

Computational Diplomacy

How Computation and Diplomacy Make Each Other

Prof. Roland Bouffanaïs[†]

[†]Global Studies Institute & Dept. of Computer Science, University of Geneva
Chair in Computational Diplomacy – Co-Director, SiDLab

Based on a chapter co-authored with Chloé Marteil, Daniela Scherer &
Didier Wernli

Plenary talk, CED-2026 – Simon Kuznets Kharkiv National University
of Economics

29 October 2026

Why Computational Diplomacy at a Kuznets Conference?

- Simon Kuznets made the economy **measurable**: national accounts turned growth from a matter of opinion into an **empirical** question (Nobel 1971)
 - Diplomacy is one of the last large social practices that has **resisted measurement**: for want of data, not of substance
 - Meanwhile, computation (loosely: “AI”) reshapes the **arenas**, the **means**, and the **stakes** of international politics: and becomes a stake itself
- ↔ This talk: a framework for studying a system **from inside it**

Two lenses, 15 minutes

Method Using computation to study diplomacy

Object Computation as a stake of diplomacy

Kuznets' own caution (1934)

“The welfare of a nation can scarcely be inferred from a measurement of national income.”

↔ Measure, but know what the measure leaves out. Keep this in mind for “AI” indicators too.

The Central Message

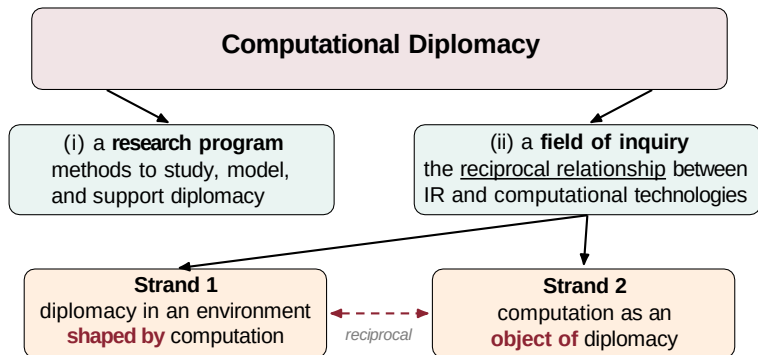
- Computation matters in geopolitics **not only** for its intrinsic added value. . .
- . . . but for the **material consequences of the expectations** surrounding it:
 - new actors emerge, others gain new importance
 - many are forced to act on anticipations (fear of missing out)
- Some of these technologies now reach maturity in **military applications** and in the **management of online information flows**

Consequence

The relationship between computation and diplomacy must be **reconsidered**

→ We propose the framework of **Computational Diplomacy**

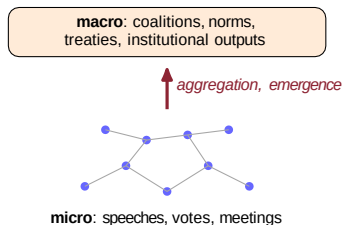
Computational Diplomacy: The Framework



↔ The two strands ultimately **converge**: the race narratives that make AI an object of diplomacy can themselves be studied computationally

Diplomacy Leaves Traces

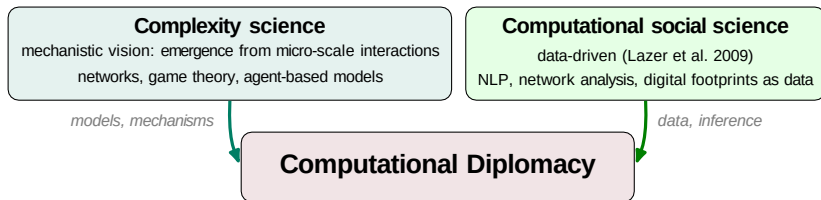
- Studying diplomacy = studying **practices**, not just outcomes (Bjola & Kornprobst); classical approaches struggle to generalize without flattening complexity
- Yet diplomacy produces abundant **observable traces**: speeches, votes, resolutions, treaties, meeting records, communiqués, institutional archives
- Interests are not pre-existing: they are **formed and transformed** through repeated interactions
 ↔ **micro-interactions are the basis of macro-outcomes**



The Kuznets move, for diplomacy

Methodological advances + growing availability of diplomatic data '→ studying diplomacy **at scale** has become feasible. Institutions are not neutral arenas: compare their design with their practice.

Two Scientific Heritages, and an Economist in Between



- “More is different” (Anderson, 1972): collective phenomena cannot always be reduced to the properties of individual components ↔ econophysics, **complexity economics**, cliodynamics
- **Schelling** (Nobel 2005), segregation model (1969): macro segregation from mild micro preferences
↔ **micro rules** ≠ **macro intentions**

The lesson for diplomacy

Collective outcomes (coalitions, escalations, races) cannot be read off individual motives: model and measure the interactions.

Three Research Themes: Conflicts, Institutions, Texts

1. Conflicts

Quantifiable, pervasive.

L. F. Richardson

(1881–1953), Quaker pacifist, modeled arms races:

$$\dot{x} = ky - ax + g$$

$$\dot{y} = lx - \beta y + h$$

→ Correlates of War, Uppsala; game theory at RAND; **conflict forecasting** (VIEWS)

2. Institutions

Conflicts **test** institutions and reveal design flaws (League of Nations).

Borrowed from AI safety: **institutional alignment**: alignment in means (rules, votes) does not guarantee alignment in outputs

3. Texts

“AI/IR” is older than you think: Hudson (ed.), 1991. Today, debates, treaties, letters become **computable material**; **Cicero** (Meta, 2022) plays Diplomacy at human level

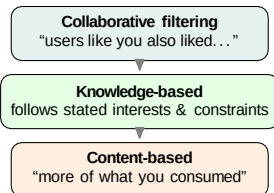
A warning before integration (Rivera et al., 2024)

In simulated wargames, off-the-shelf LLMs displayed **escalatory behavior**, up to rare simulated nuclear use. The reciprocity already: diplomatic setups become benchmarks for AI, and AI is considered for foreign-policy and military use.

Digital Diplomacy and the Landlords of the Arena

- Online platforms are now **institutionalized** spaces of diplomacy: dedicated units, strategies, mandates (Eggeling)
- Visibility no longer follows from status or authority: **the medium disciplines the message**: visually simple, emotionally framed, **computationally personalized** (Bjola & Manor)
- The arena is run by **private actors**: recommendation systems (narrow AI, deployed at scale for years) optimize **engagement**, not information quality

Three families (Aggarwal, 2016)



A complexity-science lesson

Virality is unpredictable and skewed; optimizing for reach can **erode public trust** (Narayanan). Steering such systems without reinforcing the dynamics they aim to mitigate is hard: feedback loops everywhere.

Manipulation Is Old; Computation Changes Its Economics

- **1870, the Ems Dispatch:** Bismarck edits a telegram, the mass press amplifies it
↔ Franco-Prussian War. One altered message + the new media infrastructure of the day
- **China's “50 Cent Party”** (King, Pan & Roberts, 2017): est. **448 million** messages/year, aimed not at arguing but at **strategic diversion**; known because of computational analysis of leaked archives
- What changes with computation: **cost, scale, speed, experimentation**: “lie machines” (Howard) grounded in quantitative models of circulation

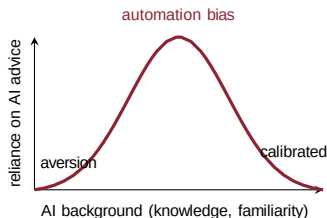


Measure before you moralize

Manipulation is **not new**, and the evidence is subtler than the rhetoric: echo chambers are less widespread than assumed (demand, not supply); GenAI in the 2024 elections mostly served ordinary campaigning. Yet the rhetoric about deepfakes has effects of its own.

Computation Reshapes Decision-Making

- Foundation models as **information gatekeepers**: $\approx 24\%$ of consumer ChatGPT messages are information retrieval (2025); $>800\text{M}$ weekly users
- Prediction was always part of governance, but today's predictors **own the infrastructure** that collects the data (Veliz) $\leftrightarrow$ self-fulfilling prophecies
- **Automation bias** vs. **algorithm aversion** (Horowitz & Kahn): a little knowledge is the dangerous amount; AI assistance to drone operators shifts attack decisions and dilutes responsibility (Salatino et al.)



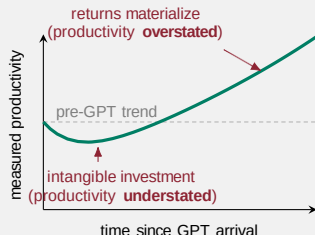
And when machines meet machines

Algorithmic pricing: learning algorithms converge to **supra-competitive prices without any explicit agreement** (Calvano et al., 2020). By analogy, interacting computational systems in diplomacy may produce unexpected **collective** outcomes.

AI as an Anticipatory Object: A General Purpose Technology?

- GPT concept (Bresnahan & Trajtenberg 1992): innovations generating **economy-wide** transformation: **pervasiveness, continuous improvement, innovation complementarities**
- Why it matters for diplomacy: a GPT restructures **several domains at once** (production, military, surveillance, governance), and its gains are **not immediately visible** ↔ incentives to invest early, without returns
- Anticipations ↔ investment: global private AI investment 2025: **\$344.7B** (+127.5%), GenAI alone \$170.9B (AI Index 2026)

The productivity J-curve (Brynjolfsson et al.)

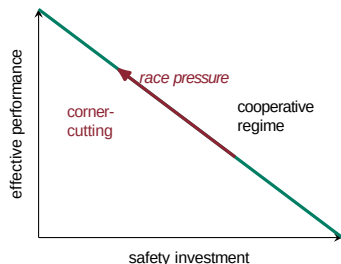


Kuznets would recognize this

The J-curve is about what national accounts miss (intangible capital): expectations run ahead of measured output, and policy runs on them.

Race vs. Competition: What Is the Difference?

- A **race** differs from competition in the belief that **falling behind has consequences of a different scale**
- Drivers of race beliefs:
 - the technology can change the **balance of power**
 - a gap would be **hard to close**
 - deep uncertainty about rivals' intentions
↔ **security dilemma**
- Emery-Xu, Park & Trager: racing creates a **performance–safety trade-off**; actors may cut safety to win, raising the probability of catastrophe
- Crucially, this depends on the **information structure**: private information ↔ excessive risk; public information ↔ depends on how close the race is

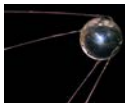


Stylized: race pressure pushes actors down the safety axis

Races Are Made, Not Found: Two Stories

Sputnik, 1957: the space race was not inevitable

- Initial US reactions were calm, even positive; until the early 1960s a **joint US-USSR Moon program** remained possible (Kennedy defended it at the UN, 1963)
- Democratic leaders **deliberately reinterpreted** Sputnik as a security crisis; the media amplified; the framing then shaped historical memory



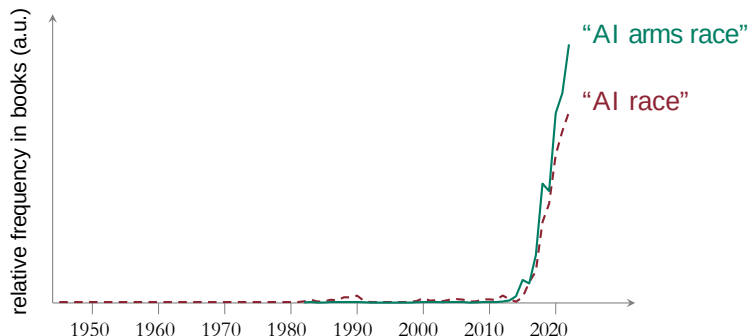
Japan's Fifth Generation, 1982: the world's response

- A national AI plan (>10 years, >\$400M); applications stayed at demonstration stage
- Responses: US DARPA Strategic Computing (\$1B+) and the MCC consortium; UK Alvey (£350M); Europe ESPRIT (>£5.5B)

Sound familiar?

A rival's national AI plan '→ crash programs, consortia, export restrictions, decoupling. The choreography of the 1980s is that of the 2020s: China in place of Japan, compute in place of mainframes. Technology and rivalry **co-produce** each other.

The Rise of the “Race” Language



Google Books Ngram: a small bump in the 1980s (Japan!), take-off after **2016–2017**. Three roots: the AI-risk community (Shulman 2009: “no alternative to a race” can be **self-fulfilling**); autonomous-weapons debates (2014); the US–China rivalry after China’s 2017 New Generation AI Plan. Now governmental rhetoric: “Winning the Race: America’s AI Action Plan” (White House, July 2025).

→ **The narrative is itself a measurable object**: this is where the two strands converge.

US–China: Two Models of AI Development

	China	United States
Approach	state-led capitalism; industrial strategy (MIC 2025, BRI)	market-driven, shifting to more active policy
Champions	Tencent, Huawei, Alibaba, Lenovo	public–private partnerships, market incentives
Style	planning of entire sectors	reactive, selective, threat-driven
Alliances	influence via infrastructure exports	alliances for alternative supply chains
Kind of AI	embodied: robots, factories, vehicles	frontier models + ~70% of global compute

- From engagement (1972–2017) to rivalry: technology is now **the core** of the rivalry, not a side effect of a trade race as with Japan ↔ 2018 trade war; 2022 CHIPS Act; 2025 “Winning the Race”
- Convergences: security-driven motivation, self-sufficiency, standards battles. New battlefield: **open-source AI**: China’s open-weight releases as soft power, the US 2025 Action Plan embracing open source in response

Story: Open Weights at the Negotiating Table

The New York Times
DealBook
With Andrew Ross Sorkin

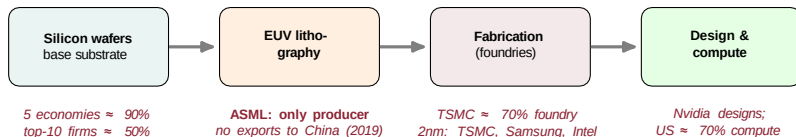
July 21, 2026

Good morning. Andrew here. Silicon Valley is locked in a heated debate over new open-weight A.I. models out of China, with critics claiming they were “distilled” — read: stolen — from U.S. counterparts. Against this backdrop, the Trump administration and Chinese officials are expected to discuss A.I. regulation ahead of President Xi’s meeting with Trump on Sept. 24.

The big question: Will Washington move to block Chinese open-weight models from the U.S.? And will this become a new chess piece in trade negotiations?

- July 2026: Chinese open-weight models accused of being “**distilled**” from US frontier models ↔ soft power meets IP conflict; AI regulation enters the **trade agenda** ahead of the Trump–Xi summit
- September 2026, Washington: the summit extends the trade truce and launches a bilateral **AI dialogue** (next round in November) ↔ AI as a **chess piece in negotiation**: Strand 2 in real time

The Material Base, and the Regulatory Loop



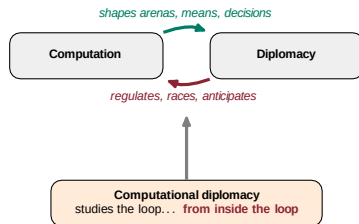
- Taiwan's “**Silicon Shield**”; the \$52B CHIPS Act + export restrictions; restrictions breed countermeasures (talent recruitment, espionage, a claimed Chinese EUV prototype, 2025)
- Regulation is a **posture** amid power struggles: a regulatory race with first-mover advantages; deregulation is equally political (Nelson)

The loop closes (Emery-Xu, Jordan & Trager)

Expect existential risk '→ early restrictions;
expect to win '→ weak governance;
intervene at **material chokepoints**.
Expectations '→ investment '→ race
rhetoric '→ regulation '→ the technology's
future '→ new expectations. **Anticipation
is the engine of the whole system.**

Conclusion: The Two Strands Converge

- **As a method:** complexity science & computational social science to study practices, institutions, conflicts **at scale:** doing for diplomacy what national accounting did for the economy
- **As a field:** a diplomatic environment **mediated by computation**, and computation as an **object of diplomacy**, charged with anticipations that shape investment, race rhetoric, regulation
- **Convergence:** races are made, not found; the narratives that make them **can be measured and modeled**



One sentence to take home

Computation and diplomacy now **co-produce** each other
↔ The expectations surrounding AI are as consequential as its capabilities: and expectations, too, can be measured.

Thank you!

Questions & discussion

Roland Bouffanais
Global Studies Institute & Department of Computer Science, University of
Geneva

roland.bouffanais@unige.ch

Chapter: Marteil, Scherer, Wernli & Bouffanais, *Computational Diplomacy: Method and Field of Inquiry*,
in *Handbook of Public Governance and Artificial Intelligence* (De Gruyter, forthcoming)