#62 SUMMER 2026

Journal

The Capco Institute Journal of Financial Transformation

The Capco Institute Journal of Financial Transformation

Journal

Recipient of the Apex Award for Publication Excellence

Table of

contents

Colliding worlds

Monetary machinations

Strategic pivots

Private markets

2026, Edition 62

Journal

Welcome to the 62nd edition of the Capco Journal of Financial Transformation, the second to be published in partnership with King's College London, a world-renowned leader in education and research.

This year marks the Journal's 25th anniversary and the beginning of **Capco30**, our strategic vision as we build towards Capco's landmark 30th anniversary.

At the heart of Capco30 is The Expert Advantage: our commitment to be the trusted partner clients turn to when the stakes are highest. We are proud to have built an exceptional business by obsessing about our deep expertise, our strong client relationships, our innovative culture and our reputation for impact.

The Journal has embodied that ambition since its launch in 2001. Over the past quarter century, it has brought together leading academics and industry practitioners – including Nobel laureates, international regulators and board members of major institutions – to examine the forces shaping financial services and the wider global economy. We are equally proud to see Capco experts sharing their perspectives alongside such distinguished contributors.

As Professor Crawford Spence, our editor from King's College, observes in his introduction, this edition might aptly have been titled 'how to survive and thrive in turbulent times'.

After decades of globalization, firms are both refocusing on local market shifts and new forms of cross-border collaboration. Technology is driving integration as trade barriers rise; financial innovation is accelerating amid monetary realignment and volatile geopolitics; and the green transition is redrawing both energy markets and fiscal priorities.

Edition 62 focuses on the **new world order** emerging from these shifts – and how leaders across financial services and energy are rethinking established assumptions, navigating new risks and identifying opportunities for growth.

The coming decade could bring one of the most consequential reconfigurations of global business and economic systems in generations.

This environment reinforces the importance of expertise, judgement and the ability to turn insight into action. At Capco, we help clients connect vision models and strengthening resilience to reimagining customer experiences and the future of work.

We bring together industry depth, technology and data capabilities, and disciplined execution to help clients navigate the moments that matter most with confidence.

The Journal is one expression of that Expert Advantage: convening leading perspectives, challenging established thinking and helping shape the agendas that will define the future of our industries.

I hope you will find the articles in this edition both thought-provoking and valuable for informing your organization's future vision and strategic decisions.

My sincere thanks and appreciation to our contributors, readers and clients.

Annie Rowland, Capco CEO

2026, Edition 62

Editor's note

KING'S College LONDON

KING'S BUSINESS SCHOOL

We could equally have labelled this volume of the Journal of Financial Transformation 'how to survive and thrive in turbulent times', given that each of the articles effectively attempts to tackle this issue in their own unique way. Times are indeed turbulent. Recent months and years have borne witness to increased military conflict, heightened economic shocks and the emergence of new political forces.

These disruptions are often captured in popular discourse by reference to, variously, the decline of the rules based international order, the rise of authoritarianism and the re-emergence of nationalism and protectionism as challenges to globalization. While putting clear conceptual parameters around what is a very unsettled geopolitical period is difficult, there is at the very least little doubt that the world is changing and that a new order is emerging. To paraphrase an old political aphorism: the old is dying and the new is struggling to be born.

This geopolitical reshuffling is the backdrop to the current issue of the Journal. Contributors were asked to think about how financial services and energy might respond to, and indeed themselves help reshape, the new global order that is emerging. What they have presented here can be grouped into four conceptual clusters, each of which makes up a distinct sub-theme in this issue of the journal: colliding worlds; monetary risks; strategic pivots; and private markets.

The first group of articles we refer to here as colliding worlds. The worlds that collide here specifically are those of finance and energy. Each of these forces us to zoom out from financial transactions and investments and think about the wider environmental and material basis for these. For example, Ragland's fascinating look at the arms race for computing power shows how the rapid expansion of data centers is at odds with the way in which existing energy infrastructure is built. Those who will be successful in rolling out advanced computing power are those who will be able to finance and build out alternative energy infrastructure at the same time.

This has strong parallels with Xu and Preda's article looking at the infrastructure of bitcoin mining, which demonstrates in some detail how supposedly digital assets have heavy materiality in terms of both the physical infrastructure that supports them and the energy consumption that goes into their mining and processing. Finally, Savignon takes the energy transition head on by exploring the mechanics and possibilities of carbon credits in Brazil, a country which is characterised by a dominance of renewables in its energy mix. Showing real heterogeneity in how carbon credits are deployed in different sectors in Brazil, this article concludes by outlining how those who are able to integrate thinking around finance, data and energy will be able benefit from and contribute to the energy transition in the Brazilian market.

Monetary risks refer to the various ways in which currencies are embroiled with risk at both the national and transnational levels. As the fundamental means via which deals are transacted, it is perhaps no surprise that we find money, in its various guises, to

have so many political inflections. Cryptoassets are often heralded as some kind of safe haven ('digital gold') during times of economic and political turmoil and there are many central banks across the globe who are now pursuing digital currency projects of their own. These have significantly lagged the development of private crypto initiatives. One of the reasons for this might be the perceived financial crime risk that is associated with digital currencies in general, something which Tjeransen's article outlines in detail.

Fiat currencies are also deeply imbricated in geopolitical factors. This is the subject of Hou et al's article which draws attention to the ways in which currency risk premia are associated with trade imbalances between nations. Such factors are of increasing importance to, not only forex traders, but any organization engaging in currency hedging. Finally, nowhere is the importance of currency to geopolitics more evident than in Zayim's article looking at how central banks engage in currency swaps with other central banks. Focusing on both China and the US, Zayim shows how each has sought to build up spheres of influence with different counterparties as part of larger attempts to accrue power and minimize geopolitical risk for themselves.

Our third conceptual 'bucket' looks at how individual organizations respond to geopolitical shifts. In other words, these articles look at the various strategic pivots that organizations make in order survive and thrive in turbulent times. These strategic pivots can take different forms depending on the geography and industry in question. For example, Schnyder shows how banks in East Central Europe can mitigate risk in response to rising anti-FDI policies in that region. In Scriven's article the author shows how software delivery needs to be reconfigured in order to overcome increasing regulatory divergence and export controls.

In times of geopolitical uncertainty, capital markets can become less efficient in allocating capital and so firms have to focus more on their internal efficiencies, which is what Gkikopoulos and Stathopoulos illustrate in their article exploring the relationship between firms and banks; such relationships have implications for tax planning and productivity. Political turbulence also impacts hiring strategies, which Nguyen and Sila demonstrate in relation to corporate hiring practices during times of election uncertainty. Finally, Confalone makes a strong case for a strategic pivot towards developing 'enterprise intelligence' in the age of AI, a shift which has governance implications for financial services organizations.

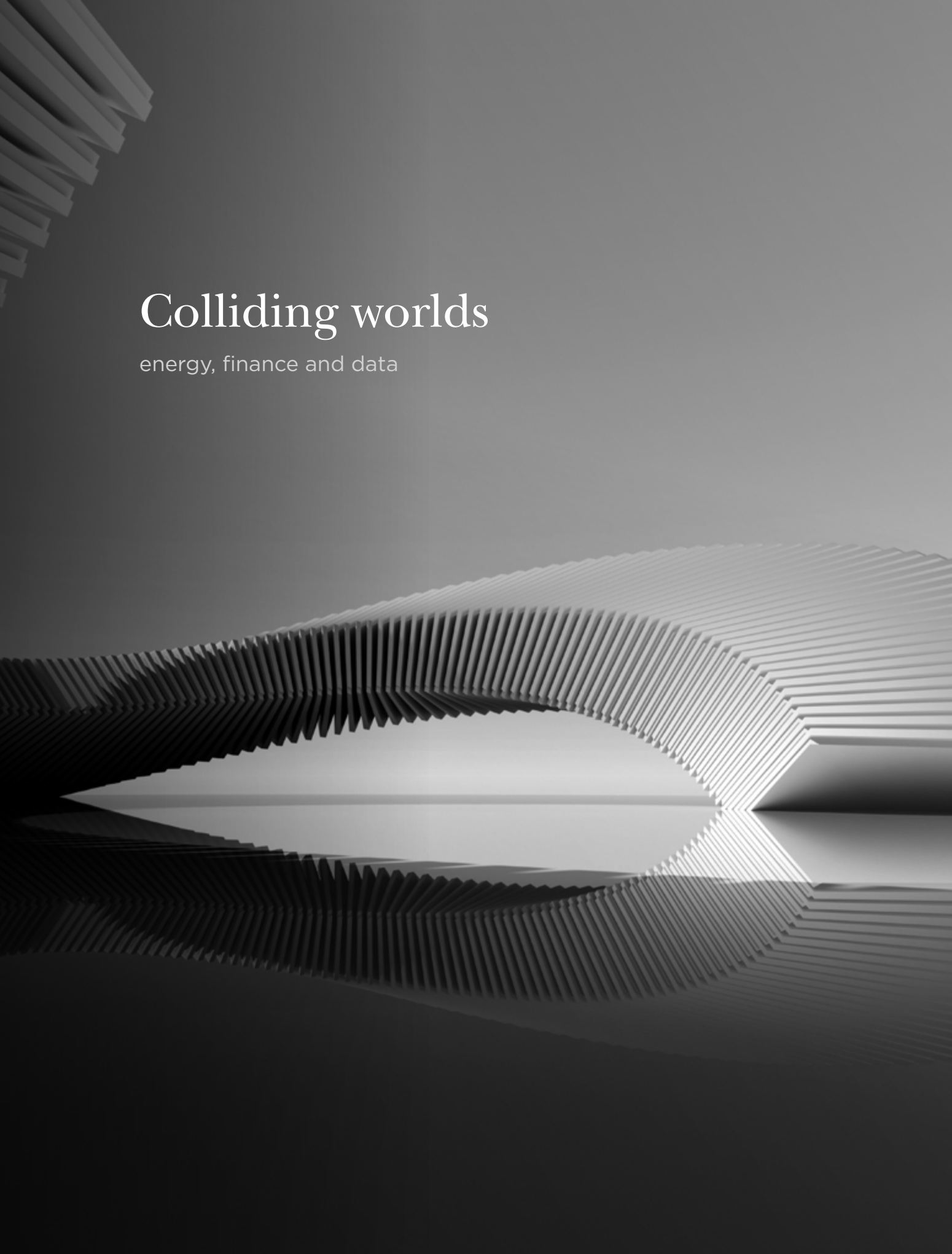
The final sub-theme of the issue reflects an ongoing interest that we have had in private markets. The shift from public to private markets is one of the big secular shifts of modern finance, which was dealt with explicitly in issue 61. In this issue, we build on this with two further articles. One of these, by Phalippou, tackles the shift of private equity towards retail investors, outlining the potential risks in selling private equity products that were designed with institutional investors in mind to a less informed and less powerful constituency. Preda and Chen's article pursues another tack, showing how venture capital funding in Asia and elsewhere has to be cognisant of surrounding government policy imperatives if it is to be successful.

Taken together, these articles argue that financial services are not simply being buffeted by geopolitical change. They are becoming one of the arenas in which that change is negotiated. Through payment systems, liquidity backstops, energy and computing infrastructure, corporate strategy and the evolving boundary between public and private markets, we can see how geopolitical forces manifest themselves and adapt to existing institutional constraints and imperatives.

Professor Crawford Spence
King's College London

Colliding worlds

energy, finance and data



Compute is global. Power is local.

Energy, capital and the structural tensions defining the new global order

Author | **Glen T. Ragland** | Partner, Capco

Abstract

This article contends that the emerging global order will be shaped more by the physical systems powering compute than by compute capacity itself. The rapid, global expansion of hyperscale data centers driven by artificial intelligence contrasts with the local constraints of power systems limited by permitting, transmission capacity, fuel supply and market design. Hyperscale compute can be deployed globally within months, while the energy systems required to power it are constrained by local physical, regulatory and temporal limits that unfold over years or decades. This structural tension – compute is global but power is local – is already reshaping energy strategy, national competitiveness and capital allocation. In response, hyperscalers are pursuing behind the meter generation through integrated ownership, independent power producer partnerships and joint ventures, transforming power from a managed cost into a strategic asset. These shifts carry implications for energy companies, regulators and financial services, including new infrastructure finance structures, revised risk models and localized grid relief over time.

1. Introduction: the world is entering a power-constrained digital era

The world faces one of the most significant structural shifts since the creation of the modern internet. Artificial intelligence (AI) and hyperscale computing are expanding at an unprecedented pace. Multi gigawatt data centers are being planned or built across North America, Europe, the Middle East, Asia and Latin America. Many of these centers will require more power than today's largest industrial plants. What was once a predictable and gradual load is now an exponential curve [S&P Global (2025)].

This rapid expansion reveals a fundamental truth about the global economy's future. Compute can scale globally, but the power needed is inherently local. Digital systems expand instantly across borders, but electricity does not. Electricity must be generated, transmitted and delivered through physical networks limited by geography, permitting, fuel supply, land use and regulation. Consequently, power systems are emerging as the limiting factor for digital growth rather than compute capacity [Ember (2025)].

This tension is evident in the strain on grids worldwide. Interconnection queues often last for years. Transmission corridors are congested.

Approvals for new generation are slow and often contentious. In many regions, including Northern Virginia, Dublin, Frankfurt, Osaka, Singapore, Sydney and São Paulo, the local grid cannot meet concentrated new demand at the needed scale [S&P Global Market Intelligence (2025)].

While this strain is reshaping the energy landscape, another dimension is the response by financial markets. NVIDIA is now among the most highly valued companies in history, reflecting extraordinary expectations for AI-driven growth and digital value creation [Bloomberg (2025)].

By contrast, the U.S. energy sector's combined market capitalization is materially smaller, depending on how the sector is defined. While market capitalization reflects publicly listed companies rather than the entire physical energy system, the gap remains a useful signal of how markets are valuing compute relative to the industries that must supply the power enabling that growth [StockAnalysis (2025)]. This divergence reveals a deeper structural issue: digital value creation is increasingly limited by physical energy, while capital markets still price compute as if power were unlimited.

This disconnect affects energy and technology companies, and it also affects the financial institutions that allocate capital, price risk and influence that regions can scale.

The emerging global order will be defined by how nations and companies respond to this constraint. Those who can deliver reliable, low carbon, always available power will lead. Those who cannot will fall behind. The story of the next decade is the story of energy.

2. AI scale data centers as the new global infrastructure backbone

Artificial intelligence is now central to global infrastructure, and it requires a vast physical footprint. Data centers for training and inference consume large amounts of electricity, water, land and specialized equipment. They are becoming as vital as ports, railroads and highways once were [McKinsey (2025)].

These facilities are being planned or built across every major region, highlighting the global nature of this trend. In the United States, large scale projects are concentrated in Texas, Arizona, Ohio and Northern Virginia. Europe is seeing significant investment in Ireland, the Netherlands, Germany and Scandinavia as countries compete to attract digital infrastructure. The Middle East is pursuing clusters in Saudi Arabia and the United Arab Emirates through large integrated energy and compute developments. Asia is advancing rapidly in Japan, South Korea, Singapore and India. Latin America is growing in Brazil, Chile and Colombia [Reuters (2025)].

The scale of these facilities is difficult to overstate. A single training campus may require several hundred megawatts or more of always available power, and some multi-site developments can exceed one gigawatt of load. This is comparable to major industrial plants and, in some cases, the electricity use of a city-sized population [McKinsey (2025)]. Data centers cannot tolerate intermittent availability. Even minor disruptions can interrupt training cycles or real time inference. As a result, these centers require power that is constant, predictable and insulated from local grid volatility.

This growth is already straining regional grids. Many of the most desirable locations for development are constrained by limited transmission capacity, slow permitting for new lines or insufficient local generation. Interconnection queues in the U.S. stretch out for years. Europe faces congestion and planning challenges in multiple markets as data center demand rises faster than grid expansion [S&P Global Market Intelligence (2025); Ember (2025)].

These constraints are not temporary. They are structural and they will shape where artificial intelligence can grow. As AI becomes central to global productivity, data centers shift from private facilities to national infrastructure. Countries and companies that can scale power for this load will attract investment and talent. Those that cannot will lose out. The competition to build and host artificial intelligence is now a competition for reliable energy.

3. Behind the meter generation: the strategic response to global grid constraints

As global artificial intelligence demand accelerates, the single most important question becomes how to supply the power needed to support it. In many regions, the grid cannot expand fast enough to meet this new load [S&P Global Market Intelligence (2025)].

These constraints lead to a simple conclusion that many hyperscale operators are now reaching. If the grid cannot deliver the required power on the required timeline, then the only viable option is to build power directly for the facility itself.

Behind the meter generation provides this path. Behind the meter refers to energy generation located on site or adjacent to the data center rather than sourced from the traditional utility grid. This approach offers a level of control over power availability, reliability and carbon intensity that the grid cannot match under current conditions. Instead of waiting for long interconnection processes, hyperscalers can pursue their own development schedules and align generation with commissioning timelines [McKinsey (2025)].

The architecture of behind-the-meter generation varies by region, but the direction is consistent. Many developments are designed as hybrid systems that combine natural gas or liquefied natural gas (LNG) fired generation with renewable energy and storage. Gas provides the reliability and stability that artificial intelligence workloads require. Renewable energy contributes to decarbonization goals and reduces overall emissions intensity over time. Storage helps manage ramping and variability. Together, these assets create a system that is more controllable and more resilient; it also aligns with aggressive climate commitments adopted by many hyperscalers, including initiatives to pursue 24/7 clean power matching [McKinsey (2024)].

In regions with insufficient grid capacity, behind the meter projects can also reduce stress on existing infrastructure. Large loads do not need to be served by the grid, which can free capacity for other industries and customers. This benefit is often overlooked, but it has long-term implications for economic development and grid stability.

Behind the meter generation is more than a technical fix for power shortages. It marks a structural shift in powering digital infrastructure. In the past, companies used grid power to run compute. Now, power and compute are built as a single integrated system. This shift will shape energy strategy for decades.

4. Structural tensions defining the new global order

The rapid expansion of AI has exposed a set of structural tensions between how digital systems scale and how power systems are built and governed. These tensions are not isolated operational challenges or temporary bottlenecks. They reflect fundamental mismatches in speed, geography, governance and valuation between global compute growth and locally constrained energy systems. Together, they help explain why behind the meter generation is accelerating, why traditional grid and market frameworks are under pressure and why energy strategy, financial markets and national policy are being reorganized around power availability.

While these tensions can be summarized succinctly, they arise from concrete physical and institutional realities. Power systems are constrained by geography, fuel availability, land use, permitting timelines and regulatory structures that vary by region. Generation and transmission assets are capital intensive, slow to build and deeply embedded in local markets and governance frameworks. These characteristics fundamentally limit how quickly energy systems can respond to the pace of global, AI-driven demand growth.

These core tensions are summarized in Table 1.

Table 1: Structural tensions defining the new global order

Structural tension	AI and compute reality	Power systems reality
Structural tension	AI and compute reality	Power systems reality
Global vs local	Workloads and models scale globally and can shift across regions	Generation, grids, and fuel supply are tied to local geography and regulation
Speed of scaling	Capacity can expand in months	New generation and transmission take years or decades
Capital mobility	Capital for AI is globally deployable and highly liquid	Energy infrastructure is capital intensive and location bound
Climate ambition vs reliability	Strong pressure for net zero and renewable matching	AI workloads require firm, dispatchable power for uninterripted operation
Market valuation	Compute firms command extraordinary growth valuations	Energy systems are valued as mature, regulated infrastructure
Governance models	Centralized and rapid technology decision making	Fragmented regulatory and permitting frameworks
Forecasting assumptions	Exponential demand growth expectations	Linear infrastructure planning and investment models

This table summarizes the structural tensions between global artificial intelligence compute growth and locally constrained power systems.

5. Global archetypes: how nations are responding

The rapid growth of artificial intelligence and the rise of behind the meter generation are not occurring in isolation. Every nation faces its own mix of energy resources, policy constraints, climate goals and industrial priorities. These differences are leading to distinct national responses. Together, they form the early structure of the new global order.

Four archetypes are emerging worldwide.

5.1 Energy rich but policy constrained nations

Countries such as the United States, Canada, Australia and Qatar have abundant natural gas and strong potential for renewable power. Their challenge is not resource availability, but the governance frameworks that shape how energy infrastructure is developed. In market-based economies, development is often constrained by policy, permitting and transmission approval processes that can slow the pace of new generation. In state-directed systems such as Qatar, constraints are less about permitting and more about strategic prioritization, capital

Table 2: National archetypes in the emerging energy-AI order

Attribute	Energy rich, governance constrained	Energy-constrained, compute heavy	Climate ambitious, physically limited	State directed integrated systems
Representative regions	U.S., Canada, Australia	Singapore, Japan, South Korea, Ireland	Germany, Netherlands, U.K.	Saudi Arabia, U.A.E., China
Primary energy resource base	Abundant gas and renewables	Limited domestic energy	Growing renewables, limited firm capacity	Abundant state-controlled energy
Core constraint	Governance, permitting, transmission approval	Land, imports, grid congestion	Physical limits, grid expansion timelines	Strategic prioritization and capital allocation
Planning and execution speed	Medium to slow	Medium	Medium	Fast
Grid flexibility	Moderate	Limited	Limited	High
Role of behind-the-meter power	Accelerating to bypass delays	Essential to enable growth	Pragmatic bridge for reliability	Integrated by design
AI growth potential	High	High but constrained	Medium to high	High
Defining competitive factor	Policy reform and investment coordination	Energy access and imports	System integration and flexibility	Centralized planning and execution

National archetypes illustrate how differences in energy resources, governance models and physical constraints shape countries' ability to support large-scale artificial intelligence growth.

allocation and national energy policy objectives. As a result, these countries have the natural ability to support artificial intelligence scale growth, but the pace and direction of development depend on governance models and investment priorities rather than resource limits.

5.2 Energy constrained but compute heavy nations

Regions such as Singapore, Japan, South Korea and Ireland have become major digital and financial hubs, but they lack the domestic resources or land to support large scale new power generation. Singapore depends on imported fuel and faces land constraints. Japan and South Korea rely on LNG imports and have limited siting options for renewables. Ireland faces grid congestion and limited local generation expansion [Reuters (2025); Ember (2025)]. These regions have strong demand for artificial intelligence development but must turn to LNG imports, behind the meter solutions and regional energy partnerships to support growth.

5.3 Climate ambitious but physically limited nations

Countries such as Germany, the Netherlands and the U.K. have strong climate commitments and aggressive renewable targets. At the same time, they face physical limits in terms of land, transmission capacity and permitting for new large scale energy projects, even as renewable capacity continues to expand. These regions are pursuing integrated planning, offshore wind and flexible generation to balance renewable variability. However, they face a structural tension between climate ambition and the physical limits of their grids. Behind the meter generation that incorporates gas and renewables is becoming a practical bridge [Ember (2025)].

5.4 State directed integrated energy and compute systems

Saudi Arabia, the United Arab Emirates and China represent a different model. These countries are integrating energy and compute development into national industrial strategy. They can plan energy and artificial intelligence infrastructure together as one coordinated system. They can direct capital, develop large integrated projects and align long-term energy policy with digital growth objectives.

This model reduces friction and allows for rapid development of multi-gigawatt energy and compute clusters, particularly in environments where energy supply, land use, capital allocation and permitting authority are centrally coordinated.

These archetypes illustrate a simple reality. Nations will not compete on artificial intelligence alone. They will compete on the ability to deliver energy that supports artificial intelligence. Energy availability, policy speed and system integration will determine which countries lead in the global digital economy. The emerging order is not defined by who adopts artificial intelligence fastest. It is defined by who can power it.

6. The inflection point: data center load is rewriting the entire energy and financial system

The rise of artificial intelligence scale data centers marks a turning point not only for the energy sector but for the entire architecture of the global economy reshaping energy systems, financial markets and national strategy. What began as a technology trend is now driving a broad and far-reaching shift across national energy strategies, financial markets and supply chains. Data center load is no longer a marginal consideration. It is triggering a system-level transformation.

The first part of this transformation is the recognition that the traditional model of building digital infrastructure atop the existing power grid no longer works. The grid was not designed for multi-gigawatt digital campuses that require constant power without interruption. When hyperscalers build large loads faster than transmission, generation and substations can be added, the constraint moves from compute capacity to energy capacity. This is the moment when the digital economy becomes directly dependent on physical energy investment.

The second part of the transformation is the impact on national energy policy. Governments that once focused on balancing renewable targets, reliability requirements and industrial growth are now confronted with a new priority. AI development is becoming a national strategic asset, and energy policy must support it. Countries are revisiting generation mixes, rethinking LNG strategies, expanding renewable portfolios and accelerating permitting reform.

As a result, energy security is no longer defined solely by access to fuel or generation capacity, but increasingly by the ability to finance, build and sustain energy systems capable of supporting high-density digital infrastructure at scale.

Financial markets are also being reshaped. Many of the largest digital and compute companies rely on long-term power contracts, firm capacity rights and dedicated infrastructure to support their workloads. These contracts behave like long-term financial instruments with predictable cash flows and multi-decade horizons. As hyperscalers invest in their own power systems, new forms of energy-backed financing emerge. Capacity commitments, renewable integration and firm supply arrangements turn power into a financial asset rather than a simple operating cost.

This convergence places financial institutions at the center of the transformation. Banks, insurers and asset managers are increasingly required to underwrite long-dated energy assets whose economics are directly tied to digital demand. Decisions around credit allocation, risk appetite and project financing now shape not only energy infrastructure but also the pace and geography of artificial intelligence expansion.

The energy supply chain is also changing. The shift to behind the meter hybrid systems increases demand for natural gas, LNG, renewable assets, storage, transformers, power electronics and mineral inputs such as copper and lithium. These supply chains introduce their own geopolitical dependencies. Countries with strong gas resources or access to LNG may gain strategic advantage. Countries with limited access must negotiate long-term supply to secure their development.

The most important point is that data center load has become a forcing function that links the energy and financial systems. Compute cannot scale without power, and power cannot scale without capital. The global economy is entering a period in which the physical constraints of the energy system will determine the speed and direction of digital growth.

This is the inflection point. From a financial services perspective, it calls for a shift in how infrastructure, technology and energy risk are assessed together rather than in isolation. Institutions that continue to evaluate data center growth, power generation, grid investment and fuel supply as separate domains risk mispricing both opportunity and risk. By contrast, financial institutions that develop the capability to underwrite integrated energy and digital infrastructure, including behind the meter generation and long-term power commitments, will be better positioned to allocate capital and support clients navigating this transition.

The future shape of the global order will be driven by the ability to build, finance and sustain energy systems that support AI. The countries and companies that understand this connection will lead the next decade.

7. From energy producer to system actor

For many years, data centers functioned as large electricity customers. They purchased power from utilities, negotiated supply arrangements, and focused on efficiency behind the meter. That model is now changing. As artificial intelligence scale growth outpaces grid expansion, the largest technology companies are no longer only consumers of energy. They are becoming active participants in power generation and system design, and this shift has broad implications for how power markets, grids and investment decisions evolve.

This transition does not always require direct ownership of generation assets. Hyperscalers are pursuing a range of arrangements, including direct development, long-term partnerships with independent power producers and joint ventures with energy companies. What matters is not the specific structure, but the outcome. Hyperscalers are securing firm, long-duration power outside traditional grid expansion pathways, and in doing so they gain influence over reliability, cost and carbon performance.

As a result, power stops being a passive operating expense and becomes a strategic system asset. Hyperscalers begin to shape regional fuel demand, renewable deployment, grid planning assumptions and market dynamics. This shift is the mechanism that produces the second-order effects explored in the following sections, including changes in energy company strategy, financial services behavior, regulatory frameworks and the potential transition from grid shortage to localized surplus.

> *For the first time in decades, energy companies are no longer just supplying growth. They are setting its limits. The pace of artificial intelligence and digital infrastructure expansion will be determined by who can deliver reliable, low carbon power at scale.*

8. What this means for energy companies

The rise of AI scale data centers and the shift of hyperscalers into power development has major implications for energy companies. For the first time in decades, energy companies are becoming both the constraint and the enabler of global economic growth. The future of artificial intelligence, cloud computing and advanced digital infrastructure depends on energy companies' ability to deliver reliable, low-carbon, large-scale power.

The first implication is that energy companies regain strategic centrality. For many years, the narrative held that electricity would become a commodity and that the power grid would follow a mature and predictable growth pattern. AI has challenged that assumption. The ability to supply firm power at scale is now a competitive advantage. Energy companies that can deliver gas-fired generation, LNG supply, renewable energy and storage in integrated configurations will shape where hyperscalers invest and where new digital clusters emerge.

The second implication is the opportunity to form long-term partnerships with hyperscalers. As described in the previous section, the integrated, IPP and joint venture models require deep energy expertise. Hyperscalers may bring capital, technology and demand. Energy companies bring

operational capabilities, regulatory experience and construction knowledge needed to develop and run large generation assets. This creates durable relationships that can span decades and yield stable contracted revenue.

The third implication is a shift in portfolio strategy. Energy companies will need to expand beyond traditional generation or supply roles. They will be expected to create hybrid systems that combine gas for reliability with renewables and storage for decarbonization. They will also need to navigate supply chains for fuel, equipment, minerals and long lead time components such as transformers. The companies that can coordinate these inputs will have an advantage in supporting hyperscale development.

The fourth implication is influence. As hyperscalers expand their energy footprint, they will work closely with energy companies to shape permitting, policy and market rules. Energy companies will play a critical role in advising on fuel security, renewable integration, grid stability and long-term planning.

In the decade ahead, the line between energy companies and technology companies will continue to blur. The companies that can operate effectively at this intersection will lead the next phase of global economic growth.

9. Implications for the financial services industry

The convergence of AI scale compute and energy infrastructure has significant implications for global financial services. Banks, insurers, asset managers and market intermediaries are no longer observing this transition from the sidelines. They are becoming central actors in how the new global order is financed, priced and governed.

The first implication is asset revaluation. For decades, energy assets have been treated as mature, cyclical and capital-intensive investments with modest growth expectations. AI driven demand challenges this assumption. The requirement for reliable, large scale, always available power introduces a new growth dynamic into the energy sector. Financial institutions will need to revisit valuation frameworks, discount rates and long-term demand assumptions for generation assets, gas infrastructure, LNG supply and hybrid systems that support digital infrastructure.

The second implication is the emergence of new financing structures. Behind the meter generation, long-term power purchase agreements and energy-backed compute contracts behave more like infrastructure assets than traditional utility supply arrangements. These instruments offer predictable cash flows, long durations and embedded demand growth. As hyperscalers partner with energy companies through integrated, IPP or joint venture models, financial institutions will play a critical role in structuring project finance, balance sheet financing and risk-sharing arrangements that align energy and digital investment horizons.

The third implication concerns risk and capital allocation. AI driven energy demand introduces new forms of commodity exposure, supply chain dependency and geographic concentration risk. Financial institutions must account for fuel availability, grid independence, regulatory treatment and carbon intensity when underwriting projects or extending credit. Climate related disclosures, stress testing and capital adequacy frameworks will increasingly need to incorporate energy reliability and infrastructure concentration as material risk factors.

The fourth implication is the evolving role of financial services as system enablers. Banks, insurers and asset managers are becoming intermediaries between technology companies, energy producers and governments. They influence which regions can attract hyperscale investment, which projects move forward, and how capital flows align with national energy and digital strategies. In this sense, financial institutions are not only pricing risk. They are shaping the structure of the new energy driven digital economy.

In the emerging global order, financial services will play a decisive role in determining how quickly and where AI infrastructure can scale. The institutions that understand the intersection of energy systems, digital demand and long-term capital formation will be best positioned to lead.

10. Regulatory tension and global market design challenges

As hyperscalers take on a larger role in energy production, regulators and grid operators worldwide face new and complex challenges. Electricity market structures, cost recovery mechanisms, planning processes and reliability frameworks were designed for a world in which large loads purchased power from the grid. That world is changing. The rise of behind the meter generation introduces regulatory questions that no region has fully resolved.

The first challenge is cost recovery. Power grids require significant investment to maintain transmission lines, substations and balancing capabilities. When large loads move off the grid, those costs must be recovered across a smaller customer base. Regulators must decide how to allocate costs fairly while preserving incentives for private investment. Some regions may adopt access charges, reliability fees or standby arrangements to maintain grid stability.

The second challenge is market design. Electricity markets differ across the world. The U.S. uses regional market operators. Europe relies on transmission system operators and distribution system operators coordinated through integrated day-ahead and intraday markets. Japan, Singapore and Australia use hybrid structures. The Middle East and China use state-directed systems. In all cases, a new question emerges. How should behind the meter assets participate in or support these systems? This includes pricing, capacity contributions, flexibility programs and the rules that govern the sale of surplus power [Ember (2025)].

A third challenge is planning and permitting. Governments must determine how behind the meter generation fits into national energy strategies. This includes coordination with renewable deployment, gas supply planning, and transmission expansion. As hyperscalers take on a larger share of power production, they become part of the national energy planning process, whether formally or through market influence.

For financial institutions, these regulatory and market design shifts introduce a new category of policy driven uncertainty. Load forecasts, regulated returns and capital recovery assumptions that have historically underpinned utility finance become less predictable as large customers self-supply. Lenders, insurers, and investors will need to reassess how they evaluate grid assets, regulatory risk and long-term infrastructure investment in a more distributed and evolving demand environment.

These regulatory tensions reflect a broader shift. Power systems that were built for traditional industrial loads must now evolve to support large scale digital infrastructure. This evolution will determine how quickly regions can attract AI development and how well they manage the transition to a more complex, distributed energy landscape.

11. The long-term shift from grid shortage to potential grid surplus

The global narrative surrounding AI growth has focused almost entirely on power shortages. Regions that host large data center clusters face overloaded substations, long interconnection queues and limited transmission headroom. This shortage has become a defining headline and an accepted constraint. Yet another, and more surprising, dynamic is beginning to emerge. As hyperscalers build behind the meter generation to serve their own facilities, major loads that would have been placed on the grid are removed from the system. Over time, this can create the conditions for localized grid surplus.

The scale of this effect is significant. A single campus may require several hundred megawatts or even exceed one gigawatt of load. When that load is served through dedicated onsite generation rather than the grid, it frees a large amount of capacity. Transmission lines that would have been congested gain room to operate. Substations that would have required upgrades remain within their limits. Grid operators gain more flexibility to support industrial electrification, residential development and commercial expansion [McKinsey (2025)].

A clear example is Northern Virginia, one of the largest data center markets in the world. The region has faced severe congestion and multi-year delays in expanding substations and high voltage lines. Several hyperscale operators are pursuing behind the meter generation to support new artificial intelligence facilities. If even a portion of expected load shifts off the grid, the region could move from persistent constraint to an environment with meaningful headroom for grid planners [Mordor Intelligence (2025)].

This shift can create new opportunities for regional economic growth. Cities that once struggled with limited power availability may regain room for new industries. Electrification of transport, heating and manufacturing becomes more feasible. Renewable integration becomes easier when the system is not pushed to its limits. In some regions, these changes may alter long-term planning and reduce the need for certain high-cost grid upgrades.

For financial institutions, this shift complicates traditional infrastructure planning and valuation. Long-term demand forecasts, revenue models and capital recovery assumptions that have historically underpinned grid investment and utility finance become less predictable when the largest new loads self-supply. Lenders, insurers and investors will need to reassess how they evaluate grid assets, regulated returns and long-term infrastructure risk in an environment where demand growth is increasingly shaped by private generation decisions rather than centralized load expansion.

Although some hyperscalers maintain a physical connection to the grid, the trend is toward full redundancy and operational independence. The largest operators are building behind the meter systems that include firm generation, onsite fuel supply, storage and microgrid controls. These systems allow data centers to operate in island mode without depending on the grid for backup or stability. The grid becomes optional rather than essential and may be used only for administrative simplicity or to sell excess energy when market rules allow it.

This shift introduces new questions for regulators and utilities. If major loads no longer rely on the grid for service, how should long-term grid planning be adjusted? How should costs be recovered when large customers no longer contribute to traditional rate structures? How should utilities forecast load when the biggest new facilities operate independent systems? How should regulators balance fairness, reliability and investment when the nature of demand is changing? These questions reflect a new planning environment in which behind the meter generation must be integrated into long-term system design, even if the grid is not providing primary service to the facility.

The transition from shortage to surplus will not occur everywhere, nor will it occur quickly. It depends on siting patterns, local resources, regulatory decisions, and the pace of behind the meter development. But the possibility itself is important. It shows that private generation not only reduces stress on power systems, but it can create new room for growth and new opportunities for regions that adapt to this change.

12. Clean energy transition acceleration driven by hyperscalers

The rise of behind the meter generation does not weaken the global transition to cleaner energy. In many ways, it strengthens it. Hyperscalers have some of the most ambitious climate and renewable energy commitments of any sector in the world. They are under sustained pressure from shareholders, customers, regulators and the public to reduce carbon intensity and demonstrate credible progress toward net zero goals [McKinsey (2024)].

Although most behind the meter systems begin with gas or LNG to provide firm power, they are designed with the expectation that the share of renewables will grow over time. Hyperscalers are pursuing configurations that combine gas for reliability with solar, wind and storage for decarbonization. They are entering long-term renewable procurement agreements and exploring hourly matching of renewable supply and demand [McKinsey (2024)].

The transition will not occur overnight. It requires new infrastructure, new market rules and new supply chain capacity. But the key point is that the pressure to pursue hybrid and clean systems begins on day one. A significant portion of future data center growth will be powered by systems that blend firm power with renewable energy and storage. This dynamic turns what might have been a purely fossil fuel driven increase in load into a catalyst for clean energy investment.

Hyperscalers are not only building the physical infrastructure for artificial intelligence. They are shaping the next phase of the global clean energy transition.

13. The physical constraints that shape AI: water, minerals and fuel supply

The expansion of artificial intelligence scale data centers is limited not only by power. It is also shaped by physical constraints essential to operating modern digital infrastructure. Water, minerals and fuel supply chains influence where facilities can be built and how reliably they can operate. These constraints add another layer of complexity to the global competition for artificial intelligence leadership.

13.1 Water as a local limiting factor

Water availability is emerging as one of the most important and least understood constraints on artificial intelligence growth. High-density computing systems produce significant heat, and many data centers depend on water-intensive cooling. Regions such as Phoenix, Dublin, Singapore and parts of Spain face water scarcity or regulatory limits on industrial consumption [Ceres (2025)].

Hyperscalers are exploring air cooling, immersion cooling and water recycling, but no technology eliminates the fundamental fact that water is a local resource that cannot be scaled globally. Compute may be global, but water is local.

13.2 Minerals and fuel supply chains as strategic variables

Behind the meter generation and hybrid systems also depend on global supply chains for gas, LNG, renewables, storage equipment and critical minerals. Renewable energy depends on copper, lithium, cobalt and other materials that are concentrated in specific regions. Transmission and grid equipment require steel, transformers and power electronics produced through long and complex manufacturing networks [U.S. Department of Energy (2023)].

These supply chains introduce geopolitical risk and create dependencies for hyperscalers and national governments. The reliability of these inputs will influence the reliability of the energy systems that support artificial intelligence infrastructure.

Together, these constraints show that artificial intelligence does not rely only on compute and power. It relies on water, minerals, fuel and equipment supply chains that are global in scale but local in impact. The future of digital development will be shaped as much by these constraints as by advances in technology.

14. Societal and economic impacts of energy driven digital clusters

The growth of artificial intelligence scale data centers and the rise of behind the meter generation have effects that go far beyond the energy system. They are already reshaping certain regions and will reshape many more in the decade ahead. Energy driven digital clusters are becoming a new form of industrial development that influences where people live, where companies invest and how governments plan for the future.

Regions that can provide reliable and affordable power are attracting new waves of digital investment. This investment supports construction, engineering and long-term operations. It also stimulates secondary industries such as advanced manufacturing, logistics and service providers that support data center operations. Cities that once competed on tax incentives or land availability now compete on energy availability and the ability to support high-density digital loads.

Labor markets are also changing. The rise of energy driven digital clusters requires a workforce that blends digital skills with energy system knowledge. Engineers, technicians, grid specialists and operations personnel are becoming essential components of digital infrastructure. This shift creates new career paths and demands new training programs that connect the energy and technology sectors.

Local communities face both opportunities and challenges. New data centers bring jobs and economic activity. They also bring increased demand for land, water and infrastructure. Regions that plan effectively can create balanced and sustainable growth. Regions that do not may struggle with mismatched land use, public concerns or infrastructure strain.

The rise of these clusters reflects a broader trend. Digital economic activity is becoming anchored to physical energy systems. The ability of regions to manage this connection will shape their economic future and their role in the global digital economy.

15. Predictions: what happens next

The forces reshaping the global energy and digital landscape are structural and long lasting. They are not temporary market conditions or short-term trends. They represent the next phase of economic development. Based on the direction of investment, the pace of artificial intelligence adoption and the constraints of the physical world, several outcomes are becoming increasingly clear.

First, hyperscalers will become some of the largest private power owners in the world. As loads grow, they will continue to build or co-develop generation assets that provide firm, reliable and controllable energy. In many regions, this will include gas, LNG, renewable energy and storage in hybrid configurations.

Second, energy assets will be revalued by global capital markets. For years, the energy system has been priced as a mature industry with modest growth expectations. That assumption no longer holds. Artificial intelligence growth will increase demand for firm power, gas supply, renewable integration and long duration storage. Capital markets will eventually reflect the central role of energy in digital growth, and energy companies will be priced accordingly.

Third, LNG and flexible gas will become essential stabilizers of global artificial intelligence growth. Countries that have access to gas or LNG import capacity will be able to support high-density digital infrastructure. Those without access will face pressure to secure long-term supply arrangements.

Fourth, carbon intensity will become a priced or price-differentiating attribute of compute. Hyperscalers and their customers will increasingly measure the carbon content of each unit of compute, and this will add a new dimension to competition among data centers and regions.

Finally, new financial structures will emerge. Energy backed compute contracts will become a distinct asset class that blends the stability of long-term energy agreements with the growth of digital demand. These contracts will define how hyperscalers and energy companies share risk, allocate capital and plan long-term investments.

Financial institutions that develop the capability to assess energy reliability, carbon intensity and digital demand together will gain a strategic advantage. Those who continue to evaluate energy, technology and infrastructure risk in isolation will struggle to price the realities of the new global order. These shifts mark the beginning of a new order. Power and compute will be developed together as one integrated system, and the world will organize around this reality.

16. Conclusion: energy will define the next global order

The world is entering a new era in which the physical constraints of energy shape the future of digital growth. Artificial intelligence, cloud computing and high-density data infrastructure are expanding at a pace that exceeds the ability of traditional power systems to support them. This shift has revealed a simple but powerful truth. Compute may scale globally, but the power required to run it is local. The new global order will be built on this foundation.

Throughout this article, we have explored how this tension drives new investment models, forms of collaboration and competition. Hyperscalers are building their own power systems because the grid cannot expand on the required timeline. Energy companies are becoming strategic partners and, in many cases, the essential enablers of artificial intelligence development. Nations are organizing themselves into new archetypes based on energy resources, regulatory structures and the ability to support artificial intelligence scale growth.

Financial markets have not yet fully aligned with this reality. Capital continues to price digital growth as if power can scale without constraint, while pricing much of the energy system as a mature industry with modest growth expectations. Over time, these assumptions will converge as investors, lenders and policymakers recognize that artificial intelligence growth depends on large scale investment in reliable generation, fuel supply, renewables, storage and supporting infrastructure.

The coming decade will be shaped by those who can deliver reliable, low cost, low carbon power at scale. Regions that align energy strategy with digital investment will attract the next generation of economic activity. Regions that cannot will face structural limits. The future belongs to nations and companies that understand that power and compute are now a single integrated system.

Energy, not compute, will define the next global order.

References

Bloomberg, 2025. Nvidia Hits $4 Trillion Value as Stock Rallies. Bloomberg, July 9, accessed July 7, 2025.

Ceres, 2025. Drained by Data: The Cumulative Impact of Data Centers on Regional Water Stress. September 23, accessed December 23, 2025.

Ember, 2025. Grids for Data Centres in Europe. June 19, accessed August 14, 2025.

McKinsey & Company, 2024. How Hyperscalers Are Fueling the Race for 24/7 Clean Power. December 18, accessed November 7, 2025.

McKinsey & Company, 2025. AI Power: Expanding Data Center Capacity to Meet Growing Demand, accessed November 7, 2025.

Mordor Intelligence, 2025. Northern Virginia Data Center Market, accessed December 12, 2025.

Reuters, 2025. Nvidia's Stock Market Value Hits $4 Trillion on AI Dominance. July 9, accessed July 9, 2025.

Reuters, 2025. Poor Grid Planning Could Shift Europe's Data Centre Geography, Report Says. June 18, accessed July 15, 2025.

S&P Global, 2025. Data Center Grid Power Demand to Rise 22% in 2025, Nearly Triple by 2030. October 14, accessed October 15, 2025.

S&P Global Market Intelligence, 2025. Grid Congestion Remains Key Issue as Data Center Load Growth Stresses System. October 16, accessed October 16, 2025.

StockAnalysis, 2025. Energy Sector Stocks, accessed November 7, 2025.

U.S. Department of Energy, 2023. 2023 Critical Materials Assessment. July 2023, accessed November 21, 2025.

Carbon credits and the energy sector in Brazil:

energy transition, implicit carbon pricing, risk management and technological innovation

Author | **Livia Savignon** | Principal Consultant, Capco

Abstract

The global energy transition constitutes one of the greatest structural challenges faced by the energy sector in the 21st century. The need to reduce greenhouse gas emissions, in line with international climate goals, imposes increasing restrictions on energy production, conversion, and consumption, without compromising security of supply, economic competitiveness, and the stability of energy systems. In this context, carbon credits emerge as a complementary economic instrument, capable of enabling gradual decarbonization trajectories, internalizing environmental externalities, and directing capital towards low-carbon solutions. In Brazil, a country characterized by a predominantly renewable electricity matrix, the relevance of the oil and gas sector, and a leading position in biofuels, carbon credits assume a unique strategic role. This article analyzes, from an academic and sectoral perspective, the relationship between energy transition and carbon credits in the Brazilian energy sector, discussing economic fundamentals, market architecture, sectoral applications, financial implications, integrity and governance challenges, as well as technological and innovation trends that shape the evolution of this market. The study argues that carbon markets are not merely complementary instruments to decarbonization, but increasingly strategic mechanisms capable of influencing investment allocation, competitiveness, and long-term energy planning. By examining regulatory developments, market dynamics, and technological enablers, the article concludes that the consolidation of a credible, transparent, and technologically robust carbon market will be essential for aligning Brazil's energy transition with economic resilience and global climate commitments.

1. Energy transition: global context and structural challenges

The contemporary energy transition differs from previous historical processes in that it is driven by global environmental constraints and not just by technological factors or resource scarcity. The consolidation of scientific consensus on anthropogenic climate change, as summarized in the reports of the Intergovernmental Panel on Climate Change [IPCC (2023)] has established clear targets for reducing greenhouse gas emissions as a necessary condition for limiting global warming.

In this scenario, the energy sector occupies a central position. According to the International Energy Agency, the global energy system accounts for the largest share of CO_2 emissions, either directly, through the combustion of fossil fuels, or indirectly, through the electrification of still carbon-intensive end sectors [IEA (2021)]. At the same time, the energy sector itself is the main technological driver of the transition, enabling renewable sources, energy efficiency, electrification, and carbon capture and removal solutions.

The complexity of this process stems from the fact that energy systems are critical infrastructures, with long-lived assets, high capital costs, and strong interdependence with economic activity. Therefore, the energy transition cannot occur abruptly, at the risk of compromising security of supply, industrial competitiveness, and macroeconomic stability.

2. Carbon credits as an economic instrument for the energy transition

In this context of environmental constraints and economic imperatives, carbon credits emerge as an economic instrument for transition. Based on the principles of environmental economics, these instruments allow for the internalization of negative externalities associated with greenhouse gas emissions, assigning monetary value to their reduction or removal [OECD (2022)].

A carbon credit essentially represents the certified reduction of one ton of carbon dioxide equivalent (tCO_2e) relative to a methodologically defined baseline. Its economic function is not to replace structural decarbonization policies, but to reduce the aggregate cost of the transition, allowing reductions to occur where they are most economically efficient [World Bank (2024)].

For the energy sector, this logic is particularly relevant. Different segments, such as electricity generation, refining, oil and gas production, and bioenergy, have quite distinct marginal costs of emission abatement. Carbon credits allow these differences to be exploited, creating intersectoral and intertemporal flexibility in the decarbonization trajectory.

3. Marginal abatement costs: empirical evidence and observed data

The theoretical analysis of marginal costs of emissions abatement can be strengthened by observing consolidated empirical data. Recent World Bank reports indicate that the cost of emissions reduction varies significantly between technologies and sectors, reinforcing the economic rationale of carbon markets as instruments of allocative efficiency [World Bank (2024)].

According to the State and Trends of Carbon Pricing [World Bank (2024)], energy efficiency measures and operational emissions reductions in the energy sector have average costs of less than U.S.$20/$tCO_2e$, while renewable generation projects generally range between U.S.$30 and U.S.$60/tCO_2e, depending on the regulatory context and the degree of technological maturity. In contrast, carbon removal technologies, such as direct air capture (DAC), still have high costs, often exceeding U.S.$300/$tCO_2e$ [World Bank (2024)].

This empirical evidence reinforces the logic that carbon credits allow for prioritizing lower-cost reductions in the short and medium term, while more capital-intensive technologies mature and reduce their learning curves, an aspect that is particularly relevant in the energy sector.

4. Architecture of carbon markets: regulated and voluntary

4.1 Regulated markets: evidence from the North Sea

The North Sea experience offers a concrete example of the application of regulated carbon markets and their interaction with carbon capture and storage (CCS) technologies. Countries such as Norway and the United Kingdom have developed CCS projects integrated with carbon pricing policies, creating clear economic incentives for mitigating industrial and energy emissions. Projects such as Northern Lights and Net Zero Teesside demonstrate how the existence of a stable regulatory framework, combined with explicit carbon prices, enables long-term investments in CCS, even in the face of relatively high marginal costs [IEA (2021)]. In these cases, carbon credits or permits act as a mechanism for internalizing environmental costs, making projects economically viable that would be unfeasible in an environment without pricing.

4.2 RenovaBio as a Brazilian case study of a hybrid market

In the Brazilian context, the RenovaBio program constitutes a relevant example of a hybrid instrument that combines elements of regulatory policy, a carbon credit market, and economic incentives. Although not a carbon market stricto sensu, RenovaBio establishes mandatory decarbonization targets for fuel distributors and creates decarbonization credits (CBIOs), which are traded on the market. RenovaBio demonstrates, in practice, how market mechanisms can be used to internalize environmental externalities in the energy sector, valuing the low-carbon attributes of biofuel production [MME (2023)]. The underlying economic logic is analogous to that of carbon credits: agents with lower abatement costs generate credits, while those with higher costs acquire them to meet their targets. This Brazilian case exemplifies the strategic use of market instruments to enable the energy transition in specific sectors.

These experiences demonstrate that different institutional architectures can coexist, provided they are aligned with criteria of environmental integrity, regulatory predictability, and economic efficiency.

5. Sectoral applications of carbon credits in the energy sector

The use of carbon credits in the energy sector is heterogeneous, reflecting the structural, technological, and regulatory differences between its segments. Oil and gas, electricity generation, and bioenergy face distinct challenges in the energy transition, which influences the role played by credits as complementary economic instruments.

5.1 Oil and gas: carbon credits as a transition bridge

The oil and gas sector faces structural challenges in the energy transition. Although it remains essential for global energy security, it is an inherently emissive segment. In Brazil, however, offshore production has a lower carbon intensity than the international average, due to technological gains and operational efficiency, especially in the pre-salt layer [IEA (2021)].

Even so, regulatory, financial, and reputational pressures demand clear mitigation strategies. In this context, carbon credits play a bridging role in the transition, being used to offset residual emissions, enable flaring reduction projects, mitigate methane leaks, electrify platforms, and, in some cases, capture and store carbon [Blaufelder et al. (2021)].

From an academic point of view, it is crucial to emphasize that the use of credits in the oil and gas sector should be understood as

complementary and transitory, not replacing the need for structural transformation of the energy portfolio over time [IPCC (2023)].

5.2 Electric sector: systemic reliability and strategic use of credits

Despite the high participation of renewable sources, the Brazilian electricity sector faces growing challenges related to the intermittency of sources and the expansion of demand. The need for firm power and operational flexibility maintains the relevance of thermal sources, especially natural gas, to ensure system stability [EPE (2023)].

In this context, carbon credits assume a complementary role by allowing the offsetting of emissions associated with thermal generation necessary for system reliability. In parallel, financial resources from the carbon market can be directed to low-carbon structuring projects, such as bioenergy, forest restoration, and energy storage technologies.

This approach contributes to an orderly energy transition, avoiding abrupt solutions that compromise security of supply or excessively increase systemic costs [World Bank (2024)].

5.3 Biofuels and bioenergy as structuring vectors of the Brazilian transition

Brazil holds a prominent global position in biofuels, such as ethanol, biodiesel, biogas, and bioelectricity. These sources have a high potential for reducing emissions when compared to conventional fossil fuels, especially when considered throughout their entire lifecycle [MME (2023)].

Carbon credits play a central role in the economic viability of these chains, monetizing climate benefits that are not fully captured by traditional energy prices. For the energy sector, this represents an opportunity for diversification, strengthening national solutions, and reducing regulatory and commercial risks associated with carbon intensity [OECD (2022)].

Together, these segments demonstrate that the role of carbon credits in the energy sector is inherently heterogeneous, varying according to technological structure, emissions profile, and degree of regulatory maturity.

6. Carbon credits as a risk management and capital allocation tool

With the advancement of carbon pricing policies and climate disclosure requirements, carbon credits are becoming integrated into risk management strategies in the energy sector. They act as a hedge against regulatory uncertainties, reputational risks, and potential financial impacts arising from emissions-intensive assets.

Financial institutions are increasingly incorporating carbon metrics into credit risk assessment, cost of capital, and eligibility for sustainable financing instruments. In this context, the ability of energy companies to structure robust and transparent strategies for the use of carbon credits becomes a relevant competitive differentiator [World Bank (2024)].

7. Environmental integrity, MRV and institutional governance

The effectiveness of carbon credits fundamentally depends on their environmental integrity. Questions about additionality, permanence, double counting, and measurement represent significant risks to market credibility, especially in countries with large territorial extensions such as Brazil [IETA (2023)].

In the energy sector, where projects involve critical assets and systemic impacts, robust measurement, reporting, and verification (MRV) systems are an indispensable condition for

regulatory acceptance and investor confidence. The absence of adequate governance can compromise not only individual projects but the legitimacy of the market [Energy Transitions Commission (2023)].

8. Regulatory development and integration with public policies

The advancement of the regulatory debate on the carbon market in Brazil represents a fundamental step towards the consolidation of credits as a structuring instrument of the energy transition [MME (2023)].

In this context, the holding of COP 30 in Belém do Pará marks an institutional turning point, with the announcement of the creation of a carbon credit certifier led by BNDES. The initiative seeks to provide credibility, standardization and transparency to a market historically marked by methodological uncertainties.

BNDES's role as a certifying entity directly addresses the challenges of environmental integrity and MRV, by monitoring the entire credit cycle, from formulation to retirement, reducing greenwashing risks and attracting long-term international capital for restoration, bioenergy and nature-based solutions projects.

9. Case study: the ProFloresta+ program as a model for sectoral decarbonization

The materialization of the theoretical concepts of flexibility and allocative efficiency can be observed in the ProFloresta+ program, a joint initiative between BNDES and Petrobras. This program exemplifies how major players in the energy sector use the carbon market as a 'transition bridge,' allowing them to maintain investments in long-life assets while offsetting operational emissions through nature-based solutions.

9.1 Market scale and dynamics

The first call for proposals for ProFloresta+, which closed on January 9, 2026, demonstrated a market receptiveness that exceeded initial expectations, attracting 16 proposals for forest restoration projects in the Amazon. The immediate goal is the acquisition of 5 million carbon credits (where each credit represents one ton of CO_2e removed), distributed across five strategic contracts. In the long term, the program projects restoring up to 50,000 hectares of degraded areas, generating approximately 15 million credits, a volume equivalent to the annual emissions of approximately 8.94 million gasoline-powered vehicles.

9.2 Economic rationality and risk management

Petrobras' strategy in this call for proposals aligns with the logic of the lowest marginal cost of abatement, since the company will select projects that present the lowest total cost for the amount of credits contracted. In addition to financial efficiency, the initiative addresses reputational and financial risk management, improving the company's ESG metrics in the eyes of international investors. For applicants, the program offers security and economic viability through access to differentiated BNDES financing lines, such as those from the Climate Fund, specifically aimed at restoration with native species.

9.3 Environmental integrity and governance

A critical point highlighted in the project is environmental integrity. Submitted proposals undergo a rigorous technical evaluation that analyzes environmental quality and compliance with socio-environmental safeguards. This methodological rigor is essential to avoid the risk of greenwashing and ensure that the generated credits have liquidity and real value

in the voluntary market and the future regulated Brazilian market. With a potential investment of over R$ 1.5 billion, ProFloresta+ not only fosters the bioeconomy in the Amazon, but also consolidates the use of carbon credits as part of the economic infrastructure of the energy transition in Brazil.

10. Structural limitations on the use of credits in the energy sector

Despite their relevance, carbon credits have clear limits. Their excessive use can delay investments in low-carbon structural technologies and prolong dependence on emitting assets. The academic literature converges in stating that credits should be used in a complementary, temporary way, aligned with scientific decarbonization trajectories [IPCC (2023)].

In the energy sector, this implies establishing clear criteria for use, transparency, and temporality, preventing credits from becoming permanent substitutes for the necessary structural transformations.

11. Regulatory development in Brazil: the Carbon Market Bill

The advancement of the regulatory debate on the carbon market in Brazil represents a fundamental step towards consolidating carbon credits as a structuring instrument for the energy transition. The Bill that establishes the Brazilian Emissions Trading System seeks to create a regulated framework, with the definition of covered sectors, reduction targets, and rules for monitoring, reporting, and verification [MME (2023)].

The connection between this regulatory development and the market architecture discussed in Section 4 is direct. The institutionalization of a regulated market tends to reduce uncertainties, raise environmental integrity standards, and expand the energy sector's participation in compliance mechanisms. At the same time, the regulatory design will need to engage with existing instruments, such as RenovaBio, to avoid overlaps and economic distortions.

From an academic point of view, the central challenge lies in balancing environmental rigor and economic flexibility. An excessively rigid market can increase transition costs and compromise the competitiveness of the energy sector, while an excessively flexible market can compromise its credibility and climate effectiveness [World Bank (2024)].

12. Technologies, innovation and trends in the carbon and energy market

The future of the integration between carbon credits and the energy sector will be largely determined by the convergence of technological advances, regulatory evolution, and the sophistication of financial markets. Far from being merely environmental instruments, carbon credits are increasingly taking on the role of economic infrastructure for the energy transition, influencing investment decisions, asset planning, and long-term corporate strategies.

12.1 Digitalization, advanced MRV, and reduction of information asymmetries

The digitalization of measurement, reporting, and verification (MRV) systems represents one of the most relevant transformations in the contemporary carbon market. The combined use of remote sensing, artificial intelligence, geospatial modeling, and digital records allows the transition from ex-post and sample models to near-continuous monitoring systems, reducing operational costs and information asymmetries [Blaufelder et al. (2021)].

> *The consolidation of the integration between carbon credits and the energy sector will depend on the convergence of technological innovation, regulatory enhancement and the increasing maturity of financial markets.*

In the energy sector, this evolution is particularly relevant, given the systemic nature of emissions and the high financial value associated with credits. Advanced digital MRV not only strengthens the environmental integrity of credits but also increases their liquidity by reducing the risks of double counting, methodological questions, and regulatory uncertainties. In this context, technology ceases to be merely an operational tool and begins to act as a market enabling condition.

12.2 Carbon removal technologies and redefining the role of the energy sector

The advancement of carbon removal technologies tends to structurally redefine the role of the energy sector in carbon markets. Solutions such as direct air capture (DAC), bioenergy with carbon capture and storage (BECCS), and carbon capture and storage applied to energy assets (CCS) broaden the scope of the sector, which goes from a mere mitigating agent to a potential net provider of removals [IPCC (2023)].

From an economic point of view, these technologies have high marginal costs at the current stage, but play a strategic role in the long term, especially in climate neutrality and negative emissions scenarios. For the energy sector, this implies a gradual repositioning, in which existing assets can be reconfigured as deep decarbonization platforms, provided they are inserted into appropriate regulatory and pricing frameworks.

12.3 Integration between energy, data and finance: carbon as an economic asset

A central trend is the growing integration between energy, data, and finance markets, in which carbon credits are treated as strategic economic assets, incorporated into energy planning decisions, project valuation, and financial structuring. Financial institutions already incorporate carbon metrics in risk analyses, cost of capital, and eligibility for sustainable financing instruments [World Bank (2024)].

In this context, carbon credits cease to be merely a compensatory mechanism and begin to directly influence capital allocation, the attractiveness of energy projects, and the relative competitiveness between technologies. This integration reinforces the role of credits as a bridge between climate objectives and economic rationality, especially in capital-intensive sectors with long investment horizons.

12.4 Regulatory standardization, market convergence, and institutional legitimacy

The consolidation of carbon credits as a structuring instrument of the energy transition will depend on regulatory evolution and convergence between voluntary and regulated markets. There is a clear trend towards higher environmental integrity standards, with increasing demands for additionality, permanence, traceability, and transparency [IETA (2023)].

For the energy sector, this standardization is fundamental. Regulatory predictability reduces legal risks, increases investor confidence, and

enables the integration of credits into formal energy and climate policy instruments. At the same time, convergence between voluntary and regulated markets tends to redefine corporate strategies, requiring greater alignment between the use of credits, decarbonization targets, and long-term planning.

12.5 Systemic implications for the energy transition

Taken together, these trends indicate that carbon credits are likely to consolidate as the economic infrastructure of the energy transition, linking technology, regulation, and finance. In the energy sector, their future relevance will not only be associated with the volume of credits traded, but also with their ability to:

- reduce systemic decarbonization costs
- enable technologies that are not yet competitive
- mitigate regulatory and financial risks
- and align private decisions with long-term climate objectives.

Thus, innovation in the carbon market should not be interpreted as a peripheral phenomenon, but as a central element of the economic and institutional reorganization that underpins the contemporary energy transition.

12.6 Technological convergence and implications for conclusive analysis

The technological, institutional, and financial trends discussed in this section reinforce the idea that carbon credits are undergoing a conceptual transformation: from a peripheral offsetting instrument to a structural element of the economic architecture of the energy transition. The digitalization of MRV systems, the advancement of carbon removal technologies, and the increasing integration between energy, data, and finance indicate that the economic value of carbon is now determined not only by traded volumes but also by criteria of integrity, traceability, and regulatory alignment.

In the energy sector, this convergence redefines the role of carbon credits throughout the entire decision-making cycle, from asset planning to financial structuring, from risk management to public policy formulation. At the same time, the distinction between transitory and structural uses of credits is widening, reinforcing the need for differentiated and temporally conditioned sectoral approaches.

In this way, the discussion on technologies and innovation does not end in itself but provides the analytical basis for understanding the limits, opportunities, and institutional constraints explored in the conclusion of this article. It is from this integrated perspective, technology, economics, and governance, that it becomes possible to critically evaluate the role of carbon credits in the Brazilian energy transition.

13. Conclusion

Carbon credits play a relevant economic and institutional role in the energy transition of the Brazilian energy sector, not as a substitute mechanism for structural decarbonization, but as a catalyst capable of reducing systemic costs, mitigating risks, and increasing the allocative efficiency of climate capital. In a sector characterized by long-lived assets, high capital intensity, and strong interdependence with security of supply, the existence of flexible mechanisms for implicit carbon pricing proves to be particularly strategic.

From an economic point of view, the heterogeneity observed in the marginal abatement costs, evidenced by empirical data from the World

Bank, reinforces the rationality of carbon markets as transition instruments. By allowing lower-cost reductions to be prioritized in the short and medium term, carbon credits enable gradual decarbonization trajectories, while creating economic space for the technological maturation of more capital-intensive solutions, such as CCS, BECCS, and direct removal technologies.

Sectoral analysis has shown that the role of carbon credits in the energy sector is inherently heterogeneous. In the oil and gas segment, credits operate predominantly as a technological and financial bridge, allowing for the mitigation of residual emissions while technical and economic constraints persist in the complete replacement of assets. In the electricity sector, they function as a mechanism complementary to systemic reliability, by mitigating emissions associated with thermal generation necessary for the stability of the energy matrix. In biofuels and bioenergy, credits take on a more structural character, monetizing environmental attributes and reinforcing Brazil's comparative advantages in the global energy transition.

From an institutional perspective, the effectiveness of carbon credits critically depends on high standards of environmental integrity, robust measurement, reporting and verification (MRV) systems, and predictable regulatory frameworks. Analysis of international experiences, such as CCS projects in the North Sea, and national experiences, such as RenovaBio, ProFloresta+, and the certification initiative led by BNDES in the context of COP 30, indicates that institutional credibility is a necessary condition for carbon credits to assume a structuring role in the energy sector, avoiding greenwashing risks and increasing their acceptance by global investors.

In addition, the advancement of the regulatory debate in Brazil, with the proposal for a national emissions trading system, signals a gradual transition from a predominantly voluntary model to hybrid and potentially regulated arrangements. This movement tends to redefine the strategic role of carbon credits in energy planning, requiring careful coordination between market instruments, sectoral policies, and long-term climate objectives. The absence of this coordination may compromise the economic efficiency and environmental legitimacy of the system; on the other hand, its proper integration can transform the carbon market into a central component of Brazilian energy and industrial policy.

Finally, the results indicate that, for financial institutions, the future of carbon credits in the energy sector will be closely tied to technological advancement and the integration of energy, data, and finance. The digitalization of MRV systems, the application of artificial intelligence, the development of structured market platforms, and the consolidation of carbon removal technologies are likely to reshape risk pricing, asset valuation, and investment structuring. As a result, carbon credits should not be viewed merely as environmental instruments, but as an emerging asset class whose strategic relevance will depend on regulatory stability, technological credibility, and institutional robustness within the broader energy transition.

References

Blaufelder, C., C. Levy, P. Mannion and D. Pinner, 2021. A blueprint for scaling voluntary carbon markets to meet the climate challenge, report, McKinsey & Company, accessed on 19 December, 2024 at https://www.mckinsey.com/capabilities/sustainability/our-insights/a-blueprint-for-scaling-voluntary-carbon-markets-to-meet-the-climate-challenge

BNDES, 2025. BNDES, Bradesco e Fundo Ecogreen anunciam certificadora brasileira de carbono, Banco Nacional de Desenvolvimento, accessed on January 11, 2025 at https://agenciadenoticias.bndes.gov.br/cop30/BNDES-Bradesco-e-Fundo-Ecogreen-anunciam-certificadora-brasileira-de-carbono/

Empresa de Pesquisa Energética (EPE), 2023. Balanço Energético Nacional [Brazil's energy balance], accessed on September 1, 2025 at https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-748/topico-687/BEN2023.pdf

Energy Transitions Commission, 2023. Financing the Transition: How to Make the Money Flow for a Net-Zero Economy Energy, Transitions Commission, London, accessed on December 15, 2024 at https://www.energy-transitions.org/wp-content/uploads/2023/08/ETC-Financing-the-Transition-MainReport_update.pdf

IEA, 2021[AM1] . Net Zero by 2050: A Roadmap for the Global Energy Sector, International Energy Agency, Paris, accessed on September 12, 2024 at https://tinyurl.com/4vby3yrh

IETA, 2023. The Evolving Voluntary Carbon Market, International Emissions Trading Association, Geneva, accessed on December 17 at https://tinyurl.com/mr426dzn

IPCC, 2023. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 35-115, doi: 10.59327/IPCC/AR6-9789291691647, accessed on July 12, 2024 at https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_LongerReport.pdf

Ministério de Minas e Energia (MME), 2023. PNE 2050: Plano Nacional de Energia, Government of Brazil, accessed on July 1, 2025 at https://www.gov.br/mme/pt-br/assuntos/secretarias/sntep/publicacoes/plano-nacional-de-energia/plano-nacional-de-energia-2050/relatorio-final/relatorio-final/relatorio-final-do-pne-2050.pdf/@@download/file

OECD, 2022. Effective Carbon Rates 2021: Pricing Carbon Emissions through Taxes and Emissions Trading, OECD Publishing, Paris, accessed on December 13, 2024 at https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/05/effective-carbon-rates-2021_02959cc3/0e8e24f5-en.pdf

Petrobras, 2024. Petrobras e BNDES lançam primeiro edital do ProFloresta+ para adquirir 5 milhões de créditos de carbono, Rio de Janeiro, accessed on January 15, 2025 at https://agencia.petrobras.com.br/w/petrobras-e-bndes-lan%C3%A7am-primeiro-edital-do-profloresta-para-adquirir-5-milh%C3%B5es-de-cr%C3%A9ditos-de-carbono

UNFCCC, 2025. Belém Political Package, United Nations Framework Convention on Climate Change, Belém do Pará, accessed on January 19, 2025 at https://unfccc.int/cop30/belem-political-package

World Bank, 2024. State and Trends of Carbon Pricing 2024, World Bank, accessed on November 12, 2024 at http://hdl.handle.net/10986/41544

‘No meal without rice’:

industry alignment and infrastructure in Chinese bitcoin mining

Authors | **Ruowen Xu** | Assistant Professor, Warwick Business School, University of Warwick
Alex Preda | Professor, Department of Sociology and Social Policy, Lingnan University

Abstract

This paper examines the infrastructure of bitcoin mining, the proverbial ‘rice’ that anchors the production of crypto assets. While debates about bitcoin have focused on value, speculation, and energy consumption, far less attention has been paid to the material infrastructures that make mining possible. Drawing on interviews with early miners, follow-up interviews with Chinese miners who later migrated to the United States, and a two-week ethnographic study of site construction in Ohio, we analyze how mining infrastructure has been built, reorganized, and industrialized over the past fifteen years. We argue that this evolution has been shaped by four forces: geopolitical realignments, opportunity-seeking across mining and adjacent sectors (especially energy and hardware), intensifying competition for speed and efficiency, and the consolidation of mining into a capital-intensive industrial sector. Together, these dynamics have transformed mining from a flexible, experimental activity into an increasingly corporatized form of industrial production. The paper shifts attention from cryptocurrencies as financial objects to the infrastructures that sustain them, showing how mining is now organized around the industrial constraints of construction, energy, logistics, and capital coordination. In doing so, it foregrounds the persistent temporal tension between intersecting material markets, speculative expectations of rapid returns, and the slow, costly work of building and sustaining durable infrastructural systems.

1. Introduction

There is an old Chinese proverb, ‘no meal without rice’: even the cleverest cook cannot make a meal without rice – the modest rice anchors the flavors of most meals. Similarly, not even the cleverest businessperson can mint and trade crypto assets without the infrastructure provided by data centers, an infrastructure that supports a variety of blockchains. One of the best-known chains is the bitcoin blockchain, anchored in a network of data centers that enable mining operations – that is, competitively handling cryptographic problems that, if successfully and timely solved, will lead to a new bitcoin being minted. This network provides the ‘rice’ to the ‘meals’ coveted by institutions and individuals alike. While the value and nature of bitcoin have been subject to much academic debate, less attention has been paid to the ‘rice’, the mining network that supports bitcoin. When this has happened, conversations have focused mostly on the energy consumption required by proof of work (PoW), the main protocol of bitcoin. How has this infrastructure evolved over time and which factors have influenced this evolution? As many a cook knows, there are many flavors of rice. What are the flavors of the bitcoin ‘rice’?

In this paper, we argue that the infrastructure of bitcoin mining has evolved significantly over the past fifteen years or so, primarily under the influence of the following factors: (1) geopolitical realignments that created barriers but also opportunities; (2) opportunity searches within both bitcoin mining and within adjacent industries, primarily the energy sector and hardware manufacturing; (3) a competition for speed not dissimilar from what we have observed in high frequency trading; and (4) consolidation of bitcoin mining as an industrial sector, characterized by scaling up, technological innovation, and response to profitability pressures. As bitcoin mining has become more of a standard industrial sector, we see it confronted with familiar issues, such as project management, timely delivery, search for cheap energy sources, and balancing financial with manufacturing constraints.

We draw on interview data with miners active in the early phases of the industry, as well as with miners who were active in Asian mining but later migrated to the U.S. We build on a two-week ethnographic study shadowing Chinese miners as they built, designed, and reconfigured mining sites in Ohio. We show how technological infrastructures have been reorganized over time and how mining has become an increasingly capital-intensive and corporatized activity. We show that mining is no longer a mere technical activity, but an infrastructure-heavy form of asset production that increasingly resembles hardware, energy, data-center, and industrial real estate projects. For financial practitioners, the case illustrates the gap produced by standard valuation models that systematically underestimate infrastructural and material time and shows why crypto finance now operates not only as a market for assets, but as a practice of managing multiple, intersecting material and financial timelines.

2. The role of Chinese miners: configuring the infrastructural work of crypto mining

In the early stages, for a short period of time, bitcoin mining was relatively diffuse, in the sense that individuals owning a working laptop could participate in this process too. However, this changed rapidly around 2010, with the switch from CPU (central processing units) to GPU (graphic processing units). Mining activities became heavily concentrated in mainland China: by 2015, China's share of bitcoin mining was already 50%, while the U.S. share declined to 10% (see Figure 1). In the mid-2010s, the regulatory environment in the U.S. was favorable to miners – for instance, in 2014 the IRS started treating virtual currency as property (meaning that miners had to declare income from mining), while miners were not treated as a money service providers (meaning that they were not subject to the respective regulations). The Chinese regulatory environment was largely indifferent in the 2010s, allowing Chinese miners to exploit some competitive advantages provided by access to cheap electricity and to already existing manufacturing ecosystems. Starting in 2013, Chinese miners began the rapid scaling of computational infrastructure, with the introduction of ASIC miners (application-specific integrated circuits), specifically designed for bitcoin mining. According to the Cambridge Bitcoin Electricity Consumption index, by 2019, China accounted for approximately 74% of the global bitcoin hashrate, effectively functioning as the infrastructural backbone of what was often represented as a decentralized system.[1]

This concentration was not accidental. It reflected the convergence of several material conditions: proximity to ASIC manufacturing, improvement of ASIC technology, access to cheap power, the availability of buildable infrastructure, and a regulatory environment that, for a long time,

[1] https://ccaf.io/cbnsi/cbeci/mining_map

remained ambiguous or indifferent, rather than explicitly prohibitive. Together, these conditions allowed China to host the durational and energy-intensive work required to sustain the bitcoin network at scale. These spatial alignments expose how 'borderless', decentralized digital finance is nonetheless grounded in place-specific material conditions.

This infrastructural centrality came to an abrupt end when the Chinese government moved to shut down domestic crypto-mining operations in 2021. China's share of the global hashrate fell dramatically following the regulatory crackdown, dropping from dominant levels to a much smaller fraction within a short period of time. Despite this, China's infrastructural contribution remains significant to this day, pointing to mining having moved under the radar. Estimates are that, in 2024, China took between 14-21% of the global hashrate.[2] Even as more mining rigs were shutting down in Xinjiang province, Chinese mining pools continue to be among the largest contributors to the global hashrate.[3,4] At the same time, the majority of specialized ASIC mining rigs are designed and manufactured by Chinese companies, embedding Chinese hardware and operational influence deeply in the global network, underscoring the ongoing material presence of Chinese actors and technologies in the industry.

After the 2021 regulatory ban, geopolitical reconfigurations forced large parts of the industry to migrate, most notably to the U.S. [Onat et al. (2024)], but also to Kazakhstan and to Russia. Kazakhstan achieved a share of 18% of the global hashrate in 2021. Paradoxically, the war in Ukraine has not prevented an increase in Russia's share of bitcoin mining from 11 to 16%, and bitcoin has been legalized in Russia in 2024.[5] It is notable that, among these developments, Europe's share in bitcoin mining has been reduced to 4%, while the U.S. share is almost back to where it was in 2009 (40% today versus 50% in 2009). We need to insert here the caveat that the use of VPNs makes accurate estimates of some countries' share of the hashrate difficult (notably China's and Russia's).

Overall, the redistribution of mining capacity that followed did not simply represent a geographic relocation of computational hashing. It marked a shift in the geopolitical regime of mining, as operations were forced to be reorganized across a different assemblage of energy markets, regulatory environments, infrastructural forms, and capital constraints. Yet this mobility is constrained by the physicality of equipment and energy systems, and more so by different temporal regimes and construction timelines. Mining, which emerged as a distinct economic sector in the late 2010s, has not been developed in isolation from other sectors, most notably the energy and the hardware ones. This has forced miners and other adjacent actors into opportunistic searches for profit generation, to which we turn in the next section.

3. When machines follow rivers: opportunistic searches for profit generation

Cheap electricity and stable power sources are the primary conditions supporting bitcoin mining. Mining relies on getting machines online and plugged into the network to start contributing computing power to process information into blocks. Whenever a block is successfully mined,

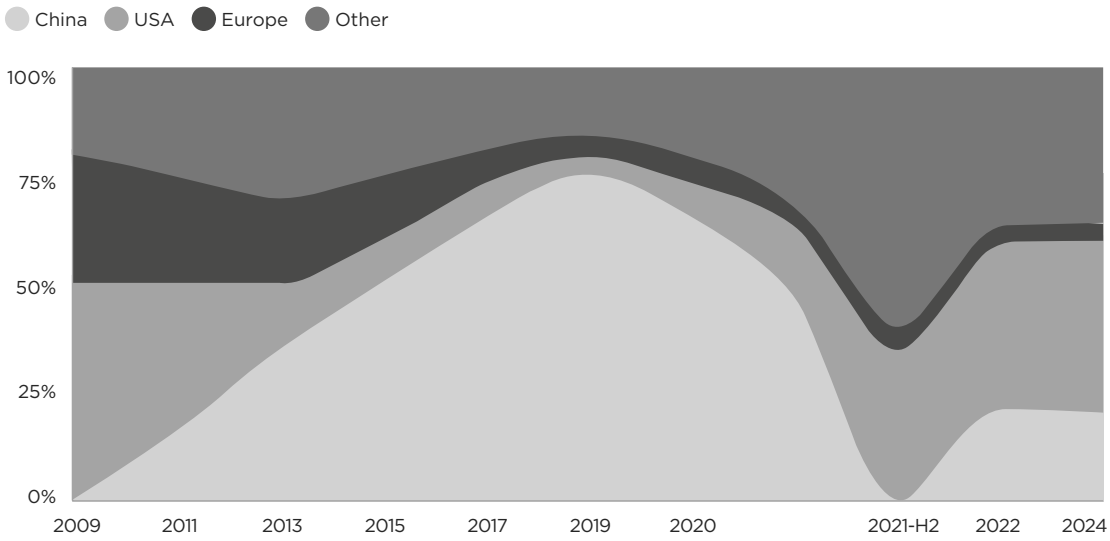
[2] https://www.coindesk.com/markets/2025/11/24/china-returns-as-third-largest-bitcoin-mining-hub-with-a-14-share-reuters?utm_source=chatgpt.com

[3] https://www.coindesk.com/markets/2025/12/15/bitcoin-hashrate-sees-sharpest-post-halving-drop-since-2024-amid-china-mining-machine-shutdowns

[4] A crypto-mining pool is a collaboration of individual cryptocurrency miners who combine and aggregate their computational power (hashrate) over a network to increase their collective chances of successfully mining a block and earning rewards with others. It functions like a cooperative 'super-miner' union to provide more frequent and predictable outcomes.

[5] https://atlas21.com/the-growth-of-bitcoin-mining-in-russia-an-overview-of-the-sector

Figure 1: The geographic distribution of the bitcoin hashrate over time



(This figure has been produced using Opus 4.5. Data sources are at the end of this article.)

the miner who wins the competition receives the block reward in bitcoin [Lo and Medda (2019)]. In the 2010s, the network was not yet highly competitive, and the intensity of competition was much lower than today. Yet, while miners had to engage in competitions for obtaining the bitcoin reward, they also had to pay attention to the cost of their operations, a significant part of which was given by electricity prices.

Bitcoin miners had always had to look for cheap electricity in order to sustain profitability. This partly explains why in the early days (2010 to 2021) their activities were located in mainland China, where they had access to comparatively cheap coal and hydroelectric power in inner Mongolia, Szechuan, and Yunnan, respectively. Moreover, they could benefit from the fact that not all hydroelectric power plants were connected to the national grid: this enabled miners to negotiate preferential, private electricity rates that deviated from the electricity prices charged in the national grid. On their side, power plant operators saw the arrival of bitcoin miners as an opportunity to recoup their investment in a shorter period of time:

"In the old days, if you built a water dam, it took eight years to repay your investment, and now if we built a crypto-asset mining center next to it, it could take half of what it took to return the investment" – bitcoin miner.

Seen from the perspective of a hydroelectric dam operator, having a crypto-asset mining center next to it and supplying electricity to such a mining center means that electricity could be converted into revenue far faster than through the conventional grid. In China, it takes one to three years to build a dam, depending on scale, and recouping the investment through selling electricity to the grid typically takes around eight years. By contrast, once mining machines are installed and brought online, returns can begin almost immediately, and the time to recoup the investment is halved. Miners effectively aligned two markets and, in doing so, altered the temporal structure of returns.

Szechuan's geography made this alignment especially viable. The province is rich in water resources and mountainous terrain, and the climate is relatively cool, often eliminating the

need for elaborate cooling systems (improving the efficiency of mining rigs). Electricity generated by small and medium hydropower plants (not all connected to the national grid) could thus be directly translated into hashing capacity. Hydropower dams rely on gravitational energy, spinning turbines to generate electricity; bitcoin mining, in turn, converts this electricity into hashing power and into bitcoin.

Setting up a mining center means purchasing ASIC machines (manufactured in China and provided by a Chinese supplier), installing containerized mining units, each capable of hosting up to 300 machines, and installing cooling systems (also provided by a Chinese manufacturer). In this sense, bitcoin mining as an industry cannot be seen in isolation from the hardware manufacturing and energy industries. Opportunistic searches for profit were not limited to actors from the energy sector; hardware manufacturers, on their side, saw the emergence of bitcoin mining too as an opportunity to generate new profits. Oral accounts point to adjacent industries as entering mining as a way of increasing sales but also of generating new profits:

"In the early days, a lot of the miners [who invest in owning ASICs] actually did not understand the value of bitcoin, but a lot of the infrastructure builders or hardware producers did invest in it. They, for example, are socket makers, plug sellers. They saw that people were busy setting up machines and data centers and ordering large quantities of sockets and plugs from them. They would send their sons to ask what these people were doing, and that's how they found their way into bitcoin mining. This was their quick way to pivot from hardware provider to crypto-asset mining" – bitcoin miner.

Many of the early participants in bitcoin mining did not come from finance or speculative trading, but from hardware supply, electrical equipment, and infrastructure-related businesses. They were already embedded in the same material world of sockets, cables, power equipment, cooling, and construction. For these actors, entering mining was not initially a move into a financial market, but an extension of existing hardware and infrastructure trading to a new application. Bitcoin mining successfully aligned a number of players that were in search of profit opportunities.

4. The emergence of a migratory industry

As mining operations are tightly coupled to energy infrastructures, disruptions in power supply (e.g., breakdowns, price volatility, unstable supply) introduce further pressures on operators. Miners were often notified in advance of planned outages or seasonal adjustments in electricity provision. Their response has been threefold: migrate in search of cheaper power sources; build your own power sources or secure preferential contracts; and innovate technologically. Let's look first at migration.

Notices of power disruptions trigger a cascade of logistical actions: machine owners must rapidly secure alternative hosting arrangements, negotiate access to other data centers, and physically relocate equipment in order to preserve continuity of operation. Here, the work of mining extends beyond computation to include the coordination of multiple material markets – hardware, energy, hosting, and logistics, each governed by different rhythms and constraints.

"When there is an electricity shortage on one site, you can see miners will arrange vehicles and carts of horse carriages to carry their machines to another site to host. As this is usually a remote mountain area, you can see there are also traffic jams in these small rural passages" – mining site manager and mining machine technician.

When the Chinese government started cracking down on bitcoin mining, many miners migrated to Kazakhstan, which was attractive not only because of geographical proximity and territorial size (about as large as Western Europe), but also because of cheap power resources. However, migration itself is costly: one could perhaps load mining rigs on horse drawn carts when a mine leaves a remote mountain location but, as operations grow, so does the cost of migrating them, and horse-drawn carts become impractical. An alternative to migration is to build one's own power plant. Firms may collaborate with energy companies to secure remote locations with existing power supplies and reconfigure them into higher-voltage sites, or they may invest in building and operating their own power infrastructure, including hydropower facilities, nuclear plants, wind farms, or solar installations. These plans involve not only substantial upfront capital investment and long-term commitments, shaping both operational flexibility and financial risk, but also the specialized expertise of energy generation and land acquisition associated with proximity to natural resources.

A third way in which bitcoin miners have addressed efficiency pressures is technological innovation. Owning an ASIC mining rig entails a substantial upfront investment, with prices varying according to computing power and energy efficiency. Newer generations of machines offer higher hashrates and improved thermal performance, enabling them to solve cryptographic puzzles more efficiently while reducing the risk of overheating and forced shutdowns. Yet the economic value of such machines is inseparable from their continuous operation. Mining revenue is generated only when machines remain online; any interruption immediately halts income generation, while the hardware continues to age and depreciate. The drive to improve efficiency has contributed to a high rate of technological obsolescence, with new machines being brought to the market every two to four years. Usually, spatial migration in search of cheaper energy sources has paired with technological migration to improved models:

"You can see a lot of machines were just thrown away and piling up on the site, which got abandoned. When the ASIC machines were too old, they normally did not provide any hashrate. These machines are not worth anything other than just broken bronze and wasted iron" – mining site manager.

The above quote indicates a distinctive regime in which profitability depends on continuous calculation, which depends on sustained uptime across extended durations. Time, in this context, is not something to be accelerated or deferred, but something that must be continuously endured and maintained. Periods of downtime are doubly costly: they represent both an immediate loss of revenue and an irreversible consumption of the machine's productive lifespan.

Mining requires technological investment on a scale that becomes clearer when miners talk about specific generations of ASIC machines. Earlier models such as the Antminer S9 released in 2016, which produces around 13–14 TH/s, were once widely used but are now effectively obsolete and often worth little more than scrap metal.[6] By contrast, newer generations such as the Antminer S17 (around 50–60 TH/s), which was released in 2019, and especially the S19 series (often exceeding 95–110 TH/s) released in late 2020 represent a completely different level of capital investment. At the time of our fieldwork, a single S17 machine could cost tens of thousands of RMB (U.S.$1,600 to over U.S.$2,500 per unit), depending on market conditions and availability.

[6] The Antminer S9, manufactured by Bitmain, was once one of the most widely-used ASIC miners for bitcoin and other SHA-256–based cryptocurrencies. As the first mass-market miner to adopt 16nm chips, it set an industry standard and dominated mining operations for a period. However, due to its relatively low hashrate and high electricity consumption by current standards, it is now widely regarded as economically obsolete under most contemporary power cost conditions.

For a container hosting two to three hundred machines, the hardware investment alone therefore easily runs into several million RMB. The latest models provided by Antminer are S21 and the S21XP Hydro, launched in 2025, which uses water cooling instead of air cooling. Today, the market is dominated by three large machine providers, namely Bitmain, Canaan, and MicroBT, with a fourth, Bitdeer, as an emerging competitor. All of them are Chinese companies.

As the network continuously upgrades to faster and more efficient ASICs, miners are forced to reinvest in new machines simply to remain competitive. Machines that fall below a baseline level of hashrate effectively drop out of the market: they can no longer 'do work' and therefore no longer produce value. In this sense, miners are not only coordinating across sites and infrastructures, but also across hardware generations, synchronizing their operations to the shifting temporal pace of the network itself.

5. The competition for speed

In the mid-2010s, miners had to pay attention primarily to infrastructural construction, hardware installation, and project delivery. Building a mining site resembled other industrial projects. ASIC machines could be ordered from manufacturers in China, cooling systems did not need to be especially sophisticated in mountain areas, and much of the work consisted of wiring, installation, and basic site preparation. Mining, at this point, was organized in the same temporal framework as hardware retail and infrastructure building, one oriented toward completion, installation, and gradual payback rather than rapid financial speculation. However, participation in mining also gradually exposed these actors to a different temporal regime, namely that of financial speculation. In the beginning, many miners did not treat bitcoin as a speculative asset at all.

"In the early days, as these miners did not see the value of these crypto assets, they would sell their bitcoins immediately to pay for electricity, and you can see that the bitcoin price will go down for a bit, because there would be a lot of sales going on."

Miners were largely operating according to cash-flow constraints, seeing bitcoin simply as a way to cover operating costs and repay electricity bills. They were not yet coordinating their actions with the speculative cycles of the crypto market, nor were they arbitraging between crypto and fiat. Only later did some miners begin to learn to operate within financial temporalities, holding coins, lending them out, or staking them to market makers who needed inventory to provide liquidity [Preda et al. (2023)]. In doing so, they gradually shifted from a purely infrastructural time horizon to one that also incorporated financial cycles and speculative expectations.

The reorientation to financial cycles increased competitiveness among miners and pressure to seek more efficient mining machines. These pressures were magnified by bitcoin's own halving cycles, which reduces the reward for miners every four years (or 210,000 blocks). This made miners compete for speed to become operational within set profitability windows:

"There is still a window to enter the bitcoin mining market in May or June this year, as the site will take more than three months before the next halving cycle happens. When the next halving cycle arrives, it will be much more difficult to mine bitcoins, as the building of the mining site has become increasingly costly, and it will no longer be profitable."

The above excerpt is from a financial practitioner who continued to pitch bitcoin mining to institutional investors in 2022, at a time when entry costs were already high [Delgado-Mohatar et al. (2019)]. He relied heavily on financial models

to explain the opportunity to investors, projecting cyclical returns and long-term trends in bitcoin's valuation. The competition for speed, reminiscent of high-frequency trading, requires in high rate of technological innovation (new mining rigs every two to four years, more efficient cooling, race to complete data centers) has also impacted the attempts to attract institutional capital.

6. Scaling up across geopolitical infrastructures

The factors discussed above – geopolitical pressures, migration in chasing efficiencies, interest alignment with the energy and hardware manufacturing sectors – have led over the course of the past decade to scaling up and consolidating mining operations. However, it did not necessarily result in the expected increased efficiencies and profitability. Scaling up brought its own challenges.

Following the 2021 mining ban in China, many Chinese miners began migrating to the U.S., particularly to areas around Detroit, where large tracts of available industrial land and high-voltage infrastructure seemed to offer favorable conditions for building large-scale data centers. Yet, the coordination required in this new setting proved far more demanding compared with their previous situation. As mining moved to the U.S. and became increasingly industrialized, infrastructural conditions resulted in a concentration of power among large operators, who had the capacity to secure energy contracts, deploy containerized data centers, and absorb the delays and risks of construction. Mining was no longer a small-scale entrepreneurial or hobbyist activity. It now required substantial capital, long planning horizons, and the ability to survive long periods without returns.

> *"However sophisticated financial models, valuation techniques or projective expectations may be, speculative projects cannot escape the slow, costly and materially enforced temporality of infrastructure.*

This shift became especially visible in the fieldwork with a group of Chinese financial practitioners who attempted to pivot into bitcoin mining by building their own site in the U.S. Coming from finance and investment backgrounds, they approached mining with strong expectations of speed and return. The project lead and site manager in the team had previous experience in mining. Yet, they quickly found themselves entangled in a web of slow, fragmented, and interdependent infrastructural timelines, with waiting times steadily accumulating.

"The main investors have arrived at the site that we would like to build, but for two weeks, we could not do anything but wait. The plan for tomorrow wasn't clear either. We may go to the office, but the main investor may want to go to the water testing lab. As I have told him many times, the water testing lab asks us to wait for results. Our boss keeps saying we should visit them tomorrow and see if we can speed things up" – fieldnote excerpt.

When the ethnographer was asking the schedule for the next day, miners showed concern that they had been put on hold to wait for the different pieces of their project to fall into place. A water test was required because the site planned to use local water for cooling, and its quality had to be verified in advance to ensure it would not clog or damage the circulation system. Until the results were confirmed, neither the cooling system nor the broader site design could be finalized.

Likewise, time was no longer something that could be accelerated through decision making or capital injection. Instead, the project became subordinated to a series of waiting periods imposed by contractors, laboratories, regulators, and landowners. The infrastructure builders faced labor shortages and would not start work until finishing previous projects. The group had to wait for a real estate agent to secure land from a landlord who owned most of the industrial property in the area and had little incentive to lease it to them. Water testing alone could take weeks.

At the same time, technical incompatibilities required extra time for configuration and further delays. All ASIC machines were manufactured in China and required different voltage standards. This meant transformers had to be installed, and new power lines extended across fields. Locally sourced transformers could take up to three months to procure. Faced with this delay, the miners considered ordering one from China and shipping it to the U.S., trading one form of waiting for another.

Energy procurement introduced yet another temporal regime. Electricity had to be secured through contracts in a volatile, market-based pricing system. Locking in a fixed price required large prepayments of millions U.S. dollars, further increasing upfront capital requirements. However, if the price was not locked in beforehand, electricity prices fluctuated seasonally, especially in summer when household consumption rose. If prices spiked, mining could instantly become unprofitable, forcing operations to shut down or scale back. In this sense, the maintenance team was not only keeping machines running but also learning to operate within the specifics of the electricity market itself.

On the one hand, miners were confronted with halving cycles, price expectations, and return models. On the other hand, they were confronted with construction time, permitting time, procurement time, and grid connection time. Project managers attempted to bridge this gap using financial models, projecting cyclical returns and long-term price trends with a linear temporality. Bitcoin training sessions for investors emphasized the halving cycle with a cyclical perspective. However, these abstract temporal projections struggled to align with the day-to-day realities of keeping a site buildable, connected, and operational.

Migrant miners arrived as speculators, expecting machines to be plugged in and producing returns quickly. In practice, however, they depended entirely on local infrastructural builders, grid connections, permits, and construction schedules. The gap between financial expectations and material timelines became a persistent source of friction. The amount of bitcoin earned per unit of hashrate had declined sharply as network difficulty increased, and returns were no longer generous enough to tolerate long delays or offline ASIC machines. With thinner margins and higher capital intensity, mistakes in timing became extremely costly. A site that came online too late or with electricity at too high a rate could easily become unviable before it ever recovered its investment.

The cyclical and projective temporality of financial models sat uneasily alongside the day-to-day work of building and operating a site: securing land, installing transformers, waiting for permits, testing water quality, and keeping machines online. The shift from abstract future-oriented valuation to the linear, interrupted, and delay-prone temporality of construction and maintenance repeatedly disrupted project timelines and investor expectations.

7. Discussion and conclusion

However sophisticated financial models, valuation techniques, or projective expectations may be, speculative projects cannot escape the slow, costly, and materially enforced temporality of infrastructure. They ultimately depend on energy systems, hardware, land, buildings, and organizational arrangements whose construction, maintenance, and amortization unfold in durational and enforced path-dependent time.

What the case of Chinese bitcoin mining shows, from early hydropower sites in Szechuan to large-scale industrial projects in the U.S., is not simply a story of technological change or regulatory displacement. It is a story about temporal misalignment across markets: between speculative financial time and infrastructural time, between cyclical valuation models and linear construction processes, and between expectations of rapid returns and the endurance required to build and sustain material systems. When resource-based markets are reorganized through financialized expectations, coordination does not fail simply because of uncertainty, but because of fundamentally different temporal orders, speculative, political, geological, and infrastructural, which cannot be fully reconciled.

Taken together, the findings show how the impatience of speculative valuation collides with the costly and time-consuming work of aligning energy, infrastructure, and hosting markets, producing a persistent tension between building durable data-center capacity and relying on temporary, opportunistic hosting arrangements to capture immediate speculative returns. Aligning expectations between finance and the intersecting markets for energy, hardware, and infrastructure is always partial and frictional, and this friction appears not as an accident but as a constitutive feature of mining at scale.

More broadly, the financial mindset of 'bridging the gap' in search of profit repeatedly underestimates the material time required to build something durable. In several cases observed in the field, this produced infrastructures that were hastily assembled, fragile, or difficult to maintain over the long term. As one miner reflected, earlier efforts to convert hydropower into faster returns already revealed this tension: while a hydropower facility might take many years to repay its investment through conventional electricity sales, mining promised much quicker returns, but only by leaning heavily on infrastructures whose own temporalities could not, in fact, be accelerated.

References

Delgado-Mohatar, O., M. Felis-Rota and C. Fernández-Herraiz, 2019. "The Bitcoin mining breakdown: Is mining still profitable?" Economics Letters, 184, https://doi.org/10.1016/j.econlet.2019.05.044

Lo, Y. C. and F. Medda, 2019. "Bitcoin mining: Converting computing power into cash flow," Applied Economics Letters, 26(14), 1171-1176, https://doi.org/10.1080/13504851.2018.1540841

Onat, N., R. Jabbar, M. Kucukvar, T. Wakjira, A. Kutty and N. Fetais, 2025. "Carbon footprint of global Bitcoin mining: Emissions beyond borders," Sustainability Science, 20(1), 173-189, https://link.springer.com/article/10.1007/s11625-024-01576-5

Preda, A., R. Xu and J. Valk, 2023. "Purity and dangers: Market making, structural uncertainty, and circuits of exchange in the cryptoeconomy," Economy and Society, 52(3), 399-420, https://doi.org/10.1080/03085147.2023.2242187


Monetary machinations
banks and currencies

50. **Central bank digital currencies in an era of financial fragmentation: assessing the potential financial crime risks**
Joshua Tjeransen, PhD Candidate, Kings College London

60. **The trade imbalance network and currency fluctuations**
Ai Jun Hou, Professor of Finance, Stockholm University
Lucio Sarno, Professor of Finance, University of Cambridge and CEPR
Xiaoxia Ye, Professor of Finance, University of Nottingham

70. **Bilateral central bank swap lines and their implications for the global financial order**
Ayca Zayim, Associate Professor of Sociology, Mount Holyoke College

Central bank digital currencies in an era of financial fragmentation:

assessing the potential financial crime risks

Author | **Joshua Tjeransen** | PhD Candidate, King's College London

Abstract

Central bank digital currencies (CBDCs) are an emerging financial innovation. Developed and issued by central banks, their purpose is to act as digital cash, ensuring the uniformity of money amid declining cash usage. Initially designed for domestic use, international applications have been debated and piloted. However, in an era of financial fragmentation and increasing monetary sovereignty, several financial crime risks need to be considered by CBDC developers, policymakers, regulators and financial institutions. Anticipating these risks is crucial if the associated harms are to be minimized and the success of CBDCs is to be ensured.

1. Introduction

Financial fragmentation, which is the disintegration of financial markets or systems into isolated segments, is often caused by differing national regulations and supervisory practices that limit the free flow of capital, liquidity, or services across borders [Catalan et al. (2023)], and is a longstanding issue. Financial fragmentation can arise for various reasons, including geopolitical tensions, changes in international trade, supply chain vulnerabilities, restrictions on capital flows, and regulatory divergence [Alfonso et al. (2024); BIS (2025); Claessens (2019)]. An early example of financial fragmentation would be countries reasserting monetary independence after WWI by abandoning the gold standard in favor of domestic monetary policy. This led to competitive devaluations and capital controls, which fragmented global finance and deepened the Great Depression [Meltzer and Fiedman (2025)].

Other examples include the breakdown of the Bretton Woods system in the early 1970s following the United States' suspension of the dollar's convertibility into gold to safeguard domestic interests [Office of the Historian (n.d.)]. During the Asian Financial Crisis of 1997 and 1998, several East Asian countries reintroduced capital controls and exchange restrictions to protect domestic markets from speculative attacks [Tamirisa (2001)]. This reinforced the regional fragmentation of capital markets that endured for years. Similarly, despite the single market in the European Union (E.U.), the 2008 Global Financial Crisis led member states to ring-fence capital and liquidity within their national borders. Variations in the national implementation of post-crisis regulatory reforms cemented this fragmentation [FSB (2019)].

In recent years, geopolitical tensions have led to increased financial fragmentation, driven by greater monetary sovereignty, the idea of not being beholden to any one system or currency [Gianviti (2008)]. Monetary sovereignty, understood in this context, has been exacerbated following the annexation of Crimea in 2014 and the 2022 invasion of Ukraine and the systematic use of sanctions.

Russia's response has been to develop its own domestic payment system and card network to lessen dependence on Western infrastructure [BBC Monitoring (2024)], and also explore alternative payment mechanisms, including CBDCs. The financial decoupling between the United States and China has followed a similar trajectory in response to tariffs and global power politics, with increased restrictions on foreign investment and the creation of alternative payment systems, including CBDCs, to minimize reliance on the U.S. dollar. This dynamic is evident in the digital asset space, with national approaches to regulating cryptocurrencies varying considerably and contributing to financial crime issues. Conversely, we are seeing very similar issues emerge with the governance of artificial intelligence and how differing interpretations and competing national agendas limit unified approaches and exacerbate problems [Hendrix (2026)].

2. CBDCs and financial crime risks

CBDCs pose a distinct challenge because they are a government-backed initiative. On initial consideration, this may seem like a simple pretext. Yet, if CBDCs do emerge as a cash equivalent, meaning they offer all the benefits while lacking only a few obvious differences, such as tangibility or complete anonymity, we could see mass adoption. When this is contrasted with today's financial sector, which has a checkered past in the court of public opinion, a new way to hold and spend money that is separate from established channels may resonate with a large share of consumers. Due to the way current payment systems operate, there are obvious air gaps that limit the large scale of criminal activity or general malfeasance. However, CBDCs designed or envisioned to be interoperable and international will create a pool of potentially hundreds of millions of victims. A parallel example could be the use of email as a communication method. It did not take long for a seemingly insignificant means of electronic communication to consume the entire world. Today, email addresses are a necessity with their origin founded in interest rather than need.

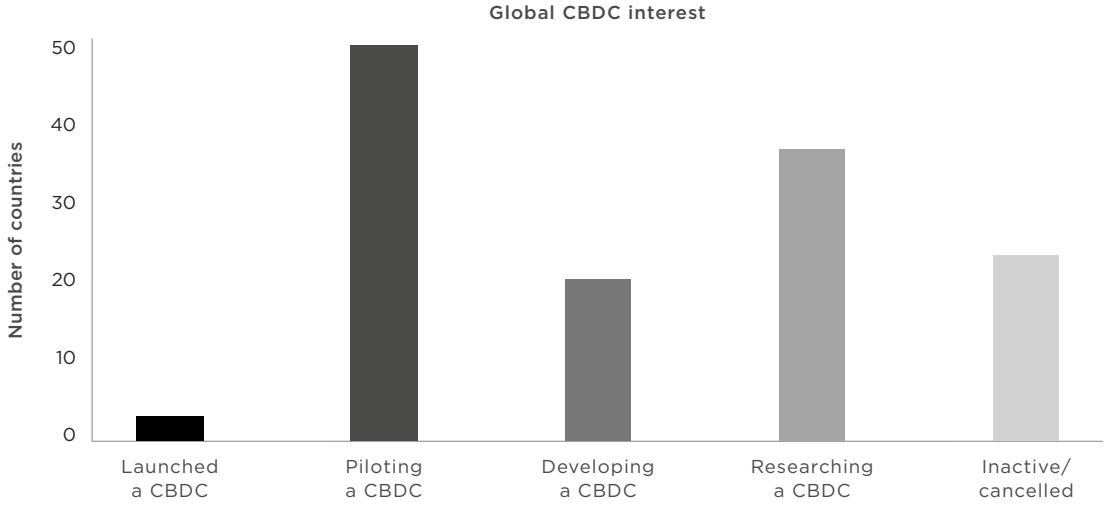
Financial fragmentation could also shape CBDC adoption as models compete for customers, offer perks such as cash back or interest, the latter of which is now a reality in China [Chhangani (2026)]. In turn, we may, and in many ways already are, see countries respond in kind. The E.U. and the U.K. are pursuing their own CBDCs, which may be heavily promoted to avoid having to ban or restrict a foreign CBDC. The requirement to offer an alternative as a method of amelioration is a lesson likely learned from the backlash of banning TikTok and restricting Huawei 5G [Friis and Lysne (2021); Gray (2021)]. Nonetheless, if CBDCs are introduced via a walled-garden approach, we may see fewer issues, as the pool of potential victims is smaller. Yet, the circumvention of sanctions is considered a primary reason for CBDC projects in Russia and China [Prokopenko (2024)], suggesting that internationalization will be a core tenet for future use. This argument ties in with China's development of its CBDC in 2014 in the aftermath of the Crimea invasion and the ensuing raft of financial sanctions levied on Russia.

The 2008 Global Financial Crisis illustrates how financial fragmentation can amplify financial crime risks. In the U.S., light-touch regulation and limited oversight of mortgage lenders and investment banks enabled the unchecked growth of subprime lending, including 'no income, no job, no assets' loans, which were supported by inflated property appraisals. Unlike more stringent supervisory frameworks elsewhere, the fragmented and permissive U.S. regulatory landscape allowed high-risk loans, often containing fabricated borrower information, to be packaged into mortgage-backed securities and collateralized debt obligations that were misrepresented to investors [Financial Crisis Inquiry Commission (2011)]. The absence of coordinated oversight enabled regulatory arbitrage, allowing the trade of extremely high-risk financial products. Although a lack of stringent regulation contributed, the situation was exacerbated by financial crimes that exploited it, including mortgage, securities, and accounting fraud.

Cryptocurrencies also show how financial fragmentation can amplify exposure to financial crime. Without harmonized standards, inconsistent and light-touch regulation has led to fraudulent initial coin offerings, market manipulation tactics such as pump-and-dump scams, and exchange fraud. Crypto assets have also been widely used for money laundering, storing illicit proceeds, and evading sanctions [Europol (2025)]. Regulatory divergence has caused some jurisdictions to adopt permissive frameworks, while others impose partial restrictions or total bans. This patchwork approach fosters regulatory mismatch, allowing offenders to move funds through the least restrictive or capable jurisdictions.

Despite increased global enforcement and understanding, the rapid pace of innovation and regulatory disparity continue to be exploited. Building on the rise of cryptocurrencies, nations continue to recognize the strategic importance of digital assets for economic sovereignty and technological progress. Though nationalist agendas bolster financial fragmentation and

Figure 1: Differing national interests in CBDCs and the respective stages1



[1] Information source – https://www.atlanticcouncil.org/cbdctracker/

undermine the efficiency and stability that a coordinated approach to digital assets could offer [Bilotta and Passos (2025)].

This divergence is already evident, with several countries adopting contrasting strategies, such as implementing comprehensive digital asset regulations, advancing national CBDC projects, or opting for fewer or more lenient rules. The United States has chosen to promote a U.S.$-backed stablecoin while banning the creation of a domestic CBDC [The White House (2025)], primarily citing privacy concerns and the need to safeguard national sovereignty. China prohibited cryptocurrency use in 2021 [BBC (2021)], citing concerns over illegal activity, financial stability, and financial security. Concurrently, China launched its own CBDC, the e-CNY, which has undergone successful pilot programs and international testing with other central banks [Bai et al. (2025); PBC (2021)]. The E.U. has adopted a mixed approach, introducing stringent regulation of crypto assets and developing an E.U.-wide CBDC project, both aimed at modernizing a layered and fragmented payment system and strengthening the E.U.'s strategic and financial independence [ESMA (2025); ECB (2023)].

Although nationalistic approaches and limited multilateral policies contribute to financial fragmentation, ultimately hampering growth and economic security, these issues often take a backseat to the illicit exploitation of misalignment, specifically financial crime. Scholarship has extensively examined various aspects of this problem: the misalignment of regulation and the gaps exploited by financial criminals [Levi and Reuter (2006)], the role that improved and targeted regulation can play in reducing financial crime [Ferwerda (2008)], and the rate at which financial crime evolves in the absence of preventative measures [Leuprecht et al. (2023)]. Existing research argues that siloed approaches and differing risk appetites across jurisdictions facilitate financial crime [Button et al. (2025); Longa (2025)], and in an increasingly globalized financial world, this dynamic continues to evolve, generating new challenges.

3. Transcendence

CBDCs present a novel way to engage with and use money. Driven by changing social dynamics, CBDCs have become a key part of an era marked by declining cash usage and growing financial digitalization. Originally designed for domestic use, the concepts behind CBDCs have expanded to include interoperability with other CBDCs or international exchange [BIS (2024); VISA (2025)]. Although this development is positive, the internationalization and greater interoperability of CBDCs introduce various financial crime risks, especially when financial fragmentation is taken into account. This section will examine how anonymity, offline functionality, instantaneous settlement, and holding limits may be affected by financial fragmentation, underscored by monetary sovereignty, and outline the potential risks.

3.1 Anonymity

Like cash, CBDCs aim to provide users with the ability to transact anonymously. However, many designs incorporate due diligence checks (verification of identity) to prevent issues such as money laundering. This contrasts with cash, which requires no such checks and allows users to interact completely anonymously. The reasons for introducing due diligence vary, but the digital nature of CBDCs, existing anti-financial crime norms, and the fact that CBDCs will be interoperable and potentially exchangeable are key considerations. The level of anonymity offered by CBDCs varies, however. The Bahamas Sand

> *Despite increased global enforcement and understanding, the rapid pace of innovation and regulatory disparity continue to be exploited.*

Dollar, the first CBDC to become operational, provides an initial tier with no due diligence requirements. However, it imposes a monthly transaction limit of U.S.$1,500 [Central Bank of the Bahamas (2019)]. As CBDCs evolve and their use becomes more widespread, challenging cash and electronic money, debates about anonymity are likely to intensify. Concerns about data privacy and the desire for an anonymous CBDC have already garnered interest, with many opposing additional perceived oversight [ACLU (2022); Chapple (2025)].

If CBDCs become interoperable internationally as envisioned, then differing levels of anonymity may present financial crime risks. If some CBDCs opt for a completely anonymous model with high transaction limits, we might see a similar pattern to that of cryptocurrencies, where users transfer funds to jurisdictions with minimal regulation, checks, or oversight to conceal illicitly obtained funds. Moreover, if a cash-out function is available, financial criminals could employ an anonymous CBDC as a conduit for money laundering or terrorist financing.

Traditionally, transferring funds overseas has been impeded by due diligence requirements, lengthy correspondent banking relationships, and high costs. This is why criminals continue to exploit secrecy jurisdictions, cryptocurrencies, or cash smuggling. A related example, which clearly highlights the issue, is the use of prepaid cards in the E.U. before 2018, when the E.U.'s 5th Anti-Money Laundering Directive was introduced. Prior to that, anonymous prepaid cards were widely utilized for illicit financial activities, prompting the E.U. to lower the customer due diligence threshold from €250 to €150 when uploading funds, with a cap of €50 for online transactions [Chapple (2025); Königreich (2019)]. If CBDCs become interoperable, financial crime may increase as criminals exploit those that emulate cash or offer similar levels of anonymity. Furthermore, weak regulation or challenges in investigative procedures, owing to anonymity, may aggravate these issues.

3.2 Offline functionality

CBDCs, like cash, will need to enable transactions when there is no internet connection. To achieve this, value can be stored locally on a device (such as a phone, smart card, or hardware wallet) and transferred directly between the payer and the payee via cryptographically signed messages. Both devices update their balances instantly, and reconciliation occurs when they reconnect to the central ledger. While this design supports resilience during network outages and promotes financial inclusion for those without constant connectivity, it also carries significant risks of financial crime.

Offline payments may enable transactions to occur without real-time monitoring or identity verification. This creates opportunities for money laundering and terrorist financing, especially if high-value offline transactions are permitted or if interoperability allows value to move across borders undetected. In areas with inconsistent internet and mobile coverage, criminals have exploited vulnerabilities during network outages by engaging in double spending or false reversals. Such dynamics show how offline-

enabled payment instruments, while supporting inclusion and resilience, can replicate cash-like vulnerabilities in the digital realm. In a fragmented international environment, different approaches to offline functionality could magnify these risks, allowing criminals to route funds through jurisdictions with more permissive offline models, as they do today with cryptocurrencies.

3.3 Instantaneous settlement

A core element of CBDC design is the ability to process payments instantly, with transactions cleared and settled immediately, as with cash. This improves efficiency and reduces counterparty risk by removing the delays found in current payment systems. However, this rapid processing also raises the potential for financial crimes. The lack of settlement delay narrows the timeframe for detecting and blocking suspicious activity, effectively removing a key layer of friction that currently helps identify and prevent such issues.

The 2016 Bangladesh central bank heist underscores the risks tied to nearly instant settlement. North Korean hackers exploited the SWIFT real-time messaging system to initiate fraudulent transfers from the bank's account at the Federal Reserve Bank of New York. Although multiple attempts were thwarted, approximately U.S.$81 million was successfully transferred to accounts in the Philippines and laundered through casinos before authorities could act [Hammer (2018)].

While this is nowhere near the speed of CBDC settlements, the example demonstrates how quickly illicit funds can disappear in a 'real-time' system. In a fragmented global environment, varying abilities to monitor and respond to real-time CBDC flows could lead to uneven enforcement, allowing offenders to exploit jurisdictions with limited capacity.

3.4 Holding limits

Many CBDC designs include holding limits to protect financial stability by preventing banking disintermediation. These limits also aim to reduce the appeal of CBDCs for illicit activities by restricting the amounts that can be accumulated or transferred. However, these restrictions could create new crime risks. High holding limits might encourage misuse by organized crime groups or lead to a new group of victims – those who choose to hold larger CBDC balances. Conversely, low limits may attract more users since lost or stolen balances are minimal. Still, this could lead criminals to focus more on CBDC users overall if widespread adoption occurs. Additionally, if there are significant jurisdictional differences in holding limits, illicit actors might move funds to more accommodating systems. Nonetheless, if designed and governed effectively, CBDCs could limit illicit financial activity as described above. Yet it is important to remember that, with all financial innovation, a retrospective approach to financial crime is the established approach. Anticipation of financial crime risks does not feature in the design stage, an issue seen across a whole host of financial innovations, from checks to credit cards. Hence, the current regulatory and supervisory paradigm focuses on post facto fixes to final crime issues.

4. Conclusion: policy recommendations and practical implications

CBDCs represent an innovative and vital form of exchange, supporting shifts towards a more cashless society, reducing dependence on commercial bank money, and enabling governments to assist citizens during crises. They also enable cash functionality, combined with improved interoperability and safety features. However, like many financial innovations before

them, CBDCs enter a landscape where financial crime seeks to exploit any and all loopholes. Additionally, if CBDCs are used internationally as anticipated, they will encounter diverse risk appetites, regulations, and public perceptions.

As noted above, the fundamental features that set CBDCs apart could also be their undoing. To ensure the success of CBDCs, policymakers must consider how financial crime could exploit design elements that initially attract users. Therefore, issues such as the misuse of anonymity, offline functionality, instant settlement, and holding limits should be central to all decisions on the internationalization and interoperability of CBDCs. Furthermore, continued assessment of the financial crime risks associated with CBDCs is essential. As we stand on the brink of a new digital financial era, we must avoid the obvious financial crime risks that could impact the vast majority. We missed this opportunity when introducing coins that were clipped, checks that were forged, or cash that was counterfeited.

In practical terms, CBDCs will pose immediate challenges that, if left unaddressed, could exacerbate the concerns mentioned above. The main issue is how existing anti-financial crime measures interact with CBDCs. Many financial institutions implement these measures dynamically, following a risk-based approach. This is important because it reflects the evolving nature of financial crime, allowing resources to be allocated where necessary. However, current approaches also consider the due diligence conducted by other financial institutions and entities involved in the transaction process. Given that CBDCs are likely to become interoperable with existing payment methods, they will introduce a new layer of risk and opacity, despite many governments advocating strict due diligence. Consequently, the unique aspects of CBDCs, as mentioned above, will require financial institutions to update or review their existing policies to assess how the potential financial crime risks linked to CBDCs might intensify existing threats. This broader issue has been observed before with the emergence of 'new payment methods' (NPMs). In a 2010 FATF report, the influence of NPMs and financial crime is discussed, noting that since 2006, the amount of funds moving through NPMs, such as prepaid cards and mobile payments, has increased, as has the related financial crimes [FATF (2010)]. However, the report concludes by stating that policymakers should aim to strike a balance between practical and comprehensive anti-financial crime measures and deterring NPM usage, out of concern that restricting NPMs could push users towards unregulated services or providers [FATF (2010)], which conflicts with the current association of NPMs and crime [Deloitte (n.d.)].

Thus, not only will policymakers need to consider the unintended consequences of CBDCs by anticipating how their uniqueness may create new opportunities for financial crime, but financial institutions will also need to evaluate how their anti-financial crime procedures must be updated and adapted. Translating the above into clear implications for practitioners, the author sets out the following assertions:

Financial fragmentation materially increases CBDC-related financial crime risk when CBDCs are interoperable or internationally accessible. As CBDCs move beyond domestic use, divergent regulatory standards, enforcement capacities, and risk tolerances will create exploitable asymmetries. Criminals will route value through CBDC systems that are more permissive.

CBDC design choices are not financially crime-neutral and must be treated as regulatory decisions, not purely technical ones. Core features such as anonymity tiers, offline functionality, instantaneous settlement, and holding limits directly shape financial crime risk exposure. Treating these features as engineering choices rather than as policy instruments risks embedding systemic vulnerabilities that cannot be fully mitigated ex post through supervision or enforcement.

CBDCs will externalize new compliance and operational risks onto financial institutions, even where central banks retain core control. Despite CBDCs being public money, financial institutions and/or payment providers will remain responsible for customer due diligence, transaction monitoring, and risk assessments at access points and interoperability layers. Existing risk-based AML/CFT frameworks will require recalibration, as reliance on upstream controls conducted by other intermediaries becomes less reliable in CBDC-enabled payment chains.

Failure to integrate anticipatory financial crime risk assessment into CBDC development threatens long-term adoption and legitimacy. Historical experience demonstrates that retrofitting controls after widespread misuse damages trust and imposes higher economic and reputational costs. If CBDC-related financial crime incidents emerge at scale, public confidence may erode rapidly, undermining policy objectives related to financial inclusion, monetary sovereignty, and payment system resilience.

References

Afonso, A., J. Alves, and S. Monteiro, 2024. "Sovereign risk dynamics in the EU: The time varying relevance of fiscal and external (im)balances," Journal of International Financial Markets, Institutions and Money, 94, 102026, https://doi.org/10.1016/j.intfin.2024.102026

American Civil Liberties Union, 2022. "Anonymity is not negotiable when it comes to digital cash," https://www.aclu.org/sites/default/files/field_document/cbdcs_fed_reserve_-_money_and_payments_comment_letter.pdf

Bai, H., L.W. Cong, M. Luo, and P. Xie, 2025. "Adoption of central bank digital currencies: Initial evidence from China," Journal of Corporate Finance, 91, 102735, https://doi.org/10.1016/j.jcorpfin.2025.102735

BBC, 2021. "China declares all crypto-currency transactions illegal," BBC News, September 24, https://www.bbc.com/news/technology-58678907

BBC Monitoring, 2024. "Briefing: Russia puts on brave face as allies snub its Mir payment system," https://monitoring.bbc.co.uk/product/b0001jc2

Bilotta, N. and N. Passos, 2025. "Digital asset regulatory fragmentation undermines global financial interconnectedness," Florence School of Banking and Finance, July 21, https://fbf.eui.eu/digitalasset-regulatory-fragmentationundermines-global-financialinterconnectedness/

BIS, 2025. "Sustaining stability amid uncertainty and fragmentation," Bank of International Settlements, https://www.bis.org/publ/arpdf/ar2025e1.pdf

Button, M., B. Hock, J.B. Suh, and C.S. Koh, 2025. "Policing cross-border fraud 'above and below the surface': Mapping actions and developing a more effective global response," Crime, Law and Social Change, 83(1), 5, https://doi.org/10.1007/s10611-024-10186-2

Catalan, M., F. Natalucci, Q. Mahvash, and T. Tsuruga, 2023. "Geopolitics and Fragmentation Emerge as Serious Financial Stability Threats," International Monetary Fund, https://www.imf.org/en/blogs/articles/2023/04/05/geopolitics-and-fragmentation-emerge-as-serious-financial-stability-threats

Central Bank of the Bahamas, 2019. "Project Sand Dollar," https://www.centralbankbahamas.com/viewPDF/documents/2019-12-25-02-18-11-Project-Sanddollar.pdf

Chapple, M., 2025. "Your data privacy is slipping away – here's why, and what you can do about it," The Conversation, https://doi.org/10.64628/AAI.pw5rqc3h6

Chhangani, A., 2026. "What to watch as China prepares its digital yuan for prime time," Atlantic Council, https://www.atlanticcouncil.org/blogs/econographics/what-towatch-as-china-prepares-its-digitalyuan-for-prime-time/

Claessens, S., 2019. "Fragmentation in global financial markets: Good or bad for financial stability?" BIS Working Papers 815, https://www.bis.org/publ/work815.pdf

Deloitte, n.d. "Payments Digitisation and Financial Crime Risk – Banking or Technology?" Deloitte Australia, retrieved September 25, 2025 from https://www.deloitte.com/au/en/services/consulting-financial/blogs/payments-digitisation-financialcrime-risk.html

ESMA, 2025. "Markets in Crypto-Assets Regulation (MiCA)," https://www.esma.europa.eu/esmasactivities/digital-finance-andinnovation/markets-crypto-assetsregulation-mica

European Central Bank, 2023. "Digital euro," https://www.ecb.europa.eu/euro/digital_euro/html/index.en.html

Europol, 2025. "Internet Organised Crime Threat Assessment (IOCTA) – Strategic, policy and tactical updates on the fight against cybercrime," https://www.europol.europa.eu/publications-events/mainreports/iocta-report

FATF, 2010. "Money Laundering Using New Payment Methods," Financial Action Task Force, https://www.fatf-gafi.org/content/dam/fatf-gafi/reports/ML%20using%20New%20Payment%20Methods.pdf

Ferwerda, J., 2008. "The Economics of Crime and Money Laundering: Does Anti-Money Laundering Policy Reduce Crime?", Review of Law & Economics, 5(2), 903-929, https://doi.org/10.2202/1555-5879.1421

Financial Crisis Inquiry Commission, 2011. "The Financial Crisis Inquiry Report," https://fcicstatic.law.stanford.edu/cdn_media/fcic-reports/fcic_final_report_full.pdf

Financial Stability Board, 2019. "FSB Report on Market Fragmentation, https://www.fsb.org/uploads/P040619-2.pdf

Friis, K. and O. Lysne, 2021. "Huawei, 5G and Security: Technological Limitations and Political Responses," Development and Change, 52(5), 1174-1195, https://doi.org/10.1111/dech.12680

Gianviti, F., 2008. "Current Legal Aspects of Monetary Sovereignty," in Current Developments in Monetary and Financial Law 4, International Monetary Fund, https://www.elibrary.imf.org/display/book/9781589065079/ch001.xml

Gray, J.E., 2021). "The geopolitics of 'platforms': The TikTok challenge," Internet Policy Review, 10(2), https://doi.org/10.14763/2021.2.1557

Hammer, J., 2018. "The Billion-Dollar Bank Job," The New York Times, May 3, https://www.nytimes.com/interactive/2018/05/03/magazine/money-issue-bangladeshbillion-dollar-bank-heist.html

Hendrix, J., 2026. "The Policy Implications of Grok's 'Mass Digital Undressing Spree'," The Tech Policy Press, podcast, https://player.captivate.fm/episode/d7d950cb-5a8c-4948-985f-78f74f62cdf8

Königreich, V., 2019. "The EU's Fifth Anti-Money Laundering Directive: A regulatory compliance perspective," https://www.nortonrosefulbright.com/en-gb/knowledge/publications/8f84c163/the-eus-fifthanti-money-laundering-directive-aregulatory-compliance-perspective

Leuprecht, C., C. Jenkins, and R. Hamilton, 2023. "Virtual money laundering: Policy implications of the proliferation in the illicit use of cryptocurrency," Journal of Financial Crime, 30(4), 1036-1054, https://doi.org/10.1108/JFC-07-2022-0161

Levi, M. and P. Reuter, 2006. "Money Laundering," Crime and Justice, 34(1), 289-375, https://doi.org/10.1086/501508

Longa, F.E.A., 2025. "Cryptocurrency and Money Laundering," American Journal of Industrial and Business Management, 15(02), 362-371, https://doi.org/10.4236/ajibm.2025.152017

Meltzer, A. and M. Fiedman, 2025. "Britannica Money," https://www.britannica.com/money/money

Office of the Historian, n.d. "Milestones in the History of U.S. Foreign Relations, 1969-1976," https://history.state.gov/milestones/1969-1976/nixon-shock

People's Bank of China, 2021. "Digital Currency Institute of PBC Joins m-CBDC Bridge Project," http://www.pbc.gov.cn/en/3688110/3688172/4157443/4196114/index.html

Prokopenko, A., 2024. "Can the Digital Ruble Shield Russia From Western Sanctions?", https://carnegieendowment.org/research/2024/10/russia-digitalruble-development?lang=en

Tamirisa, N., 2001. "V Capital Controls in Response to the Asian Crisis," in Malaysia: From Crisis to Recovery, International Monetary Fund, https://doi.org/10.5089/9781589060470.084

The White House, 2025. "Strengthening American Leadership in Digital Financial Technology," https://www.whitehouse.gov/presidential-actions/2025/01/strengthening-american-leadershipin-digital-financial-technology/

The trade imbalance network

and currency fluctuations[1]

Authors | **Ai Jun Hou** | Professor of Finance, Stockholm University
Lucio Sarno | Professor of Finance, University of Cambridge and CEPR
Xiaoxia Ye | Professor of Finance, University of Nottingham

Abstract

In a rapidly evolving global economy, understanding how global trade imbalances shape currency movements is increasingly important for investors and policymakers. This paper develops a new framework linking countries' positions in the global trade network to exchange rate fluctuations and currency risk premia. Extending standard two-country models to a multi-country setting, we show that a country's centrality within the global trade imbalance network determines how much currency risk financial intermediaries must bear. This leads us to construct the Centrality-Based Characteristic (CBC), a novel measure that captures how a country's trade interconnections influence its currency's riskiness. Using data on 41 currencies from 1995 to 2021, we find that CBC is a powerful and robust predictor of future currency returns. High-CBC currencies systematically earn higher risk premia, and a strategy that goes long high-CBC and short low-CBC currencies delivers an annualized Sharpe ratio of 0.65, outperforming traditional approaches such as carry trades and bilateral trade measures. Counterfactual analyses of the U.S.–China trade war and the Russia sanctions illustrate how geopolitical shocks reshape trade networks and spill over to third-party currencies. For practitioners, especially financial intermediaries involved in international transactions, CBC offers a new, theory-based signal to enhance currency allocation and risk management. For policymakers, it provides a tool to assess external vulnerabilities and systemic importance. Overall, the framework highlights the critical role of global trade interconnections in shaping currency dynamics.

1. Introduction

As the global economic order undergoes profound shifts driven by geopolitical realignments, technological transformations, and changing trade patterns, understanding the mechanisms linking international trade and currency dynamics has become increasingly vital. International trade and trade imbalances significantly influence global macroeconomic and financial outcomes. In recent years, concerns over global financial fragmentation have grown amid rising geopolitical tensions, shifting trade alliances, and the weaponization of financial flows. The Ukraine war, U.S.–China rivalry, and moves toward 'de-risking' supply chains have exposed the fragility

[1] This is a reprint version of the VoxEU Column article at: https://cepr.org/voxeu/columns/trade-imbalance-network-and-currency-fluctuations

of global trade and financial networks. In this context, understanding how trade imbalances affect currency dynamics is more urgent than ever.

Global trade drives exchange rate fluctuations and currency risk premia, as highlighted in various exchange rate determination theories [e.g., Gourinchas and Rey (2007); Gabaix and Maggiori (2015); Colacito et al. (2018); Maggiori (2022); Clayton et al. (2024)]. Empirical studies have established a predictive link between external imbalances and currency excess returns across countries [e.g., Della Corte et al. (2012); Della Corte et al. (2016); Riddiough et al. (2016); Della Corte and Krecetovs (2021)]. However, a key challenge arises when applying two-country theoretical models to multi-country empirical settings, as bilateral relationships may not generalize without additional assumptions. Hou et al. (2025) extend the theoretical framework of Gabaix and Maggiori (2015) to a multi-country context, incorporating the global trade imbalance network to uncover new insights into currency risk premia. The study provides a framework to capture the indirect, network-driven channels through which trade imbalances shape currency risk premia, especially under financial market imperfections.

2. Theoretical framework

The Gabaix and Maggiori (2015) model posits that countries with trade imbalances operate in imperfect financial markets, where financiers bear currency risk by buying the currency of deficit countries and shorting that of surplus countries. The financiers' risk-bearing capacity, limited by the riskiness of their balance sheets, influences exchange rates. Deficit countries' currencies must offer high return in the form of currency appreciations to compensate financiers for the risk they bear. The model identifies two critical factors for currency premia: financiers' risk-bearing capacity and individual countries' external imbalances.

Extending this to a multi-country setting, we incorporate the global trade imbalance network, which captures bilateral imbalances and interdependencies among countries. This network provides richer information about financiers' balance sheets than a two-country model. We use trade imbalance network centrality to measure financiers' risk-bearing capacity, reflecting the complexity of their global balance sheets. The centrality of a country in this network indicates the investment options available to financiers, as a country's net deficit or surplus creates opportunities for long or short positions in its currency, balanced by opposite positions in other currencies.

The complexity of these balance sheets increases with a country's network centrality, as it offers more investment options. A country's net deficit only partially reflects these options, as it induces deficits or surpluses in other countries, further shaping investment opportunities based on network proximity. Centrality quantifies this complexity, providing a more comprehensive measure of financiers' risk-bearing capacity.

The core mechanism here is that financial intermediaries (like large banks) act as the 'connective tissue' of global trade. When countries have trade imbalances, these intermediaries must hold the resulting currency risk on their balance sheets. This mechanism predicts that expected currency returns are positively associated with the trade imbalance network centrality.

3. Empirical analysis

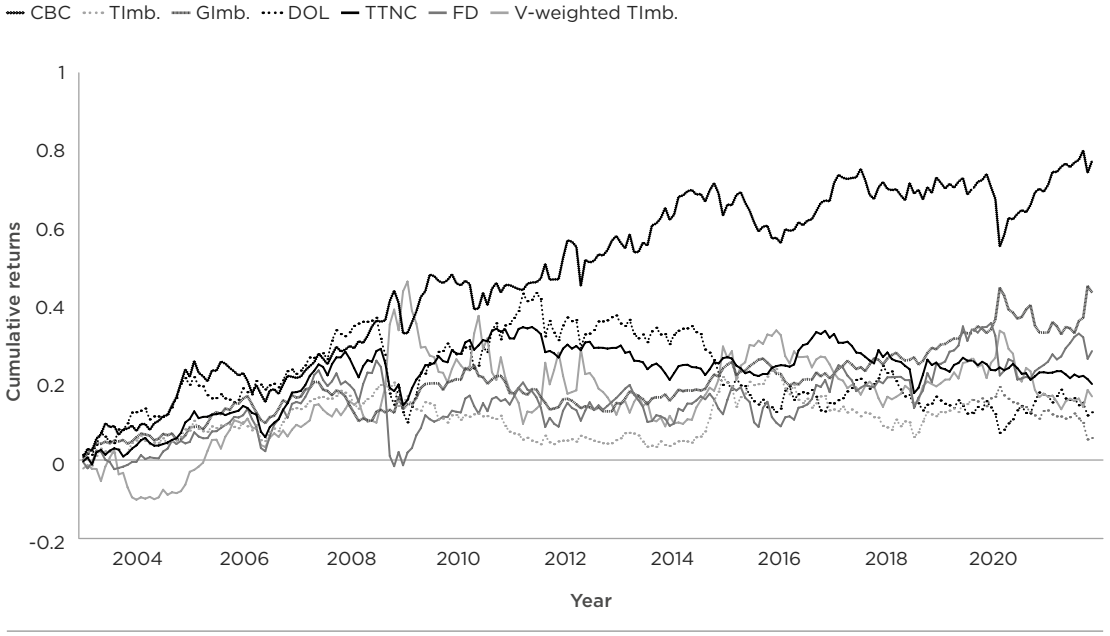
Using data for 41 currencies from 1995 to 2021, Hou et al. (2025) empirically test the model, using a Centrality-Based Characteristic (CBC) for each currency that combines network centrality and exchange rate covariances. Portfolio sorts based on CBC reveal that currency returns increase monotonically from low- to high-CBC countries. A high-minus-low portfolio (long high-CBC, short low-CBC) yields an annualized Sharpe ratio of 0.65, outperforming investment strategies based on total trade network centrality [Richmond (2019)], global imbalance measures [Della Corte et al. (2016)], and carry trades.

The CBC factor, defined as the time series of excess returns from the CBC-sorted strategy, captures unique information not explained by standard currency factors [e.g., Lustig et al. (2011)] or intermediary asset pricing factors [e.g., Adrian et al. (2014); He et al. 2017)]. Figure 1 plots the long-short portfolios' cumulative returns from 2003 to 2021. Several characteristics have strong performance before 2008, but after 2008, CBC clearly dominates relative to other sorting variables.

Figure 1 plots the cumulative returns of buying G4 and selling G1 in the portfolio sorting using as characteristics: CBC, total trade network centrality (TTNC), trade imbalance (TImb), forward discount (FD), global imbalance (GImb), and variance-covariance weighted trade imbalance (V-weighted TImb), as well as the average currency excess return (DOL).[2]

Figure 2 plots of average currency excess returns against CBCs for different subsample periods show a clear positive relationship. High-CBC countries like Mexico and New Zealand exhibit higher currency premia, while low-CBC countries like Japan and Thailand show lower premia. The figure nicely summarizes the insight of this exploration on the global imbalance trade network.

Figure 1: Cumulative returns over time in portfolio sorting comparison



[2] Source: Hou et al. (2025)

Figure 2: CBC versus currency risk premia

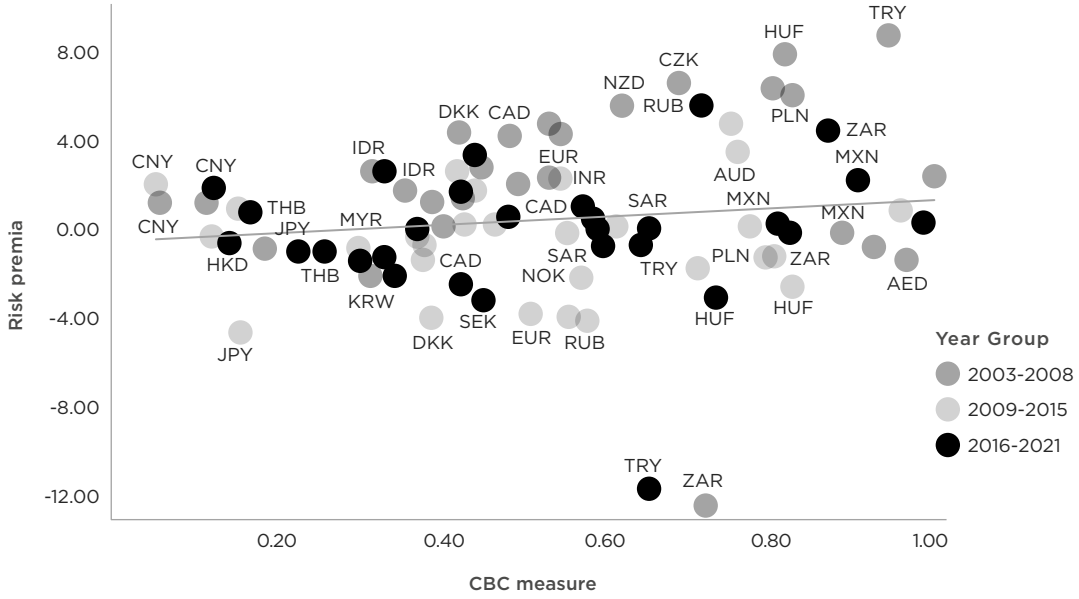


Figure 2 is the scatter plot of the global trade CBC measure of 41 currencies versus the average of annualized risk premia.[3]

Currency risk premia can be decomposed into three components: total imbalance, individual country importance, and neighborhood importance. A variance decomposition indicates that the neighborhood component accounts for approximately 68% of the variation in cross-sectional currency premia, underscoring the trade imbalance network's critical role. Figure 3 illustrates the decomposition in various parameter settings and the dominance of the neighborhood importance component in the figure highlights the essential role of trade imbalance network centrality.

In summary, our analysis from 1995 to 2021 confirms that this network position is a powerful predictor of currency performance:

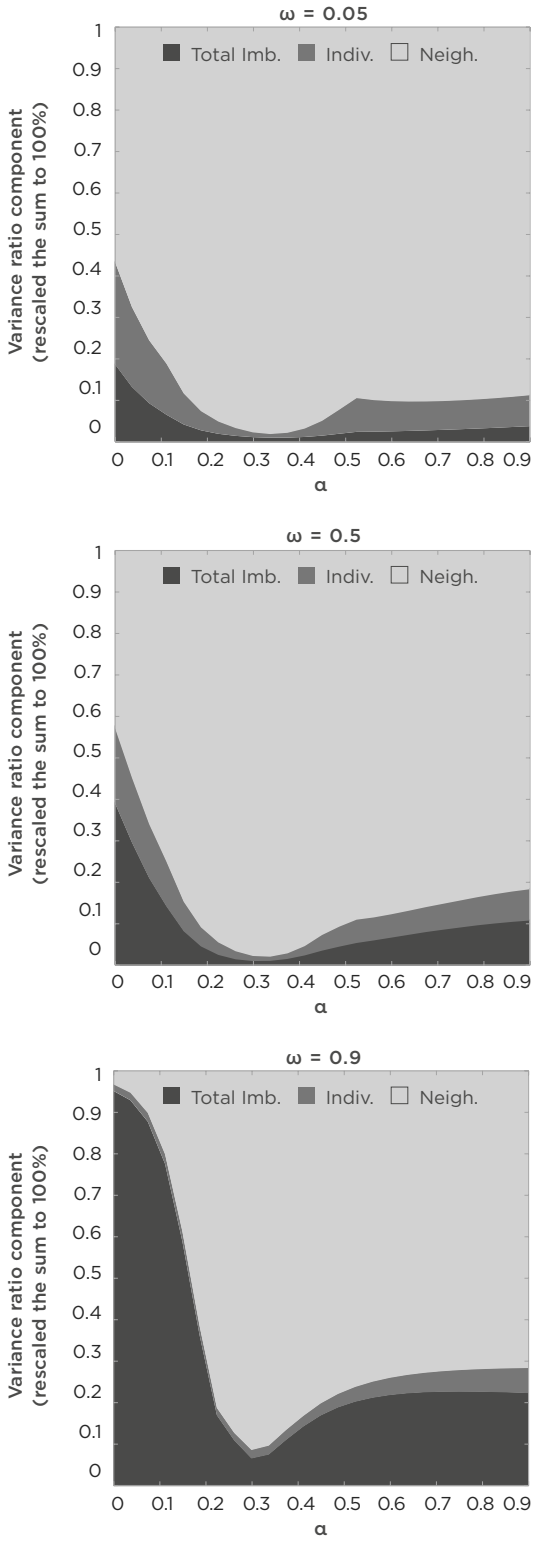
Superior returns: A strategy buying high-CBC (high-risk/high-reward) currencies and selling low-CBC currencies achieved an annualized Sharpe ratio of 0.65, significantly outperforming traditional carry trades and standard trade measures.

The 'neighborhood' effect: We found that roughly 68% of the variation in currency returns is driven not by a country's own trade balance, but by its 'neighborhood' – the trade imbalances of its closest trading partners.

Post-2008 resilience: While many currency strategies faltered after the 2008 financial crisis, the CBC-based approach remained robust, dominating other traditional indicators.

[3] Source: Hou et al. (2025)

Figure 3: Variance decomposition



The panels in Figure 3 plots the (re-scaled) variance ratios of total imbalance (total Imb.), individual importance (indiv.), and neighborhood importance (neigh.) given different values of α and w. α goes from 0 to 0.9 in all three panels. The left panel shows the variance ratios components for w = 0.05, the center panel for w = 0.5, and the right panel for w = 0.9.[4]

4. Applications

Hou, Sarno and Ye (2025) also investigate the dynamic behaviour of the CBC around major geopolitical and economic shocks, including the 2018–2019 US–China trade war and the 2022 sanctions imposed on Russia following its invasion of Ukraine. These case studies provide important insights into the endogenous response of trade networks to shocks and how such responses affect currency pricing.

In Figure 4, Panel (a) plots the global trade imbalance network based on 2017 pairwise bilateral imports and exports, Panel (b) plots the same network with pairs CNY and USD replaced by the data in 2019.[5]

A counterfactual analysis shows that during the 2018 U.S.–China trade war, there was a marked shift in CBC rankings: the effect of the U.S.-China trade war clearly goes beyond impacting these two countries, as bystander countries increased their trades with the U.S. and China. We find that the rank of CNY has not changed and the nine currencies noticeably affected are THB (3% rise in the percentage ranks), RUB (3% drop), GBP (3% rise), IDR (3% drop), PHP (3% rise), SEK (3% drop), ZAR (3% rise), SAR (3% rise), and MXN (6% drop). Figure 4 shows the significant impact of the 2018 U.S.–China trade war on global currency premia via the model's counterfactual analysis.

[4] Source: Hou et al. (2025)
[5] Source: Hou et al. (2025)

Figure 4: Counterfactual analysis: the effect of the U.S.-China trade war

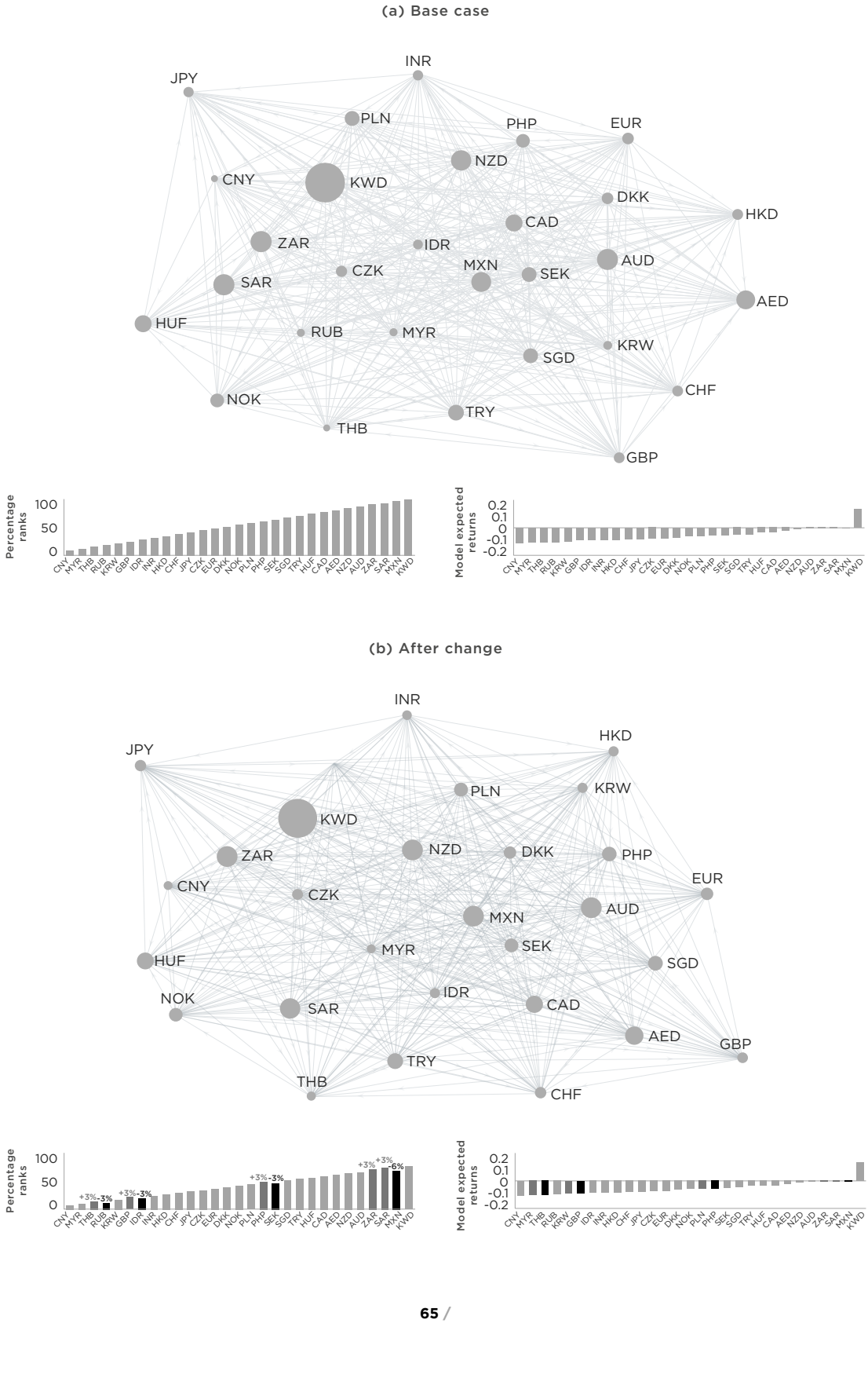
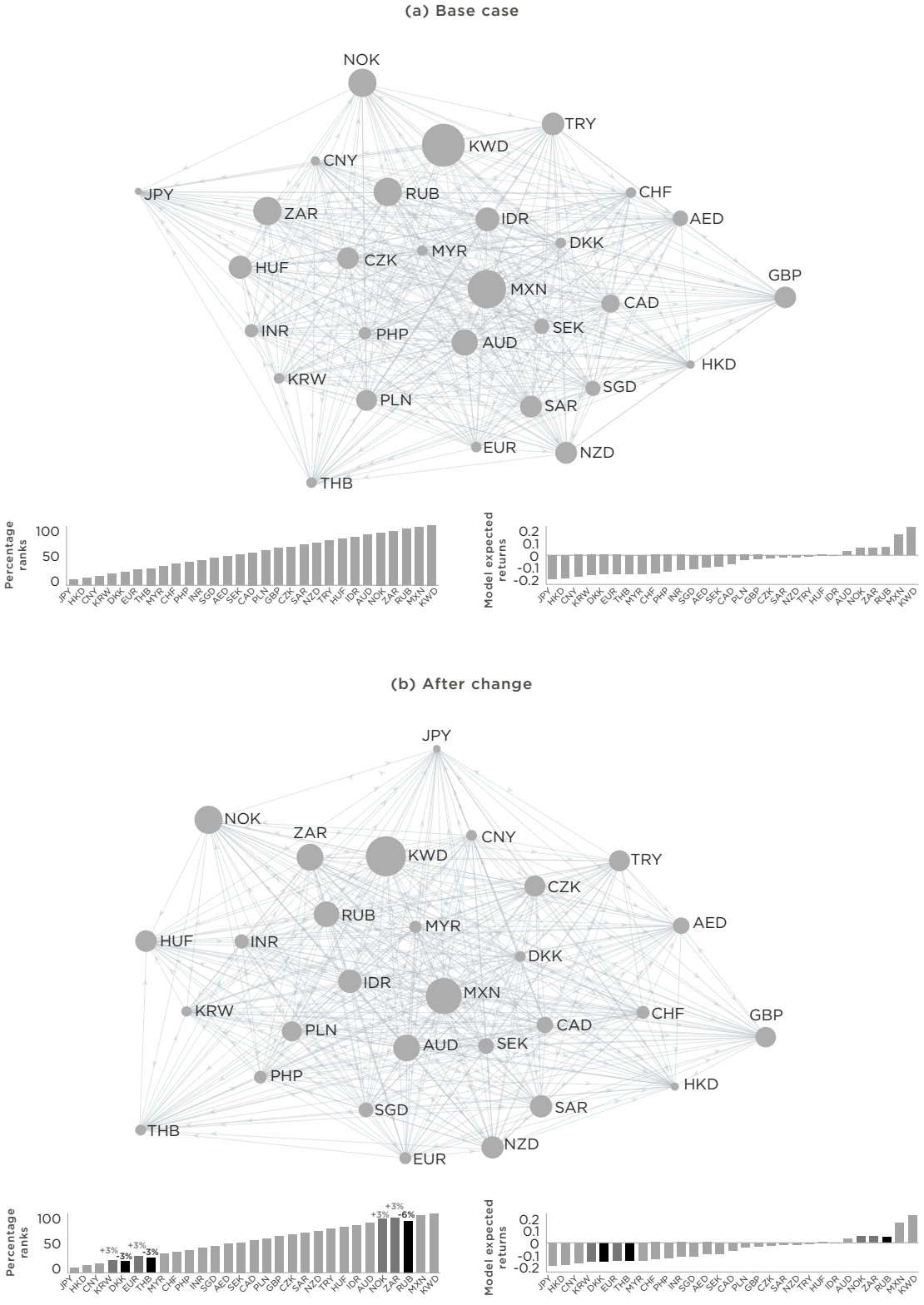


Figure 5: Counterfactual analysis: the effect of collective sanctions on Russia



The Russia sanctions episode provides another compelling illustration as shown in Figure 5. As Western countries imposed sweeping trade and financial sanctions on Russia, the country's CBC plummeted. This sharp decline in centrality coincided with a collapse in the ruble's value and a dramatic tightening of capital controls. Interestingly, DKK (3% drop), THB (3% drop) and ZAR (3% rise) are not in the group of countries that imposed severe sanctions on Russia but display noticeable changes, while GBP, AUD, JPY, CAD, CHF, and PLN, which are in the group, do not display much of a change. This analysis again shows that the global trade imbalance network brings complexity in assessing the effect of significant international events on currency premia.

In Figure 5, Panel (a) plots the global trade imbalance network based on 2021 pairwise bilateral imports and exports, Panel (b) plots the same network with the values of elements in the adjacency matrix related to RUB and a group of currencies shrink by 90%.[6]

5. Contributions and conclusions

This research advances our understanding of exchange rates by placing them squarely within the new global economic order. We are moving away from a world of seamless global integration toward one defined by geopolitical fragmentation, sanctions, and strategic decoupling. Our framework shows that currency risk is not just about local inflation or interest rates; it is an endogenous reflection of how trade networks are being reshaped by these geopolitical forces. By introducing CBC, we have a theoretically grounded characteristic for designing profitable currency investment strategies and for understanding the impact of trade shocks on currency movements. The outcome is a refined understanding of the link between international trade and currency risk that describes how the trade imbalance network generates both direct and indirect impacts on currency premia. The trade imbalance network serves as a transmission mechanism for geopolitical shocks, meaning that a 'decoupling' between two major powers like the U.S. and China creates ripples that redefine the riskiness of bystander currencies. Ultimately, our work proves that in a fragmented world, a country's currency value is inseparable from its strategic position within the global trade network.

We are moving away from a world of seamless global integration toward one defined by geopolitical fragmentation, sanctions and strategic decoupling.

6. Implications for investors and policymakers

The above findings carry important implications for both currency investors and policymakers. For investors, CBC offers a novel and effective predictor of currency risk premia. Incorporating CBC into currency allocation models may enhance return forecasting and improve risk-adjusted performance by capturing information about a country's indirect exposure within the global trade imbalance network, something that traditional metrics generally omit. For policymakers, CBC

[6] Source: Hou et al. (2025)

provides a powerful tool for assessing external vulnerability beyond bilateral trade statistics. Countries with high trade network centrality tend to bear greater systemic exposure, and may therefore require stronger buffers, such as larger foreign exchange reserves, regional swap lines, or targeted macroprudential tools.

The counterfactual analyses illustrate how geopolitical shocks can propagate through trade networks and affect third-party countries' currencies. This underscores the need for policymakers to consider indirect spillovers when designing sanctions, tariffs, or supply chain interventions.

Finally, regulators should be aware that network-driven risks can be amplified when intermediaries face balance sheet constraints. As the world trends toward de-globalization and the formation of distinct trade blocs, the CBC dynamics are likely to become even more polarized. We expect that as supply chains reconfigure around 'friend-shoring', currency risk will increasingly cluster within these regional nodes, potentially leading to higher volatility for countries that sit on the fault lines of competing blocs. Policymakers must move beyond bilateral statistics and use network-informed stress testing to monitor these evolving systemic vulnerabilities. More broadly, as power balances and trade linkages evolve in this changing global economic order, frameworks that account for the networked nature of international imbalances will be indispensable for interpreting currency movements and anticipating future financial vulnerabilities.

References

Adrian, T., E. Etula, and T. Muir, 2014. "Financial intermediaries and the cross-section of asset returns," The Journal of Finance 69, 2557-2596, https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12189

Clayton, C., M. Maggiori, and J. Schreger, 2024. "A framework for geoeconomics," National Bureau of Economic Research, Working Paper 31852, https://www.nber.org/papers/w31852

Colacito, R., M.M. Croce, F. Gavazzoni, and R. Ready, 2018. "Currency risk factors in a recursive multicountry economy," The Journal of Finance 73, 2719-2756, https://onlinelibrary.wiley.com/doi/10.1111/jofi.12720

Della Corte, P. and A. Krecetovs, 2021. "Current account uncertainty and currency premia," Management Science, 70 (9), https://pubsonline.informs.org/doi/10.1287/mnsc.2023.4949

Della Corte, P., S.J. Riddiough, and L. Sarno, 2016. "Currency premia and global imbalances," The Review of Financial Studies 29, 2161-2193, https://academic.oup.com/rfs/article-abstract/29/8/2161/2583738

Della Corte, P., L. Sarno, and G. Sestieri, 2012. "The predictive information content of external imbalances for exchange rate returns: how much is it worth?", The Review of Economics and Statistics 94, 100-115, https://www.jstor.org/stable/41349163

Fang, X. and Y. Liu, 2021. "Volatility, intermediaries, and exchange rates," Journal of Financial Economics 141, 217-233, https://www.sciencedirect.com/science/article/abs/pii/S0304405X21000891

Gabaix, X. and M. Maggiori, 2015. "International liquidity and exchange rate dynamics," The Quarterly Journal of Economics 130, 1369-1420

Gourinchas, P.O. and H. Rey, H., 2007. "International financial adjustment," Journal of Political Economy 115, 665-703, https://www.jstor.org/stable/10.1086/521966

He, Z., B. Kelly, and A. Manela, 2017. "Intermediary asset pricing: New evidence from many asset classes," Journal of Financial Economics 126, 1-35, https://www.sciencedirect.com/science/article/abs/pii/S0304405X1730212X

Hou, A.J., L. Sarno, and X. Ye, 2025. "The trade imbalance network and currency returns," Journal of Financial Economics, 172, 104112, https://www.sciencedirect.com/science/article/pii/S0304405X25001205

Jiang, Z. and R. Richmond, 2023. "Origins of international factor structures," The Journal of Financial Economics 147, 1-26, https://www.sciencedirect.com/science/article/abs/pii/S0304405X22002082

Lustig, H. and R.J. Richmond, 2020. "Gravity in the Exchange Rate Factor Structure," The Review of Financial Studies 33, 3492-3540, https://doi.org/10.1093/rfs/hhz103

Lustig, H., N. Roussanov, and A. Verdelhan, 2011. "Common risk factors in currency markets," Review of Financial Studies 24, 3731-3777, https://doi.org/10.1093/rfs/hhr068

Maggiori, M., 2022. "International macroeconomics with imperfect financial markets," in Handbook of International Economics, vol. 6, 199-236, https://www.sciencedirect.com/science/chapter/handbook/abs/pii/S1573440422000144

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf, 2012. "Currency momentum strategies," Journal of Financial Economics, 106, 660-684, https://www.sciencedirect.com/science/article/abs/pii/S0304405X12001353

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf, 2017. "Currency value," The Review of Financial Studies, 30, 416-441, https://doi.org/10.1093/rfs/hhw067

Richmond, R.J., 2019. "Trade network centrality and currency risk premia," The Journal of Finance, 74, 1315-1361, https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12755

Riddiough, S.J., L. Sarno, and P. Della Corte, 2016. "Global imbalance risk and exchange rates", VoxEU.org, February 29, https://cepr.org/voxeu/columns/global-imbalance-risk-and-exchange-rates

Verdelhan, A., 2018. "The share of systematic variation in bilateral exchange rates," Journal of Finance, 73, 375-418, https://onlinelibrary.wiley.com/doi/10.1111/jofi.12587

Bilateral central bank swap lines

and their implications for the global financial order

Author | **Ayca Zayim** | Associate Professor of Sociology, Mount Holyoke College

Abstract

Are we moving toward a post-neoliberal global financial order? This article discusses the implications of the rise of China and bilateral central bank swap agreements for global financial governance. It begins by surveying the workings of swap lines and the key role of the U.S. and China in the emergent international swap network. It shows that China's rise as a swap partner has enabled national governments to bypass the IMF and other multilateral institutions of the neoliberal order and pursue policies that diverge from neoliberal orthodoxy. In some cases, China's liquidity support has even enabled authoritarian or illiberal regimes to sustain themselves. I join others in arguing that the rise of illiberalism around the world, including the recent developments in the U.S., suggests a discontinuity with neoliberalism. Nonetheless, it is premature to pronounce neoliberalism dead. On the one hand, China poses a limited threat to U.S. hegemony, given the uncertainty surrounding the long-term sustainability of China's bilateral swap agreements. On the other hand, it is reasonable to expect the U.S. to continue supporting a liberal global financial order, given its long-standing financial interests and the U.S. dollar's continued hegemony.

1. Introduction

Bilateral central bank swap agreements have become the most significant source of emergency liquidity since 2008. According to IMF estimates, the swap network among central banks, excluding those covered under regional financial agreements (RFAs), was worth U.S.$1.4 trillion at the end of 2019. As a key pillar of the "global financial safety net" [IMF (2016)], the lending capacity of swaps has exceeded the volume of short-term financing available through RFAs and the IMF. Although not all swap lines have been utilized since 2008, their role has expanded during the COVID-19 pandemic.

A central bank currency swap refers to a contractual agreement between two central banks. It allows central banks to swap their currencies with a commitment to reverse the transaction at a future date. Unlike long-term project finance provided by the World Bank and multilateral development banks, swaps are a source of short-term liquidity and effectively function as collateralized loans. They provide immediate, short-term access to foreign exchange at high volumes, bypassing lengthy decision making and disbursement processes. When used during emergencies, central bank swaps bolster the capacity of monetary authorities to provide liquidity to their banking systems; therefore, they are critical in calming financial investors' nerves.

In recent years, two economies have emerged as key players in the emergent international swap network: the U.S. and China. The actions of the U.S. are not surprising from a historical perspective. Since the end of World War II, the U.S. has been the leader and guarantor of the "rules-based international liberal order" [Ikenberry (2018)]. As early as the 1960s, the U.S. provided credit unilaterally to select countries in line with its economic and financial interests [McDowell (2017)]. U.S. policymakers have also intervened alongside the IMF, which has been the primary international lender of last resort since World War II, to prevent financial crises from derailing an increasingly globalizing economy [Helleiner (1995)]. For instance, during the 1980s debt crisis, the U.S. spearheaded an emergency lending program involving commercial banks, central banks, and the IMF [Sims and Romero (2013)]. In return for implementing what came to be known as 'Washington Consensus' or 'neoliberal' policies, developing economies received debt rescheduling and loans from the IMF. Similarly, during the 1990s East Asian crisis, the U.S. and the IMF took centerstage in rescuing countries facing currency and balance-of-payments crises. The U.S. Fed arranged a bridge loan for Thailand and persuaded private banks to roll over their short-term debts to South Korea [Carson and Clark (2013)]. The IMF provided around U.S.$35 billion to Thailand, Indonesia, and South Korea, exceeding all other multilateral and bilateral loans combined [IMF (2000)]. After the worst of the East Asian crisis was over, Time magazine (1999) pronounced Fed Chair Greenspan, Treasury Secretary Rubin, and Deputy Secretary Summers "The Committee to Save the World" and the IMF "the Committee's fire brigade."

Those days seem to be gone. In March 2023, the same Time magazine featured an article entitled, How China Became a Global Lender of Last Resort [Rajvanshi (2023)]. Aside from becoming the world's largest official creditor under its Belt and Road Initiative [Horn et al. (2021); Malik et al. (2021)], what has been remarkable is China's increasing role as a source of emergency lending – a role that has traditionally been played by the IMF with support from the U.S. For instance, according to The New York Times, China gave U.S.$40.5 billion to countries in distress in 2021 while the IMF lent U.S.$68.6 billion [Bradsher (2023)]. A significant part of China's emergency lending was disbursed through bilateral swap lines organized by China's central bank, the People's Bank of China (PBoC), with the remaining coming through as loans and deposits from Chinese state-owned banks and commodity enterprises [Horn et al. (2023b)].

This article explores the implications of the rise of China and bilateral swap lines for global financial governance. In particular, it explores whether these shifts suggest the emergence of a post-neoliberal global financial order. Neoliberalism refers to an ideology that upholds the market as "the main arbiter of economic outcomes" [cf. Mudge (2008)]. Spearheaded by Ronald Reagan and Margaret Thatcher in the 1980s, the institutionalization of neoliberalism required restructuring national economies in line with free market principles, thereby giving increasing power and authority to private individuals and institutions to govern market-led economic outcomes. In the Global South, this restructuring often occurred through policy-based lending and interventions imposed by the IMF and World Bank. Neoliberal reforms have led to the global integration of domestic markets, enabling multinational financial and non-financial corporations to expand and their profits to multiply. As private business has become the key driver of economic growth in the neoliberal era, policymakers have increasingly been subject to the discipline of domestic and Western-based foreign investors. This has often presented a paradox. While the gap between the

haves and have-nots has widened under financial globalization, governments have had limited room to interfere with the workings of markets to promote social goals, refraining from any policy interventions that challenge private business interests.

The rise of China and the increasing prominence of bilateral swap networks complicate the governance of the neoliberal order in two ways. First, as bilateral agreements, swaps are selective and highly discretionary. They operate alongside the well-established, Western-based international financial institutions and legal infrastructures of the neoliberal order; yet, they diverge from the norms and ideals that underlie this order, such as multilateralism, the rule of law, and inclusive membership [Ruggie (1992)]. Second, swaps offer an alternative source of funding outside the IMF and do not involve ex-post conditionality typically associated with multilateral institutions. This has led some observers to conclude that national economies now enjoy greater policy space to achieve development and social goals that have been ruled out under the neoliberal order [Grabel (2019, p. 49)]. Indeed, China does not require neoliberal conditionalities, such as sound monetary policy and trade liberalization.

Do these developments suggest the decline of neoliberalism and the role of the U.S. as its guarantor? This article joins others in arguing that we increasingly see challenges to the once firm stronghold and legitimacy of the neoliberal order. The rise of China as a swap partner, for instance, has allowed national governments to bypass the IMF and other multilateral institutions and to escape their neoliberal conditionalities. In some cases, China's liquidity support has even enabled authoritarian or illiberal regimes to sustain themselves. In that sense, the rise of China has relaxed constraints on implementing policies and practices that diverge from neoliberalism. Nonetheless, questions remain around the long-term sustainability of China's bilateral swap agreements. Coupled with the continued U.S. dollar hegemony, this article views China as a limited threat to the U.S. in its role as a hegemon and guarantor of the global financial system. It claims that it is premature to pronounce the death of neoliberalism and the emergence of a post-neoliberal order.

2. U.S. swap lines and 'virtual apartheid'

To better understand the implications of the rise of China as a source of emergency lending, a brief overview of the U.S. Federal Reserve's swap network is helpful.

The U.S. Fed is a key player in the international swap network. Scholars and policymakers highlight the Fed swap lines as a significant mechanism to backstop the dollar-based international monetary order [Mehrling (2015); Tooze (2018); McDowell (2019a); Bordo (2021)] and argue that the Fed had become the world's de facto international lender of last resort since 2008 [McDowell (2012); Helleiner (2014); Bordo et al. (2015); Broz (2015); Duran (2015); Destais (2016)]. During the 2008 crisis, the Fed established swap lines with 14 central banks to resolve liquidity shortages in offshore markets and to stabilize the global banking system [Bahaj and Reis (2022)]. The Fed's swap partners included the central banks of ten advanced and four emerging economies (Brazil, Mexico, Singapore, and South Korea). The Fed provided liquidity amounting to U.S.$288 billion in September 2008, U.S.$534 billion in October, and a peak of U.S.$554 billion at the end of December, accounting for over 25% of the Fed's total assets [Bertaut and Pounder (2009)]. When the COVID-19 pandemic hit, the Fed resorted to swap lines again. The Fed provided unlimited funding to five central banks through its

standing lines (Bank of Canada, Bank of England, Bank of Japan, European Central Bank, and Swiss National Bank) and established temporary swap lines to nine other central banks (central banks of Australia, Brazil, Denmark, South Korea, Mexico, New Zealand, Norway, Singapore, and Sweden). In May 2020, swap usage reached U.S.$449 billion [Choi et al. (2022)]. During both periods, Fed liquidity promoted the stability of the global financial system by alleviating funding pressure, reducing dollar borrowing costs and shortages, and bolstering confidence among market participants [Bahaj and Reis (2020); Eren et al. (2020); Bordo (2021); Bahaj and Reis (2022)].

Although benefiting the global economy, it is important to note that Fed swap lines have primarily been geared to shield the U.S. economy and monetary policy from adverse developments [McDowell (2012); Helleiner (2014); Hardie and Maxfield (2016); McDowell (2017); Hardie and Thompson (2021); Aizenman et al. (2022); Cassetta (2022); Pape (2022)]. As such, U.S. policymakers have been highly selective, leading to vast inequalities in the allocation of U.S. dollar liquidity. While the Fed has prioritized systemically important European economies with financial linkages to the U.S., only a handful of developing economies were allocated swap lines during both crises, and when it did, the allocation more often reflected the U.S.'s strategic and political concerns [Chey (2012); Destais (2016); Sahasrabuddhe (2019); Bordo (2021)]. In 2017, the governor of the Reserve Bank of India starkly described the situation as a "virtual apartheid" whereby "systemic central banks protect themselves and their self-interest" while emerging economies are "systematically denied access" [Patel (2017)].

Murau et al. (2022) propose a three-layered international monetary hierarchy based on the mechanisms by which other central banks access emergency dollar liquidity. These mechanisms are swaps with the Fed, the FIMA (Foreign and International Monetary Authorities) repo facility established in March 2020, and the Special Drawing Rights system administered by the IMF. For a privileged group of advanced economies that host financial institutions with deep financial interlinkages to the U.S., swaps confer unlimited credit and spending power from the Fed. The FIMA facility limits a country's borrowing capacity to its stock of U.S. Treasuries and cannot instill the same level of confidence among investors as swaps do. However, especially constrained are those economies without access to swaps or sufficient collateral. These economies are severely constrained in ameliorating financial distress since the IMF's current quota system and lending practices limit borrowers' room to maneuver during financial crises [Gallagher and Kozul-Wright (2022)].

3. China's swap lines at the service of debt repayment

Given persistent difficulties in accessing dollar liquidity, several emerging economies have turned to China. China has built an extensive swap network since 2008. In 2020, it had 31 bilateral swap lines out of 91 lines globally [Perks et al. (2021)]. This number increased to 38 in 2022, amounting to around 4 trillion renminbi (RMB) [Perez-Saiz and Zhang (2023)]. China's swap partners include several emerging economies, such as Brazil, Indonesia, Thailand, and Turkey, as well as a few advanced economies, such as Canada, the U.K., and the E.U. While many of these swap agreements were signed as early as 2009, several remained inactive until 2015.

According to the AidData dataset on China's emergency lending, the PBoC has provided U.S.$185 billion between 2000 and 2021 to 17 countries [Horn et al. (2023c, 2023b)]. This figure includes 'rollovers' whereby short-term

swap loans, with de jure maturities of three to twelve months, are "extended again and again" to refinance maturing debts, "resulting in a de facto maturity of more than three years, on average" [Horn et al. (2023b, pp. 3 and 9)]. In addition to swaps, China has provided other forms of rescue lending, including short- and medium-term loans from Chinese state-owned banks and large cash payments from Chinese state-owned oil and gas companies. These rescue payments have amounted to U.S.$70 billion, less than half of the total bilateral swap drawings [Horn et al. (2023b)].

China's bilateral swaps are often viewed as a tool of "financial statecraft" [Armijo and Katada (2015)] whereby China employs swap lines strategically to achieve its foreign policy objectives [McDowell (2019b)], most significantly, to promote RMB internationalization [Prasad (2016); Eichengreen et al. (2018)]. Bilateral swaps also serve China's economic and financial interests. Specifically, Horn et al. (2023b) find that most recipients of PBoC swap liquidity (13 out of 17 central banks) have resorted to swap lines during periods of economic and financial distress, such as balance-of-payments crises or imminent sovereign defaults on external debt. These countries are heavily indebted to China under its Belt and Road Initiative (BRI) [Horn et al., (2021); Wooley (2023)], and they have drawn on U.S.$170 billion in emergency liquidity from the PBoC [Horn et al. (2023b)]. Although these debts have remained largely 'hidden' because of the secrecy surrounding the Chinese government's lending arrangements [Horn et al. (2021); Malik et al. (2021); Gelpern et al. (2023)], developing economies have increasingly faced payment difficulties since 2019 [Mosley and Rosendorff (2023)].

China's response to what has been called 'hidden defaults' has been twofold [Horn et al. (2022); Horn et al. (2023a)]. On the one hand, China has rolled back its expansive lending strategy, limiting new lending by Chinese banks and enterprises [Horn et al. (2023a)]. On the other hand, China has restructured the outstanding debt of low-income countries and offered emergency liquidity to middle-income countries to avoid any potential defaults. Middle-income countries account for 80% of China's total overseas loans; thus, their repayment difficulties or defaults pose significant balance sheet risks to Chinese banks [Horn et al. (2023a)]. To ensure that these economies are "sufficiently liquid to continue servicing their outstanding debts" [Horn et al. (2023a, p. p113)], China has provided bilateral swaps and other types of emergency loans. As Carmen Reinhart argues, "Beijing is ultimately trying to rescue its own banks. That's why it has gotten into the risky business of international bailout lending" [quoted in Wooley (2023)]. In this regard, China is similar to the U.S. in defending its economic and financial interests. However, unlike Fed swaps, which have targeted temporary liquidity shortages, PBoC swaps are repeatedly used for protracted balance of payments problems, making them similar to IMF bailouts [Horn et al. (2023b, p11-13)].

It is critical to highlight that because China's role as a lender of last resort is closely tied to its BRI lending, it is uncertain whether China will play a similar role in the future. China's lending to governments in the Global South has turned negative for the first time during 2019-2021 [Horn et al. (2023a)]. As current recipients gradually service their existing debts and new lending to emerging economies remains negative, the financial risks to Chinese banks would dissipate, which has been the main driver of Chinese rescue lending [Horn et al. (2023a)]. Therefore, China might no longer act as a lender of last resort.

4. Rise of China and illiberal regimes

While China's future role as a source of emergency lending remains uncertain, there is enough data on the effects of Chinese lending. Using Turkey as an example, this section shows how China's rise as a lender of last resort has enabled governments to bypass Western-led multilateral institutions of the neoliberal order and even created opportunities for illiberal regimes to sustain themselves.

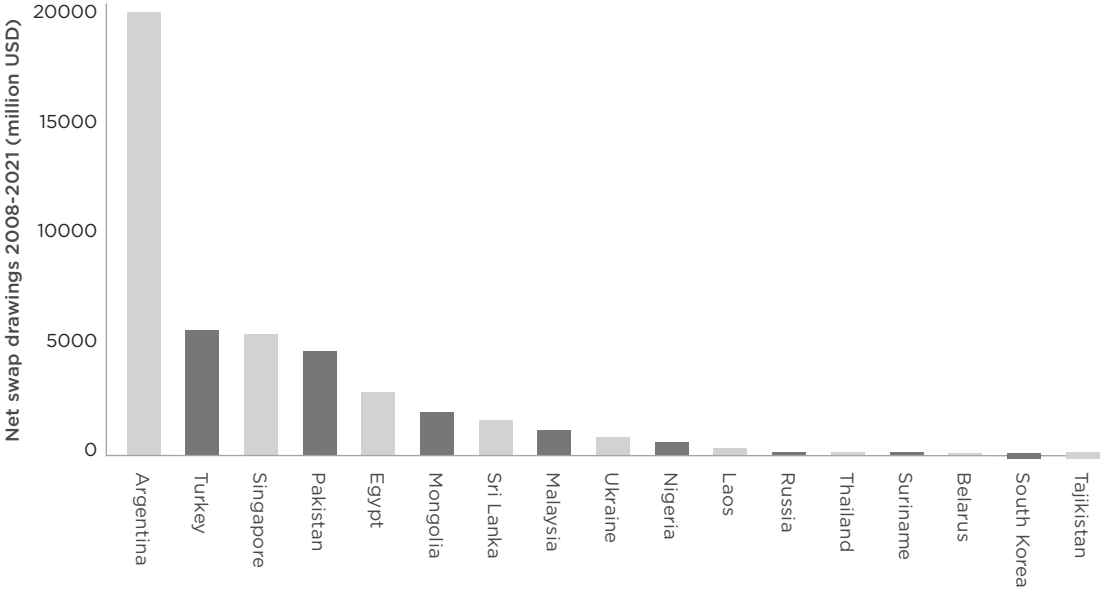
4.1 Buying time: Erdogan's reliance on PBoC swaps

China is one of Turkey's key swap partners, alongside Qatar, South Korea, and the United Arab Emirates. Turkey and China signed the first bilateral swap agreement in 2012 for 10 billion RMB (3 billion Turkish lira, TRY) [PBoC (2021)]. The agreement was then renewed in 2015 and 2019 [PBoC (2021)], and its funding capacity was increased to 35 billion RMB in June 2021 (approximately 46 billion TRY, equivalent to U.S.$5.5 billion) [PBoC (2021)].

Based on the AidData dataset, Turkey ranked second during 2008-2021 in terms of net PBoC swap drawings, which excludes rollovers (see Figure 1). In addition to swap liquidity, Turkey has received investment under China's Belt and Road Initiative since 2015. BRI investments have increased rapidly since 2017, accompanying Turkish President Erdogan's calls for building a strong economic and political relationship with China, amidst the growing anti-E.U. and anti-U.S. public sentiment after the coup attempt in 2016 [Chen (2020); Kutlay and Öniş (2021)].

Based on publicly available data, Turkey has drawn on its swap line with China during at least two periods of financial distress: November 2016 and June 2021. Both periods were marked by currency depreciation, declining central bank reserves, and a looming balance-of-payments crisis. The first drawing in November 2016 occurred amid portfolio outflows driven by heightened economic and political risks, following the coup attempt in July and credit rating downgrades. Capital flight caused significant distress for domestic banks

Figure 1: Total net PBoC swap drawings, 2008-2021 (U.S.$ million)



> *China cannot be relied upon to provide emergency lending to countries that are not indebted to its banks and enterprises, or whose financial instability does not adversely or directly affect Chinese creditors.*

and non-financial corporations with high levels of foreign-currency-denominated debt and heightened the risk of a balance-of-payments crisis, given already low foreign-currency reserves. The second drawing came in June 2021, amid Turkey's mounting economic problems stemming from Erdogan's growing control over the economy. Equipped with new powers and facing few checks and balances following Turkey's shift from a parliamentary to a heavily centralized presidential system in 2017, Erdogan gradually established control over Turkey's formally independent central bank. Between 2019 and 2022, Erdogan changed the central bank governor three times, pressuring the central bank to lower interest rates at all costs. A commitment to low interest rates was not only ideological. It supported Erdogan's macroeconomic populism and the political party AKP's clientelistic relationships; thus, it was critical to Erdogan's political survival. Erdogan's interventions fueled inflation and currency depreciation. To combat the depreciation of the lira without raising interest rates, policymakers resorted to different measures, including foreign exchange interventions, attracting foreign investment, and securing external finance through swap lines. As one Istanbul-based economist described to me, an "unsustainable economic system" was only "possible thanks to help from Qatar, Russia, and this and that" (interview, March 2023, Istanbul).

The limitations of publicly available data do not allow us to discern exactly how the Turkish central bank used the local currencies received through its swap agreements. However, as revealed during my interviews with financiers in 2023, the widespread conviction was that the Turkish central bank used the swap liquidity to boost its foreign exchange reserves, which was hoped would instill confidence in the economy and ensure against capital flight and financial instability. The Turkish central bank's gross reserves increased from around U.S.$87 billion in March to U.S.$111 billion in December 2021 [World Bank (2022)]. Erdogan commented on the rising reserves in September 2022, saying, "Many friendly countries are supporting us. Their lending strengthens our central bank" [Yenicag (2022)]. However, the net reserves – assets minus liabilities – remained well below U.S.$50 billion, and when swaps were excluded, they remained persistently negative throughout 2020 and 2021 [World Bank (2022)]. This raised concerns within the financial community.

Turkey's swap drawings, including the one with China, should be seen as part of a broader government effort to cushion the lira and avert a balance-of-payments crisis. Swaps have accompanied several other measures during this period to limit the lira's depreciation, most notably, various regulations that restricted access to foreign exchange for domestic financial and non-financial corporations. These regulations effectively served as "backdoor capital controls," in the words of one interviewee. To the extent that swap agreements have been successful in containing the collapse of the lira, it is reasonable to assume that they have helped Turkey's creditors. This includes China's banks and enterprises. At the same time, though, they have helped Erdogan's authoritarian regime stay in power, which, unsurprisingly, has been a reliable partner of the Chinese government

since the mid-2010s. Erdogan was able to avoid turning to the IMF despite the economic turmoil over the past few years. The IMF not only represents Western dominance for Erdogan. More critically, the IMF involvement would likely bring intrusive conditionalities, undercutting extensive clientelistic relations between the AKP and its supporters and exposing the government to greater criticism from the public [Dreher et al. (2022); Cormier (2023)]. Instead of going to the IMF, Erdogan has grasped every financial opportunity to stabilize the economy. These 'non-Western' funding sources, including bilateral swaps, have expanded Erdogan's room to maneuver by easing Turkey's funding constraints. They have, therefore, helped prolong his populist authoritarian rule.

5. Implications for a post-neoliberal order

Examining the implications of the rise of China and bilateral swap lines for global financial governance, we can discern two clear dynamics:

1. The shift from multilateral to bilateral arrangements reflects that the rules by which the global financial system is governed are becoming more opaque, fragmented, and contested. The neoliberal order was governed by multilateral financial institutions and arrangements (IMF, World Bank, WTO). While these institutions unequivocally reflected the geopolitical and economic interests of their main shareholders, especially the U.S., they often upheld a set of rigid rules and regulations that prioritized the market in economic governance and restructured national economies according to its logic. By contrast, bilateral swap agreements are highly discretionary, and there is a lack of transparency regarding any tacit conditions that might be attached to them.
2. PBoC swaps have provided an alternative to IMF funding. To the extent that China's swap network allows governments to escape IMF conditionality, it has increased emerging economies' room to maneuver and their capacity to diverge from neoliberal orthodoxy through policies such as capital controls, government subsidies, and tariffs. While policymakers can use this emerging policy space to pursue development goals, it can also strengthen and prolong the rule of illiberal or authoritarian leaders. In either case, these experiments diverge from the neoliberal order, challenging its multilateral, rule-based, free market principles. When viewed in light of the recent developments in the U.S., it clearly amounts to a discontinuity with the neoliberal order.

Does this rupture imply the decline of U.S. hegemony? Not necessarily. The first issue relates to the long-term sustainability of China's bilateral swap agreements. As mentioned earlier, China cannot be relied upon to provide emergency lending to countries that are not indebted to its banks and enterprises, or whose financial instability does not adversely or directly affect Chinese creditors. In that sense, the PBoC is similar to the Fed in that its emergency lending is shaped by the financial interests of its domestic economy. The second issue relates to the continued U.S. dollar hegemony. Despite China's economic gains in the global economy and its efforts to internationalize the RMB, U.S. dollar hegemony has not (yet) been eroded. This limits the usefulness of PBoC swaps for their recipients. Unless there are agreements in place allowing local currencies to be converted into U.S. dollars, recipients are limited in terms of their ability to pay their debts or intervene in foreign exchange markets.

Where does this leave us? Despite the recent disruptions to the neoliberal order, including the rise of illiberalism in the U.S., the U.S. has historically had an interest in constructing and maintaining a liberal financial order. This was, in large part, informed by its foreign financing needs since the 1970s. Given its long-standing financial interests and the U.S. dollar's continued hegemony, it is reasonable to expect the U.S. to maintain support for a liberal global financial order in the near future. In this context, a major restructuring of the IMF's lending arrangements, including the issuance of new Special Drawing Rights, may strengthen the IMF's role as a lender of last resort. These efforts will be bolstered by the continued support of the U.S. financial sector and those of other European economies for financial liberalism. Until U.S. policymakers change course, though, the financial sector will have to develop strategies to deal with current illiberalism. Unless we see careful, coherent, and consistent actions to institute a post-neoliberal order, it seems premature to pronounce neoliberalism dead.

References

Aizenman, J., H. Ito, and G. K. Pasricha, 2022. Central Bank Swap Arrangements in the Covid-19 Crisis, Journal of International Money and Finance, 122, 1-20

Armijo, L. E. and S. N. Katada, 2015. Theorizing the Financial Statecraft of Emerging Powers, New Political Economy, 20, 42-62

Babb, S. and A. Kentikelenis, 2018. International Financial Institutions as Agents of Neoliberalism, The SAGE Handbook of Neoliberalism, SAGE Publications, 16-27

Bahaj, S. and R. Reis, 2020. Central Bank Swap Lines During the Covid-19 Pandemic, Covid Economics, 1-12

Bahaj, S. and R. Reis, 2022. Central Bank Swap Lines: Evidence on the Effects of the Lender of Last Resort, The Review of Economic Studies, 89, 1654-1693

Bertaut, C. C. and L. Pounder, 2009. The Financial Crisis and U.S. Cross-Border Financial Flows, Federal Reserve Bulletin, 95, accessed October 21, 2022 at https://www.federalreserve.gov/pubs/bulletin/2009/pdf/bulletin_article_november_2009a1.pdf

Bordo, M. D., 2021. Monetary Policy Cooperation/Coordination and Global Financial Crises in Historical Perspective, Open Economies Review, 32, 587-611

Bordo, M. D., O.F. Humpage, and A. J. Schwartz, 2015. The Evolution of the Federal Reserve Swap Lines since 1962, IMF Economic Review, 63, 353-372

Bradsher, K., 2023. After Doling out Huge Loans, China Is Now Bailing out Countries, The New York Times, March 27

Broz, J. L., 2015. The Politics of Rescuing the World's Financial System: The Federal Reserve as a Global Lender of Last Resort, The Korean Journal of International Studies, 13, 323-351

Carson, M. and J. Clark, 2013. Asian Financial Crisis, Federal Reserve History, accessed at https://www.federalreservehistory.org/essays/asian-financial-crisis

Cassetta, J. M., 2022. The Geopolitics of Swap Lines, M-RCBG Associate Working Paper Series, 181, accessed at https://www.hks.harvard.edu/centers/mrcbg/publications/awp/awp181

Chen, Y., 2020. Developments in China–Turkey Relations: A View from China, Critical Sociology, 46, 777-787

Chey, H.-k., 2012. Why Did the Us Federal Reserve Unprecedentedly Offer Swap Lines to Emerging Market Economies During the Global Financial Crisis? Can We Expect Them Again in the Future?, GRIPS Discussion Paper, 11-18, National Graduate Institute for Policy Studies

Choi, M., L. S. Goldberg, R. Lerman, and F. Ravazzolo, 2022. The Fed's Central Bank Swap Lines and Fima Repo Facility, Economic Policy Review, 28, accessed on October 21, 2022 at https://www.newyorkfed.org/medialibrary/media/research/epr/2022/epr_2022_fima-repo_choi.pdf

Cormier, B., 2023. Chinese or Western Finance? Transparency, Official Credit Flows, and the International Political Economy of Development, The Review of International Organizations, 18, 297-328

Destais, C., 2016. Central Bank Currency Swaps and the International Monetary System, Emerging Markets Finance and Trade, 52, 2253-2266

Dreher, A., A. Fuchs, B. Parks, A. Strange, and M. J. Tierney, 2022. Banking on Beijing: The Aims and Impacts of China's Overseas Development Program, Cambridge University Press

Duran, C. V., 2015. The International Lender of Last Resort for Emerging Countries: A Bilateral Currency Swap?, GEG Working Paper, 108, accessed on October 21, 2022 at https://www.geg.ox.ac.uk/sites/default/files/WP_108%20-%20The%20International%20Lender%20of%20Last%20Resort%20for%20Emerging%20Countries%20-%20Camila%20Duran_0.pdf

Eichengreen, B., A. Mehl, and L. Chitu, 2018. How Global Currencies Work: Past, Present, and Future, Princeton University Press

Eren, E., A. Schrimpf, and V. Sushko, 2020. US Dollar Funding Markets During the Covid-19 Crisis – the Money Market Fund Turmoil, BIS Bulletin, accessed on 21 October 2022 at https://www.bis.org/publ/bisbull14.pdf

Fritz, B. and L. Mühlich, L., 2019. Regional Financial Arrangements in the Global Financial Safety Net: The Arab Monetary Fund and the Eurasian Fund for Stabilization and Development, Development and Change, 50, 96-121

Gallagher, K. P. and R. Kozul-Wright, 2022. The Case for a New Bretton Woods, Polity Press

Gelpern, A., S. Horn, S. Morris, B. Parks, and C. Trebesch, 2023. How China Lends: A Rare Look into 100 Debt Contracts with Foreign Governments, Economic Policy, 38, 345–416

Ghosh, A. R., J. D. Ostry, and C. G. Tsangarides, 2012. Shifting Motives: Explaining the Buildup in Official Reserves in Emerging Markets since the 1980s, IMF Working Papers, accessed on October 21, 2022 at https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Shifting-Motives-Explaining-the-Buildup-in-official-Reserves-in-Emerging-Markets-Since-the-25683

Grabel, I., 2013. Global Financial Governance and Development Finance in the Wake of the 2008 Financial Crisis, Feminist Economics, 19, 32-54

Grabel, I., 2019. Continuity, Discontinuity and Incoherence in the Bretton Woods Order: A Hirschmanian Reading, Development and Change, 50, 46-71

Hardie, I. and S. Maxfield, 2016. Atlas Constrained: The US External Balance Sheet and International Monetary Power, Review of International Political Economy, 23, 583-613

Hardie, I. and H. Thompson, 2021. Taking Europe Seriously: European Financialization and US Monetary Power, Review of International Political Economy, 28, 775-793

Helleiner, E., 1995. Explaining the Globalization of Financial Markets: Bringing States Back In, Review of International Political Economy, 2, 315-341

Helleiner, E., 2014. The Status Quo Crisis: Global Financial Governance after the 2008 Meltdown, Oxford University Press

Horn, S., C. M. Reinhart, and C. Trebesch, 2021. China's Overseas Lending, Journal of International Economics, 133, 103539

Horn, S., C. M. Reinhart, and C. Trebesch, 2022. Hidden Defaults, AEA Papers and Proceedings, 112, 531-535

Horn, S., B. Parks, C. M. Reinhart, and C. Trebesch, 2023a. Debt Distress on China's Belt and Road, AEA Papers and Proceedings, 113, 131-134

Horn, S., B. Parks, C. M. Reinhart, and C. Trebesch, 2023b. China as an International Lender of Last Resort, World Bank Policy Research Working Paper, WPS10380, accessed at https://tinyurl.com/34vy43zz

Horn, S., B. Parks, C. M. Reinhart, and C. Trebesch, 2023c. China as an International Lender of Last Resort Dataset, Version 1.0.1, AidData, accessed at: https://www.aiddata.org/data/china-as-an-international-lender-of-last-resort-dataset-version-1-0

Ikenberry, G. J., 2018. The End of Liberal International Order?, International Affairs, 94, 7-23

International Monetary Fund (IMF), 2016. Adequacy of the Global Financial Safety Net, IMF Policy Papers, accessed at https://www.imf.org/en/Publications/Policy-Papers/Issues/2016/12/31/Adequacy-of-the-Global-Financial-Safety-Net-PP5025

International Monetary Fund (IMF), 2016. Recovery from the Asian Crisis and the Role of the IMF, Washington, DC: International Monetary Fund

Kapur, D. and R. Webb, 2007. Beyond the IMF, G-24 Discussion Paper Series, 43, accessed on May 13, 2023 at https://casi.sas.upenn.edu/sites/default/files/bio/uploads/Beyond%20the%20IMF.pdf

Kring, W. N. and K. P. Gallagher, 2019. Strengthening the Foundations? Alternative Institutions for Finance and Development, Development and Change, 50, 3-23

Kutlay, M. and Z. Öniş, 2021. Turkish Foreign Policy in a Post-Western Order: Strategic Autonomy or New Forms of Dependence?, International Affairs, 97, 1085-1104

Laruffa, F., 2023. Making Sense of (Post)Neoliberalism, Politics & Society, 52(4), accessed at: https://journals.sagepub.com/doi/10.1177/00323292231193805

Malik, A. A., B. Parks, B. Russell, J. J. Lin, K. Walsh, K. Solomon, S. Zhang, T.-B. Elston, and S. Goodman, 2021. Banking on the Belt and Road: Insights from a New Global Dataset of 13,427 Chinese Development Projects, accessed on February 28, 2024 at https://docs.aiddata.org/ad4/pdfs/Banking_on_the_Belt_and_Road__Insights_from_a_new_global_dataset_of_13427_Chinese_development_projects.pdf

McDowell, D., 2012. The US as 'Sovereign International Last-Resort Lender': The Fed's Currency Swap Programme During the Great Panic of 2007–09, New Political Economy, 17, 157-178

McDowell, D., 2017. Brother, Can You Spare a Billion? The United States, the IMF, and the International Lender of Last Resort, Oxford University Press

McDowell, D., 2019a. Emergent International Liquidity Agreements: Central Bank Cooperation after the Global Financial Crisis, Journal of International Relations and Development, 22, 441-467

McDowell, D., 2019b. The (Ineffective) Financial Statecraft of China's Bilateral Swap Agreements, Development and Change, 50, 122-143

Mehrling, P., 2015. Elasticity and Discipline in the Global Swap Network, International Journal of Political Economy, 44, 311-324

Mosley, L. and B. P. Rosendorff, 2023. The Unfolding Sovereign Debt Crisis, Current History, 122, 9-14

Mudge, S. L., 2008. What Is Neo-Liberalism?, Socio-Economic Review, 6, 703-731

Mühlich, L., B. Fritz, and W. N. Kring, 2022. No One Left Behind? Assessing the Global Financial Safety Net Performance During Covid-19, Journal of Globalization and Development, 13, 123-147

Murau, S., F. Pape, and T. Pforr, 2022. International Monetary Hierarchy through Emergency US-Dollar Liquidity: A Key Currency Approach, Competition & Change, , 27, 3-4

Ocampo, J. A., 2018. Resetting the International Monetary (Non)System, Oxford University Press

Pape, F., 2022. Governing Global Liquidity: Federal Reserve Swap Lines and the International Dimension of Us Monetary Policy, New Political Economy, 27, 455-472

Patel, U. R., 2017. Financial Regulation and Economic Policies for Avoiding the Next Crisis, 32nd Annual G30 International Banking Seminar (Inter-American Development Bank, Washington, D.C.)

People's Bank of China (PBoC), 2021. Rmb Internationalization Report', accessed on October 22, 2023 at http://www.pbc.gov.cn/en/3688241/3688636/3828468/4434999/4433239/2021122809344466286.pdf

Perez-Saiz, H. and L. Zhang, 2023. Renminbi Usage in Cross-Border Payments: Regional Patterns and the Role of Swaps Lines and Offshore Clearing Banks, IMF Working Papers, 2023/077, accessed on March 1, 2024 at https://www.imf.org/en/Publications/WP/Issues/2023/03/31/Renminbi-Usage-in-Cross-Border-Payments-Regional-Patterns-and-the-Role-of-Swaps-Lines-and-531684

Perks, M., Y. Rao, J. Shin, and K. Tokuoka, 2021. Evolution of Bilateral Swap Lines, IMF Working Paper Series, accessed on October 21, 2022 at https://www.imf.org/en/Publications/WP/Issues/2021/08/06/Evolution-of-Bilateral-Swap-Lines-463358

Prasad, E., 2016. Gaining Currency: The Rise of the Renminbi, Oxford University Press

Rajvanshi, A., 2023. How China Became a Global Lender of Last Resort, Time, accessed March 1, 2024 at https://time.com/6266658/china-emerges-major-global-lender/

Ruggie, J. G., 1992. Multilateralism: The Anatomy of an Institution, International Organization, 46, 561-598

Sahasrabuddhe, A., 2019. Drawing the Line: The Politics of Federal Currency Swaps in the Global Financial Crisis, Review of International Political Economy, 26, 461-489

Shea, P. E., B. Reinsberg, and A. Kern, 2024. China Lending and the Political Economy of Leader Survival, Journal of Conflict Resolution, 69(4)

Sims, J. and J. Romero, 2013. Latin American Debt Crisis of the 1980s, Federal Reserve History, accessed on 25 February 2024 at https://www.federalreservehistory.org/essays/latin-american-debt-crisis

Time, 1999. The Committee to Save the World [cover], accessed on February 11, 2025 on https://content.time.com/time/covers/0,16641,19990215,00.html

Tooze, A., 2018. Crashed: How a Decade of Financial Crises Changed the World, Allen Lane

Wooley, A., 2023. Belt and Road Bailout Lending Reaches Record Levels, Raising Questions About the Future of China's Flagship Global Infrastructure Program, accessed at https://www.aiddata.org/blog/belt-and-road-bailout-lending-reaches-record-levels

World Bank, 2022. Sailing against the Tide, Turkey Economic Monitor, Washington, DC, World Bank

Yenicag, 2022. Erdoğan: Dost Ülkeler Sağ Olsun Destekliyor. Borçlanmayla Merkez Bankası Güçleniyor, accessed at https://www.yenicaggazetesi.com.tr/erdogan-dost-ulkeler-sag-olsun-destekliyor-borclanmayla-merkez-bankasi-gucleniyor-576886h.htm

Strategic pivots

organizational responses to geopolitical turbulence

Multinationals in stormy waters:

lessons from Central Europe to navigate the populist wave

Author | **Gerhard Schnyder** | Professor of International Management & Political Economy, Loughborough University

Abstract

Tariffs, trade wars, (re-)nationalizations, and other nationalist political interventions in the economy create huge costs for firms, especially those who do business across borders. The second Trump administration is at the forefront of such nationalist economic policies, leaving companies reeling and scrambling to change their business models and restructure their supply chains. In East Central Europe (ECE), national populist governments have implemented such policies for over a decade. Therefore, the region can provide important lessons on the impact of populists on multinational companies (MNCs). One thing that emerges clearly from academic research is that MNCs in the financial services sector are particularly exposed to national populists' anti-foreign direct investment (FDI) policies. This article seeks to explain why financial services firms face high levels of political risk and how they can mitigate it. To do that, the article analyzes the ambiguous policies of ECE national populists towards foreign investors and identifies sectoral- and country-level factors that explain variation in political risk that foreign companies face. It also discusses which political strategies firms in the region have adopted to mitigate political risk work and which ones do not, particularly in the financial services industry.

1. Introduction

Since Donald Trump's return to power, businesses have anxiously watched the deteriorating relationship between the U.S. and China, the ups and downs in the U.S.'s approach to Russia, and most importantly perhaps the Trump administration's policy on tariffs and trade. As a result of the tariffs on Chinese exports to the U.S., Apple, for instance, has moved parts of its iPhone production from China to India. Trump's economic nationalism is not an isolated case though, but representative of a wider phenomenon of national populist economic policies that have been spreading across the globe. East Central Europe (ECE) has experienced such policies for over a decade and therefore provides a useful laboratory for us to better understand how national populists in power may impact business activity and how business – in turn – can react to the increased political risk.

2. National populists' Janus-faced FDI policies

When it comes to foreign direct investment (FDI) policy, national populists in East Central Europe face a fundamental dilemma: On the one hand, these countries have extensively relied – more than most others – on a steady inflow of FDI for technology transfer and upgrading of their post-Soviet productive capacities. This has led to a situation of a strong 'dependence' on foreign multinationals [Nölke and Vliegenthart (2009)]. On the other hand, their nationalist political strategy crucially relies on anti-foreigner rhetoric focused on 'taking back control.' This slogan does not only concern physical borders and immigration, but also the inflow of capital. Attacks on financial investors like George Soros and the Central European University he founded have made headlines outside the region. But multinational companies (MNCs) can also become targets of populists' economic nationalism. Indeed, anti-FDI policies are a common tool used in the region to try to reduce the dependency on foreign capital. A case in point is the partial re-domestication of the banking sectors in Hungary and Poland. The next section shows that national populists in Hungary and Poland use a range of subtle tools to implement their anti-FDI policies in the financial sector.

2.1 The re-domestication of finance in Hungary and Poland

The financial industry has become the target of national populists in ECE countries where finance was particularly strongly dominated by foreign capital. In the case of Poland, foreign-owned banks dominated the industry from the late 1990s onwards. After 2015, progressive re-domestication took place under the Law and Justice (PiS) government. By 2018, Polish domestic investors had replaced foreign owners as dominant investors, holding 54.1% of aggregate banking assets.

Similarly, under the Fidesz government after 2010, the largely foreign-dominated banking industry in Hungary saw the proportion of banking sector shares held by domestic owners rise to nearly 50% by 2018. The state owned close to two-thirds of them.

As a result, inward FDI into finance fell sharply both in Hungary (from 17.7% of GDP in 2015 to 9.1% in 2020) and in Poland (from 20.5% of GDP in 2017 to 12.4% in 2020).

These trends were not the result of foreign banks divesting from the region voluntarily. Rather, they resulted from Poland's and Hungary's national populist governments' anti-FDI policies that aimed at reducing dependency on foreign multinationals and transferring ownership to nationals.

The tools used to achieve this goal sometimes included outright nationalization of foreign assets. More often than not, however, they were much more subtle – and harder for companies to anticipate: The strategic use of special laws that disadvantage foreign companies is a prominent tool. Strongly progressive taxes based on company size, for instance, typically are targeted at (large) multinational companies to the advantage of (smaller) domestic firms. Thus, a special banking levy introduced in Hungary in the 2010s taxed the first 50 billion forint of assets at 0.15%, but any additional income at 0.5%, therefore hitting foreign banks harder than domestic ones [Johnson and Barnes (2015)]. Similarly, beyond direct subsidies, public procurement and government advertising are used to privilege domestic firms over foreign MNCs.

These economic strategies often aim at weakening foreign companies and ultimately force their owners to transfer ownership to domestic investors. In the most extreme cases, this is achieved with what one of our respondents called 'mafia-like methods,' which we called elsewhere 'forced buyouts' (FBOs) [Sallai and Schnyder (2021)]. For instance, the use of unannounced tax inspections to put pressure on foreign companies' share price allows the government to force the sale of the business to domestic investors at discounted prices. In other cases, the threat of adverse government action is enough to 'convince' owners to transfer ownership to domestic actors.

Such methods are not limited to the financial sector. In retail, the Hungarian government adopted a series of very creative laws targeting foreign retailers like Carrefour and Spar. Thus, one law prohibited non-family-owned businesses from opening on Sundays – reducing the attractiveness of large supermarkets compared to smaller, family-owned businesses. Moreover, a progressive retail tax was introduced in 2020. The tax rates increase steeply with turnover. Large foreign retail chains – who report consolidated turnover – are taxed at 4.5%, while domestic retailers, who operate via franchise systems and do not consolidate turnover for tax purposes, face tax rates of 0-1%. Importantly, foreign-owned firms are not allowed to restructure into a franchise system that Hungarian retailers use to avoid the tax, which has led to complaints of foreign chains and governments about the discriminatory effect of the retail tax regime [Nemeth (2024)].

In other cases, however, national populists roll out the red carpet for foreign companies, e.g., in the form of generous tax rebates and cheap land deals. German car manufacturers, for instance, are generally very happy with populist economic policies in Hungary and Poland [Sallai et al. (2024)]. Political risk stemming from national populists therefore varies. The next section explores the determinants at both the country and the sector levels.

3. Factors determining national populist political risk

3.1 Country-level determinants of populist risk

Our research identified three main factors that influence the extent to which foreign MNCs are exposed to political risk in national populist governed countries: Firstly, the degree of authoritarianization of the country's political system; secondly, the degree to which a country's growth model is dependent on FDI; and thirdly, the political alignment between the MNC's home and the host country's government [Sallai et al. (2024)].

Populist parties in government tend to use their power to fundamentally reshape the democratic setup of the country. Examples include reducing the independence of the judiciary (e.g., by replacing judges with new ones close to the government), reshaping electoral rules, and changing political procedures to concentrate power in the hands of the executive [Bartels (2024)]. In that context, the larger the parliamentary majority of a national populist party and the stronger its grip on power, the higher the likelihood of authoritarianization. For instance, Fidesz's two-thirds majority in the Hungarian Parliament allowed Victor Orbàn to adopt constitutional changes that help his chances of staying in power [Sallai and Schnyder (2021)].

Therefore, countries with longer national populist tenure and stronger populist control over parliament are riskier for MNCs. A higher degree of authoritarianism allows populist governments to more easily favor their allies in business and will therefore increase political risk for foreign multinationals. For instance, large parliamentary majorities and reduced opportunities for stakeholder input into the legislative process has allowed the Fidesz government to adopt anti-FDI legislation, like special taxes, without opposition.

A second country-level factor concerns the country's reliance on an FDI-led growth model. After the fall of state socialism, ECE countries generally were very successful in upgrading their manufacturing base relying on inward FDI from Western MNCs [Nölke and Vliegenthart (2009)]. The downside of that growth model is the dependence it creates and the fact that economic decisions are taken outside the country's borders in the headquarters of foreign MNCs. Therefore, this so-called Dependent Market Economy (DME) model increasingly has come under pressure. National populist governments have been particularly active in adopting economic policies aimed at shifting away from FDI-led to more consumption-led growth models. Natalist policies for instance are part of that strategy. For foreign MNCs operating in countries adopting such policies of reducing FDI dependence, political risk may be heightened.

A third country-level factor concerns the political proximity between the MNC's home country and the ECE host country. National populists tend to seek alternative sources of FDI from illiberal countries to reduce dependence on Western sources. Hungary's increasing efforts to attract FDI from China, for instance, can be seen in that light. More generally, national populists often make a point of supporting ideologically aligned governments, as is illustrated by Donald Trump's support for Javier Mielei's government in Argentina when facing a fiscal crisis [Nugent et al. (2025)]. For MNCs from a country whose government is politically and ideologically distant from the populist host government, political risk will be increased.

The fact that political risk stemming from national populist governments is influenced by country-level factors suggests a first way for MNCs to mitigate it, namely location choice. Indeed, the most basic strategy to mitigate country risk is to stay away if the risk is deemed excessive. That rather crude strategy, of course, also means missing out on potential profits that companies in some sectors can make in a politically risky country. Indeed, our research also shows that political risk does not only vary by country, but also by sector, implying that even in a country deemed risky some sectors may face acceptable levels of risk.

3.2 Sector-level determinants of populist risk

Our research shows that at the sector level political risk stemming from right wing populists in government varies on two key sectoral dimensions: the political and the economic strategic value of a sector to the regime [Schnyder et al. (forthcoming)].

The distinction of sectoral strategic value stems from Rosie Hsueh's (2016) work. She defines the political dimension of strategic value as considerations of external and internal security that make some sectors critical for political regime stability.

Economic strategic value, on the other hand, refers to a "sector's contribution to the national technology base and the growth and competitiveness of other sectors and the rest of the economy" [Hsueh (2016, p. 89)].

This has implications for the political risk exposure of foreign firms active in different sectors (see Table 1). The riskiest sectors are those with high political strategic value but (relatively) low economic strategic value. The media sector is an example of that. National populist parties rely particularly strongly on generating popular support through their media strategy. As a result, they may consider control over the media as a key element of regime survival. Therefore, firms active in the media sector face high levels of political risk in national populist governed countries [Schnyder et al. (2024)].

In sectors where political strategic value is low, but economic strategic value high, political risk is relatively low, because nationalist populists benefit from technology transfer via inward FDI and do not gain directly in terms of political stability from taking over such sectors. That is the case of car manufacturing, explaining its relative immunity to national populist political risk.

Sectors with low economic and political strategic value may still experience considerable risk, because in those sectors national populists may choose to implement anti-FDI policies to signal to voters the government's commitment to 'taking back control' – and can do so without causing significant economic damage. An example is the retail sector in Hungary, where domestic retailers could replace multinational retail chains relatively easily, notably due to the relatively low-tech nature of the sector and therefore the absence of dependence on inward technology transfer via MNCs.

Finally, cases of high political and economic strategic value pose a real dilemma for national populists. The financial sector is a case in point. This sector is important to achieve regime stability (e.g., by controlling interest rates, lending to households, etc) and other political goals (e.g., industrial policy by channeling credit into specific industries). However, ECE countries are also reliant on capital from outside the country. Indeed, in these countries, financial resources are notoriously scarce, leading to a high dependence on foreign banks [Piroska and Epstein (forthcoming)]. Financial services also increasingly rely on high tech controlled by multinational banks' headquarters. Retaking control of the finance sector is hence politically valuable for a national populist government but comes at a certain risk of disrupting capital and technology transfers. Here, an important indicator for financial market firms to consider is the national populist government's economic strategy. Many populist governments do not have particularly cohesive economic strategies [Benczes and Szabó (2023)]. Yet, some do develop sophisticated industrial policies, which has led scholars to describe countries under their control as 'conservative developmental states' [Bluhm and Varga (2020)]. In such cases, the political strategic value of the financial sectors increases because control over financial flows is crucial for the implementation of the national populist industrial policy. This may lead to a situation where the political gain from re-nationalizing or re-domesticating financial services firms outweighs the risks associated with losing access to foreign capital and technology in the short term. Poland under the Law & Justice (PiS) government may illustrate this case most clearly.

Our research on Hungary and Poland suggests that two additional risk factors contribute to enhanced political risk for financial service firms: The relatively long tenure in government and strong grip on power, which creates more leeway for bold action against foreign investors.

Table 1: Sectoral characteristics and economic and political strategic value and exemplary sectors

		Political strategic value	
		High	**Low**
Economic strategic value	**High**	Banking and finance	Manufacturing
	Low	Media	Agriculture and forestry Retail

Source: Schnyder et al. (forthcoming)

The next sections explore what strategies foreign companies operating in countries with national populist governments have used – successfully or otherwise – to mitigate political risk.

4. Political risk mitigation strategies

How can MNCs – especially those in the exposed financial sector – mitigate their exposure to populist political risk? The next two sections explore the role of political ties and country embedding in international institutions.

4.1. Political ties: risk mitigation or trojan horse?

Corporate political strategy has become increasingly sophisticated. Large multinational companies have developed government affairs departments that use professional political strategies such as lobbying, campaign financing, and the provision of information to governments and parliaments.

However, national populist governments' disregard for formal parliamentary procedure and the rule of law makes such sophisticated political strategies less effective [Schnyder and Sallai (2020)]. During 'democratic backsliding,' consultations about legislation are shortened (in some cases reduced from months to days), purely symbolic, or completely abolished. Parliamentary debate is replaced with governmental control over the legislative process. For companies this means the number of access points to influence legislation is drastically reduced and the timeframe to do so shortened. One of our interviewees put it in the following way:

> *"In Hungary it is not worth representing interests, because today in the parliament a 'voting machine' operates.*

"In Hungary it is not worth representing interests, because today in the parliament a 'voting machine' operates" [Sallai and Schnyder (2021)].

In such situations where formal channels of political activity are closed down, companies are increasingly forced to revert to political ties – one of the oldest and most basic political strategies. Indeed, appointing politically well-connected figures – such as former ministers or members of parliament – to your board of directors has long been used as a way of getting a sitting government's ear or curry its favors.

However, research on corporate political activity has shown that political ties are a double-edged sword. Ties with politicians from a given political party may be an asset when the party is in power but turn into a liability when it is voted out or – worse – removed in a revolution [Darendeli and Hill (2016)].

That risk is heightened still by the fact that national populist leaders tend to increasingly centralize power over the course of their tenure. In the extreme case of long-ruling populist leaders, the power elite may become so restricted that only ties with the leader and their immediate entourage will be useful, which increases risk further in times of regime change.

Moreover, national populists' economic strategies to 'take back control' means that political ties pose additional risks for MNCs. A case in point is the one of prominent Polish bankers who turned on their foreign owners.

Jan Krzysztof–Bielecki was Prime Minister of Poland in 1991 and became president of the Italian-owned Bank Pekao between 2003 and 2010. He was first economic adviser to Prime Minister Donald Tusk's liberal government from 2010 to 2014. However, during the PiS majority government, he helped the government re-polonizing of Bank Pekao in 2016 [Naczyk (2022)].

Similarly, Mateusz Morawiecki held the post of chairman of Bank Zachodni WBK (BZ WBK) between 2007 and 2015. He too was adviser to the liberal Tusk government, but joined the PiS government after 2015 to become PM from 2017 to 2023. He was actively – albeit ultimately unsuccessfully – involved in trying to orchestrate a Polish takeover of BZ WBK who was owned by the Irish AIB [Naczyk (2022)].

The two cases indicate that high-level political ties – which may be deemed necessary to mitigate political risk in uncertain national populist contexts – can easily become Trojan horses when these politicians join the nationalist movement and use their inside knowledge to help implement nationalist strategies.

The cases of the financial sector in Hungary and Poland show the difficulty multinational firms may face to manage political risk in cases where a bold and strongly established national populist government is in place. Even strong political ties may not be sufficient to mitigate the risk. Slovenia constitutes a contrasting case, which hints at another important factor in understanding how national populism affects foreign firms. Namely, the countries embedding in international institutions such as the E.U.

4.2 E.U. sheltering of Western banks in Slovenia

In Slovenia, Janez Janša's national populist SDS party was in power for ten of the past 21 years. He pursued similar nationalist economic policies to the ones adopted in Hungary and Poland. Yet, these strategies have been less successful than in the other two ECE countries. Partly, this is explained by their shorter periods in government, compared to Poland and especially Hungary, which – as mentioned above – is an important factor determining the extent to which national populists can subject the economy to their control.

There is another factor, however, which explains why even when they were in power, Slovenia's national populists were less successful in implementing their nationalist economic policies, namely Slovenia's stronger embedding in E.U.

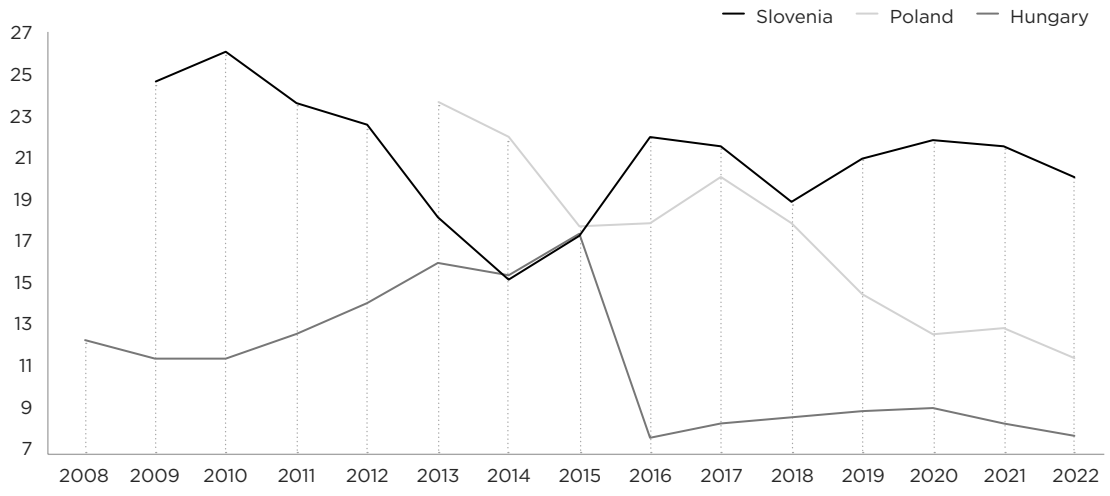
institutions compared to Hungary and Poland. Contrary to Hungary and Poland, Slovenia joined the European Banking Union (EBU) after its establishment in 2014.

The 2008 Global Financial Crisis hit Slovenian banks hard, with many of them registering a large number of non-performing loans. The Slovenian government undertook a series of capital injections, which led to a strong increase in public indebtedness. In 2013, the ECB successfully demanded a comprehensive restructuring of the Slovenian banking sector, leading to the nationalization of the two largest banks, NKBM and NLB. The new Single Supervisory Mechanism (SSM) implied that EBU member Slovenia's 'significant' banks were subject to the ECB capital adequacy standards and to regular reviews. As a result, the government's leeway to show leniency on capital requirements was curtailed and capital injections became impossible. This meant that ultimately the privatization of the banks became inevitable.

Privatizations were very unpopular and triggered massive protests and public mobilization by political parties, trade unions, civil society organizations, and leading economists. Privatization also implied increases in inward FDI into the Slovenian banking sector at a time where Hungary and Poland went the other way. The hedge fund Apollo Global Management acquired 80% of NKBM shares. NLB's share ownership became dispersed, but the European Bank for Reconstruction and Development (EBRD) became one of the largest minority shareholders before Bank of New York Mellon took control.[1]

It would, therefore, appear that national populist anti-FDI strategies were prevented in Slovenia by its membership in the EBU, which conferred the ECB the power to enforce privatization [Schnyder et al. (forthcoming)]. EBU membership meant that PiS's bank repolonization strategy or the Hungarian banking nationalism was not an option for Slovenia. Overall, contrary to Hungary and Poland, inward FDI into Slovenia has increased following the establishment of the EBU in 2014 (see Figure 1).

Figure 1: FDI inflow in financial services and insurance as a percentage of FDI between 2005-2021 in Hungary, Poland and Slovenia



Source: Schnyder et al. (2025)

[1] https://www.nlb.si/shares, last accessed August 23, 2024.

The resolute intervention of the ECB in the Slovenian banking sector case contrasts with the remarkable reluctance of the European Commission to intervene in Hungary and Poland to oppose policies that may be seen as discriminatory against E.U. multinationals, such as special and sectoral taxes. In the latter case, E.U. institutions – including the European Court of Justice (ECJ) – have generally shown remarkable leniency towards national populists' anti-FDI policies. This leniency may be surprising at least from the point of view that the E.U. is a 'liberalisation machine,' which primarily defends the interests of Western multinationals [Buch-Hansen and Wigger (2010)].

Our research suggests that a key reason why the ECB has been less lenient towards nationalist economic policies has to do with the difference in its institutional setup compared to the Commission and the ECJ [Schnyder et al. (2025)]. With its extensive supranational powers and wide-ranging isolation from politics, the ECB does not directly rely on member state governments' support to carry out its mandate. The Commission, on the other hand, has increasingly become reliant on member state support for its integration agenda, leading to what some have called 'vertical bilateralism' [Höpner (2024)] where the Commission tends to negotiate compliance with directives rather than punishing member states. Similarly, the ECJ has been considered to become increasingly politicized. Michael Ovádek's research finds that the court "systematically adjusts its rulings in response to Member States's preferences" [Ovádek (2021)]. The ECB, on the other hand, remains largely isolated from politicization and pursues its technocratic mandate more independently from member state preferences than E.U, institutions that are less isolated from politics [Epstein and Rhodes (2016)]. As our own research shows, supranationalism and technocracy help explain the ECB's greater effectiveness in challenging economic nationalist policies in Slovenia [Schnyder et al. (2025)].

For multinational banks this hints at another important country-level factor to consider when assessing political risk stemming from national populists: the extent of technocratic and supranational international organizations that a national populist country is member of. While technocracy and supranationalism are often considered to undermine democratic control over the economy, in the case of national populists, supranationalism and technocracy may constitute a roadblock to illiberal interventions in the economy. This may be interpreted as an attack on national sovereignty and the preferences of democratically-elected governments, but in the case of increasingly authoritarian and kleptocratic tendencies in national populist countries [Sallai and Schnyder (2021)], it may put a halt to the further concentration of economic power in the hands of the populist elite.

5. Conclusion

Overall, then, countries run by national populist governments in ECE have constituted particularly challenging contexts for foreign MNCs and banks in particular.

Given the extensive periods of national populist governments, the region provided a laboratory and training ground for firms to learn from various political risk mitigation strategies. Our research provides a roadmap towards successful political risk management strategy for companies facing similar challenges in other parts of the world as well.

Location choice – deciding to invest or not, 'stay or go' in a given country – based on country-level risk factors is one possibility. For financial services firms, one key country-level factor is whether national populist governments pursue a "conservative developmental state strategy" [Bluhm and Varga (2020)]. Those who do tend to be particularly risky for financial market firms because the sector becomes politically strategically valuable, potentially outweighing the government's concern with cutting off transfer of capital and technology from outside the country. Where national populist governments have more haphazard economic strategies, financial services' high economic strategic value that relies considerably on capital and knowledge transfer from abroad will be less exposed. We can draw the following practical lesson from this insight:

Lesson 1. For financial services firms, it is worth analyzing a country's economic policy closely to assess the extent to which a government is likely to aggressively intervene in the financial sector to achieve its strategic development goals. Countries with a very ambitious developmental strategy are more likely to nationalize banks and other financial services firms, which should be factored into any country risk assessment.

This also suggests more broadly that country-level assessments of political risk alone are too crude to make sound investment decisions. Our research shows large variations in risk exposure depending on the economic and political strategic value a host government associates with a given sector. Political risk is highest in sectors that are considered as high political strategic value, but that add relatively little value or technology to the country's economy. It is lowest in sectors that add much value and technology to the economy, but control over which is not considered politically important. Risk is more difficult to assess in sectors where both political and economic value are high, or both are low. As a result, a second lesson can be drawn:

Lesson 2. In their location choice, beyond country-level risks, investors should consider political risk at the sectoral level. A thorough analysis of a national populist government's economic strategy can help assess which sectors the government may consider particularly politically valuable (and hence may seek to control) and which ones it may consider economically valuable (and hence may consider too risky to disrupt). Which sectors fall into which category will also depend on the country's economic structure and its place in the world economy. These factors, therefore, need to be taken into account as well.

Yet, the rather crude location choice approach may mean missing out on opportunities that exist in risky countries. Here, the risk mitigation strategies of foreign MNCs in ECE can provide more nuanced insights for companies beyond the region. I have particularly highlighted the importance of political ties as a way of mitigating political risk in contexts where more professional corporate political strategies lose value. However, I also argued that such ties can be used against the firm in question. These insights suggest a third lesson.

Lesson 3. In countries governed by national populists, ties with politically well-connected individuals – e.g., through managerial and board positions – are an important complement to professional, market-based forms of political activities. However, very careful selection and due diligence will be required to avoid appointing political actors that enhance rather than reduce political risk. Here, a thorough understanding of the political power relations and allegiances in the country, possible electoral trends and social

movements leading to regime changes, as well as an analysis of individual actors' past and present political affiliations and incentives are crucial to make the right choices.

Such strategies may not completely shelter MNCs from populist anti-FDI policies, but they may help them weather the stormy waters of the national populist wave better than would otherwise be the case.

Funding note: This work was supported by the NORFACE-funded research program 'Democratic governance in a turbulent age (Governance)' grant no. 462-19-080 (POPBACK project – popback.org).

References

Bartels, L.M., 2024. Democracy erodes from the top: leaders, citizens, and the challenge of populism in Europe. Princeton University Press

Benczes, I. and K. Szabó, 2023. An Economic Understanding of Populism: A Conceptual Framework of the Demand and the Supply Side of Populism. Political Studies Review, 21(4), 680–696. https://doi.org/10.1177/14789299221109449

Bluhm, K. and M. Varga, 2020. Conservative Developmental Statism in East Central Europe and Russia. New Political Economy, 25(4), 642–659. https://doi.org/10.1080/13563467.2019.1639146

Buch-Hansen, H. and A. Wigger, 2010. Revisiting 50 years of market-making: The neoliberal transformation of European competition policy. Review of International Political Economy, 17(1), 20–44

Darendeli, I.S. and T.L. Hill, 2016. Uncovering the complex relationships between political risk and MNE firm legitimacy: Insights from Libya. Journal of International Business Studies, 47, 68–92. https://doi.org/10.1057/jibs.2015.27

Epstein, R.A. and M. Rhodes, 2016. The political dynamics behind Europe's new banking union. West European Politics, 39(3), 415–437. https://doi.org/10.1080/01402382.2016.1143238

Höpner, M., 2024. Lässt sich das europäische Demokratiedefizit beheben? [Can the European democratic deficit be eliminated?] Merkur 78 (896), 55–63

Hsueh, R., 2016. State Capitalism, Chinese-Style: Strategic Value of Sectors, Sectoral Characteristics, and Globalization. Governance, 29(1), 85–102. https://doi.org/10.1111/gove.12139

Naczyk, M., 2022. Taking back control: Comprador bankers and managerial developmentalism in Poland. Review of International Political Economy, 29(5), 1650–1674. https://doi.org/10.1080/09692290.2021.1924831

Nemeth, M., 2024. Hungary's retail tax under fire: EU investigates discrimination claims by Austrian government and SPAR, Daily News Hungary, https://dailynewshungary.com/hungary-government-spar-eu/

Nölke, A. and A. Vliegenthart, 2009. Enlarging the Varieties of Capitalism: The Emergence of Dependent Market Economies in East Central Europe. World Politics, 61(04), 670–702. https://doi.org/10.1017/S0043887109990116

Nugent, C., K. Duguid, and M Stott, 2025. US Treasury intervenes in Argentina's currency market. Financial Times October 9. https://www.ft.com/content/778e4609-772b-4edd-83d2-d8915f1c4628

Ovádek, M., 2021. Supranationalism, constrained? Locating the Court of Justice on the EU integration dimension. European Union Politics, 22(1), 46–69. https://doi.org/10.1177/1465116520967414, p.46

Piroska, D. and R.A. Epstein, forthcoming. "Return to Diversity": Explaining Variation in East Central European Financial Strategies. Journal of Common Market Studies

Sallai, D. and G. Schnyder, 2021. What Is "Authoritarian" About Authoritarian Capitalism? The Dual Erosion of the Private – Public Divide in State-Dominated Business Systems. Business & Society, 60(6), 1312–1348. https://doi.org/10.1177/0007650319898475

Sallai, D., G. Schnyder, D. Kinderman, and A. Nölke, 2024. The antecedents of MNC political risk and uncertainty under right-wing populist governments. Journal of International Business Policy, 7(1), 41–63. https://doi.org/10.1057/s42214-023-00154-3

Schnyder, G. and D. Sallai, 2020. Between a rock and a hard place: Internal and external institutional fit of MNE subsidiary political strategy in contexts of institutional upheaval. Journal of International Management, 26(2), 1–19. https://doi.org/10.1016/j.intman.2020.100736

Schnyder, G., M. Radl, F. Toth, M. Kucukuzun, T. Turnšek, B. Çelik, and M. Pajnik, 2024. Theorizing and mapping media ownership networks in authoritarian-populist contexts: a comparative analysis of Austria, Hungary, Slovenia, and Turkey. Media, Culture and Society, 46(1), 38–59. https://doi.org/10.1177/01634437231179366

Schnyder, G., A. Nölke, and D. Sallai, forthcoming. Centimanes v. Titans: Right-wing Populist Governments' Treatment of Foreign Multinationals in East Central Europe (under review)

Schnyder, G., A. Nölke, D. Kinderman, and D. Sallai, 2025. Technocracy, Supranationalism, and Right-wing Populism: The variegated sheltering of Western assets in East Central European countries. Journal of Common Market Studies. https://doi.org/10.1111/jcms.70073

From global factories to sovereign AI pods:

how geopolitics and GenAI rewire software delivery

Author | **Gerhardt J. Scriven** | Executive Director, Capco

Abstract

While geopolitics has always shaped the global order, it is now exerting a far more direct and constraining influence, transforming regulatory divergence, data sovereignty, sanctions and export controls (including restrictions on technology transfer), and critical-infrastructure resilience into binding constraints on cross-border service delivery in financial services and energy. In parallel, GenAI is reshaping global service delivery models by compressing routine tasks into AI-assisted workflows and deepening technical integration even as jurisdictions fragment. Using the software development lifecycle (SDLC) as a high-signal microcosm, this paper shows how geopolitical constraint and AI intensity shift delivery from labor arbitrage factories toward smaller, higher seniority, jurisdiction bounded AI augmented crews, while reshaping education pathways and high skill mobility. A 2×2 framework maps four plausible futures – global factory, augmented global delivery, localized duplication and sovereign AI pods – and the conditions that produce each. It closes with instrumentable indicators and a sequenced response to avoid localized duplication by hardening AI-aware SDLC controls, redesigning sourcing around data sovereignty, and rebuilding talent pipelines so that systems thinking and secure engineering become early-career defaults. This enables staffing for the control plane – practitioners who can validate AI outputs and hold accountable decision rights in regulated and critical environments – while also supporting selective mobility for scarce, high-accountability roles.

1. Introduction

After decades in which globalization expanded cross-border operating models and reinforced large-scale labor arbitrage, the operating environment for financial services and energy firms is tightening in two directions at once.

First, geopolitical contraction is becoming operational. In practical terms, this increasingly manifests as regulatory divergence across jurisdictions, data localization and sovereignty expectations, export controls and sanctions regimes, intensified third-party scrutiny as well as renewed emphasis on critical infrastructure protection [Airaudo et al. (2025), Gopinath et al. (2024), PIFS (2024)]. For example:

- For the financial services domain – under the E.U.'s Digital Operational Resilience Act (E.U. DORA) banks and insurers cannot treat cloud/tooling merely as procurement; they must be able to evidence resilience, governance, and concentration-risk controls across their critical providers [European Union (2022)].

- For the energy domain – in the U.S., critical infrastructure protection reliability standards translate directly into measurable control obligations that shape what operational telemetry exists, who can access it, and how changes are governed – all of which feeds back into delivery geography and operating models (NERC CIP-010-4; NERC CIP-013-2).

Note that financial services and energy are treated together here because both are regulated, resilience-critical sectors where technology delivery is inseparable from auditable controls, accountable decision rights, and third-party risk – making them structurally sensitive to the same 'how/where' pressures.

These regulatory pressures do not merely add compliance overhead; they constrain where sensitive work can be performed, how systems are architected, and which vendors and locations remain viable for delivery at scale. It should be noted that fragmentation does not necessarily suggest a collapse of globalization in aggregate: flows can remain substantial overall, while being reweighted and re-routed along geopolitical alignment and policy constraints.

Second, generative AI (Gen AI) is changing the mechanics of technology delivery. In practical terms, GenAI accelerates and partially automates routine knowledge-work tasks embedded in the software development lifecycle (SDLC): translating requirements into artefacts; drafting and refactoring code; generating tests and documentation; and improving incident triage and runbook execution [DORA (DevOps Research and Assessment) (2024), Cui et al. (2025)]. For example, reporting on Meta's internal coding assistant describes capabilities including analyzing failed tests and submitting proposed fixes for human review [Dixit (2025)]. Similarly, Google Cloud documentation describes how Gemini Cloud Assist can be used directly inside Cloud Logging to summarize log entries during incident triage [Google Cloud (2025)]. Within Capco Brazil's AI lab, a publicly available requirements document was analyzed from seven distinct viewpoints (inspired by the Systems Engineering Methodology) using GenAI, and the corresponding resultant assets were then used to generate a working and bug-free application prototype using Agentic AI – all within a single day.

Notably, this evolution is not suggesting a wholesale replacement of the engineering discipline, but rather a reconfiguration of how work is produced, reviewed, and governed – often compressing cycle times while increasing the premium on verification, risk controls, and senior technical judgment.

The combination of geopolitical contraction and AI-infused delivery creates a perfect storm in both how work is done (team composition, controls, cadence, and accountability) and where work is done (the geography of delivery, the viability of offshore factory models, and the demand for jurisdiction-specific expertise) for the financial and energy sectors. These particular sectors act as bellwethers for how geopolitical constraint and GenAI-driven software delivery compression translate into binding operating constraints on tooling, data access, and accountable decision rights. The SDLC itself provides a precise lens for analyzing these shifts because it sits at the intersection of productivity, risk, and cross-border delivery – and it is structurally central to both financial services and energy, where complex technology estates must meet stringent resilience, safety, and regulatory expectations.

Thesis: The combined impact of regulatory fragmentation and GenAI is reshaping the operating environment for the new global order, shifting delivery away from large labor-arbitrage models toward smaller, higher-seniority,

jurisdiction-bounded AI-augmented crews, with material knock-on effects for education pathways and high-skill mobility. This thesis is demonstrated through changes in SDLC mechanics and the resulting reconfiguration of delivery geography.

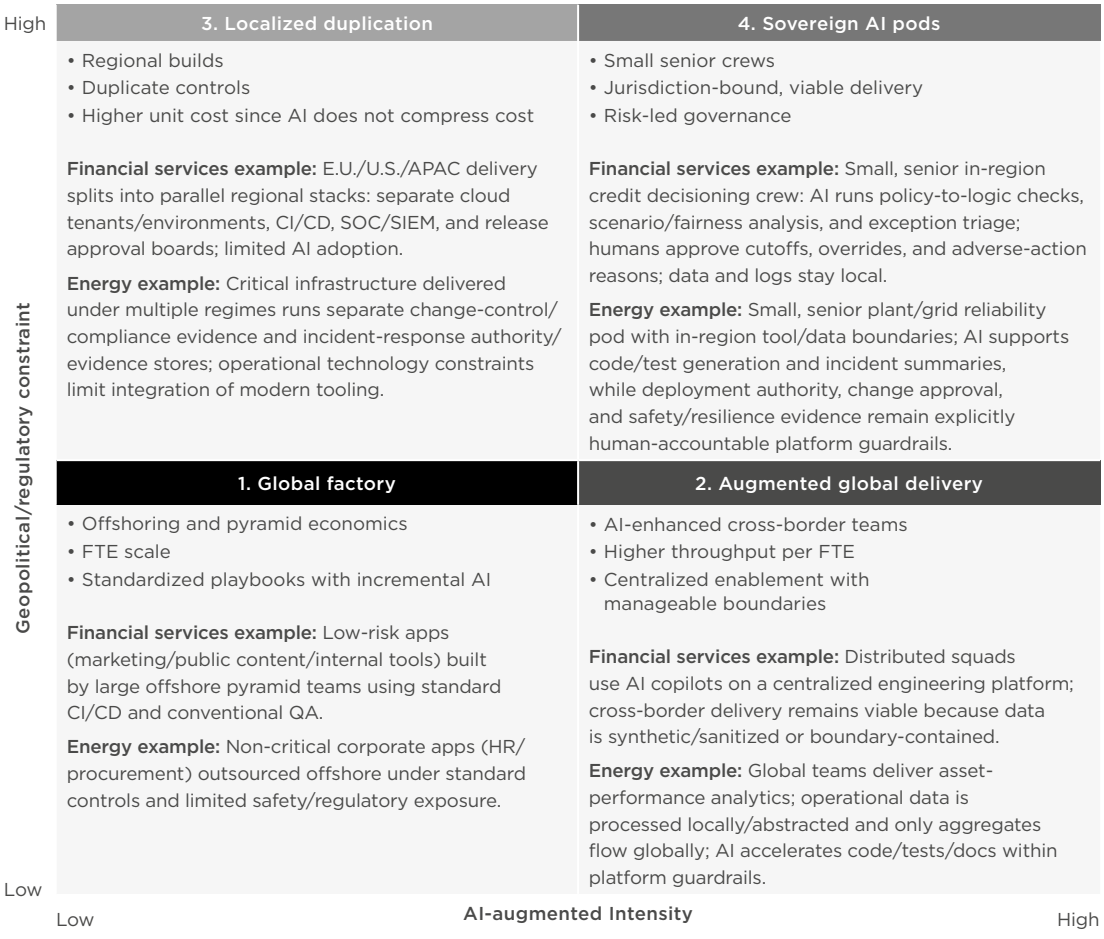
The remainder of this paper translates these two forces into a practical operating-model lens: a 2×2 framework that distinguishes four delivery archetypes and a small set of indicators that show which archetype an organization is drifting towards. The SDLC is used throughout as the microcosm to show where these forces manifest and what a sequenced response might look like in practice.

2. A simple framework: global delivery archetypes by constraint and AI intensity

Towards establishing a compact model for this complex operating environment, we utilize a simple 2×2 framework, as is illustrated in Figure 1.

In Figure 1, the horizontal axis represents AI augmentation intensity (low → high): the extent to which delivery work is routinely accelerated or partially automated by GenAI across SDLC activities. The vertical axis represents geopolitical/regulatory constraint (low → high): the degree to which delivery is bounded by jurisdictional rules, supervisory expectations, and critical-infrastructure risk posture. Note that in energy

Figure 1: Global delivery archetypes by regulatory constraint and AI intensity



– and increasingly in systemically important financial services – critical-infrastructure reliability and resilience requirements translate directly into enforceable control expectations, making delivery design (tooling, access, evidence, and change authority) a first-order operating constraint rather than an implementation detail.

The 2x2 framework yields four operating-model archetypes:

1. **Global factory** (low constraint, low AI) reflects classic offshoring and pyramid economics optimized for scale and unit cost, i.e., business as usual.
2. **Augmented global delivery** (low constraint, high AI) retains cross-border teaming but retools it around AI copilots and centralized platforms, typically increasing throughput while preserving global integration at the workflow layer.
3. **Localized duplication** (high constraint, low AI) emerges where jurisdictional barriers force regional builds and duplicated control stacks, while AI augmentation intensity is low (or adoption is constrained by policy). It results in multiple mini-factories, one per jurisdiction, each with its own tooling, governance, and risk control raising cost and reducing change velocity.
4. **Sovereign AI pods** (high constraint, high AI) combine tight jurisdictional boundaries with AI-enabled productivity, resulting in smaller, higher-seniority crews, stronger governance, and tooling/data that is explicitly bounded to a regulatory perimeter.

Crucially, this framework is not a forecast of a single future but describes the conditions under which different delivery models become economically and operationally rational – and it provides a consistent lens for assessing implications for workforce formation, education pathways, sourcing strategy, and high-skill mobility.

3. The SDLC case: what changes in how work is done

The SDLC is a structured set of activities organizations use to translate an idea or requirements into working software that is released into day-to-day business use [ISO/IEC/IEEE (2017)]. While several delivery methodologies exist (e.g., sequential waterfall, iterative agile, or hybrid variants), the SDLC exists to make change repeatable and defensible by establishing clear handoffs, decision rights, and quality controls across planning, design, build, test, deployment, and operations phases. While SDLC operations and governance practices may differ across institutions, the underlying lifecycle principles and disciplines, especially those tied to risk, traceability, and operational resilience, remain persistent. As such, the SDLC is an appropriate micro-foundation for understanding the operational impact of GenAI because it is where productivity, risk management, and organizational accountability intersect.

At an industry level, DORA's 2024 research explicitly frames AI as having a significant impact on software development practices and outcomes [DORA (DevOps Research and Assessment) (2024)]. At the level of measured productivity, large-scale randomized field experiments with software developers show that access to an AI coding assistant can increase completed tasks, with the combined analysis across three real-world trials reporting a 26.08% increase in weekly completed tasks among developers using AI [Cui et al. (2025)].

Note that this does not eliminate the SDLC; it changes the balance of work inside it – shifting effort from first-draft production toward verification, governance, and higher-order engineering judgment.

What follows is a pragmatic stage-by-stage decomposition: SDLC stage, what AI changes, what it does not change, implications for team

shape and controls. The intent is to make the AI-augmented crew concrete and operational rather than aspirational.

3.1 Plan/discover

What AI changes. In planning and discovery, GenAI can improve the speed and breadth of discovery artifact synthesis, such as summarized stakeholder inputs, initial requirement specifications, as well as comparable solution options. In practice, this increases throughout into the delivery pipeline – more candidate options, faster initial articulation, and quicker iteration on scope.

What AI does not change. In regulated, safety- and availability-critical environments, GenAI cannot own prioritization, accountability, or regulatory interpretation. The decision rights around what to build (and what not to build), why it is justified, and how it aligns with risk appetite remain inherently human because they are governance decisions, not text-generation tasks.

Implications for team shape and controls. The AI-augmented crew typically shifts from requirements authorship towards requirements stewardship. Controls tighten around (i) traceability (what sources informed the requirement), (ii) approval (who signs off), and (iii) reproducibility (whether the same artifact can be regenerated and validated). This naturally pushes teams toward smaller groups with clearer accountable owners, supported by AI for drafting and analysis.

3.2 Design

What AI changes. In design, GenAI accelerates pattern selection and option exploration: feeding off reference architectures, producing structured design documents, and prompting threat modeling questions that broaden coverage (e.g., common misconfigurations and data handling paths). Qualitative evidence from software engineering professionals suggests GenAI can increase productivity but also surfaces adoption barriers that are sociotechnical (how tools fit into routines and organizational practices), not merely technical [Russo (2024)].

What AI does not change. Architectural decisions remain bounded by operational realities (latency, resilience, observability, audit evidence, and integration complexity), none of which can be dismissed by faster drafting. In energy-sector delivery, this validation burden is often amplified by the interface between IT delivery and operational technology environments, where safety and availability expectations make failure modes and change impact analysis materially less forgiving.

Implications for team shape and controls. Design becomes more explicitly governed. An AI-augmented crew benefits from standardizing design checkpoints (architecture review, security review, data-handling review) and instrumenting design artifacts for accountability. This is also where trustworthy AI disciplines begin to appear as delivery controls: NIST's Generative AI Profile, for example, emphasizes governance, pre-deployment testing, and incident disclosure as key considerations for managing risks novel to or exacerbated by generative systems [Autio et al. (2024)].

3.3 Build

What AI changes. In build activities, GenAI speeds scaffolding, refactoring, and translation across languages/frameworks. It can propose implementations, generate boilerplate code, and assist in navigating legacy estates by explaining unfamiliar code paths. The empirical signal here is increasingly difficult to dismiss: the large-scale field experiments cited earlier found not only higher completed task volumes but also increases in intermediate production indicators such as code updates (commits) and compilation frequency – consistent with acceleration of iterative development work [Cui et al. (2025)].

What AI does not change. Correctness, secure-by-design decisions, and dependency hygiene remain human responsibilities. The faster code can be produced, the more important it becomes to ensure it is not merely plausible, but correct, maintainable, compliant with enterprise standards, and resilient to security weaknesses.

Implications for team shape and controls. This is the point in the SDLC where the economics of the junior-heavy engine starts to change. When first-draft code production becomes cheaper, the bottleneck shifts to expert review, integration, and risk controls. The pragmatic control response is a stronger review and validation spine: code review standards that explicitly account for AI assistance, secure SDLC automation (e.g., scanning and policy gates), and stronger traceability of generated artifacts. NIST's emphasis on lifecycle risk management provides a useful external anchor for why velocity must be paired with control [Autio et al. (2024)].

3.4 Test

What AI changes. GenAI expands test generation capacity: proposing unit and functional tests together with regression scenarios and generating synthetic test data derived from requirements and prior defect patterns. It can also draft test documentation and translate between business acceptance criteria and technical assertions, improving alignment between intent and verification.

What AI does not change. Test strategy, quality gates, and evidence suitable for audit remain human-owned. In financial services and energy, testing is not simply about coverage – it is about demonstrating controlled change, reliability, and defensible operation under stress. That requires judgment about what must be proven, at what rigor, and how to document that proof.

Implications for team shape and controls. AI pushes teams toward a more formalized notion of evaluation as a first-class artifact. In addition to code tests, teams increasingly maintain evaluation suites for AI-assisted workflows (e.g., prompt libraries, retrieval configurations, and acceptance thresholds for model-assisted outputs). This aligns with risk-management emphasis on pre-deployment testing and governance for generative systems [Autio et al. (2024)].

3.5 Deploy/operate

What AI changes. In operations, GenAI improves triage and diagnosis workflows: summarizing incidents, proposing likely root causes, generating runbook steps, and accelerating knowledge retrieval across fragmented documentation or directly from the source code itself. This can reduce mean time to understand (MTTU) and make operational teams more effective – particularly in complex estates.

What AI does not change. Production risk decisions and change approval logic remain human responsibilities, even if workflow automation is extensive. High-impact operational actions – those affecting safety, systemic stability, or regulatory reporting – require explicit accountability and controlled decision rights. In energy and other critical infrastructure domains, this constraint is often reinforced by operational separation and formalized change authority for high-impact environments, where evidence of safe operation and controlled change can be as material as functional delivery velocity.

Implications for team shape and controls. Human-in-the-loop becomes a design principle, not a slogan. AI outputs are treated as decision support; approval rights and escalation paths are clarified; and incident response is instrumented for audit and disclosure expectations. NIST's profile explicitly includes incident disclosure as

a key consideration in GenAI risk management, reinforcing the need to connect operational AI use to governance and reporting disciplines [Autio et al. (2024)].

3.6 The central 'how' impact

The SDLC is not dead; it is reconfigured. As GenAI compresses routine production work, delivery value shifts toward architecture, verification, provenance, security, and accountable approvals, i.e., a thicker control plane and higher seniority at decision points. This is because speed without controls magnifies risk and reduces defensibility. This is the operating-model logic behind the move from traditional pyramid delivery (junior-heavy execution) through diamond models (more experienced blended onshore/offshore teams) toward AI-augmented crews: smaller teams, higher effective seniority in decision points, and a thicker control plane embedded throughout delivery.

4. The delivery geography case: what changes in where work is done

When GenAI reduces the quantity of routine build work and simultaneously increases the importance of governance, auditability, and data controls, delivery location becomes more constrained and more strategic as a direct outcome. Put differently, AI shifts the binding constraint from "how many developer-hours can we buy in the cheapest market?" to "where can we legally, securely, and accountably execute an SDLC that now depends on sensitive data, model usage controls, and regulator-facing evidence?". This is consistent with evidence that geoeconomic fragmentation is increasingly expressed through policy-driven frictions, including a wider use of interventions that shape trade and financial flows across geopolitical lines [Airaudo et al. (2025)].

4.1 From labor arbitrage to risk and jurisdiction arbitrage

In classic global factory delivery, geography decisions optimize labor cost and scale. Under higher-constraint conditions (upper half of Figure 1), the dominant optimization changes: the highest-value SDLC decisions – architecture, security posture, model/tool approvals, and release accountability – move closer to (i) the data, (ii) the regulated or critical-infrastructure entity, and (iii) the executives who carry institutional accountability.

This is not simply cultural preference. In financial services, cross-border data movement is increasingly constrained by localization requirements and supervisory expectations that shape viable cloud architectures and operating models [PIFS (2024)]. In energy and critical infrastructure industrial contexts, enforceable cybersecurity and supply-chain requirements similarly harden expectations around configuration change control, vulnerability assessment, and vendor access – placing practical limits on where production-adjacent work and evidence generation can occur [NERC CIP-010-4; NERC CIP-013-2; ISA/IEC 62443].

Once GenAI tooling becomes embedded in SDLC execution, the control plane (data access paths, telemetry, provenance, and evidence) becomes integral to delivery itself – not an overlay.

4.2 From 'factory scale' to 'crew leverage'

The shift from pyramid economics toward smaller, higher-seniority AI-augmented crews changes the geometry of delivery – spanning how software is developed as well as how it is implemented and then maintained over a longer period. A ten-person pod led by a senior engineer, architect, and risk/security counterpart can be localized (or duplicated regionally) far more easily than

a 200-person offshore factory that depends on high volumes of junior build capacity. In the language of Figure 1, this represents one way in which organizations can migrate from global factory toward sovereign AI pods when constraints rise: AI compresses routine throughput, making smaller teams viable without necessarily losing productivity.

It should be noted, however:

- that duplication remains expensive without strong platform foundations (common pipelines, shared controls, reusable reference architectures) – which is why the Augmented Global Delivery quadrant often depends on centralized platforms and standardized governance patterns that can be deployed across regions;
- while AI intensity remains low – because tools cannot be adopted at scale or governance prevents sustained use – organizations can land in Localized Duplication: regional builds and duplicated control stacks, with higher cost and slower change.

4.3 What happens to offshore delivery

Offshoring does not end; it changes form. As routine build work becomes less economically central, offshore delivery is more likely to concentrate in:

- **Specialized engineering centers** (platform engineering, security engineering, test automation, SRE/observability, tooling enablement), where capabilities can be industrialized and reused; and
- **Tightly bounded operational roles,** including follow-the-sun operations, but with stricter segmentation: hardened toolchains, controlled datasets, and auditable handoffs rather than open-ended access to production-adjacent environments.

The practical effect is a more explicit partitioning of SDLC activities into sovereign work (must be local) versus portable work (can be distributed), and that partition often maps cleanly onto data access and change authority.

4.4 Vendor model implications

As AI changes unit economics, FTE-based contracting weakens: the buyer is no longer purchasing linear capacity; they are purchasing outcomes delivered through a mix of humans, automation, and vendor platforms. As point in case, Deloitte's 2024 Global Outsourcing Survey reports broad adoption of AI within outsourced services and highlights increased emphasis on value/outcome-based relationships, alongside heightened governance needs for the extended workforce ecosystem [Deloitte Insights (2024)].

Market data also indicates continued enterprise spend shifting towards cloud-first platforms in the context of AI adoption [ISG (2025)]. The governance burden rises in parallel: model usage policies, provenance logging, audit trails, and evidence packages become contractually material, because they determine whether the delivery model is regulator-acceptable and operationally defensible.

This is one of the clearest ways financial services and energy actively shape the new global order: by embedding jurisdictional accountability and evidentiary standards into procurement and delivery contracts, they indirectly reshape vendor roadmaps, managed-service designs, and the practical rules of engagement for cross-border digital work. Where integration used to be driven primarily by cost and platform scale, it increasingly becomes conditional on auditability, provenance, and control planes. In the new world a vendor or platform provider often cannot realize outcomes through a remote factory alone; it needs senior engineers embedded close to the domain, stakeholders, and operational constraints.

5. Second-order effects

5.1 Education and the career pipeline

The fundamental implication for education is not simply that skills must change, but that the SDLC learning ladder changes when routine tasks are automated. Historically, early-career software engineers progressed by producing high volumes of first-draft work (boilerplate code, straightforward tickets, basic test cases, and documentation) while learning standards and risk instincts through repetition and review. As GenAI tools measurably increase throughput on routine software tasks [Cui et al. (2025)] and become embedded in delivery practices [DORA (2024)], that rung of the learning ladder compresses. The same is true for non-technical finance professionals: in many finance roles, early-career progression has historically depended on high-repetition tasks – building spreadsheets, drafting memos, preparing reconciliations, compiling management reporting, writing policy/control narratives, etc. GenAI can now generate credible first drafts of many of these outputs. That changes the learning ladder in the same way it does for engineers: less time spent producing, more time required supervising, validating, and exercising judgment.

The result is an acute training challenge: how to develop sound engineers when the practice surface area shifts away from first drafts and toward validation, integration, and control. This pushes a concrete rebalancing of curricula and early-career development toward four emphases:

1. **Systems thinking and architecture fundamentals.** Graduates must reason across dependencies, service boundaries, data lifecycles and failure modes, i.e., systems thinking – and develop skills that remain non-trivial even when code is faster to draft [World Economic Forum (2025)]. In safety- and availability-critical contexts, systems thinking is not only a mindset; it is frequently operationalized through systems engineering and model-based systems engineering (MBSE) disciplines aligned to lifecycle standards (INCOSE, Systems Thinking).
2. **Verification and validation discipline.** Testing strategy, reasoning about edge cases, and evidence generation for auditability become core competencies, not specialist add-ons [Autio et al. (2024)].
3. **Secure coding and dependency risk.** As code generation accelerates, so does the risk of propagating insecure patterns or fragile dependencies; security must be taught as an engineering default, not a downstream gate [Autio et al. (2024)].
4. **Working effectively with AI tools.** Prompting is secondary. The primary skill is supervision and evaluation: assessing output quality, detecting failure modes, enforcing policy, and maintaining provenance and accountability [Autio et al. (2024), UNESCO (2023)].

Apprenticeship models therefore need to be redesigned. Juniors can learn by validating AI output, expanding tests, improving guardrails, and conducting structured reviews – work that directly builds judgment while producing defensible artifacts for regulated environments. At a policy level, this aligns with broader calls for more agile, data-driven skills and lifelong learning as skill demand shifts faster than traditional education cycles are able to adapt [OECD (2025a)], alongside concerns that training supply may lag emerging AI skill needs [OECD (2025b)].

There is also a need to understand the feasibility of workers acquiring newly valuable expertise as AI changes task composition [National Academies of NASEM (2025)]. Ultimately, the objective is not to advocate labor substitution, but to ensure

the talent pipeline remains viable as the SDLC reorganizes around AI-augmented production and control.

5.2 Migration and talent mobility

The 'where' implications of SDLC reconfiguration are ultimately expressed through people flows, but in this paper, they are best framed as a workforce composition and location problem driven by operating constraints rather than as a general debate about migration. As delivery models shift away from large labor-arbitrage factories and toward smaller, jurisdiction-bounded AI-augmented crews, the center of gravity moves toward roles that carry explicit accountability for risk, controls, and regulatory defensibility. Three implications follow:

- First, organizations should expect increased demand for high-accountability roles in-region – including architects, security and resilience engineering leadership, AI governance leads, and risk owners who can sign off controls and evidence packages. This follows from the way GenAI changes the SDLC: it accelerates production, but raises the premium on governance, provenance, pre-deployment testing, and incident disclosure disciplines, which are inherently tied to accountable decision rights [Autio et al. (2024)]. Where data localization and supervisory expectations constrain architectures and operating models, accountable roles predictably cluster closer to regulated entities, sensitive data, and primary regulators [PIFS (2024)].
- Second, continued demand for global talent is to be expected, but with more selective profiles. As routine work compresses, firms will disproportionately value adaptable technology generalists, senior engineers, and IT professionals who combine domain context with risk fluency – capabilities aligned toward analytical, systems, and creative thinking, technology literacy, and continuous learning [World Economic Forum (2025), OECD (2025a)]. The practical effect is not less global talent, but a reweighting toward roles that can supervise AI-assisted workflows, validate outputs, and operate within tighter governance constraints.
- Third, a plausible organizational response is the emergence of talent corridors: formalized mobility pathways for scarce, high-accountability roles as firms rebalance where accountability must sit. Recent OECD analysis describes how OECD countries are increasingly tailoring labor migration channels to specific labor market needs, adjusting eligibility thresholds, and expanding bilateral and regional cooperation mechanisms that facilitate labor mobility [OECD (2025c)]. In practice, these corridors are likely to be explicitly aligned to roles that sit at the intersection of engineering, governance, and regulated delivery.

6. Conclusion: what leaders should watch and do

The analysis in the previous sections translates a complex macro narrative into operational mechanics where two forces are colliding inside delivery organizations:

- GenAI is compressing the SDLC and shifting the work mix toward higher-order judgment. As a result, delivery value moves from sheer volume of build capacity to control-plane capabilities: verification, provenance, security, architecture, and accountable approvals.
- Jurisdictional friction is increasing. Re-concentration of control activities in-region is consistent with data localization and supervisory constraints, and with broader evidence that fragmentation increasingly operates through policy-driven frictions.

Importantly, financial services and energy do not merely adapt – through themselves hardening evidentiary expectations for digital operations, third-party dependence, and critical infrastructure resilience, they actively influence the operating standards and commercial designs that will govern cross-border digital delivery in the new world order.

The 2×2 framework (Figure 1) makes the operating-model implications concrete. AI creates a rightward pull that can increase throughput and standardize workflows while regulation creates an upward push that limits where work, tooling, data, and decision rights can reside. However, this is the new global order contradiction in delivery form: technical integration can deepen even as legal and political perimeter fragment. Depending on where a firm sits on these two axes, it will rationally converge on one of four archetypes (Global Factory, Augmented Global Delivery, Localized Duplication or Sovereign AI Pods) – each with distinct economics, risk posture, and delivery geometry, i.e., how delivery is physically and organizationally arranged which, in practical terms, captures:

- **Footprint and distribution:** how many delivery sites exist, where they are, and whether delivery is centralized or dispersed.
- **Team topology:** how work is split across onshore/nearshore/offshore; the mix of local versus global teams; and whether teams are pods, centers, or factories.
- **Flow of work and handoffs:** the number and type of handoffs across borders, time zones, and legal entities; and whether work is sequential (assembly-line) or end-to-end (pod-based).
- **Placement of decision rights:** where approvals, accountability, and risk ownership sit (e.g., local sign-off versus centralized governance).
- **Data/tooling residency constraints:** where data is processed, where environments run, and how tightly they are partitioned by jurisdiction.

Second-order effects follow operational mechanics directly. As routine drafting of SDLC deliverables becomes cheaper and faster, the SDLC learning ladder compresses: junior progression shifts from producing first drafts to validating AI output, expanding test evidence, and strengthening guardrails. In parallel, the talent location equation changes: accountable roles (risk, security, architecture, model governance) concentrate closer to regulated entities and critical environments, while global talent demand persists but becomes more selective in terms of the type of work that is done and credential heavy.

For leaders, the central question is not whether GenAI will change delivery, but whether the organization can industrialize GenAI without drifting into Localized Duplication – a state where rising constraints force regional replication of tooling and controls while productivity gains fail to offset the cost and complexity. The practical objective is to land in a stable operating model – typically Augmented Global Delivery and/or Sovereign AI Pods by explicitly designing the control plane, the sourcing footprint, and the talent pipeline. Towards this, a viable approach is to instrument a small set of observable indicators that reveal which quadrant the organization is moving towards and then act in a sequenced way to steer toward the desired operating model.

6.1 Signals to watch

1. **Delivery seniority mix.** Track the senior-to-junior ratio in delivery teams, especially in decision roles (architecture approvals, security sign-off, release accountability). If this ratio rises steadily, it indicates that AI is reducing routine production work.
2. **Speed versus quality debt.** Instrument cycle time using lead time for change and deployment frequency and pair this with rework and stability measures (e.g., defect escape rate, rollback rate, and change failure rate). Acceleration is only net-positive if it does not convert into downstream incident load or audit-defensibility gaps.
3. **Contract construct shift.** Track the share of spend and critical workloads delivered under FTE/time-and-materials constructs versus outcome-, platform-, or managed-service constructs. Separately, track whether contracts explicitly require AI usage controls, which become material as governance expectations rise.
4. **Geographic redistribution of control versus execution.** Track where control plane work is performed: risk acceptance, security exceptions, model/tool approvals, change authority, incident disclosure decisions, and audit evidence production – versus where execution plane work is performed (build, automation, lower-sensitivity operations).

Using these signals, a simple scoring system (that turns the four signals into indicators) can be devised to calculate AI augmentation versus constraint scores and then correlate this back to the 2x2 framework. This datapoint – a snapshot in time – can be refreshed quarterly.

6.2 What to do

Actions that steer the organization toward its chosen delivery posture (i.e., the intended mix of archetypes in the 2x2 framework) by workload sensitivity and jurisdictional constraint):

1. **Redesign SDLC controls for AI-augmented workflows**

   Embed AI-related evaluation gates, provenance, secure usage policies, and incident disciplines into delivery rather than treating them as compliance overlays.

   How this helps:

   This step builds the control plane that makes higher AI augmentation scalable and defensible – these disciplines convert AI use from tool adoption into an auditable operating capability. Without this, organizations may see speed gains but will struggle to sustain them under audit, resilience scrutiny, or supervisory challenge – especially in high-criticality environments.

2. **Redesign sourcing footprint and vendor models around jurisdictional constraints**

   Once controls and evidence requirements are explicit, partition delivery into sovereign work (must sit in-region due to data, accountability, or critical-infrastructure constraints) versus portable work (can remain distributed).

   How this helps:

   Geography becomes a design constraint rather than an emergent problem. Furthermore, it aligns vendor commercials with the new unit economics, i.e., contracting for outcomes as opposed to raw capacity.

3. **Redesign talent pipelines**

   With the target operating model defined through steps 1 and 2 above, retool

apprenticeship, curricula partnerships, and internal academies toward supervision, verification, and systems thinking – so that the pipeline can reliably produce engineers who can operate AI-augmented workflows under regulated constraints.

How this helps:

This rebuilds the pipeline around the new early-career practice surface area so the organization can staff AI-augmented workflows without overpaying for scarce senior labor indefinitely.

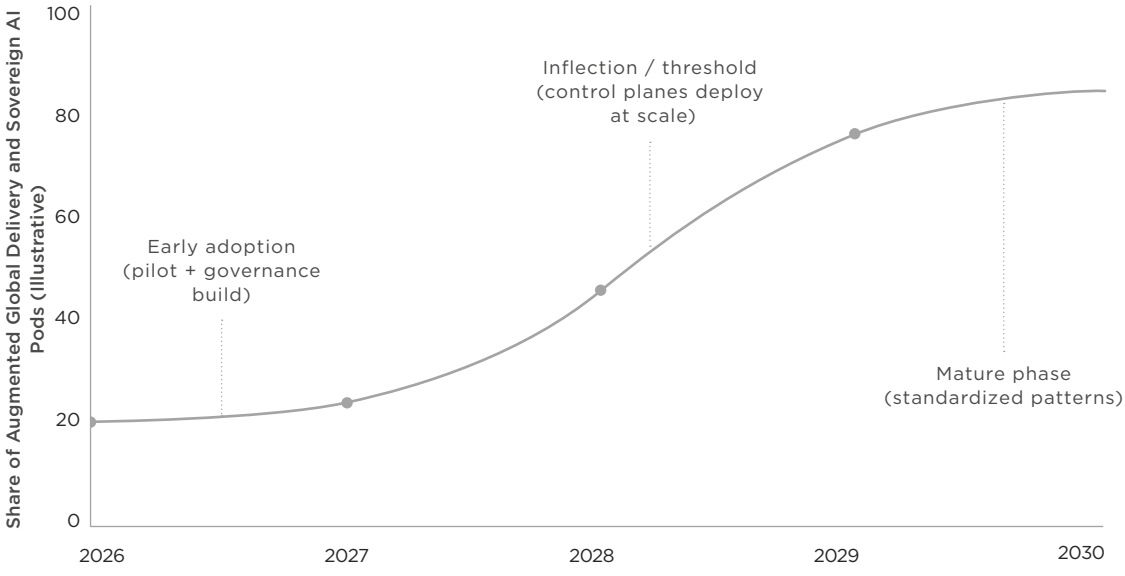
Over time, the industry will not converge on a single archetype; rather, the market share of the four operating models will shift as geopolitical constraint and AI intensity evolve. In the near term (through ~2026), the distribution is likely to remain weighted toward the left-hand quadrants: Global Factory persists for low-sensitivity work, while Localized Duplication expands as organizations respond to sovereignty and policy frictions faster than they can industrialize GenAI control planes and obtain governance/assurance comfort.

By approximately 2030, the center of gravity is likely to move rightward, as GenAI becomes embedded in delivery practice and as evaluable, auditable control patterns (provenance, pre-deployment testing, incident discipline) become standardized and productized into platforms. Augmented global delivery should grow where data can be bounded technically; sovereign AI pods should become the default for high-sensitivity and critical environments.

This transition is unlikely to be linear: expect an S curve (see Figure 2 as illustrative example) – incremental adoption while controls, permissions, and skills mature, followed by more rapid reconfiguration once defensible control planes are deployable at scale.

Ultimately, organizations in tightly regulated sectors such as finance and energy are best positioned for the dual shift in geopolitical contraction. GenAI reshaping delivery models will be those that treat GenAI not as a productivity add-on, but as a catalyst to re-architect delivery – aligning controls, geography, and talent into a coherent, auditable operating model that remains scalable under fragmentation.

Figure 2: Illustrative share of right-hand archetypes in the 2x2 framework between 2026 and 2030



References

Airaudo, F., F. De Soyres, K. Richards and A.M. Santacreu, 2025. Measuring Geopolitical Fragmentation: Implications for Trade, Financial Flows, and Economic Policy, International Finance Discussion Papers No. 1408, Board of Governors of the Federal Reserve System

Altman, S. A. and C.R. Bastian, 2024. DHL Global Connectedness Report 2024, DHL Group (in partnership with NYU Stern)

Autio, C., R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall,and K. Roberts, 2024. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1), National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Cui, Z. (Kevin), M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz, 2025. The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers, working paper/preprint), Microsoft Research

Deloitte, 2024. Global Outsourcing Survey 2024: Multidimensional sourcing – Orchestrating the extended workforce ecosystem, Deloitte Insights

Dixit, P., 2025. Meta is using rival AI models like Anthropic's Claude for better coding efficiency, Business Insider, June 13, https://www.businessinsider.com/meta-anthropic-ai-llama-coding-efficiency-2025-6

DORA (DevOps Research and Assessment), 2024. Accelerate State of DevOps Report 2024, Google Cloud

European Union, 2022. Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (Digital Operational Resilience Act – DORA), Official Journal of the European Union

Google Cloud, 2025. Gemini Cloud Assist / Cloud Logging documentation: log summarisation during incident triage, Google Cloud Documentation

Gopinath, G., P.O. Gourinchas, A.F. Presbitero, and P. Topalova, 2024. Changing Global Linkages: A New Cold War?, IMF Working Paper WP/24/76, International Monetary Fund

Information Services Group (ISG), 2025. AI Drives Tech Services, Software Spending to New High in Q3: ISG Index™, press release / ISG Index), ISG

INCOSE (International Council on Systems Engineering), n.d. INCOSE Systems Engineering Handbook (Version 5) – product page, INCOSE

ISA (International Society of Automation), n.d. ISA/IEC 62443 series of standards (Industrial Automation and Control Systems Security)

ISO/IEC/IEEE, 2017. ISO/IEC/IEEE 12207:2017 – Systems and software engineering – Software life cycle processes, International Organization for Standardization / International Electrotechnical Commission / IEEE

ISO/IEC/IEEE, 2023. ISO/IEC/IEEE 15288:2023 – Systems and software engineering – System life cycle processes, International Organization for Standardization, ISO

National Academies of Sciences, Engineering, and Medicine (NASEM), 2025. Artificial Intelligence and the Future of Work, National Academies Press https://doi.org/10.17226/27644

NERC. (CIP-010-4). Cyber Security – Configuration Change Management and Vulnerability Assessments

NERC. (CIP-013-2). Cyber Security – Supply Chain Risk Management.

OECD (Organisation for Economic Co-operation and Development), 2025a. OECD Skills Outlook 2025, OECD Publishing

OECD (Organisation for Economic Co-operation and Development), 2025b. Bridging the AI skills gap: Is training keeping up? OECD Publishing

OECD (Organisation for Economic Co-operation and Development), 2025c. International Migration Outlook 2025, OECD Publishing, https://doi.org/10.1787/ae26c893-en

Program on International Financial Systems (PIFS), 2024. Data Localization, Cloud Adoption, and the Financial Sector, PIFS International.

Russo, D., 2024. Navigating the Complexity of Generative AI Adoption in Software Engineering, ACM Transactions on Software Engineering and Methodology, 33(5), Article 135, https://doi.org/10.1145/3652154

UNESCO, 2023. Guidance for generative AI in education and research, United Nations Educational, Scientific and Cultural Organization

World Economic Forum (WEF), 2025. The Future of Jobs Report 2025, World Economic Forum.

Corporate tax planning:

a productivity lever in a low-growth era

Authors | **Spyridon Gkikopoulos** | Alliance Manchester Business School, The University of Manchester
Konstantinos Stathopoulos | Alliance Manchester Business School, The University of Manchester

Abstract

Persistently sluggish productivity growth has become a defining challenge for advanced economies, threatening long-term prosperity. In this article, we explore corporate tax planning as an underappreciated lever to boost firm-level productivity amid these headwinds. Drawing on recent research, we find that firms that actively manage their tax affairs – thereby optimizing their effective tax rates – achieve measurably higher productivity. The gains are especially pronounced for businesses facing financial constraints or investing heavily in knowledge and innovation, which depend on internal funding. Our analysis highlights that increased after-tax cash flows from tax planning fuel growth in capital expenditures, hiring, and R&D, translating into more output per unit of input. Crucially, we identify banking relationships as a conduit for effective tax planning: firms that partner with tax-savvy banks see significant productivity bumps relative to peers. These findings carry important implications for corporate leaders, banks, and policymakers. With careful design and execution, tax planning can serve as a strategic tool to alleviate financial frictions in capital markets and spark productivity growth – a much-needed tailwind in an era of economic stagnation. In what follows, we outline the conceptual framework linking tax planning to productivity, examine the mechanisms at work, discuss the facilitating role of banks, and offer actionable insights for practitioners and regulators.

1. Introduction

Western economies are grappling with a prolonged productivity slowdown. Over the past two decades, productivity growth – particularly total factor productivity (TFP), the efficiency with which firms turn labor and capital into output – has decelerated across most advanced nations [Bhatt (2024)]. This stagnation is not a mere statistic; it undermines economic dynamism, erodes living standards, and even poses risks to financial stability. Sluggish productivity also undercuts the competitiveness of Western businesses in an increasingly globalized market. As the global economic order shifts and populations age, reversing the productivity slump has become a paramount policy and business challenge. Yet consensus on how to reignite productivity remains elusive.

In this context, corporate leaders and policymakers are searching for levers to boost productivity growth. One often-overlooked lever lies in the realm of corporate taxation. Traditionally, corporate tax planning – the strategic management of tax obligations to

optimize effective tax rates – is discussed in debates on tax fairness and compliance. But emerging evidence suggests it may also have important real effects on business performance. In particular, our research finds that prudent tax planning can enhance a firm's productivity by freeing up resources for productive investment. At a time when external economic forces are squeezing margins and output, tax planning represents a tangible lever under management's control to improve operational efficiency.

In this article, we adapt and synthesize findings from the academic study, Does Corporate Tax Planning Affect Firm Productivity? by Gkikopoulos et al. (2025) into actionable insights for senior financial executives. We frame our discussion in practical terms, focusing on why and how optimizing a firm's effective tax rate can unlock productivity gains, and what enablers – notably banking relationships – make these gains achievable. We deliberately avoid technical econometric jargon, emphasizing instead the economic intuition and strategic implications of the findings.

The remainder of the article is organized as follows. In Section 2 (Framework), we lay out the conceptual links between tax planning, financial frictions, and productivity, defining key concepts such as total factor productivity and financial constraints. Section 3 (Mechanisms) examines the channels through which tax planning influences productivity, highlighting why the effect is strongest for financially constrained firms and those rich in knowledge capital. Section 4 (Banking as enabler) discusses the critical role of banks in facilitating effective tax strategies, presenting evidence that firms establishing relationships with tax-savvy banks subsequently realize higher productivity. In Section 5 (Policy takeaways), we outline practical implications for business strategy and public policy – how firms can harness tax planning for growth, how banks can support this process, and how regulators might balance tax enforcement with the goal of fostering productivity. Section 6 (Conclusion) then summarizes our key findings and recommendations, positioning corporate tax planning as a feasible tool to help address the productivity malaise afflicting many economies.

2. Framework: tax planning, financial frictions and productivity

Productivity broadly measures how efficiently a firm turns inputs into outputs [Syverson (2011)]. A common metric is TFP, which captures the portion of output not explained by the amount of capital and labor used. In plain terms, TFP reflects the effectiveness of a firm's processes, technology, and know-how in creating value. When a firm's TFP rises, it means the business is getting more output from the same resources – a sign of improved efficiency and often innovation. At the macro level, robust TFP growth is the cornerstone of sustained economic expansion and rising living standards. Conversely, persistent weakness in productivity growth – as currently seen in many Western economies – acts as a drag on GDP growth and prosperity.

What causes firms to fall short of their productivity potential? Research points to several culprits, including financial frictions [Duval et al. (2020)]. Financial frictions refer to impediments in raising or allocating capital – for example, credit constraints, high costs of borrowing, or information asymmetries between borrowers and lenders. These frictions can misallocate resources by preventing funding for high-productivity projects. Firms facing steep financing hurdles often cut back on investments in equipment, employees, and innovation, which in turn hampers their productivity growth. For instance, studies find that after the 2008 financial crisis, companies

with constrained credit access saw marked drops in productivity, linking tighter finance directly to efficiency losses [Duval et al. (2020)]. In short, when firms cannot readily obtain external funds or must pay dearly for them, they are less able to invest in productivity-enhancing initiatives.

Two types of companies are especially vulnerable to financial frictions. Financially constrained firms – typically smaller or highly leveraged companies with limited access to capital markets – face high external financing costs or outright credit rationing [Levine and Warusawitharana (2021)]. These firms often have to forego profitable projects due to lack of funds. Meanwhile, knowledge-intensive firms (such as tech startups or R&D-driven companies) rely heavily on intangible assets like research, software, or intellectual property. Such assets usually have limited collateral value, so lenders are reluctant to finance them, and investors demand high returns for the risk [Wang (2017)]. As a result, even fundamentally strong innovative firms can be starved of capital, constraining their growth and productivity.

Our core premise is that corporate tax planning can relax these financial constraints and thereby improve productivity. By tax planning, we mean any legal strategy a firm uses to maximize its after-tax returns – examples include optimizing the corporate structure, taking advantage of tax credits and deductions, strategic transfer pricing or timing of expenditures. Such strategies increase the firm's after-tax cash flows, effectively acting as an additional source of internal financing. Every dollar shielded from taxation is a dollar that can be reinvested in the business. To the extent that internal funds are cheaper and more readily available than external capital (especially for constrained firms), tax savings can finance projects that would otherwise go unfunded.

Moreover, effective tax planning can improve a company's access to external financing in indirect ways. A healthier cash flow and balance sheet could lower the firm's perceived credit risk and make lenders more willing to extend credit. There is evidence that firms engaging in tax planning enjoy a lower cost of equity capital, particularly if they were financially constrained to begin with [Goh et al. (2016)]. A study reports that tax savings can induce a preference for equity financing (issuing stock) because the firm's improved financial metrics make raising equity more attractive [Lee et al. (2023)]. In essence, by boosting internal cash and improving financial ratios, tax planning can ease both internal and external financial frictions that hold back investment.

The distinctive insight is that tax planning affects productivity not because it simply increases internal resources, but because it compensates for failures in how external capital is normally allocated. In well-functioning financing environments, tax planning would have little effect on productivity; its impact emerges precisely because high-productivity projects often remain unfunded due to imperfections in external capital markets.

It is important to note that tax planning is not without potential downsides. Aggressive or overly complex tax schemes can introduce information frictions – opacity that makes it harder for investors or other stakeholders to assess the firm's true performance [Balakrishnan et al. (2019)]. Elaborate tax-driven structures might also cause management to allocate resources in a suboptimal way, for example, investing in a low-tax jurisdiction for tax purposes rather than for operational efficiency [De Simone et al. (2022)]. These factors could, in theory, hurt productivity. However, the crux of our findings – detailed

below – is that on average, the positive effects of tax planning dominate, especially when tax savings are productively deployed. By mitigating financial constraints and funding growth, smart tax planning tends to raise, not lower, firm-level productivity in practice.

3. Mechanisms: why tax planning boosts productivity

We next delve into the two main mechanisms through which corporate tax planning translates into higher productivity: relaxing financial constraints and fueling knowledge capital formation. Our empirical analysis finds that both channels are at work, and importantly, that the productivity payoff from tax planning is concentrated in firms where these mechanisms matter most – i.e., firms that are financially constrained and those that are knowledge-capital-intensive.

3.1 Financial constraints mechanism

When a firm is starved for funds, an extra dollar of internally generated cash can have an outsized impact. We find that tax planning provides precisely such a boost for constrained firms by augmenting their internal financing. In our study of U.S. public companies, we observe that the positive link between tax savings and productivity is significantly stronger for financially constrained firms. In other words, firms that have difficulty raising external capital gain the most efficiency from each dollar saved in taxes. This aligns with economic intuition: if a firm cannot easily borrow or issue stock, then cash freed up via tax planning may be one of the few resources available to invest in new technology, upgrade equipment, or hire skilled workers – all important drivers of productivity.

To illustrate, consider a hypothetical constrained firm that manages to optimize its effective tax rate through careful planning. The additional cash could be used to purchase a new piece of machinery that speeds up production, or to expand the sales team to reach new markets. Such investments would boost output relative to input, raising the firm's TFP. Our findings generalize this intuition. We document that across firms, a reduction in the effective tax rate is associated with subsequent increase in both capital and employment growth. This indicates that tax savings are being channeled into real business expansion – companies are buying more equipment and hiring more staff – rather than just sitting as idle cash. These real investments contribute directly to higher productive capacity and efficiency in operations.

Crucially, prior research supports the notion that tax savings get reinvested. For example, Green and Kerr (2022) find that roughly 20 cents of every $1 of cash tax savings is reinvested into real activities like R&D. In other work, Edwards et al. (2016) show that the boost to investment from tax planning is greatest for firms under financial duress. These studies reinforce our mechanism: tax planning augments the scarce internal funds of constrained firms, and a meaningful portion of those funds goes into productive uses, rather than, say, sitting offshore or being purely returned to shareholders. The net result is improved operational efficiency. In short, tax planning helps loosen the binding financial constraints that often force firms to cut R&D budgets or delay capital projects, thereby enabling those firms to maintain or increase their productivity trajectory.

3.2 Knowledge capital mechanism

The second mechanism relates to firms that rely heavily on innovation and intangible assets – what we refer to as knowledge capital. These firms (common in sectors like technology, pharmaceuticals, or digital services) need substantial, continual investment in research and development, employee training, and intellectual property creation to stay competitive. Yet, as discussed, external funding for such investments is hard to come by, because intangible assets usually cannot be collateralized to issue debt and their future returns are uncertain. Such companies, thus, depend on internal cash flows to fund innovation, a point emphasized by economists Hall and Lerner (2010) and Wang (2017), among others.

Our study finds clear evidence that tax planning disproportionately benefits firms with greater knowledge capital intensity. In firms that have large stocks of intangibles (measured, for example, by the amount of capitalized R&D on the balance sheet), the association between lower taxes and higher productivity is particularly strong. This suggests that tax savings are enabling these innovation-driven businesses to invest more in their core knowledge-generating activities. The additional after-tax cash might fund a new research project, allow the firm to hire additional engineers or data scientists, or otherwise expand its innovative output – all of which ultimately improve productivity through technological progress or improved processes. In fact, separate research by Li et al. (2021) provides a telling inverse case: when opportunities for tax planning are curtailed (for instance, by regulations that limit tax-motivated profit-shifting), firms produce fewer patents and innovations, indicating that reduced tax savings can discourage innovative output. This underscores how important internally generated funds are for sustaining innovation.

Another way to think about the knowledge capital mechanism is that tax planning effectively acts as an internal R&D financer. Companies often complain that high taxes drain funds that could be used for innovation. Our findings validate that notion – by cutting a firm's tax bill, more cash is left on the table to invest in productivity-enhancing innovation, which is especially crucial for firms whose business model relies on continual R&D. We also note that in our sample, the positive effect of tax planning on productivity remains robust even after accounting for various firm characteristics and broader tax policy changes. This gives us confidence that what we are capturing is indeed the productive use of tax savings, not some spurious correlation or purely cyclical effect.

Before moving on, it is worth acknowledging a balance: while our evidence tilts positive, tax planning should be well-governed to translate into productivity gains. If tax bill reduction strategies become so aggressive that they introduce excessive complexity or risks, they could backfire. For instance, extremely convoluted tax shelters might distract management or create unforeseen liabilities, undermining operational efficiency. Effective tax planning, as a productivity tool, therefore, hinges on prudence and purposeful reinvestment. Companies that treat tax savings as a windfall to be squandered will not see productivity benefits, whereas those that reinvest the savings into growth will. One recent study on small and medium enterprises (SMEs) in Europe underscores this point: it found that tax avoidance boosts productivity in SMEs, and managers should channel the cash saved through tax planning into productive uses such as investing in new capabilities [Sánchez-Ballesta and Yagüe (2024)]. The same study emphasizes maintaining high-quality financial reporting and governance so that tax strategies do not exacerbate agency

conflicts or opaqueness, which can dilute the gains. In summary, the mechanisms we identify assume that tax planning is done in the service of the business – funding expansion and innovation – rather than as an end in itself. When that condition holds, the evidence strongly indicates that effective tax practices lead to higher firm productivity, especially via relieving financial tightness and fueling innovation.

4. Banking as enabler: the role of banks in tax planning

How do firms actually improve their tax planning capabilities? Sophisticated tax strategies often require specialized knowledge and advice. This is where the banking sector comes into play. Banks – particularly large commercial and investment banks – have emerged as key intermediaries in corporate tax planning. Many banks offer tax advisory services or design financial products that can help clients achieve tax efficiencies. In fact, research has provided large-scale evidence that certain banks specialize in assisting corporate clients with tax planning, effectively serving as tax planning intermediaries [Gallemore et al. (2019)]. These banks leverage their informational advantages and expertise to help firms structure transactions or reorganize in tax-efficient ways.

Our study sheds light on how forming a relationship with a tax-savvy bank can directly impact a firm's productivity. Using a difference-in-differences approach (a statistical method to isolate causality by comparing changes between a treated group and a control group), we examined firms that initiated new lending relationships with banks known to be providing client tax planning services. We compared their subsequent performance to similar firms that formed relationships with more traditional banks. The results are striking: firms that commenced loan facilities with tax-planning intermediary banks experienced a 1.8%–3.3% increase in productivity in the period after the relationship began, relative to the control group. This productivity bump suggests a causal effect – the new banking relationship helped unlock better tax strategies, which in turn boosted the firm's efficiency.

The intuition is straightforward. Banks often possess 'soft information' about their corporate clients' finances and operations, gained through the lending and due diligence process. A bank that is skilled in tax matters can use this information to identify tailored tax-saving opportunities for the client. For example, the bank might recommend an optimal financing structure (debt versus equity) or the use of certain financial instruments (like leases or derivatives) that carry tax advantages. It might also connect the firm with specialized tax consultants or help navigate cross-border tax rules if the firm is expanding overseas. Through these channels, the bank enables the firm to reduce its tax burden more effectively than it could on its own.

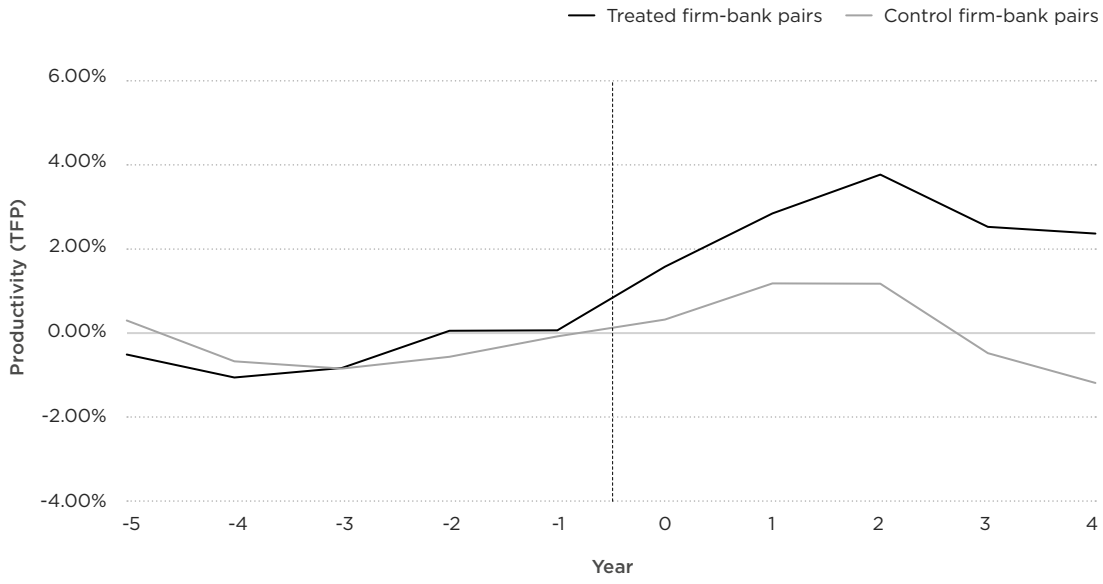
John Gallemore and coauthors (2019) documented that some banks consistently help their clients achieve lower effective tax rates, indicating a sort of 'tax planning premium' for clients of those banks. Our findings build on this by demonstrating that the benefit is not merely lower taxes but extends to higher productivity growth for the clients. In concrete terms, if a manufacturing company switches to a bank known for tax expertise, that company is likely to implement new tax strategies (perhaps exploiting credits for capital investment or improving transfer pricing practices) under the bank's guidance. The tax savings then finance upgrades in the factory or the deployment of advanced robotics, for instance, which raise output per worker. The data capturing the 1.8%–3.3% TFP increase align with precisely this kind of story.

Notably, the banking channel can be especially advantageous for mid-sized firms or those without large in-house tax departments. Large multinationals often have internal tax planning teams; smaller firms may not. By engaging a bank with the right capabilities, the latter can essentially 'outsource' some of the sophisticated tax planning function and reap similar benefits. This has implications for the banking sector too: it suggests a competitive differentiator for banks in winning corporate business. Those that cultivate tax advisory prowess can deliver more value to clients, not just in terms of financing but in improving client performance. In turn, helping clients become more productive and financially healthy also benefits banks by enhancing the clients' creditworthiness and business growth (a true win-win).

As an additional piece of evidence, our study also leverages a quasi-natural experiment unrelated to banking: the introduction of the U.S. Check-the-Box regulations in 1997. This tax regulation change inadvertently made it easier for U.S. multinationals to engage in certain tax planning, specifically, it simplified how firms could designate subsidiaries for tax purposes, enabling more profit shifting to low-tax jurisdictions. We compared the productivity of multinational firms to purely domestic firms before and after 1997. The result was that multinationals (who could utilize the new tax flexibility) saw a relative productivity gain of about 2.2%–3.1% after the rule change. In essence, a policy that reduced tax planning costs led to an observable productivity boost for firms positioned to take advantage. We mention this finding here to reinforce the point: whether through an external advisor like a bank or through regulatory change, when firms gain access to effective tax planning tools, their productivity tends to increase.

For financial executives in banking, these findings highlight an important role that banks play in the broader economy's productivity puzzle. By acting as conduits of tax planning knowledge, banks can indirectly foster real economic growth. This may encourage banks to invest in their tax advisory services or form partnerships with tax experts, as the demand from corporate clients

Figure 1: Productivity changes surrounding new relationships with tax-savvy banks



for such value-added services is likely to grow. It also underscores that from a lending perspective, understanding a client's tax strategy might be integral to evaluating their investment capacity and growth prospects. In summary, our research identifies banking relationships as a key enabler: the right banking partner can amplify the productivity benefits of corporate tax planning.

5. Policy takeaways and practical implications

The evidence that effective corporate tax planning can enhance firm productivity carries several actionable insights for different stakeholders – corporate managers, financial intermediaries and policymakers.

For corporate executives: The foremost implication is that tax planning should be viewed as a strategic financial management tool, not merely as a compliance or cost-minimization exercise. In an era of thin margins and slow top-line growth, improving productivity is one of the few levers within a firm's control to boost competitiveness. Our findings suggest that actively managing your firm's tax position – for instance, by structuring operations to lawfully take advantage of tax incentives, credits, and deductions – can free up capital to invest in productivity-enhancing initiatives. CFOs and finance VPs, in particular, might consider increasing collaboration between the tax function and operational units. For example, tax departments can work closely with R&D and operations to ensure the company is maximizing use of R&D tax credits, investment allowances, or accelerated depreciation for new equipment. The resultant tax savings should be transparently reinvested into those same areas to generate a virtuous cycle of innovation and efficiency gains. The key practical takeaway for firms is to reinvest tax savings wisely – whether in upgrading technology, training workers, or expanding capacity – rather than using tax avoidance purely to boost short-term accounting profits. When tax planning is integrated with business strategy, it effectively becomes a self-funding source of competitive improvement.

At the same time, corporate leaders should maintain strong governance around tax strategies. As noted, overly aggressive tax schemes can introduce risks and complexities. It is crucial to strike a balance between optimizing taxes and maintaining clarity and credibility in financial reporting. Ensuring high-quality information flow to investors and stakeholders can mitigate any negative perceptions associated with tax planning. Put simply, transparency and prudent risk management must accompany tax planning. Firms that achieve this balance can enjoy the dual benefits of lower tax outflows and higher productivity inflows, with lower likelihood of regulatory or reputational fallout.

For banks and financial services providers: Banks should recognize their role not just as financiers but as partners in their clients' productivity journey. The evidence that certain banks help clients cut taxes and subsequently witness those clients perform better is a compelling value proposition. Banks – especially those catering to mid-market companies – might invest in developing tax advisory capabilities, either in-house or via partnerships with accounting firms and tax consultants. By doing so, they can differentiate their service offering. For example, a bank could offer an integrated product: a loan or credit line bundled with tax optimization advice related to the financed project. If a client is investing in new machinery, the bank's team might ensure the client fully utilizes available tax credits for capital investments, thus effectively reducing the net cost of the project. Such support not only strengthens client relationships but can also improve the client's financial health, which is in the bank's interest.

> *By helping nurture their clients' productivity through tax efficiency, banks also contribute to broader economic growth – aligning profitability with purpose.*

Additionally, banks should incorporate understanding of a firm's tax strategy into credit analysis. A company actively managing its tax liabilities may have more free cash flow to service debt and invest in growth. Conversely, a firm ignoring available tax breaks might be leaving money on the table, indicating less efficient management. Banks can use this insight in advising clients: pointing out, for instance, "We notice your effective tax rate is high relative to peers; let's explore if there are legal, safe tax planning opportunities – doing so could improve your cash flow and ability to expand." In markets where banks compete for corporate lending, those that act as true advisors (beyond just providing capital) will likely win more business. By helping nurture their clients' productivity through tax efficiency, banks also contribute to broader economic growth – aligning profitability with purpose.

For policymakers and regulators: The relationship between corporate tax policy and productivity revealed by our analysis suggests that regulators should take a nuanced approach in tax law design and enforcement. The traditional view often pits tax planning as a pure negative (erosion of the tax base, unfair advantage, etc.). Our findings highlight an economic upside: tax savings, when reinvested, can stimulate real growth in output and jobs. Policymakers, therefore, might consider targeted tax policies that encourage productive investment. This could include, for example, investment tax credits, enhanced deductions for R&D, or lower tax rates for income that is demonstrably linked to new capital formation and innovation. Such measures are akin to directly incentivizing the behaviors that we find firms undertake when they have more after-tax cash. Notably, international institutions have echoed this idea: experts have called for targeted tax credits and support for smaller businesses and startups to counter slow productivity growth [Bhatt (2024)]. Designing tax policy with productivity in mind could mean that any tightening of loopholes comes along with provisions that encourage firms to channel resources into productivity-boosting uses rather than simply paying more tax to the treasury without a growth payoff.

Of course, this does not imply that all forms of tax planning are desirable. Aggressive tax avoidance that purely exploits loopholes without any real investment or that shifts profits in ways unrelated to real activity can still be harmful, eroding the fairness and integrity of the tax system. The policy goal should be to distinguish between tax strategies that finance growth and pernicious tax schemes that distort the tax base. One practical implication is the importance of international cooperation, such as the OECD's BEPS (Base Erosion and Profit Shifting) initiatives, to curb the most egregious avoidance. But even within such efforts, regulators might appreciate that completely eliminating corporate tax planning could have unintended side effects on business investment. A balance can be struck by, for instance, providing safe harbors or lower tax rates for income invested in certain high-productivity projects, much like how many jurisdictions offer tax incentives for clean energy or innovation. In fact, the positive impact of tax planning on productivity in SMEs observed in research mentioned previously provides support for

special tax provisions favoring small businesses' growth. The logic is that easing the tax burden on firms that are financially constrained (like many SMEs) can enable them to invest more and become engines of job creation and innovation – outcomes beneficial for the economy at large.

Another policy takeaway is the value of financial infrastructure and advisory support. Governments and industry groups might consider programs that improve knowledge diffusion about tax incentives and efficient tax practices, especially to smaller firms that lack resources to hire top tax lawyers. Even something as simple as better outreach and education on available tax credits for R&D or investment could increase uptake and thus spur productivity. After all, if large firms benefit from banks as tax planning intermediaries, as we demonstrated, smaller firms could benefit from public or private initiatives that bridge the information gap on taxation.

In summary, for policymakers the message is: a stringent stance on corporate tax planning should be balanced with an understanding of its economic context. Policies that simultaneously aim to close artificial loopholes but also reward genuine reinvestment of earnings will likely yield the best outcomes – curbing harmful avoidance while preserving the mechanism by which tax planning can support productivity growth.

6. Conclusion

Improving productivity is paramount for firms and economies striving to thrive in today's low-growth environment. Our analysis highlights an unconventional yet practical lever to achieve that end: corporate tax planning. We find that firms which actively engage in tax planning – thereby optimizing their effective tax rates – tend to allocate the freed resources toward growth, resulting in higher output per input. The effect is most salient in firms that need it the most: those constrained by finances and those driven by innovation, for whom extra cash can immediately translate into new investments. These findings recast tax strategy not as a narrow effort to minimize expenses, but as a tool to maximize after-tax returns; it is also a means to fund expansion, technological progress, and ultimately higher productivity.

We also underscore the enabling role of the banking sector. Banks with tax expertise serve as important facilitators, helping companies realize the productivity gains from tax planning. This symbiotic relationship between financial institutions and corporates suggests that improving economic productivity is a shared endeavor – one that spans beyond internal corporate decisions to the broader financial ecosystem and policy landscape.

These insights also speak to the current environment of heightened geopolitical and policy uncertainty. Frequent changes in trade policy, tariffs, and regulations tend to weaken productivity by increasing geopolitical uncertainty around costs, supply chains, and investment returns, often leading firms to delay or scale back productivity-enhancing projects [Caldara and Iacoviello (2022)]. Exploratory, untabulated analyses from our work are consistent with this interpretation: productivity generally declines during periods of elevated geopolitical uncertainty, but the decline appears less pronounced among firms with greater tax planning flexibility. We view this evidence as indicative rather than conclusive, yet it reinforces the broader message that internal funding helps firms sustain productivity and adapt operations when external conditions tighten.

For executives, the clear call to action is to integrate tax considerations into strategic planning, ensuring that any tax savings feed into value-creating uses. For banks, it is to recognize advisory opportunities that enhance client productivity alongside traditional financial services. And for policymakers, it is to design tax regimes that encourage productive investment and acknowledge the nuanced ways in which tax policy influences on-the-ground business efficiency.

In conclusion, our work contributes to reframing the narrative on corporate taxes: from being seen purely as a cost or a moral obligation, to being understood as a potential source of competitive advantage and economic stimulus when managed wisely. In a world where Western economies are searching for productivity breakthroughs, unleashing the productive power of retained earnings via thoughtful tax planning is a tangible step in the right direction. Harnessed correctly, corporate tax planning can be a win-win – bolstering firm performance while indirectly fueling broader economic growth. As the global economic order continues to evolve, those firms and financial institutions that adeptly navigate tax strategy in pursuit of innovation and efficiency will be well-positioned to lead in the new era of value creation. The challenge of low productivity may be persistent, but the solutions – as we find – are very much within our grasp.

References

Balakrishnan, K., J.L. Blouin, and W.R. Guay, 2019. Tax Aggressiveness and Corporate Transparency, The Accounting Review, 94(1), 45-69

Bhatt, G., 2024. Sustained Economic Growth Hinges on Productivity Gains as Populations Age, Finance & Development, A Quarterly Publication of the International Monetary Fund, 61(3), 5

Caldara, D. and M. Iacoviello, 2022. Measuring Geopolitical Risk, American Economic Review, 112(4),1194-1225

De Simone, L., K.J. Klassen, and J.K. Seidman, 2022. The Effect of Income-Shifting Aggressiveness on Corporate Investment, Journal of Accounting and Economics, 74(1), 101491

Duval, R., G.H. Hong, and Y. Timmer, 2020. Financial Frictions and the Great Productivity Slowdown, Review of Financial Studies, 33(2), 475-503

Edwards, A., C. Schwab, and T. Shevlin, 2016. Financial Constraints and Cash Tax Savings, The Accounting Review, 91(3), 859-881

Gallemore, J., B. Gipper, and E. Maydew, 2019. Banks as Tax Planning Intermediaries. Journal of Accounting Research, 57(1), 169-209

Gkikopoulos, S., E. Lee, and K. Stathopoulos, 2025. Does Corporate Tax Planning Affect Firm Productivity? Management Science, forthcoming

Goh, B. W., J. Lee, C.Y. Lim, and T. Shevlin, 2016. The Effect of Corporate Tax Avoidance on the Cost of Equity, The Accounting Review, 91(6), 1647-1670

Green, D. H. and J.N. Kerr, 2022. How Do Firms Use Cash Tax Savings?, Journal of the American Taxation Association, 44(1), 93-121

Hadlock, C. and J. Pierce, 2010. New Evidence on Measuring Financial Constraints, Review of Financial Studies, 23(5), 1909-1940

Hall, B. H. and J. Lerner, 2010. The Financing of R&D and Innovation. Handbook of the Economics of Innovation, available at: https://www.sciencedirect.com/science/chapter/handbook/abs/pii/S0169721810010142

Lee, Y., T. Shevlin, and A. Venkat, 2023. The Effect of Tax Avoidance on Capital Structure Choices, The Journal of the American Taxation Association, 45(1), 91-115

Levine, O. and M. Warusawitharana, 2021. Finance and Productivity Growth: Firm-Level Evidence, Journal of Monetary Economics, 117(1), 91-107

Li, Q., M.S. Ma, and T. Shevlin, 2021. The Effect of Tax Avoidance Crackdown on Corporate Innovation, Journal of Accounting and Economics, 71(2-3), 101382

Syverson, C., 2011. What Determines Productivity? Journal of Economic Literature, 49(2), 326-365

Wang, Y., 2017. Debt Market Friction, Firm-Specific Knowledge, Capital Accumulation and Macroeconomic Implications. Review of Economic Dynamics, 26(1), 1939

When growth isn't enough:

how uncertainty reshapes hiring decisions

Authors | **Duc Duy Nguyen** | Professor in Finance, Durham University
Vathunyoo Sila | Professor and Chair of Finance, University of Edinburgh

Abstract

Hiring has slowed across many advanced economies despite continued economic growth. This article examines how political uncertainty reshapes firms' hiring decisions, using evidence from more than 10 million U.S. job postings around state governor elections. We show that uncertainty affects not only the volume of hiring but also the composition of roles firms choose to fill. Firms disproportionately scale back specialized, high-commitment positions, shift towards more flexible and temporary roles, and relax job requirements to preserve adaptability. These adjustments reflect the partially irreversible nature of hiring and the value of waiting when future regulatory and policy conditions are difficult to predict.

Worker behavior reinforces these patterns: job-to-job mobility declines under uncertainty, while matches formed during uncertain periods are less stable. Together, these dynamics help explain why labor markets can cool even when output remains resilient. The findings are particularly relevant for financial services firms operating in highly regulated environments, where hiring mistakes are costly to reverse. Rather than signaling weak fundamentals, slower hiring often reflects a strategic response to uncertainty, with implications for workforce planning, internal mobility, and talent risk management.

1. Introduction

Recent macroeconomic data present an unusual combination. While economic growth has remained resilient at around 3% in recent years, the demand for new workers is easing in many countries.[1] In several major economies, including the U.S., the U.K., Germany, Canada, and Japan, job openings per unemployed worker have fallen back toward or below their 2019 levels, despite continued growth in aggregate output. Recent U.S. data show job openings falling to one of their lowest levels in nearly five years and hiring remaining sluggish even as the economy continues to grow, underscoring the unusual decoupling between output and labor demand.[2] Figure 1 illustrates this pattern for the U.K., where firms are advertising fewer new roles than before, even as the broader economy continues to grow. This divergence suggests that factors other than weak demand may be at play.

[1] https://apnews.com/article/jobs-economy-hiring-e7b5d4c29d444ea019199d8bb759bb51
[2] ibid.

Figure 1 plots real GDP (in billion GBP) and vacancies per unemployment for the U.K. Real GDP is measured as GDP at market prices using the Office for National Statistics (ONS) chained volume measure. Labor market tightness is measured as the number of job vacancies per unemployed worker, constructed from ONS vacancy and unemployment data (VACS01). Annual GDP for 2025 is a provisional estimate extrapolated from quarterly GDP growth, as official annual GDP data are not yet available.

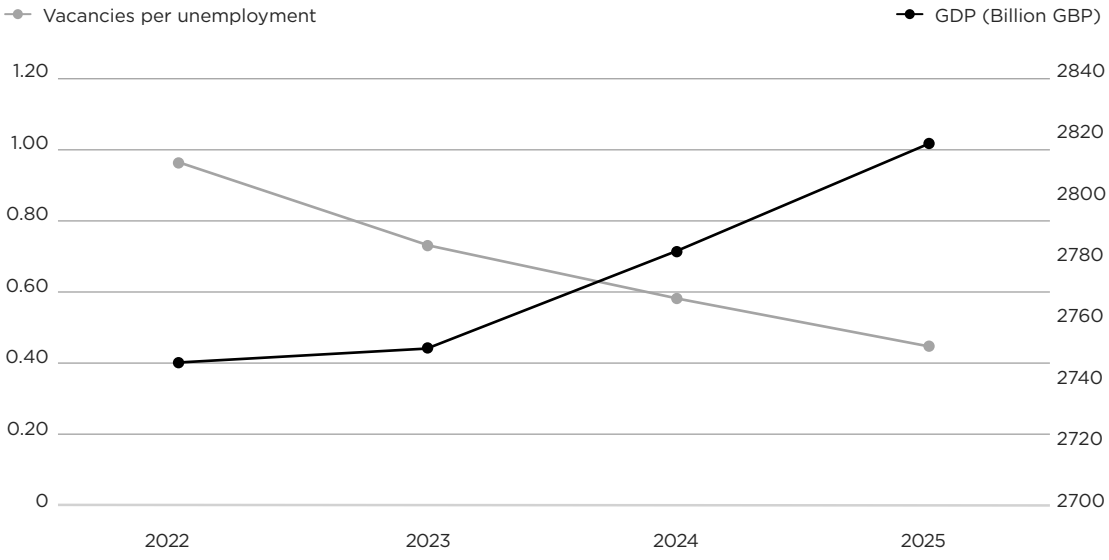
The answer may lie in the growing uncertainty and how it shapes workforce decisions. Across the global economy, firms are navigating shifting regulatory agendas, ongoing geopolitical tensions, and rapid technological change. Elections in major economies, renewed trade frictions, and uncertainty about global financial markets have all added to the sense of unpredictability. In this environment, businesses are becoming more hesitant to make long-term commitments, and hiring could be an area where such caution becomes visible.

2. Why does hiring slow under uncertainty?

To understand why firms pull back on recruitment when uncertainty rises, it helps to consider the nature of hiring itself. Adding a new employee involves advertising, screening, interviewing, onboarding, training, and integrating them into a team. All these activities require time, resources, and managerial attention. These upfront investments cannot be recovered if the role is later withdrawn or the match ends prematurely. While firms can dismiss workers if conditions deteriorate, doing so is disruptive, costly, and often reputationally sensitive. In this sense, hiring is only partially reversible, and the difficulty of unwinding these commitments makes firms more cautious about expanding their workforce when the outlook becomes uncertain.

This logic is consistent with well-established economic theory on 'real options' and irreversible investments. When an investment is costly to reverse, uncertainty encourages decision makers to wait until more information becomes available.

Figure 1: Output growth and hiring in the U.K.



Hiring follows this pattern. Unlike buying raw materials, taking on a new employee typically involves a long-term relationship and substantial sunk costs. At the same time, it is not as irreversible as building a new factory or investing in large-scale physical infrastructure. In this sense, hiring sits somewhere in the middle: partially reversible, but still costly to unwind.

As uncertainty increases, the value of the 'option to wait' also increases. Firms may not know how regulations will evolve, whether demand will remain stable, or how geopolitical or financial developments will unfold. In such situations, recruiting a permanent employee becomes riskier because the firm cannot be fully confident that future conditions will justify the hire. By delaying the decision, the firm keeps its options open and avoids making a commitment it might later regret.

> *Uncertainty affects not only the volume of hiring but also the composition of roles firms choose to fill.*

Importantly, this 'wait-and-see' behavior does not necessarily reflect pessimism about the economy. Instead, it reflects caution in the face of unpredictable conditions. Even firms with healthy order books and strong financial performance may slow their hiring if they believe that key aspects of their operating environment, such as policy direction, regulatory frameworks, or global market conditions, are unusually difficult to forecast.

3. Looking at the data: how firms adjust their hiring under uncertainty

Our research examines how firms alter their recruitment decisions when uncertainty rises, using detailed data on 10.5 million U.S. job postings between January 2010 and December 2019.[3] By tracking changes in the number and content of job advertisements, the types of roles being posted, and the timing of hiring decisions, we are able to observe how firms adjust not just the volume of job advertisements but also the composition of their workforce demand.

To study how firms respond to uncertainty in a systematic way, we use U.S. state governor (gubernatorial) elections as a source of variation in uncertainty. These elections can trigger state-level changes in tax rules, labor regulations, public spending priorities, and business conditions, even though their outcomes and implications are uncertain in advance.

We focus on the period between the global financial crisis and the COVID-19 pandemic because it allows us to study firms' hiring behavior in response to repeated episodes of policy and economic uncertainty without the confounding effects of large, economy-wide shocks to labor demand and supply. This period captures multiple election cycles and enables us to isolate the effects of uncertainty associated with elections from crisis-driven global disruptions.

While our evidence is drawn from the context of election-related uncertainty, the underlying mechanisms we identify are relevant to more recent sources of uncertainty, including heightened geopolitical tensions and the rapid adoption of artificial intelligence (AI) and its implications for business models and workforce decisions.

[3] While similar hiring patterns are visible across advanced economies, our empirical analysis focuses on U.S. firms, where we can observe recruitment decisions in granular detail.

Our analysis reveals that uncertainty influences not only how much firms hire, but also the kinds of roles they seek to fill.

Firms scale back high-skill and specialist roles more sharply. These high-skill roles, such as technical, managerial, and specialist positions, typically require more screening, longer onboarding, and deeper integration into teams, meaning they carry higher sunk costs. When uncertainty rises, firms delay or redesign such roles, often lowering experience requirements or broadening skill expectations to maintain flexibility.

Firms prefer flexibility over fixed commitments. Our analysis shows a significant shift toward contract, temporary, or project-based roles during periods of elevated uncertainty. Permanent, full-time positions decline disproportionately, especially in headquarters or central functions where staffing decisions carry greater strategic weight. This reflects a preference for roles that preserve organizational agility and minimize long-term commitments.

Hiring slows more in areas where dismissal is costlier. We observe larger hiring reductions in industries in which more workers are union members or covered by collective bargaining agreements. This pattern aligns with the economic logic of partial irreversibility: when unwinding a hiring decision is expensive, firms become more cautious about making that decision in the first place.

Job requirements become more flexible. Our analysis of required skills and qualifications reveals that firms lower their educational and experience thresholds, reduce highly-specific skill requirements, and place more emphasis on broad, general competencies. This suggests that firms are avoiding niche or tightly-defined roles and instead prefer employees who can be more easily redeployed across different parts of the organization if conditions change.

In addition to job advertisements, we also examine whether election uncertainty is associated with an increase in layoffs. We find no evidence that firms lay off more workers before a gubernatorial election. This suggests that the reduction in job postings is not driven by firms anticipating negative news or worsening economic conditions. Instead, it is the uncertainty surrounding the election itself that leads firms to scale back hiring.

4. Do workers react to uncertainty?

Uncertainty does not influence only firms. Workers also adjust their behavior, and these responses can reinforce or amplify the patterns we observe on the demand side. Although employee decisions are more difficult to observe directly, our analysis of worker flows provides some evidence that uncertainty shapes labor supply in important ways.

Workers become less willing to change jobs. We find that the number of employees joining firms declines in the quarters leading up to a gubernatorial election. Job search involves time, effort, and the risk of stepping into a role with uncertain prospects. When uncertainty rises, workers appear more inclined to stay in their current jobs rather than pursue new opportunities.

Employees do not leave firms at higher rates. Despite the slowdown in new workers joining the firms, we find no evidence that workers are more likely to leave their employer before an election. The absence of an increase in worker outflows is consistent with workers valuing job stability when the environment becomes harder to predict. Combined with lower inflows, our evidence points to a reduction in job-to-job mobility.

Matches formed under uncertainty are less stable. Even when firms hire during uncertain periods, the resulting matches tend to be less durable. Workers hired shortly before an election

are more likely to leave within their first year. This suggests that worker–firm matching is more difficult when uncertainty is high. Firms may relax their requirements or accept candidates who are not the ideal fit, and workers may accept roles that do not align perfectly with their skills or long-term plans. As conditions stabilize, these mismatches become more apparent, leading to earlier departures.

Taken together, our research shows that uncertainty reshapes hiring from both sides of the labor market. Firms post fewer vacancies, particularly for more specialized or longer-term roles. At the same time, workers become more cautious about switching jobs, reducing the flow of applicants. And among those who are hired, early turnover increases. This combination helps explain why hiring slows even when economic activity remains strong. Although our evidence is drawn from broad cross-industry data, the implications are particularly relevant for financial services firms, which operate in highly regulated environments and are often more exposed to political or policy shifts.

5. Key takeaways for financial services firms

Many of the hiring adjustments we observe in the data mirror the challenges experienced in the U.K. According to the ONS, U.K. job vacancies have been consistently declining since their peak in March to May 2022.[4] Within this broader contraction, financial services continue to grow – but the growth is highly concentrated in areas where firms face mandatory requirements or significant skill shortages such as regulatory, risk, and technology roles.[5] This sector is also particularly exposed to uncertainty, including changing regulatory expectations, requirements around operational resilience, and consumer protection. Similar to other industries, financial services also have to deal with the emerging implications of artificial intelligence and the broader geopolitical pressures.

Our research may help explain a broader shift in how financial institutions think about workforce decisions. When uncertainty rises, firms may become more hesitant to hire. Because hiring is costly to reverse and matches formed in uncertain periods tend to be less stable, financial institutions often choose to delay recruitment and distribute work across existing teams. This can lead to short-term workload pressures even when business performance remains strong.

For workers in financial services, this shift means that transferable skills and internal redeployability are becoming more important. As job descriptions broaden and firms prioritize flexibility, employees with adaptable capabilities are better positioned to move across functions. For employers, talent strategy increasingly focuses on building multi-skilled teams, strengthening internal mobility pathways, and reducing dependence on narrowly specialized roles that are harder to realign when business condition changes.

Finally, slower hiring should not be interpreted as a sign of deteriorating business fundamentals. In financial services, where regulatory, policy, and technological uncertainty is substantial, cautious hiring often reflects firms' desire to preserve agility rather than a deterioration in performance or outlook.

[4] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/jobsandvacanciesintheuk/october2025

[5] Office for National Statistics (ONS). UK Job Vacancies (thousands) – Financial and Insurance Activities, Series JP9Q. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/timeseries/jp9q/lms

6. Conclusion

Across industries, and particularly within financial services, uncertainty is becoming a defining feature of the operating environment. Our research shows that when the future outlook becomes difficult to predict, firms respond by slowing recruitment, redesigning roles, and placing greater emphasis on flexibility. Workers, too, react by reducing their willingness to move, tightening labor supply precisely when firms are trying to manage risk through more selective hiring. These adjustments help explain why labor markets can soften even when output growth remains steady.

For financial institutions, our research has direct relevance. Many of the hiring adjustments we observe in the data mirror patterns currently visible in the U.K.: slower recruitment pipelines overall, continued growth in regulatory and technology roles where skill shortages are acute, and greater reliance on more flexible staffing. In this context, workforce strategy becomes less about expanding headcount and more about building resilience through multi-skilled teams, stronger internal mobility, and role designs that allow talent to be redeployed as business needs evolve.

Crucially, a slowdown in hiring should not be misinterpreted as a signal of weakening fundamentals. In a sector shaped by regulatory change, technological disruption, and political uncertainty, cautious hiring is often a strategic response rather than a sign of stress. Understanding how uncertainty reshapes both labor demand and labor supply can help financial institutions anticipate bottlenecks, design more adaptable talent models, and make workforce decisions that remain robust even when the external environment is unsettled.

AI and institutional intelligence in the new global order:

re-architecting judgment in financial services

Authors | **Mauro Confalone** | Executive Director, Capco
Amit Kumal Goyal | Managing Principal, Capco

Abstract

Financial services are entering a structurally fragmented era in which geopolitical tension, regulatory divergence, sanctions complexity, data localization pressures and technological realignment are reshaping the economics of decision making. The central constraint is no longer digitalization alone, but the rising cost of coherent judgment across heterogeneous regimes. This paper argues that contemporary artificial intelligence (AI) does not merely automate tasks but transforms the architecture through which institutional judgment is produced. It introduces the concept of enterprise intelligence: an organizational capability in which data, models, policies, and human oversight operate as an integrated, governed system of judgment.

The paper advances two related propositions. First, although financial institutions have long relied on distributed expertise, competitive advantage increasingly depends on how well organizations structure, connect and govern that expertise across data, models, and workflows. Second, leadership in AI-enabled financial institutions centers on stewardship of institutional judgment: overseeing the systems that shape, evidence, and adapt organizational decisions under contextual variability. Drawing on examples from financial services compliance and regulatory interpretation, the analysis outlines design principles and governance implications for financial services leaders, C-suites and boards.

1. Introduction: finance in a fragmented world

For much of the post-crisis decade, the strategic challenge in financial services (FS) could be summarized as digitalize faster: modernize legacy stacks, move to cloud, automate workflows and use data more aggressively. That agenda is still unfinished. But it is no longer the hardest problem.

The harder problem is that finance is now operating inside a 'new global order' defined less by linear global integration and more by fragmentation and decentralization combined with global alignment on selected topics.

Fragmentation describes the weakening of shared rules and assumptions across jurisdictions whether in geopolitics, regulation, market access, sanctions enforcement, data governance or sustainability expectations. Decentralization reflects the shift from a single dominant center of coordination toward a multipolar arrangement, with multiple regulatory centers of gravity and regionally anchored financial infrastructures. At the same time, selective reintegration persists.

Initiatives such as the G20 cross-border payments program demonstrate cooperation around specific functional objectives even as broader convergence weakens.

For financial institutions, this is not abstract macroeconomics. It appears in everyday operational heterogeneity: differences in supervisory expectations, diverging sanctions regimes, variations in ESG disclosure frameworks, data residency requirements, and operational resilience standards such as the Digital Operational Resilience Act (DORA), which imposes specific system-level governance and resilience requirements not mirrored identically across jurisdictions. The same underlying transaction or client relationship may need to be evaluated differently depending on jurisdiction, booking location, data flows, or sectoral exposure.

The strategic consequence is frequently misdiagnosed. It is tempting to frame the moment as a renewed race toward digital acceleration; adopt AI sooner, automate more aggressively, modernize faster. Yet digitalization is not the bottleneck in a fragmented world. The bottleneck is coherent judgment: the ability to make decisions that remain consistent in intent, accountable in governance, and adaptive in execution when context is variable and changing.

Fragmentation increases the number of constraints per decision, the frequency of edge cases, and the speed at which assumptions expire. Getting decisions 'right' now requires not only analytical accuracy but contextual validity across regimes and defensibility after the fact. In other words, fragmentation raises the cost of correct judgment.

Competitive advantage therefore depends less on individual expertise and more on the institution's capacity to reliably know, reason, and act under contextual variability. This is the intuition behind enterprise intelligence: not artificial intelligence as a collection of tools, but as an organizational capability for producing and governing judgment where data, models, policies, controls, and human decision makers operate as one integrated system.

2. Knowledge and expertise in financial services: from tacit expertise to institutional intelligence

Financial services have long been characterized by their reliance on expert judgment. Across risk management, compliance, trading, operations, and client-facing activities, value has historically been created not simply through the application of formal rules or models, but through the interpretation of context, the navigation of ambiguity, and the exercise of professional discretion.

Classic theories of knowledge help explain why this has been the case. Polanyi's distinction between tacit and explicit knowledge emphasizes that "we know more than we can tell," highlighting the limits of formalization in domains requiring judgment and experience [Polanyi (1966)]. Much of what matters in financial decision making, like interpreting the intent of regulation or sensing emerging risks, has traditionally fallen into this tacit category. Expertise has, therefore, been embedded in individuals, socialized through apprenticeship, and reinforced through communities of practice rather than encoded exhaustively in systems.

Nonaka's knowledge creation framework and subsequent work on communities of practice highlight that organizational knowledge emerges through interaction rather than documentation alone [Nonaka and Takeuchi (1995)]. In financial institutions, escalation committees, peer review processes, and informal networks have functioned as vehicles for collective sense-making.

Table 1: The evolving locus of knowledge and judgment in financial institutions

	Tacit expertise	Documented knowledge	Executable institutional intelligence
Locus of knowledge	Individuals	Documents and committees	Organizations
Mode of knowing	Experience, intuition, judgment	Codified rules and referential guidance	Contextual, system-mediated judgment
Limiting factors	Scale, turnover, fragmentation	Interpretation, inconsistency	Accountability, trust and drift

Digitalization improved access to information but did not fundamentally alter how judgment was produced. Policies proliferated; governance frameworks expanded; documentation improved. Yet responsibility for interpretation and application remained human. Knowledge was referential, not executable.

Under fragmentation, this model becomes strained. The same decision increasingly demands different answers depending on jurisdiction, sanctions exposure, sustainability taxonomy, or data constraints. Expertise does not scale easily across contexts, nor does it preserve institutional memory as environments shift. Bounded rationality, Herbert Simon's (1957) insight that humans cope with complexity by simplifying, becomes a limiting constraint when contextual variability accelerates.

The traditional model of expertise, deeply human, context-sensitive, but locally bounded, struggles to absorb the pace and heterogeneity of change now confronting financial institutions. Table 1 summarizes how knowledge and judgment have historically been organized in financial institutions, and the distinct pressures that have emerged as scale, fragmentation, and regulatory complexity have increased.

The emergence of executable institutional intelligence does not eliminate tacit expertise or documented knowledge; rather, it repositions them within an organizational capability designed to scale judgment consistently and accountably across fragmented contexts.

The shift required is, therefore, not from tacit to explicit knowledge alone, but from documented knowledge to institutional intelligence: the ability to retrieve, contextualize, and apply knowledge within decision workflows in a manner that is consistent, auditable, and adaptable. The shift is further intensified by geopolitical fragmentation and regulatory divergence, which multiply interpretive variance across jurisdictions and supervisory regimes. Under such conditions, organizational judgment cannot rely on stable precedent or uniform regulatory baselines. The problem becomes not simply how to access expertise, but how to coordinate and evidence context-sensitive interpretation at scale.

This does not dissolve the role of human expertise; rather, it repositions expertise from being the primary site of decision making to becoming a source of training, validation, and stewardship for organizational judgment.

3. From technical modernization to institutional cognition: AI and the architecture of judgment

Previous technology waves transformed infrastructure and processes. Automation industrialized repetitive work where rules were stable and inputs structured. Digitalization rendered documents searchable and workflows traceable. Cloud increased scalability and resilience. Event streaming enhanced responsiveness.

These were substantive improvements. They modernized the substrate of the institution. But they did not fundamentally redesign how judgment was formed.

Latest advances in AI through language models differ because they operate indeed on language, the medium through which institutions interpret rules, construct rationales, and justify decisions. It enables systems not only to execute tasks but to participate in interpretation, recommendation, and explanation. This is not the automation of judgment; it is its augmentation.

3.1 The key move: from task automation to judgment participation

The thesis of this paper is that AI changes the architecture of judgment. That claim rests on a specific move:

Task automation replaces human effort in repetitive procedures where the rule-set is pre-defined (and stable preferably).

Judgment participation inserts AI into the formation of decisions where interpretation, context, and justification matter.

This is not a semantic shift; it is an architectural one. When AI participates in judgment, three things change simultaneously:

> *AI does not merely automate tasks but transforms the architecture through which institutional decisioning is produced. It shifts the authority from being primarily epistemic ('I know') to being architectural ('I ensure we know consistently and accountably').*

1. **The unit of design becomes the decision, not the task.** The question is no longer "How do we automate this workflow step?" but "How do we design a human-AI system that produces a coherent, accountable decision under variable context?"

2. **Governance shifts from model risk alone to judgment system governance.** Traditional controls (validation, testing, monitoring) remain necessary, but they are not sufficient. Institutions must govern how AI-generated interpretations are used, how rationales are constructed, how exceptions are handled, and how accountability is preserved when humans and machines jointly produce outcomes.

3. **Institutional memory becomes operational rather than archival.** Institutional memory must function operationally rather than archivally. Earlier knowledge management initiatives focused on preserving expertise in repositories for later retrieval. By contrast, AI-enabled enterprise intelligence integrates retrieval, synthesis, and explanation directly into decision workflows. Institutional knowledge becomes an active input into judgment formation, enabling contextualized, auditable reasoning and reducing reliance on informal 'hero experts' across teams and jurisdictions.

Table 2: Technical modernization versus institutional cognition (AI-enabled judgment systems)

Dimension	Technical modernization (automation/digitalization/ cloud/streaming)	Institutional cognition (GenAI/foundation models)
Primary objective	Increase speed, efficiency, resilience, connectivity	Improve coherence, defensibility and adaptability of judgment under contextual variability
Unit of design	Task/process step (workflow, hand-off, control point)	Decision/judgment loop (interpret, recommend, justify, escalate/act)
What gets encoded	Rules and workflows for stable, repeatable operations	Interpretation and rationale patterns grounded in policy, evidence, and context variables
Typical inputs	Structured data, known fields, deterministic triggers	Language and mixed evidence (policies, cases, correspondence, reports, incomplete signals)
Role of humans	Execute exceptions; interpret context; own reasoning	Own accountability, but AI can participate in interpretation, drafting, comparison, and justification
How 'knowledge' behaves	Mostly referential (documents exist; humans retrieve and apply)	Partly operational (retrieval + synthesis + explanation embedded in workflow)
Governance object	Systems, controls, model validation (where applicable)	Human–AI judgment systems: when/where AI is used, escalation boundaries, rationale standards, audit trails
Key risks / failure modes	Process brittleness, integration debt, outages, control gaps	Misinterpretation, plausible-but-wrong rationales, inconsistency across contexts, automation of weak practice
Auditability	Traceable steps; often weak capture of 'why'	Potentially stronger 'why' capture – if designed: cited evidence, logged assumptions, uncertainty, approvals
What 'good' looks like	Low cost-to-serve; fewer manual touches; faster cycle times	Lower cost of correct judgment; consistent decisions across regimes; faster adaptation to rule changes

This is precisely the capability gap that fragmentation exposes. What appears at macro level as regulatory asymmetry or data sovereignty constraints, for example, translates operationally into more exceptions, overlapping obligations and context-dependent decisions. Judgment participation offers a way to scale contextual reasoning while keeping it governable.

The practical implications of this shift become clearer in the discussion and Figure 1 in Section 3.3.

3.2 Proposition 1 – AI introduces a new architectural domain: design and governance of institutional intelligence

This domain sits alongside (but is distinct from) application architecture, data architecture, and infrastructure architecture. It is concerned with questions such as:

- **Where should interpretation occur?** At point of decision, upstream in policy codification, or via centralized advisory services?
- **What constitutes 'situational validity'?** Which context variables must be resolved – jurisdiction, product, client type, data location, sanctions exposure, ESG regime – before a recommendation is actionable?
- **How is rationale structured and preserved?** What evidence is cited, how is policy referenced, how is uncertainty expressed, and what is logged for audit?
- **How are escalation boundaries defined?** When can the system recommend and draft versus when must it route to humans, committees, or specialist review?
- **How is accountability maintained?** Who owns the judgment, who approves it, and how is responsibility allocated when AI contributes?
- **How is drift managed?** When rules change, guidance evolves, or geopolitical constraints shift, how does the judgment system stay aligned and defensible?

In other words, AI forces institutions to treat judgment as something that can be architected – designed, decomposed, instrumented, audited, and improved – not merely exercised.

This also reframes leadership attention. The strategic question becomes less "How quickly can we deploy AI tools?" and more "How do we build an institutional capability that produces correct judgment reliably across fragmented regimes?" That capability is what we refer to as enterprise intelligence: an organizational system in which data, policies, models, controls, and decision makers operate as an integrated judgment apparatus.

3.3 Applied example: sanctions screening under fragmentation

Consider a European bank processing a cross-border payment instruction for a multinational corporate operating in the energy sector. The transaction touches a U.S. correspondent bank, a U.K. booking entity, and an E.U.-regulated subsidiary. It must be evaluated against U.S. OFAC guidance, U.K. OFSI interpretation, E.U. consolidated sanctions lists, internal risk appetite thresholds, and potentially divergent supervisory expectations.

The underlying exposure is not clearly prohibited in any single jurisdiction, but sectoral restrictions, ownership thresholds, and recent interpretive guidance create ambiguity. The same transaction may be defensible under one regime and contestable under another.

Under traditional models, this would require manual retrieval of policies, committee decisions, consultation across compliance teams in different jurisdictions, and drafting of a rationale for approval or rejection. The process is cognitively heavy, context-sensitive, and often repeated across similar cases.

Under an enterprise intelligence architecture, the judgment loop can be partially system-mediated while remaining governed and accountable (Figure 1).

In this architecture, AI retrieves relevant policy and prior decisions, highlights regime conflicts, drafts a structured rationale citing evidence, and proposes a recommendation aligned with

defined risk tolerances. Human decision makers validate, override, or escalate as appropriate. The rationale and evidence trail are logged for audit and supervisory defensibility.

The system does not replace judgment; it structures it. It reduces the cognitive burden of contextual retrieval and cross-regime reconciliation while preserving accountability, escalation boundaries, and ultimate human authority.

4. Enterprise intelligence as an organizational capability

The preceding sections establish two related claims. First, fragmentation in the new global order increases the cost and complexity of correct judgment in financial services. Second, traditional models of expertise – centered on individuals, documents, and committees – struggle to scale judgment consistently under conditions of contextual variability. Together, these claims raise a deeper organizational question: if judgment can no longer be reliably produced through people and artefacts alone, where does it reside?

Figure 1: Hybrid sanctions judgment workflow

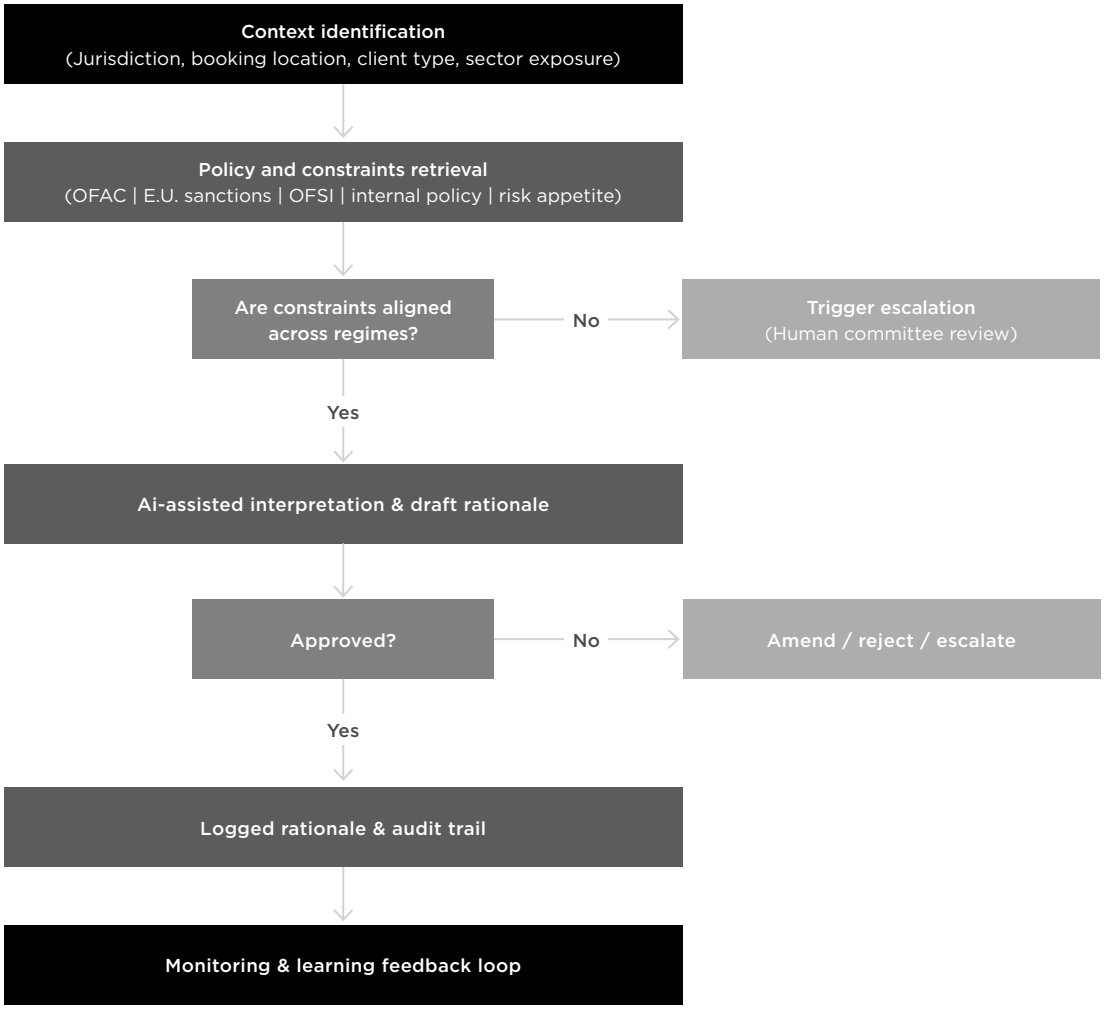
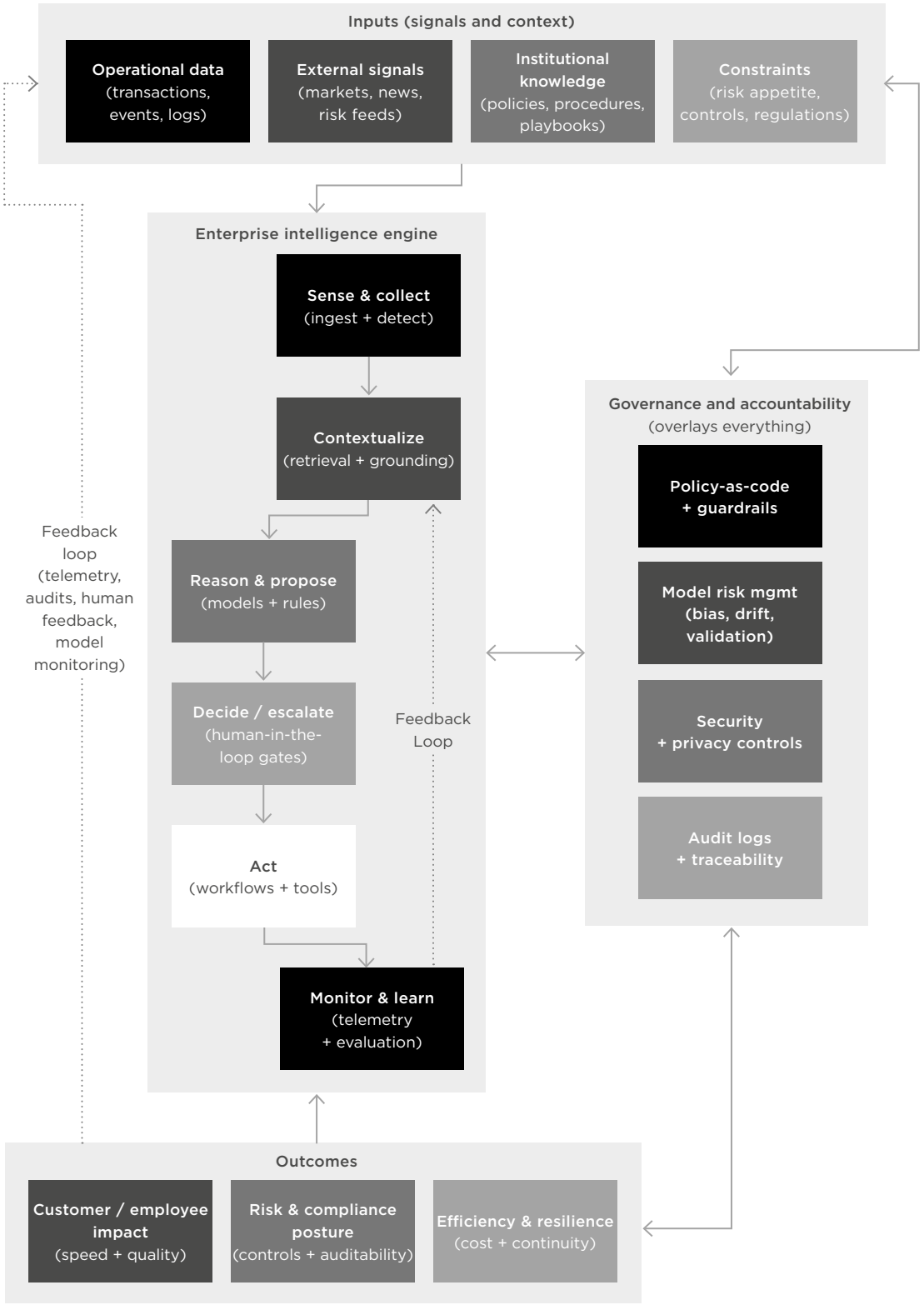


Figure 2: Enterprise intelligence as a governed judgment loop (conceptual)



This paper uses the term enterprise intelligence to describe an emerging organizational capability that addresses this question. Enterprise intelligence is not synonymous with analytics, automation, or the deployment of AI tools. Rather, it refers to the capacity of an organization to reliably know, reason, and act through integrated decision systems in which data, models, policies, controls, and human judgment are coherently aligned.

4.1 The enterprise intelligence judgment loop

Enterprise intelligence becomes concrete when we describe the unit of operation: the judgment loop. In most material decisions, institutions repeatedly move through the same sequence: interpreting a situation, forming options, justifying a recommended course of action, deciding (or escalating), acting, and learning from outcomes.

Traditional institutions execute this loop largely through people (and the artefacts they consult). Enterprise intelligence re-engineers the loop so that institutional knowledge participates directly in the decision process, while accountability stays human and governance stays explicit (Figure 2).

Returning to the sanctions example in Section 3.3, the judgment loop becomes visible in practice. The system first ingests operational signals such as transaction data, client identifiers, and booking information and retrieves relevant regulatory interpretations and internal policy constraints. It then constructs a structured recommendation, highlighting where jurisdictional rules align or diverge. Where constraints conflict, escalation pathways are triggered. If approved, the rationale is logged, not merely as documentation, but as structured institutional memory. Subsequent similar cases draw upon this recorded reasoning, while monitoring and audit processes evaluate whether patterns of exception, override, or supervisory sensitivity are emerging. The loop therefore governs not only individual decisions, but the evolution of judgment standards over time.

4.2 What enterprise intelligence is made of

To make the judgment loop operational, enterprise intelligence requires four interlocking design elements:

- **A knowledge substrate:** curated institutional sources (policies, regulatory interpretations, controls, product rules, precedents, supervisory communications) that can be retrieved and cited at the point of decision.
- **A context model:** an explicit representation of the variables that change the meaning of a decision; jurisdiction, booking location, client type, data residency, sanctions nexus, outsourcing constraints, sustainability regime, and so on.
- **A decision workspace:** a governed environment where evidence is assembled, interpretations are drafted, uncertainty is expressed, and human decision makers can approve, amend, or escalate with full traceability.
- **A governance substrate:** controls that define where AI may participate, what must be reviewed by humans, how rationales must be structured, and what is logged for audit and defensibility.

4.3 From standardization to governed differentiation

A crucial implication is that enterprise intelligence is not simply a new route to standardization. Fragmentation makes full standardization unrealistic: a single ‘global’ answer often cannot exist.

Instead, enterprise intelligence enables governed differentiation: the institution can deliver different answers in different contexts, while keeping consistency in intent, accountability, and rationale standards.

This is why enterprise intelligence should be treated as an architectural capability rather than a collection of use cases. Once the judgment loop is designed, multiple business domains can be onboarded onto the same institutional machinery, reducing duplicated interpretation effort and improving consistency across the enterprise.

5. Leadership after expertise

Financial institutions have long operated as professional bureaucracies characterized by distributed expertise across specialist communities. Authority has rarely rested in a single individual alone; rather, judgment has been formed through committees, escalation structures, and layered review processes. Senior leaders, however, were ultimately accountable because they were presumed to possess broader experience and the capacity to arbitrate complex trade-offs under uncertainty.

This model aligned with an environment in which much of what mattered remained tacit, context-sensitive, and difficult to codify. When interpretation depended on experience accumulated over years, authority naturally concentrated in individuals who had survived cycles, crises, and regulatory change. Even where expertise was plural, judgment was embodied in people and coordinated through deliberative structures. Organizational control rested, ultimately, on who had the right to decide.

However, the conditions that sustained this model are now shifting. As argued in earlier sections, fragmentation has increased the contextual variability of decisions while AI has altered the economics of knowledge mobilization. The result is not the disappearance of expertise, nor the sudden emergence of distributed leadership, but a reconfiguration of how institutional judgment is coordinated and evidenced, shifting leadership accountability toward stewardship of those judgment systems.

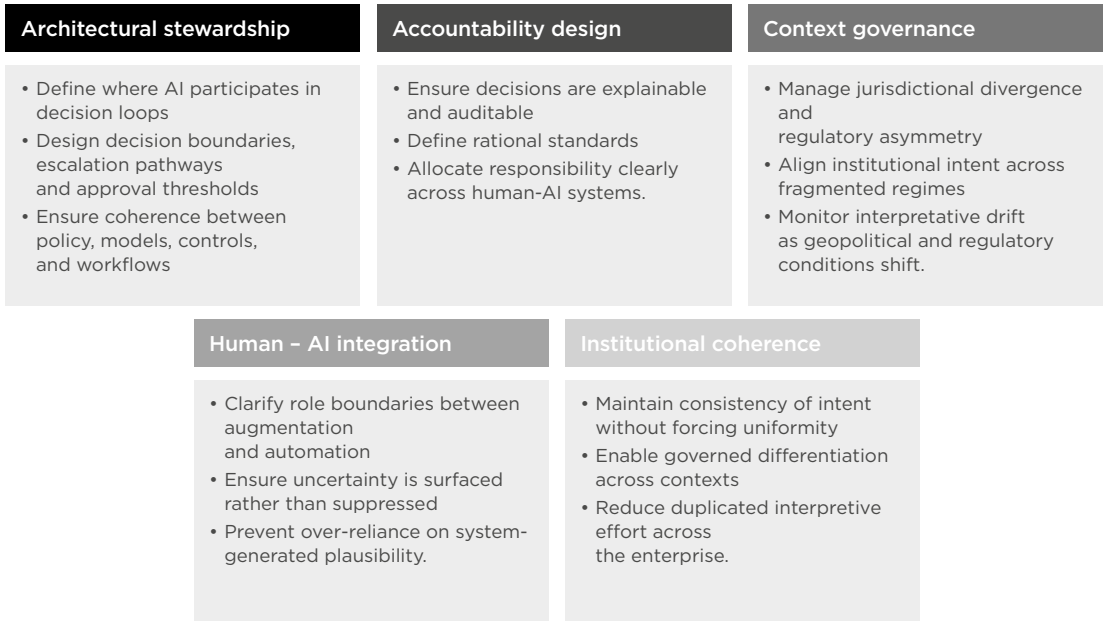
The leadership implications of contemporary AI differ fundamentally from those of traditional automation. Whereas classical task automation replaces human labor in stable, rule-based activities and primarily drives efficiency gains, modern AI, particularly language-based systems, engages directly in interpretive and reasoning processes. By analyzing policies, weighing trade-offs, comparing precedents, and drafting structured explanations, AI participates in the formation of judgment rather than simply executing predefined tasks. This shift reframes leadership: expertise no longer needs to reside solely in individuals, as judgment becomes partially system-mediated. What scales, therefore, is not the individual expert, but the institutional judgment system within which expertise operates.

5.1 Proposition 2 – leadership in AI-enabled financial institutions centers on stewardship of institutional judgment

If judgment increasingly resides within system-mediated architectures, leadership responsibility shifts accordingly. Leaders are now responsible not merely for decisions, but for the architecture through which decisions are formed. This includes defining where AI may participate, establishing rationale standards, setting escalation boundaries, and ensuring uncertainty is properly surfaced. Stewardship of institutional judgment requires specific capabilities that extend beyond technical literacy or traditional managerial authority.

Leaders in AI-enabled financial institutions must, therefore, develop and oversee five interlocking capabilities, summarized in Figure 3.

Figure 3: Leadership capabilities for stewardship of institutional judgment



The figure outlines the core capabilities required to steward institutional judgment under contextual variability. Leadership authority in this model rests not on exclusive possession of expertise, but on responsibility for the coherence, defensibility, and adaptability of the judgment systems themselves.

This shift has direct implications for boards and executive committees. Under emerging AI governance regimes, oversight bodies must ensure that AI-supported decisions are explainable, that accountability remains clear, and that systems are aligned with regulatory expectations. Fiduciary responsibility increasingly attaches to whether the judgment system itself is fit for purpose. Authority is no longer primarily epistemic ('I know'). It becomes architectural ('I ensure we know consistently and accountably').

In a fragmented environment, leadership legitimacy rests less on being the expert and more on ensuring that institutional intelligence remains coherent, defensible, and adaptive.

5.2 Accountability under hybrid, probabilistic, and distributed judgment

The most difficult implication of this shift concerns accountability. Traditional accountability models assume a clear chain from decision to decision maker. Hybrid judgment systems disrupt this assumption. Outcomes may reflect contributions from multiple humans, multiple models, evolving policies, and probabilistic inference rather than deterministic rules.

This does not eliminate accountability, but it changes its locus. Responsibility must attach not only to outcomes, but to the design and governance of the system that produced them. Leaders are accountable for:

- whether the system was appropriately scoped and controlled
- whether its use was aligned with stated intent and regulatory obligations

- whether decision rationales were explainable and auditable at the time they were made
- whether known limitations and uncertainties were responsibly managed.

In effect, accountability migrates from "Who decided?" to "Was this judgment system fit for purpose, and was it properly governed?"

This reframing aligns closely with emerging regulatory and supervisory expectations, which increasingly emphasize governance, explainability, and control over model-assisted decision making rather than prohibiting its use outright.

6. Conclusion: governing judgment in a fragmented world

Financial services are entering a structurally different era. Fragmentation across geopolitical, regulatory, and technological domains has raised the cost and complexity of correct judgment. The constraint is not simply data or processing power alone. It is coherence under contextual variability.

AI alters the architecture through which institutions can respond to this constraint. It enables institutional knowledge to participate directly in decision loops. Its value depends on how it is architected and governed.

Two propositions follow:

1. AI introduces a new architectural domain: the design and governance of institutional intelligence. Institutions must treat judgment as something that can be decomposed, structured, instrumented and improved, not merely exercised.

2. Leadership in AI-enabled institutions centers on stewardship of institutional judgment. Accountability increasingly attaches to the governance of human–AI systems rather than to isolated acts of decision.

The forward-looking implication is strategic. Institutions that treat AI as isolated productivity tools risk amplifying inconsistency inside already fragmented environments. Institutions that deliberately architect enterprise intelligence gain a different advantage: lower cost of correct judgment, stronger regulatory defensibility in supervisory dialog, and greater resilience under geopolitical volatility.

In the new global order, competitive strength may depend less on speed alone and more on the ability to remain coherent without being uniform, adaptive without being arbitrary, and accountable without relying on hero experts.

Enterprise intelligence is not a technology program. It is an organizational capability: a foundation of durable trust and strategic resilience. In a geopolitically fragmented environment, where sanctions regimes diverge, supervisory expectations vary, and digital sovereignty concerns intensify, the ability to produce consistent, defensible decisions across jurisdictions becomes a strategic asset. Institutions that cannot align judgment across regulatory blocs risk operational friction, supervisory tension, and reputational exposure.

In practical terms, this requires financial institutions to invest not only in AI tools, but in judgment governance infrastructure: explicit escalation design, rationale standards, drift monitoring mechanisms, and board-level oversight of human–AI decision systems. Without this, AI deployment risks amplifying inconsistency rather than reducing it.

References

Abbott, A., 1988. The system of professions: An essay on the division of expert labor, University of Chicago Press

Aiyar, S., A. Presbitero, and M. Ruta (eds), 2023. Geoeconomic Fragmentation: The Economic Risks from a Fractured World, IMF and CEPR Press, Paris & London, https://cepr.org/publications/books-and-reports/geoeconomic-fragmentation-economic-risks-fractured-world-economy

Autor, D., 2015. Why Are There Still So Many Jobs? Journal of Economic Perspectives, Vol 29(3): 3-30, http://bit.ly/4ljp2bv

Basel Committee on Banking Supervision (BCBS), 2013. Principles for effective risk data aggregation and risk reporting, BCBS 239, https://www.bis.org/publ/bcbs239.pdf

Board of Governors of the Federal Reserve System, 2011. SR 11-7: Guidance on model risk management, April 4, https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm

Doshi-Velez, F. and B. Kim, 2017. Towards a rigorous science of interpretable machine learning, https://arxiv.org/abs/1702.08608

Drucker, P. F., 1993. Post-Capitalist Society, HarperCollins

European Banking Authority (EBA), 2020. EBA report on big data and advanced analytics, EBA/REP/2020/01, January, https://www.eba.europa.eu/sites/default/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf

European Central Bank (ECB), 2024. Guide on effective risk data aggregation and risk reporting, May, https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guides240503_riskreporting.en.pdf

European Commission (EC), CSRD overview and timing

European Commission (EC), 2021. Standard Contractual Clauses, June 4, https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en

European Parliament and Council of the European Union, 2024. Artificial Intelligence Act (Regulation EU 2024/1689), https://artificialintelligenceact.eu/the-act/

European Parliament and Council of the European Union, 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series, 2024/1689, July 12, https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Parliament, 2020. The CJEU judgment in the Schrems II case, July 20, https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf

European Securities and Markets Authority (ESMA), n.d. Digital Operational Resilience Act (DORA), https://bit.ly/3P9ADho

Financial Action Task Force (FATF), 2025. FATF Recommendations (amended Oct), https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html

Financial Stability Board (FSB), 2024. Recommendations on alignment and interoperability across data frameworks for cross-border payments, https://www.fsb.org/2024/12/recommendations-to-promote-alignment-and-interoperability-across-data-frameworks-related-to-cross-border-payments-final-report/

Freidson, E., 2001. Professionalism: The third logic, University of Chicago Press

Gupta, S., R. Ranjan, and S. N. Singh, 2024. A comprehensive survey of retrieval-augmented generation (RAG): Evolution, current landscape and future directions, https://arxiv.org/abs/2410.12837

Hutchins, E., 1995. Cognition in the wild, MIT Press

IFRS/ISSB, IFRS S1/S2 introduction (global baseline) https://www.ifrs.org/sustainability/knowledge-hub/introduction-to-issb-and-ifrs-sustainability-disclosure-standards/

International Monetary Fund (IMF), 2023. Global Financial Stability Report, April, https://www.imf.org/-/media/files/publications/gfsr/2023/april/english/text.pdf (especially Chapter 3: Geopolitics and Financial Fragmentation)

International Organization of Securities Commissions, (IOSCO), 2019. Market Fragmentation and Cross-border Regulation (FR07/2019), report, June, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD629.pdf

International Organization for Standardization and International Electrotechnical Commission, 2023. ISO/IEC 42001:2023 Information technology – Artificial intelligence – Management system, https://www.iso.org/standard/42001

Kahneman, D. and G. Klein, 2009. Conditions for intuitive expertise. American Psychologist, 64(6), 515–526, https://psycnet.apa.org/buy/2009-13007-001

Lave, J. and E. Wenger, 1991. Situated Learning: Legitimate Peripheral Participation, Cambridge University Press

Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks, in: H. Larochelle and M. Ranzato (eds), NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems, https://dl.acm.org/doi/abs/10.5555/3495724.3496517

Lundberg, S. M. and S.-I. Lee, 2017. A unified approach to interpreting model predictions, in: U. von Luxburg (ed.), NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, https://dl.acm.org/doi/10.5555/3295222.3295230

March, J. G. and H. A. Simon, 1958. Organizations, John Wiley and Sons

Mintzberg, H., 1979. The structuring of organizations, Prentice Hall

Monetary Authority of Singapore (MAS), 2018. Principles to promote fairness, ethics, accountability and transparency (FEAT) in the use of artificial intelligence and data analytics in Singapore's financial sector, November 12, https://www.mas.gov.sg/-/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

National Institute of Standards and Technology (NIST), 2023. Artificial intelligence risk management framework (AI RMF 1.0), NIST AI 100-1, https://tinyurl.com/3ey9n6kj

Nonaka, I. and H. Takeuchi, 1995. The Knowledge-Creating Company, Oxford University Press USA

Office of the Comptroller of the Currency (OCC), 2011. Sound practices for model risk management; supervisory guidance on model risk management, OCC Bulletin 2011-12, https://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-12.html

Office of Financial Sanctions Implementation (OFSI), 2026. UK financial sanctions general guidance, updated January, https://www.gov.uk/government/publications/financial-sanctions-general-guidance/uk-financial-sanctions-general-guidance

Office of Foreign Assets Control (OFAC), 2019. Framework for OFAC Compliance Commitments, https://ofac.treasury.gov/media/16331/download?inline

Organisation for Economic Co-operation and Development (OECD), n.d., Cross-border data flows, https://www.oecd.org/en/topics/sub-issues/cross-border-data-flows.html

Parasuraman, R. and V. Riley, 1997. Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253, https://doi.org/10.1518/001872097778543886

Polanyi, M., 1966. The Tacit Dimension, University of Chicago Press

Power, M., 2007. Organized uncertainty: Designing a world of risk management, Oxford University Press

Prudential Regulation Authority (PRA), 2021. SS2/21: Outsourcing and third party risk management, Bank of England, Supervisory Statement 2/21, https://www.bankofengland.co.uk/prudential-regulation/publication/2021/march/outsourcing-and-third-party-risk-management-ss

Prudential Regulation Authority (PRA), 2023. SS1/23: Model risk management principles for banks, Bank of England, Supervisory Statement 1/23, May, https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss

Ribeiro, M. T., S. Singh, and C. Guestrin, 2016. "Why should I trust you?": Explaining the predictions of any classifier, in: KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, https://dl.acm.org/doi/10.1145/2939672.2939778

Securities and Exchange Commission (SEC), 2025. SEC Votes to End Defense of Climate Disclosure Rules, press release, March 27, https://www.sec.gov/newsroom/press-releases/2025-58

Simon, H. A., 1957. Administrative Behavior: a Study of Decision-Making Processes in Administrative Organization, MacMillan

Tsoukas, H., 2005. Complex Knowledge: Studies in Organizational Epistemology, Oxford University Press USA

Weick, K. E., 1995. Sensemaking in Organizations, Sage Publications

Weick, K. E. and K. M. Sutcliffe, 2007. Managing the unexpected, Jossey-Bass

Wenger, E., 1998. Communities of Practice: Learning, Meaning, and Identity, Cambridge University Press

Woods, D. D. and E. Hollnagel, 2006. Joint cognitive systems, CRC Press

Yu, H., A. Gan, K. Zhang, S. Tong, Q. Liu, and Z. Liu, 2024. Evaluation of retrieval-augmented generation: A survey, https://arxiv.org/abs/2405.07437

Private markets

Private equity's retail turn:

anatomy of a misselling risk

Author | **Ludovic Phalippou** | Professor of Financial Economics, University of Oxford, Said Business School

Abstract

Private equity's retail expansion has been promoted as 'democratization,' yet it primarily transfers the complexity and opacity of institutional structures to individual investors. Semi-liquid vehicles now marketed to the public rely on discretionary valuations, layered fees, and misleading performance metrics. These features, once tolerated among institutions, are now creating predictable legal and economic risks. The likely outcome is a wave of litigation that will expose hidden practices and force greater transparency. This process will make private equity cleaner, fairer, and ultimately more sustainable, but only after the reckoning has begun.

1. Introduction

Private equity's reach has expanded dramatically over the past decade. Once limited to institutional investors such as pension funds and endowments, it is increasingly offered to individuals through new fund structures and wealth platforms. This shift is often presented as 'democratization'. In principle, broader access can be beneficial. But access is not the same as suitability. Extending a model built for institutions to households without retail-grade disclosure, valuation governance, liquidity rules, and accountability is closer to putting non-licensed drivers on the road than to the democratization of transport.

A recurring rationale for opening private equity to individuals is that public markets are shrinking. Private assets are said to offer exposure to companies that are no longer listed. Yet this argument is overstated. Eckbo and Lithell (2023) show that much of the decline in the number of listed firms reflects mergers rather than a withdrawal from public ownership. Many former stand-alone companies have simply been absorbed into larger listed entities. Today's conglomerates – Google, Amazon, Microsoft, among others – each encompass multiple business lines that would previously have been separate public firms. Moreover, empirical evidence shows that the elimination of most idiosyncratic risk does not require owning 'everything on earth'. Statman (1987) shows that a portfolio of roughly 40 stocks captures the bulk of diversification benefits. More recent work [Zaimović et al. (2021)] suggests that the number may now be somewhat higher, but investors do not need hundreds of equity positions to be well diversified.

Another common justification is the claim that private equity reliably outperforms public markets. The argument is then framed as a matter of fairness: if these higher returns exist, it would

be unjust to reserve them for the wealthy. But this premise is not supported by data. Across private capital funds, performance has broadly matched that of public equity once the appropriate benchmarks are chosen [Phalippou (2020), Phalippou (2026)].

Despite weak evidence for superior returns and the shaky diversification argument, the fairness narrative has proven powerful. It has provided the political and commercial justification for a wave of product innovation aimed at bringing private equity to a mass audience. Since 2023, a growing set of semi-liquid funds – interval and tender-offer funds in the U.S., European Long-Term Investment Funds (ELTIFs) in the E.U., and Long-Term Asset Funds (LTAFs) in the U.K. – have channeled household savings into private equity and private debt. These vehicles lower investment minimums, offer some continuous subscription and redemption features, etc. Yet they retain the essential characteristics of institutional private equity: discretionary valuations, opaque and multi-layered fees, performance metrics that do not map into investor returns, and limited fiduciary protections.

The objective of this paper is to examine what happens when the institutional private equity model is extended to non-professional investors without changing its mechanics. It shows how performance presentation and pricing conventions can overstate returns and understate risk, how valuation discretion turns into a transfer mechanism, and how weak disclosure and governance create both legal and economic risks.

Retail participation in private markets is not inherently problematic. The risk comes from the way the products are currently packaged and distributed. Semi-liquid vehicles import institutional features (valuation discretion, complex fee stacks, and limited liquidity) into retail channels where investors cannot negotiate terms and often rely on intermediaries whose incentives are flow-driven. If reporting, governance, and accountability are not upgraded to retail standards, conduct risk and litigation become likely.

2. Retail vehicles

In the U.S., interval funds, tender offer funds, non-traded business development companies (BDCs), and non-traded REITs are the main retail vehicles for private equity and private credit. Industry estimates place assets in these semi-liquid structures at roughly U.S.$350 billion at the end of 2024, up from about U.S.$215 billion at the end of 2022 (around 60% total growth over two years).[1] The largest retail-facing private credit funds are managed by a relatively small set of sponsors (for example, Blackstone, Cliffwater, Blue Owl, Apollo), although concentration varies by segment and changes over time.

Interval funds are registered under the 1940 Investment Company Act. They offer mandatory quarterly redemptions, normally up to 5% of net asset value (NAV), and continuous subscriptions on a daily or weekly basis. Tender offer funds are also closed end registered funds but permit discretionary redemptions. By year-end 2024, there were approximately 118 interval funds and 113 tender-offer funds in the U.S., with about U.S.$99 billion and U.S.$80 billion in assets respectively. Business development companies held about U.S.$225 billion in net assets, putting the total for these semi-liquid and private-credit-linked vehicles near U.S.$400 billion in aggregate assets.[2]

[1] Source: https://www.advisorhub.com/private-funds-for-retail-investors-jump-to-350-billion-morningstar-says
[2] https://www.ici.org/research/closedend-funds

In addition, recent regulatory changes have both widened the investor base and loosened product restrictions. The Securities and Exchange Commission lifted its long-standing 15% cap on private asset exposure in registered funds, allowing them to hold larger allocations to private markets and to be sold to a broader pool of investors. At the same time, Congress has expanded the definition of accredited investor, and by 2025 roughly one quarter of U.S. households meet the threshold. An executive order also directed the Department of Labor to permit private market exposure within 401(k) retirement plans. These combined steps have opened the door for large asset managers to design and market semi-liquid products directly to individual investors.

In the U.K., the Long Term Asset Fund (LTAF) regime introduced by the Financial Conduct Authority in 2021 was designed to facilitate investment in long-term illiquid assets within a regulated framework. Market development has been gradual but growing. Industry summaries such as that of Morningstar put the assets in FCA-authorized LTAFs at around £5 billion, with early activity led by firms such as Schroders, Aviva, and BlackRock. From April 2026, LTAFs will also be eligible for inclusion in Individual Savings Accounts (ISAs), the U.K.'s main tax-advantaged defined-contribution retail savings wrapper, which holds over £700 billion of household assets.[3]

Europe's equivalent is the European Long-Term Investment Fund (ELTIF) 2.0, which became operational in early 2024. Under the revised framework, 55 ELTIFs were launched in 2024, more than double the previous high, and this strong momentum has continued into 2025. According to Scope Fund Analysis, the total ELTIF market volume reached €20 billion by the end of 2024, with Luxembourg and France among the leading domiciles for these vehicles. Large asset managers, including BlackRock and Amundi, have been active participants in this segment.[4]

Together these vehicles mark a structural change in how household savings reach private markets. They lower entry thresholds and create the appearance of liquidity while retaining the economic and governance features of institutional private equity. The next section examines the implications of this new architecture, focusing on how it can turn an institutional model into a retail misselling problem.

3. Issues

3.1 Performance metrics and comparability

Comparing private equity performance to public markets is more complex than most retail marketing suggests. The dominant reporting metric in private equity is the internal rate of return (IRR), a money-weighted measure that is fundamentally different from the time-weighted returns commonly used for public equities. Most importantly, IRR is not a measure of investor wealth compounding. It assumes reinvestment of interim cash flows at the same rate, an assumption that rarely holds in practice. As a result, IRRs can be high even when the growth of invested capital is modest.

This property of IRR is particularly relevant in a retail context, where reported figures are often interpreted as stable annual returns. Early distributions mechanically raise IRR and can anchor high values that change little over time, even if subsequent outcomes are weaker. This feature is visible in public disclosures. For example, KKR reports a since-inception (1976) gross IRR of 25.5% for its private equity program, a figure that has shown little variation across reporting periods (see Table 1). Apollo

[3] https://www.ajbell.co.uk/group/news/isas-unpacked-who-holds-them-and-how-much-do-they-have
[4] https://www.scopeexplorer.com/files/get/?name=news.ReportFile%2Fbytes%2Ffilename%2Fmimetype%2FScope_ELTIF_study_2025_ENG.pdf

has similarly reported since-inception IRRs of around 39% year after year. These figures are not incorrect as reported, but they do not correspond to a compound rate at which an investor's wealth would have grown. Interpreted as such, they would imply implausibly large terminal wealth levels. Outside an institutional context, these metrics are therefore easy to misinterpret and are not directly comparable to public-market return measures.

The issue is not limited to the choice of metric but extends to how performance is presented. Retail platforms frequently display 'since-inception IRR' in chart form as a smooth, upward-sloping line, visually resembling a time series of annual returns. Yet IRR is not a time-series return, and this mode of presentation can blur the distinction between money-weighted fund-level reporting and investor-level wealth compounding. The risk of misinterpretation is amplified when such charts are shown alongside public equity indices whose reported returns do represent time-weighted growth.

Once returns are measured and benchmarked on a consistent basis, the empirical evidence does not support the view that private equity has systematically outperformed public equity net of fees. Using comparable benchmarks and definitions, the aggregate net performance of private capital funds has been close to that of public equity over long horizons [Phalippou (2020); Phalippou (2026)]. This does not preclude

Table 1: KKR Since-Inception Gross IRR (1976–2024)

Reported for end of year	Gross IRR (%)	Filing
2007	26.3	S-1 (Mar 2007)
2008	26.1	S-1 (Mar 2008)
2009	25.8	S-1 (Dec 2009)
2010	25.8	Form 10-K (Dec 2010)
2011	25.7	Form 10-K (Dec 2011)
2012	25.7	Form 10-K (Dec 2012)
2013	25.7	Form 10-K (Dec 2013)
2014	26.0	Form 10-K (Dec 2014)
2015	25.6	Form 10-K (Dec 2015)
2016	25.6	Form 10-K (Dec 2016)
2017	25.6	Form 10-K (Dec 2017)
2018	25.6	Form 10-K (Dec 2018)
2019	25.6	Form 10-K (Dec 2019)
2020	25.6	Form 10-K (Dec 2020)
2021	25.6	Form 10-K (Dec 2021)
2022	25.6	Form 10-K (Dec 2022)
2023	25.5	Form 10-K (Dec 2023)
2024	25.5	Form 10-K (Dec 2024)
Note: Legacy PE Funds from 1976 to 1996 vintages pooled reported gross IRR	26.1	Form 10-K (Dec 2024)

This table reports the gross internal rate of return (IRR) since inception disclosed by KKR & Co. Inc. in its SEC filings. The IRR represents the aggregate performance of the firm's private equity funds since 1976 as reported each year.

dispersion across funds, but it does undermine claims of broad, persistent outperformance that are often implicit in retail marketing narratives.

In a retail setting, these distinctions matter. When complex institutional metrics are presented without explicit explanation, investors may reasonably infer properties that the metric does not have. Clearer disclosure, standardized performance reporting, and explicit education about the meaning and limits of IRR are therefore not optional add-ons but necessary conditions for meaningful comparability and informed participation.

3.2 Biased valuations

In traditional closed-end partnerships, NAVs are periodic accounting estimates with little immediate effect. In open-ended retail vehicles, investors enter, exit, and pay fees based on those marks, giving small valuation changes immediate cash consequences. When new investors subscribe at optimistic NAVs, wealth transfers to continuing investors, and fees based on inflated NAVs overcharge everyone.

A notable illustration appeared in 2024, when Hamilton Lane's retail fund purchased private-equity stakes at roughly 80% of stated NAV and marked them up to 100% the next day, booking large one-day gains. Such practices are routine across the industry. They transform valuation discretion into a direct source of economic transfer and litigation exposure. As explained in Phalippou and Magnuson (2025), once retail investors buy and redeem at those inflated marks, mispricing ceases to be a technical matter – it becomes a consumer-protection and contract issue.

Real-estate and credit vehicles face the same dynamic. When the Blackstone BREIT trust met only a quarter of redemption requests in 2023, investors began questioning whether reported NAVs reflected realizable prices. The same pattern has since appeared in several interval funds, where 'mark-to-model' pricing determines who exits profitably and who remains locked in.

3.3 Fee opacity and layering

The classical '2 and 20' fee model persists, but retailization adds more layers. Management and performance fees accrue at the fund level, asset-level fees at portfolio companies, and new layers – distribution, platform, and placement costs – at the retail interface. Phalippou and Gottschalg (2009) show that standard institutional fees reduce annual returns by about 7% for the average fund. Adding the three percentage points typical of retail products leaves little room for investor profit.

In addition, retail investors are unlikely to grasp the multitude of fees and expenses embedded in the underlying structures. Beyond the stated management and performance charges, private equity funds extract additional revenue through monitoring and transaction fees at portfolio companies [Phalippou et al. (2018)], affiliate consulting arrangements, and financing costs linked to subscription credit lines. Retail investors typically see only the top-level charges of the product they buy, while the economic burden of these lower-layer costs remains hidden. There is still no requirement to disclose a consolidated total expense ratio that would reveal the true cost of participation.

3.4 Governance limits and fiduciary exposure

Limited-partnership agreements in private equity routinely waive fiduciary duties and permit related-party transactions under broad indemnities. Such clauses were tolerated in institutional settings under the assumption of informed consent. Transposed into retail structures, they remove the core investor protection on which securities law rests.

As shown in Phalippou and Magnuson (2025), fiduciary waivers and indemnity clauses have become central to the industry's risk transfer: they turn what would otherwise be regulatory violations into contractual defaults. Retail investors enter through feeders and platforms that diffuse accountability.

Fund boards owe duties to the vehicle, not to investors in affiliated funds. The result is a governance vacuum where no party bears ultimate responsibility for fairness or disclosure.

The legal exposure is already visible. Continuation funds and cross-fund transactions – where managers sell assets from one fund to another they control – create textbook conflicts of interest. In 2025, investors sued H.I.G. Capital for selling a company from one fund to another at a self-set price. Such cases demonstrate how structural conflicts, once (barely) manageable within an institutional context, become significantly litigable once retail investors enter. As Phalippou and Magnuson (2025) emphasize, private equity's retailization shifts accountability from regulators to courts: what used to be a question of disclosure will now be tested through doctrines of contract, fiduciary duty, and consumer protection.

4. Litigation risk and the case for regulatory safeguards

The migration of private equity structures into semi-liquid, retail-facing formats magnifies long-standing weaknesses: valuation discretion, fee complexity, performance metrics that are easy to misinterpret, and limited governance constraints. These features were historically manageable in institutional settings, where investors had some bargaining power, contractual protections, and repeat-player relationships. In a retail context, they become sources of significant legal exposure. What was once an insider debate about benchmarking and disclosure increasingly takes the form of a consumer-protection problem that is likely to shape the next phase of private-capital regulation.

As Phalippou and Magnuson (2025) show, the same mechanisms that insulate private equity firms from public-market scrutiny can expose them to private-law liability once ordinary investors are involved. Expanding the investor base transforms limited partnerships into potential targets for contract, fiduciary-duty, and consumer-protection claims. Retail investors face no reputational or career constraints on litigation. Individual claims may be small, but they are numerous, and class-action procedures can aggregate them into material legal risk.

Two elements make litigation not merely possible but predictable. First, the presentation of performance metrics, especially IRR, creates a public disclosure record that courts can evaluate under doctrines of misrepresentation, negligent misstatement, or unfair and deceptive practices. Second, NAV-based pricing combined with layered fee structures generates identifiable cash transfers between investors, making potential damages quantifiable. Unlike disputes over strategy or skill, these claims turn on observable accounting and disclosure choices.

Legal doctrines long viewed as peripheral to private markets are therefore becoming central. Courts can scrutinize valuation practices, fee opacity, and conflicts of interest under common-law contract and fiduciary principles, as well as under consumer-protection statutes. Misleading interpretations of IRR, incomplete disclosure of layered costs, and self-priced transactions between affiliated funds fit squarely within these frameworks.

Litigation is thus likely to act as both a sanction and a disclosure mechanism. Private lawsuits will expose practices that were previously shielded by opacity or complexity. Yet reliance on ex post

enforcement is an inefficient way to discipline market design. The preferable outcome is to reduce ambiguity through ex ante safeguards that clarify performance reporting, valuation governance, and cost transparency before disputes arise.

At a minimum, retail-facing documentation should make explicit four elements that are central to investor outcomes but easy to misunderstand:

(i) how the timing of cash flows mechanically affects IRR and why it should not be interpreted as a compound return;

(ii) how subscriptions and redemptions at reported net asset values can transfer value between investors when underlying assets are illiquid;

(iii) what liquidity-management tools such as gates, queues, and repurchase caps imply under stressed conditions; and

(iv) the consolidated, all-in cost borne by the investor across all layers of the product.

Without such measures, the retailization of private equity risks reproducing a familiar cycle of financial innovation: initial enthusiasm, inflated expectations, and subsequent litigation. Access without accountability breeds disputes; access under transparent and enforceable rules can make the market sustainable.

5. Conclusion

Retail participation in private markets is not inherently problematic, but the current structure is not well adapted to it. A framework designed for institutional negotiation cannot simply be repackaged and sold to individuals. Once private equity's valuation discretion, performance metrics that are easy to misinterpret, fee layering, and governance waivers are transferred to a retail setting, they cease to be technical features and become sources of predictable conduct and litigation risk.

The path forward is no longer straightforward. The safeguards that could have prevented large-scale disputes are not yet in place, and litigation is therefore likely to play a central role in the next phase of market development. Lawsuits will expose practices that were long tolerated in institutional contexts but become contested when applied to retail investors. This process will be costly, but it may also perform a function that regulation has so far struggled to achieve: forcing greater transparency, clearer disclosure, and stronger accountability.

Over time, this reckoning is likely to reward firms that already operate with disciplined valuation practices, conservative disclosure, and restrained use of complexity, rather than those that rely on opacity or aggressive marketing. Private equity can serve retail investors, but only under conditions that recognize the limits of institutional metrics and contracts in a retail environment. Access alone does not create value; transparency, governance, and cost discipline do.

References

Eckbo, B.E. and M. Lithell, 2023. "Merger-driven listing dynamics," Journal of Financial and Quantitative Analysis, https://doi.org/10.1017/S0022109023001394

Phalippou, L., 2020. "An inconvenient fact: Private equity returns and the billionaire factory," Journal of Investing 29, 11-39, https://www.pm-research.com/content/iijinvest/30/1/11

Phalippou, L., 2026. "Apples and oranges: Benchmarking games and the illusion of private equity outperformance," Journal of Private Market Investing, https://www.pm-research.com/content/iijpriveq/24/2/28

Phalippou, L. and O. Gottschalg, 2009. "The performance of private equity funds," Review of Financial Studies 22, 1747-1776, https://www.jstor.org/stable/30225708

Phalippou, L. and W. Magnuson, 2025, "Private equity, public capital, and litigation risk," working paper, University of Oxford and Texas A&M University, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5748424

Phalippou, L., C. Rauch, and M. Umber, 2018. "Private equity portfolio company fees," Journal of Financial Economics 129, 559-585, https://www.sciencedirect.com/science/article/abs/pii/S0304405X18301478

Statman, M., 1987. "How many stocks make a diversified portfolio?", Journal of Financial and Quantitative Analysis 22, 353-363, https://www.jstor.org/stable/2330969

Zaimović, A., L. Babić-Hodović, and A. Džafić, 2021. "How many stocks are sufficient for equity portfolio diversification? A review of the literature," Journal of Risk and Financial Management 14, 551, https://www.mdpi.com/1911-8074/14/11/551

Venture capital investment

and the return of geopolitics

Authors | **Alex Preda** | Professor of Sociology, Lingnan University
David Xingyi Chen | Research Officer, Lingnan University[1]

Abstract

We set out two major ways states and geopolitics intervene in venture capital (VC) operations. First, states and geopolitics constitute uncertainties and risks that VCs need to account for, on the capital that VCs attract and deploy, and the technologies that VCs invest in. Second, governments participate in the private capital markets by providing VCs with funds, collaborating with VCs, and purchasing startup services. We observed changes in how these two avenues influence VC operations. Geopolitics are increasingly pertinent, resulting in a situation of heightened risks and uncertainties, impacting access to markets and potential returns. Governments behave more strategically in select vectors and retrench in others, resulting in a corresponding reorientation of VC strategies. These developments go against the understanding of the retreat of the state, where private capital increasingly becomes the go-to medium for funding technological innovation. Instead, the states are becoming more hands-on, acting according to geopolitical competitions and their strategic priorities. We call for this development to be recognized by academics and practitioners and offer practical suggestions.

1. Introduction

While clustered within hubs of capital, technological expertise and academic institutions (e.g., New York, Silicon Valley, London, Hong Kong, or Singapore), venture capital (VC) represents transnational capital flows, with practitioners hailing from many different nationalities, intervening and deploying in multiple geographic, political and regulatory environments. As such, VCs come into contact with state bodies and geopolitics in their operations.

One may argue that nation states and their geopolitics influence many markets. VCs are indeed not alone in having to respond to and account for national policies or trade breakdowns. The stock market is famously responsive to geopolitical tensions, as illustrated by their impact on S&P 500, FTSE 100, Nikkei and Hang Seng Index. The VC investment market is different, as they consist of primarily privately associated, large-scale capital, and are not directly controlled by states or influenced by public perceptions of geopolitical developments. Though there have been recent developments where states set up their own VC funds [Sun (2025)], the VC market remains primarily private in nature. This is not to say that VCs can safely ignore macroenvironments, more that they engage with them differently. Since the ultimate obligation of general partners (GPs) is to

[1] This work is funded by the Hong Kong Research Grant Council, the General Research Fund, grant 13608124. ChatGPT 5.2 was used for research on regulatory regimes across different countries.

return funds to primarily private investors, they are relatively autonomous and at arm's-length from geopolitical factors. This article will set out the specificities of VCs in their engagement with geopolitics in detail, as well as setting out recent changes in the field. The arguments and observations we put forward are based on 54 ethnographic interviews with practitioners in the field (VCs, startup founders, and adjacent professionals). We start by pointing to VCs' global features to establish a basis for geopolitics to intervene in VC operations, through their capital flows, cosmopolitan practitioner scenes, project visions, and navigation of regulatory regimes. We then move on to propose two major avenues where geopolitics intervene: 1) as uncertainties and risks and 2) as participants of private capital. We will then highlight how the expressions of these avenues are changing.

2. Global venture capital

Venture capital, as a category of high-risk, high-reward investment, originates from whaling ventures in the nineteenth century [Nicholas (2019)]. These ventures, from their inception, are underlined by the extreme risks within which an investment takes place – a whaling ship that one invests in might return to port empty-handed, or even not at all. In the reverse instance, a ship may yield astronomical returns if the expedition proves successful. The exposures to highly-skewed risk profiles define the modus operandi of VCs. Venture capital as an institutional investment form solidified first in the U.S., and to a lesser extent the U.K., from the post-war period onwards, engaging with ever-larger sums of capital under management [Klingler-Vidra (2018); Nicholas (2019)]. While subjected to the boom-bust cycles of the capital market at large, VCs have become, increasingly, where capital is chaneled towards technological innovation.

This sustained growth means VCs emerge from their predominantly U.S. roots towards global prominence, as expressed in capital flows, expertise, vision, and regulation. Contemporarily speaking, VCs tend to operate with transnational capital and deal flows. Investments into any given startup are likely to involve an investment syndicate consisting of cross-border investments [Bradley et al. (2019); Maula and Wunder (2023)]. As venture investment grows in both instances and sizes, more capital is required for VCs to effectively intervene in startups. This enhanced demand for deployed capital results in the attraction of capital beyond the domestic market, provided that the host nation allows the inflow of foreign capital [Klingler-Vidra (2018)].

Globalization is also present on the upstream of venture capital, as VCs need to raise funds before they can deploy them in startups. The institutions that provide VCs with liquidity (the limited partners, or LPs) can be university endowments, pension funds, sovereign wealth funds, family offices, or high-net-worth individuals, all from different countries. Practitioners in this space are equally diverse, which in turn facilitates the flow of capital and returns across national boundaries [Klingler-Vidra and Pardo (2025); Madhavan and Iriyama (2009)]. These dynamics lead to increased global dispersion of VC funds. At the turn of the century, nearly 80% of VC funds were headquartered in the U.S., this number is now less than 50% [National Science Foundation (2024)]. Other centers of VC investments, like the U.K., E.U. and China, grew in importance.

VC investments are often fundamentally transnational in their purpose, too: VCs invest in technologies that could potentially be transformative globally. In a practical sense, VCs need their startups to scale up as quickly as possible into as many target markets as possible;

this ensures a rapid growth of valuation/VC exit payout. VCs aim for their investments to transform into "unicorns" — companies valued at $1 billion and more. This is easier to achieve if the firms in question have global reach. In a rhetorical sense, VCs, through their investments into different technological projects, are proposing new imaginaries of human existence [Beckert (2016)]. The usual lingo of startups is that of radical change: solutions that will revolutionize how we view everything; products that change human life for the better; the world will never be the same with this service. Throughout these presentations, the intended audience and customer base for VCs' investments, with the notable exception of defense, are defined as large sectors of the world – in other words, global.

That VCs are global entities (based capital flows, expertise, and vision) does not mean that the field of venture capital should be understood as a space of global parity. On the contrary, VCs are extremely concentrated in select metropolitan locales, with Silicon Valley as the VC cluster par excellence [Ferrary and Granovetter (2009); Guzman and Stern (2015)]. Clusters facilitate the sourcing of available capital, the assemblage of VC syndicates, and the discoveries of promising startups. Prominent hubs are Silicon Valley and New York in the U.S., London in Europe, and Hong Kong and Singapore in Asia. This means that, instead of engaging with states and geopolitics on a philosophical level, VC funds and firms need to concretely navigate the specific regulations of these global clusters.

Sovereign regulatory regimes can be friendly to VCs or not, with differing characteristics. An example is to compare the relaxed governance in places like the Cayman Islands with stricter controls in China. Regulations on VCs and startups can also be recently established or even not established in a given jurisdiction, which enable VCs to strategically pursue business in locations with limited oversight. Further, regulatory regimes can propagate from one jurisdiction to others in a form of oversight globalization. For instance, the setup and structure of administrative guidance and legislation that enable the establishment of onshore private VC funds in Hong Kong references similar structures in the U.S. and the Cayman Islands [Eversheds Sutherland (2020)]. As such, VCs mitigate risks of overexposure to regulatory regimes of singular powerful states by electing to domicile in jurisdictions where the regulations are laxer, or where the authorities in question are unable and/or unwilling to commit to rigorous enforcement. VCs may elect to domicile in neutral locations rather than the U.S. or China, like the British Virgin Islands and the Cayman Islands. Beyond these neutral locations, each clusters have their own specificities. While the U.S. environment is conventionally understood as a neo-liberalistic [Klingler-Vidra (2018)], a closer look at the impact of regulatory principles on fund operations across their lifecycles reveals important differences that go beyond mere labels (see Table 1).

States also have a preference on what technologies are of national importance and what types of capital are allowed to intervene in the development of said technologies – with AI being the most pertinent demonstration. As the U.S. and China are the number one and two destinations of VC capital deployment [National Science Foundation (2024)], geopolitics also comes into play in more direct formats of bans, sanctions and trade wars. Due to these structures of venture investment within different states and geopolitical influences, we ask, how do states and geopolitics intervene in VC operations? What are the avenues of influence? How are the impacts of these avenues shifting with changing global politics?

Table 1: The impact of regulatory principles on venture fund hurdles across the lifecycle

Type	Formation	Fundraising and marketing	Operations	Portfolio and exit	Wind-down
Disclosure-based / market discipline (U.S.)	Medium (private contract)	**High** (offer exemptions + anti-fraud)	**High** (conflicts/ fees, controls)	Medium (MNPI, insider trading)	Medium
Manager-centric / prudential gatekeeping (E.U.)	**High** (manager authorization, governance)	**High** (passport/ NPPR constraints)	**Very high** (depositary/ oversight, reporting)	Medium	**High**
State-directed / industrial policy (China)	**Very high** (registrations, "who may manage")	**High** (who may invest/ raise)	**High** (ongoing supervision + penalties)	**Very high** (capital controls, sector rules)	Medium
Jurisdictional arbitrage / contractual freedom (Cayman)	**Very high** (vehicle engineering)	Medium (relies on "elsewhere" rules)	Medium (AML/ beneficial ownership trend)	Low	Medium
Hybrid / functional regulation (Hong Kong)	Medium	**High** (professional-investor regime and licensing trigger)	**High** (SFC conduct/ internal controls)	Medium	Medium

3. How states and geopolitics intervene

3.1 Uncertainties and risks

As discussed above, VCs deploy and operate in different national markets where the government has differing forms, regulatory regimes and priorities. Resultantly, some sovereign governments create environments conducive to VCs while others are less friendly. Regardless of whether the investment environment within a nation is accommodating or not, the states and geopolitics intervene in VC operations by creating uncertainties and risks that VCs have to account for. Given the power disparity between national governments and VCs, accounting for such uncertainties and risks are often impossible, but VCs would need to at least have an awareness of how these factors could potentially disrupt, redirect or even halt investment activities. Some of these impacts are currently known and are therefore factored in and quantified through firm valuations. These "known unknowns" (or risks) often have a probabilistic pathway towards fixed outcomes – A may have a 60% chance of happening as opposed to B's 40%, but the outcomes of the event will be either A or B. Other impacts in the future cannot be currently predicted or factored into consideration. These

impacts are often radical in nature, fundamentally unknowable in the present, yet pose the potential to hinder VC investments in startups. These are 'unknown unknowns' or uncertainties. The state and geopolitics constitute uncertainties and risks, which in turn impose costs and constraints (see Figure 1) to VC operations in turn through two primary avenues of intervention: capital influences and technological crackdowns.

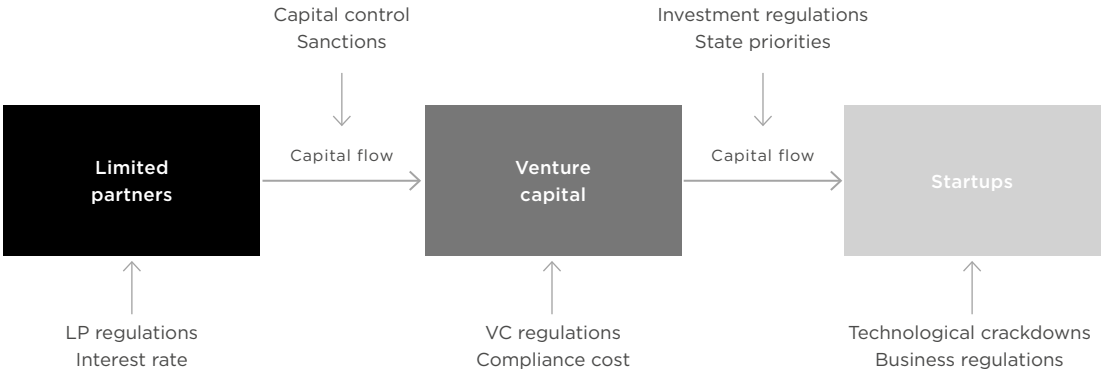
Since VC capital can be transnational, it wades into the issue of currency control. Though USD is the de facto unit of exchange in the VC space, the origins of VCs' assets-under-management are not exclusively received in dollars, but in a basket of major currencies depending on the nature and location of the LPs. Each VC fund has its specific nature of LPs and capital, thus becoming 'U.S.$ funds', 'RMB funds' and so on. Hence, the availability of capital and the return expectations are influenced by national monetary and fiscal policies impacting a single currency or a group of currencies in a geographical area. Naturally, with U.S. dollar being the global reserve currency, U.S. policies have a larger impact on more VC funds whose main currency of operation is that of U.S. dollar. A low-interest rate environment, as determined by policies from national central banks, is understood as inflating valuation as this VC analyst describes:

"In terms of valuation, in Asia it is very different from what it used to be [lower than previously – authors' note], because of the interest rate hikes… In the last 15 years, there's a lot of cheap powder in the industry. This is why the valuation figures pre-COVID is very high."

Interest rates are set through central banks to adjust inflation, yet their impacts constitute a risk for VCs as they depend on readily available capital from LPs. When the interest rate is low and capital ('dry powder') abundant, VCs can afford to commit to larger sums of investment into any given startup project, driving up the valuation of a company in a fashion that reflects the investment bull market. On the contrary, when rates are hiked, VCs need to be more conservative in their valuation lest they overpay. Since startup projects are inherently highly risky, the bets VCs make in each firm would be smaller when capital is scarcer, as LPs can locate other opportunities of capital gain at lower risks elsewhere in a high-interest rate environment.

If interest rates determined by sovereign governments contextualize startup valuation, regulations on capital flow intermediate the operations of VCs more directly. There are two aspects to consider here: capital control and

Figure 1: State-induced costs and constraints



administrative sanctions. Each state could impose in its jurisdiction a set of regulations that prevents the inflow/outflow of certain capital. China is a good example of capital control, as prior approval is needed for capital inflow and outflow through its Overseas Direct Investment framework, bringing further compliance requirements and costs for VCs. Sanctions are direct expressions of international diplomacy and government priorities. These are of rarer occurrence but more impactful, outright banning capital inflow or freezing assets already present within a jurisdiction. One need to only look at the sanctions imposed on Russian assets by the E.U. as an example.

Beyond the uncertainties and risks imposed on VC capital, the second aspect of state-related intervention is to do with the objects of VC investments – the startups. Since venture capital invests in new technologies, it is of crucial importance that these technologies are received by governments positively. A supportive government would facilitate R&D efforts and remove administrative and regulatory roadblocks for startups, whereas a hostile government could ban the development of a technology outright or impose extra costs to compliance so as to make it uneconomical. A good case in point here is blockchain and crypto. Blockchain technology makes the uncertainties and risks of technological crackdown pertinent, as it is in essence a solution against centralized authorities and intermediaries: blockchain uses decentralized networks to bypass controlled intermediaries like banks, undermining state control over transactions. As a result, states adopt different approaches to blockchain, and different parties within a state may have different views too. Given how important a friendly regulatory environment is to the growth of the startup, VCs have to take this into account:

"In the U.S., they have a party that hates cryptocurrency. Gary Gensler is heavily against this… Now that Trump is elected president, there are bullish signals [This interview was done in December 2024 – authors' note]. Donald Trump has been very pro crypto currency. We expect a 4-year bull run for crypto when Trump comes into office. We're looking forward to expand our business to America once a clearer regulatory framework emerges."

Here a VC makes a connection between national politics (the election of Trump) and the prospects of the blockchain technology ('crypto') he invests in. Trump would, in his words, provide a regulatory framework within which business can be done. Though he frames it as an opportunity, the regulatory environment is underlined by uncertainties and risks – just as Trump might create a 'bull run,' a change in electoral fortune could just as easily result in the 'crypto-hating' Democrats coming into power. VCs need to take these changes in national policy into account, as a failure in consideration could result in a substantial if not total loss of their investments. An example of a hostile jurisdiction to blockchain is that of China, who cracked down on blockchain startups, resulting in companies closing down or moving overseas. A failure to anticipate these potential policies lead to investment failure, as a VC who specializes in blockchain laments: "Our biggest failure so far is that we should have lowered investment in mainland China… The government has not been very supportive."

3.2 Private capital participants

Beyond constituting the regulatory environment within which VCs and their startups operate, governments also intervene in the cycles of investment more directly, by being the LPs, collaborative partners, or customers for startups.

> *Discourses of geopolitics seep into investment decisions. State power becomes an equivalence of startup legitimacy, on merit of the nation's strategic focus and resources rather than the merit of the project itself.*

Government-backed capital can provide capital to VCs to invest in their strategic preferences. These preferences can be centered around a specific sector, like semiconductors or AI, or geared towards attracting startups to a specific location. Instead of regarding VCs as a form of private capital devoid of state direct participation, it is important to recognize the often direct participatory inputs from governments in the formation of the venture startup market. Klingler-Vidra (2018) argues that in the U.S., the inception of the VC industry in the post-war era is, to a large extent, propped up by U.S. defense contracts that provide reliable buyers and income for startups. This dynamic is not restricted to the U.S. It is also pronounced in countries with drastically different political and regulatory environments. The Chinese government, through local governments and state-owned enterprises, has a significant capital presence in the Chinese VC industry as LPs. For VCs, this is a positive signal for their operational credibility. As VCs need to collaborate in syndicates, having the government stamp of approval is beneficial. Here, a VC discusses who he might collaborate with:

"[We consider] if [a startup's] lead investor is a legit investor, for instance we always go with XXX [VC name], because XXX is backed by the government, then we can choose to trust them."

Having the government onside is considered here as an indication of legitimacy, as it mitigates uncertainties surrounding the startup – if the government is happy to invest in this through a VC, they are unlikely to be adverse to its development. Of course, the inflow of state-backed capital comes with caveats and provisions for increased control over VCs, as an analyst who works for a VC with a government LP describes:

"They [The LP] are government owned, so they really want to prevent loss of national capital. This is very obvious… compliance and risks, they will have rigid control. For any document that we produce, the original need to stay with them, we only keep a scanned copy… They don't have expectation of how much we should make. They're also audited, quite rigorously. So an event of a loss of national capital would be a very bad thing for the one in charge."

As government LPs deploy their capital towards strategic means, the financial return is less relevant to them. However, VC agendas are challenged, since a prevention of national capital loss trumps the needs of VCs to seek risky projects for exponential returns. This goes against the division of labor in this field, where LPs provide the capital and VCs manage said capital autonomously. In so far as the government capital is concerned, they retain the ultimate right to accountability by VCs, who are expected to fulfill the strategic design of the state through their investments while not generating a loss to the capital invested, in a manner more akin to state capitalism as opposed to free-market capitalism within which VCs are usually understood to be positioned. In return, the expectations of state investments yielding substantial profits are tempered. Not all VCs would accept this trade-off, some VCs we interviewed explicitly said they would turn down government investment to retain their autonomy. Further, VCs who take government money are still required to generate a payout for their other LPs, since government capital is unlikely to be their sole source of funds, resulting in a balancing act between the agendas of the state and those of private capital.

The state also intervenes in startups before they reach the stage where they are invested in by VCs. Startup founders usually come from university laboratories, where they try to convert research results into startup companies. Before startups capture the attention of VCs, they often need to go through an 'incubation' stage, where a combination of university funding, government support, and angel investment intervenes to support the startup's inception. A good demonstration of this is Hong Kong, where the government funds major innovation support clusters – Cyberport and HKSTP – which collaborate with universities and provide incubation training and funds to promising projects. These supporting institutions operate in the same way as state-backed capital LLPs, but intervene earlier in a startup's lifecycle, as an employee laid out:

"XXX [cluster] is a government backed corporation. So, their [sic] mandate is to help Hong Kong, is to have some value add to Hong Kong economy, or to support local entrepreneurs. Of course, this is not to say that financial return is not important, because of course part of the mandate is to make a financial return. But we're not expecting 100 times return like other investment funds."

The important thing here is again the external agenda of the government – to support the economy, or to facilitate the growth of local entrepreneurs. These agendas are executed through a purposefully-built institution geared explicitly towards government priorities. This makes it more hands-on compared with intervening through VCs. However, this is not to say that direct government presence can substitute for VCs, since incubation's end-goal is to make startups VC-ready. VCs have a specific expertise in scaling companies that governments cannot replicate. All things being equal, a startup would still prefer VC investment:

"Of course, we [a government backed incubator] are not Y Combinator [a leading accelerator-VC hybrid in the industry – authors' note]. So in terms of reputation, having an investment from Y Combinator is obviously better. But it's still free money."

Startups need the cash, which is the selling point of these incubation programs. When VC expertise is available, these programs cannot compete.

From the discussions above, we propose the avenues through which state and geopolitics intervene in global venture capital: states engage with VCs at arm's length through constituting the uncertainties and risks, and directly through investments as LPs and incubators. These channels are long standing in the industry and need to be recognized as such. However, just like global geopolitics at large, these avenues are subject to change and new developments. What are these developments? What do they mean for the VC industry?

4. Recent changes in geopolitics for VCs

4.1 Heightened risks

When we first started doing fieldwork in late 2024, the talk of the town within VC circles was Donald Trump: whether he will come to power, and what that would mean to blockchain, AI, and the VC industry at large. In early 2025, Trump was sworn in as president. It did not take him long to act. He imposed sweeping tariffs on vast baskets of imports. The rationale is, in a sense, a protectionist one – tariffs serve as incentives for the revival of the U.S. domestic manufacturing industry. Following this line of logic, one can see how this renewed protectionist instinct plays out in the field of VCs and what they are investing in. The technologies VCs bet on can be regarded by

government entities as proprietary to themselves, or to hostile government overseas. The recent sale of TikTok in the U.S., a requirement imposed by the U.S. government, is a good demonstration for VCs on the situations of heightened risks. The TikTok sale is a result of negotiations not between private investors who buy and sell equity, but between American and Chinese national governments [Duffy (2025)]. As far as VCs are concerned, in this type of situation, the decision to sell is effectively taken out of their hands, leading to a fire sale of assets that cuts into profit margins. National, geopolitical considerations have become increasingly explicit in constituting investment uncertainties and risks. This is compounded by the overlapping of VC investment hotspots and government agendas, with AI being the most recent demonstration. Further, with tariffs and sanctions regularized by the U.S. administration as an effective medium to acquire negotiation leverage diplomatically, supply chains and capital flows, which startups and VCs depend on, become vulnerable. Markets can be closed to funds from a different nation. A Chinese VC describes the situation:

"Before 2021 and 2022, a lot of U.S.$ funds are deployed in China. Especially in the period after COVID, in the resurgence of China after COVID. But this is a bit of a bad vintage. Nowadays because of the U.S. China tension, the Chinese market is closed off to these U.S. funds. The US funds has [sic] a mandate to invest the money in APEC. But the Chinese market, which is 85% of the total APEC market, is closed off to them. So the viable option for them, because they have to deploy the money there, is either North Asia or India."

The VC partner proposed a direct relationship between geopolitical tension and fund access. This is a demonstration of geopolitical uncertainty wrongfooting VCs. Since the Chinese market consists of a large percentage of APEC, the outburst of tensions between U.S. and China invalidate the original design of some VC funds who are oriented towards deployment within APEC. These funds are forced to pivot to other locations with less ideal conditions and market size for the optimal generation of investment returns. The fluctuations in currency value due to these geopolitical tensions also cut into the total available capital VCs have access to:

"Trump has brought a lot of volatility to U.S. dollars and the U.S. economy. Our money comes primarily from Korea, so if you're talking about the act of exchanging Korean money into U.S. dollars, we might already lose 15% there already through currency exchange rates."

The trend for VCs is clear, they need to pay attention to geopolitics like never before, since it is becoming more centerstage to their operations and money sources. This increased importance of geopolitics in VC operations brings with it risks, but also opportunities. The recent boom of AI startups in China is, in some sense, a direct result of the competition between China and the U.S. with AI. In any case, geopolitics serves as a powerful structuring signal to VCs by heightening uncertainty and risks.

4.2 Strategic LP retrenchment

The tumultuous international market outlook also brings increased concerns for state-backed capital to protect their assets. As a result, barring selected strategic vectors of investment, we observed a retrenchment of active government deployment of capital via LP structures. This is of extra importance for startups and VCs operating in China, as the investment terms are institutionally structured to favor state capital interests. For instance, in standard term sheets in China, VCs with state-backed capital presence tend to insert buyback clauses that enable VCs

to recall their investments by imposing a duty of repayment onto startup firms, or the founders personally:

"Because national capital is involved, you cannot let it suffer losses, so you insert a buyback clause, or a repurchase clause on founders… It gets triggered and you have to pay back [your investors]… This is standard in the vast majority of cases, with the exception that you are a star team, star founder, then the terms might not be there."

This in itself is not a problem for founders. The startup industry works through processes of elimination, and if one fails to develop their firm to schedule, the investors are within their rights to call back their investments. The problem starts when state-backed capital calls back its investment due to reasons beyond that of startup performance. Several practitioners we interviewed in China described startups who took state-capital backing struggling to meet the buyback clause, which was triggered mainly by a major Chinese state owned enterprise (SOE) trying to recoup funds. Due to the power disparity between founders and their investors, founders are often unable to resist the insertion and activation of such clauses. However, this is not to say that this retrenchment impacts all investment vectors equally. Government capital continues to behave with strategic intent, imposing constraints on where deployment can be made. Here a practitioner describes how these governments' strategic impetus might play out:

"We're here to make money for investors, not to serve as a local industrial-policy tool. If I took money from, say, Shandong, and I later find a "DJI-level" company worth investing in, but it's not based in Shandong, I can't invest, because I must "reinvest" into Shandong. That's restrictive… They restrict the geographic scope. If the money is from Shenzhen government, they want you to support Shenzhen firms. Not firms in Shandong. Otherwise, the tax revenue, workforce, and industrial output all flow elsewhere."

Here the strategic consideration is the local economy, but the consideration can also be conducted on the national level. For instance, Chinese and U.S. governments directed funds into AI R&D, transforming the startup space into one akin to the Space Race in mid-20th century. In the consideration of VCs, the prospects of the startups VC invest in are then transformed into an expression of national resources:

"AI is really hot, but they're hot for people in U.S. and China and EU. If you see a Southeast Asian company that says that they're doing an LLM [large language model – authors' note]… Southeast Asia cannot come up with LLMs, only the European Union, China and U.S.A. can do that. If you see a company doing that in Southeast Asia, they are a scam. They just want to take your money."

Here we can see how the discourses of geopolitics seep into investment decisions. State power becomes an equivalent of startup legitimacy, on merit of the nation's strategic focus and resources rather than the merit of the project itself. In this sense, innovation as VCs know it – privately funded and privately owned – is fading in favor of a reconfigured model where a hybrid of state and private resources works in conjunction with each other to fund R&D costs. In the space of AI, VCs might be funding the startups, but the governments are underwriting the infrastructural costs of data centers, making chip supplies available, and creating a friendly regulatory framework. In summary, China, as participants of private capital, are increasing their footprints in the VC industry, but only for startups with strategic importance. In other vectors, state LPs are retrenching to hedge against geopolitical uncertainties.

5. Conclusion

This article delineated the two major avenues where states and geopolitics intervene in VC operations: as uncertainties/risks, and as participants of private capital. We argued that since VCs are global, they need to consider and respond to macro state-political interventions on top of financial-investment concerns. We pointed out two new developments in how geopolitics and the state intervene, where risks are heightened due to geopolitical uncertainties, while state capital pivots towards more strategic vectors. These developments have strategic implications for VCs, who need to evaluate not only how geopolitics shapes the macroenvironment in which they act (e.g., through currency risks, capital controls, tariffs, and sanctions), but also how (in)direct engagement with state actors as LPs impacts decisional autonomy and action capabilities (e.g., through state policies toward specific technologies).

Our findings indicate that states are intervening in the VC landscape in more pronounced and strategic manners, especially given the backdrop of U.S.-China trade and technology wars. As such, we primarily selected examples from the U.S., Mainland China and Hong Kong to demonstrate these dynamics. These findings go against the conventional understanding in academia, which until recently has assumed a shift away from state institutions in matters of innovation. We now see a resurgence of state-related dynamics in VC considerations and state direct interventions alongside VCs.

The reemergence of the state as a powerful mediator of VC operations require practitioners in the industry to reevaluate their positions and strategies both upstream and downstream. On the upstream side, capital availability and their associated agendas are changing. We discussed how government-backed LPs are retrenching in non-strategic sectors to safeguard national capital. For VCs, this provides both opportunities and challenges. The vector selection processes become more important for raising new VC funds. Beyond evaluating whether a vector holds technological and commercial promise, an enhanced understanding of government attitude is required: is this investment vector incorporated into geopolitical state competitions? Is government strategic funding readily available in this vector? What are the implications of accessing this funding? What avenues of oversight and control are VCs prepared to yield to government LPs in exchange for their liquidity?

On the downstream side, a situation of heightened risks and uncertainties persists over VC deployment into startups. This requires VCs to be more responsive to geopolitical shifts, either through operating more conservatively by committing to smaller bets, or through derisking attempts on operation scopes, like domiciling in neutral, low enforcement locales or diversifying holdings geographically. In situations where the states involve themselves directly in startups, VCs need to understand governments' intentions: is the state enabling private investment or competing with private investment in startups? VCs cannot compete with governments who intend to hold certain technologies as proprietary to the state and should therefore pivot. When governments collaborate with VCs, an understanding of the scope and duration of favorable government reception should be formulated by VCs: is state support conditional to external pressure? How long may these conditions of support last, and what are the factors that influence their support, like partisan politics or the replacement of the official in charge? In the event of a withdrawal of support, what are the backup plans for VCs? This developing reconfiguration and rebalancing of the startup investment market should be watched closely by practitioners and academics alike.

References

Beckert, J., 2016. Imagined Futures: Fictional Expectations and Capitalist Dynamics, Cambridge, MA: Harvard University Press.

Bradley, W. A., G. Duruflé, T. F. Hellmann, Wilson, K. E., 2019. Cross-Border Venture Capital Investments: What Is the Role of Public Policy?, Journal of Risk and Financial Management, 12(3), 112.

Duffy, C., 2025. TikTok Has Signed the Deal to Spin Off Its US Entity with American Investor Group, CNN, https://edition.cnn.com/2025/12/18/tech/tiktok-signs-us-sale-deal

Eversheds Sutherland, 2020. Part 1 – Overview of the Limited Partnership Fund: How Does Hong Kong Match Up to the Cayman Islands and Singapore?, https://www.eversheds-sutherland.com/en/asia/insights/part-1-overview-of-the-limited-partnership-fund-how-does-hong-kong-match-up-to-the-cayman-islands-an

Ferrary, M. and M. Granovetter, 2009. The Role of Venture Capital Firms in Silicon Valley's Complex Innovation Network, Economy and Society, 38(2), 326-359.

Guzman, J. and S. Stern, 2015. Where is Silicon Valley?, Science, 347(6222), 606-609.

Klingler-Vidra, R., 2018. Venture Capital and VC Policy, In the Venture Capital State: The Silicon Valley Model in East Asia. Cornell University Press pp. 48-62.

Klingler-Vidra, R. and R.P. Pardo, 2025. Startup Capitalism: New Approaches to Innovation Strategies in East Asia, Cornell University Press.

Madhaven, R. and A. Iriyama, 2009. Understanding Global Flows of Venture Capital: Human Networks as the 'Carrier Wave' of Globalization, Journal of International Business Studies, 40(8), 1241-1259.

Maula, M. and D. Wunder, 2023. Cross-Border Venture Capital Investments, In The Palgrave Encyclopedia of Private Equity, Springer International Publishing pp. 1-7.

National Science Foundation, 2024. Invention, Knowledge Transfer, and Innovation, U.S. National Science Foundation, https://ncses.nsf.gov/pubs/nsb20241/data/table/SINV-31

Nicholas, T., 2019. VC: An American History, Cambridge, MA & London: Harvard University Press.

Sun, L., 2025. China Launches Venture Capital Fund to Lead the Charge on Tech Investment, South China Morning Post, https://www.scmp.com/economy/china-economy/article/3337798/china-launches-venture-capital-fund-lead-charge-tech-investment

About Capco

Capco, a Wipro company, is a global technology and management consultancy shaping change in the financial services and energy industries. For almost 30 years, we have been trusted to help our clients adapt, transform and create long-term value across capital markets, banking, payments, insurance, wealth and asset management, and the energy and utilities sectors. Our deep industry expertise, partnership mindset and award-winning Be Yourself At Work culture are amplified by our strengths in advisory, technology, data and AI innovation. We support our clients to establish clear priorities, connect vision to value and deliver measurable impact when the stakes are highest.

Expert-led, AI-infused, impact-focused – we do not just respond to change, we help shape it. To learn more, visit www.capco.com or follow us on LinkedIn, Instagram, Facebook and YouTube.

Worldwide offices

APAC
Bengaluru – Electronic City
Bengaluru – Sarjapur Road
Bangkok
Chennai
Gurugram
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

MIDDLE EAST
Dubai

EUROPE
Berlin
Bratislava
Brussels
Edinburgh
Frankfurt
Geneva
Glasgow
London
Milan
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
Montreal
New York
Orlando
Toronto

SOUTH AMERICA
São Paulo
Rio de Janeiro

About King's Business School

King's Business School is a triple-accredited business school in the heart of London. It ranks amongst the top 10 business schools for research in the UK, based on the Research Excellence Framework 2021, and is consistently ranked in the top ten in the UK for business by the Times and Sunday Times Good University Guide, the Complete University Guide and the Times Higher Global University rankings.

The School advances world-leading education and research across four knowledge frontiers: (1) developing a modern workplace, (2) creating sustainable and socially responsible business, (3) Improving health and public services organisation and (4) strengthening economic and financial systems.

It is the ninth faculty of King's College London, leveraging interdisciplinary potential across the College. It drives pioneering thinking and engages the limitless energies of the city's businesses, policymakers, entrepreneurs and change-makers. King's Business School puts its commitment to drive positive change at the heart of its life-changing research and "London as a classroom" experiential education.

CAPCO.COM

CAPCO