphasized striking a balance between protecting vital interests and fostering innovation

Why is it so hard to regulate AI?

Policy dilemmas



Fast innovation



General Purpose Technology

The story of the battery

Invented:

1800 (Alessandro Volta)

Understood:

1860 (Maxwell equations)

AI Policy
in Barbados

How to think of AI Policy in Barbados?

1) AI is, in many ways, a global equalizer

How to seize this moment?

2) The AI value chain is global

How to benefit from Barbados's position in it?

AI Policy Preparedness Recommendations

1. **Adopt a proactive but flexible policy stance**
   - Preserve policy optionality
2. **Acknowledge and Understand Policy Trade-offs**
   - Technical and incentive trade-offs
3. **Rapidly Augment AI Expertise in Government**
   - Promote AI literacy in government, invest in skills development, and build a network of external advisors to keep pace with AI advancements.
4. **Continuously Explore New AI Capabilities**
   - Experiment with AI to improve public sector operations across the board.
5. **Develop a Procurement Framework for AI Tools**
   - Implement cost, risk, and benefit analyses while ensuring sustainability and ethical standards in AI procurement.

AI Policy Preparedness Recommendations (cont.)

6. **Adopt Modern Data Management Tools**
   - Explore cloud integrations for structured and unstructured data.
7. **Prepare for workforce disruption**
   - Reskilling, adaptability and evaluate income transfers
8. **Enhance Cross-border Cooperation**
   - Prioritize information exchange.
9. **Cybersecurity risk: act now**
   - AI is rapidly enhancing cybersecurity threats. Governments can’t afford to stay behind.

The AI supply chain is global

- **Upstream**: Large models are base or "foundation" models
  - Base models are (increasingly) costly and difficult
  - Few upstream providers
  - Provide "world knowledge"

Training infrastructure layer
(Power, compute, chips)

Foundation model layer
(US: OpenAI, Anthropic, Google, Meta / China)

- **Downstream**: Smaller, domain-specific or task specific AI models
  - "Connect" to the upstream via fine-tuning, prompting, etc
  - Downstream AI is becoming easier, much less expensive and faster
  - Expect an explosion of downstream AI in start-ups and, corporations and even governments

Inference infrastructure layer
(Power, compute, chips)

Application layer
Consumer
Enterprise
Government