Working Paper

A-Series Information:
WP-D examines governance mechanisms and institutional design in AI-mediated systems, focusing on how infrastructure, authority, and decision processes generate lock-in, coordination burden, and failure risk and how such risks can be structurally contained.

Recommended Citation:
Wu, Shaoyuan. (2026). *Why the South? Institutional Friction and the Spatial Reorganization of Data Center Infrastructure in the United States* (EPINOVA Working Paper No. EPINOVA–WP–D–2026–02). Global AI Governance and Policy Research Center, EPINOVA LLC. https://doi.org/10.5281/zenodo.18572133

GLOBAL AI GOVERNANCE RESEARCH CENTER

EPINOVA

Why the South?
Institutional Friction and the Spatial Reorganization of Data Center Infrastructure in the United States

Author: Shaoyuan Wu

ORCID: https://orcid.org/0009-0008-0660-8232

Affiliation: Global AI Governance and Policy Research Center, EPINOVA LLC

Date: February 09, 2026

Abstract

Despite common assumptions that large-scale data center infrastructure should concentrate in coastal technology hubs, low-tax jurisdictions, or energy-abundant regions, recent expansion in the United States follows a different spatial pattern. In the AI and hyperscale era, new large-scale development has increasingly clustered in the American South and selected interior regions. This paper addresses this puzzle by advancing a structural explanation centered on institutional feasibility rather than factor endowments alone.

The paper introduces a structural belt model and a three-axis framework—local institutional friction, utility buildability, and network interconnection—to explain why certain regions function as low-resistance infrastructure zones. It shows that data center expansion follows a logic of friction minimization across governance, permitting, and power delivery systems, producing contiguous corridors of growth rather than isolated point-optimal sites.

This spatial reorganization carries important governance implications. Rapid infrastructure lock-in in low-friction regions can outpace the adaptation of regulatory and political oversight, generating infrastructure–governance asymmetries. The findings contribute to debates on AI compute governance, infrastructure planning, and the political economy of large-scale digital systems.

Keywords

Infrastructure friction; Data centers; Institutional feasibility; Governance; Political economy; Artificial intelligence; Infrastructure lock-in; Spatial reorganization

1. Introduction

The rapid expansion of data center infrastructure is a defining feature of the AI and hyperscale computing era. As cloud services, large language models, and data-intensive applications proliferate, data centers have shifted from peripheral technical facilities to forms of critical infrastructure, increasingly entangled with energy systems, land-use governance, and regional political economies (Graham & Marvin 2001; Henderson et al. 2020; Crawford 2021).

Yet the geography of this expansion is increasingly counterintuitive. Conventional expectations suggest that data centers should concentrate in coastal technology hubs, low-tax jurisdictions, or regions endowed with abundant and inexpensive energy. Recent U.S. patterns diverge from these predictions. While coastal states continue to dominate demand generation, capital concentration, and corporate control, large-scale new construction has increasingly clustered in the American South and selected interior regions.

This paper addresses two related puzzles: why historically dominant technology centers such as California and New York have experienced a relative decline in new hyperscale build-out, and why the American South has emerged as a primary infrastructure-bearing region despite not being uniformly superior in energy costs, tax policy, or technological agglomeration.

We argue that these patterns cannot be explained by factor-based location theories alone. Instead, data center expansion is shaped by institutional conditions that determine the feasibility of large-scale deployment under compounded infrastructural constraints. Hyperscale facilities are capital-intensive and temporally locked-in assets whose realization depends not only on resource availability, but on the capacity of institutions to deliver approvals, power, and connectivity on predictable timelines.

This paper makes three contributions. First, it reconceptualizes data centers as lock-in infrastructure conditioned by governance arrangements as much as by economic inputs. Second, it introduces a structural belt model to explain why expansion occurs in contiguous low-resistance regions rather than isolated “optimal” sites. Third, it operationalizes this perspective through a three-axis framework, namely, institutional friction, utility buildability, and network interconnection, illustrated through comparative state-level cases in the United States. Together, these contributions provide a structural explanation for the emerging geography of hyperscale infrastructure and motivate an examination of its governance implications in the AI era.

2. Existing Explanations and Their Limits

Existing scholarship explains data center location through agglomeration economies, energy costs, and fiscal incentives. While these factors capture important dimensions of siting decisions, they do not fully explain the recent phase of U.S. expansion in which large-scale facilities have clustered across the American South and Southeast. More importantly, they offer limited leverage for explaining where capacity can be approved, built, and scaled rapidly, rather than where data centers might operate under idealized conditions.

2.1 Agglomeration Theory

Agglomeration-based accounts emphasize proximity to technology firms, financial centers, skilled labor, and global connectivity. Historically, coastal regions benefited from early investments in network infrastructure and dense corporate ecosystems, advantages that remain significant for innovation, demand coordination, and strategic control.

However, these advantages are less decisive for contemporary hyperscale facilities. Provided latency thresholds are met, large data centers no longer require proximity to headquarters, research institutions, or dense labor markets. Agglomeration effects increasingly operate at the level of demand generation and governance coordination rather than physical infrastructure siting (Henderson et al. 2020). Coastal regions continue to shape digital economies, but they no longer uniquely determine where new compute capacity is constructed.

2.2 Energy Cost and Resource Determinism

A second explanation emphasizes electricity prices and resource endowments. Given the energy intensity of hyperscale operations, regions with low-cost or abundant power are often treated as natural destinations. Yet, this view overstates marginal price effects while understating delivery and governance constraints.

Data centers are grid-dependent loads, not generators. What matters is not theoretical energy abundance, but the institutional capacity of utilities and regulators to deliver large increments of power on predictable timelines (IEA 2023; Armond & Manning 2023). Regions with low nominal prices but limited transmission capacity, prolonged interconnection queues, or regulatory uncertainty frequently fail to attract hyperscale investment. Energy availability is therefore mediated by institutional buildability rather than determined by resource endowment alone.

2.3 Incentives without Spatial Coherence

A third account highlights tax abatements and fiscal incentives. While such instruments can influence marginal siting decisions, their explanatory power is limited for hyperscale infrastructure. Incentives rarely compensate for delays, uncertainty, or governance conflict associated with permitting, zoning, and utility coordination.

Moreover, incentives are typically administered as discrete, jurisdiction-specific tools rather than as regionally continuous governance regimes. If tax competition were decisive, expansion would appear as scattered opportunistic sites rather than as geographically contiguous corridors spanning multiple jurisdictions. Incentives matter at the margin, but they do not explain the observed spatial coherence of recent hyperscale expansion.

Taken together, these explanations are not incorrect, but incomplete. They identify conditions under which data centers may operate efficiently, yet offer limited insight into where large-scale facilities can be approved, built, and scaled with minimal institutional resistance. Explaining contemporary spatial patterns therefore requires shifting attention from factor optimization to institutional feasibility.

3. Data Centers as Lock-In Infrastructure

To account for these patterns, this paper reconceptualizes data centers as lock-in infrastructure rather than flexible or footloose investment (Pierson 2000; Unruh 2000). This perspective clarifies why location decisions in this sector are unusually sensitive to institutional conditions and why expansion follows corridors of feasibility rather than isolated cost optima.

First, data centers are highly capital-intensive. Hyperscale facilities require large, sunk investments in land, power infrastructure, cooling systems, and network integration, rendering relocation or repurposing economically infeasible once built.

Second, they exhibit pronounced temporal lock-in. Designed to operate over multi-decade horizons, facilities embed long-term commitments to local energy systems, regulatory regimes, and surrounding communities. Early siting decisions thus shape regional infrastructure trajectories well beyond initial investment cycles.

Third, data centers are deeply governance-dependent. Their feasibility hinges on zoning classifications, permitting timelines, utility coordination, and regulatory stability. Unlike mobile capital, hyperscale facilities cannot easily exit or renegotiate unfavorable institutional environments once construction begins.

Together, these properties make data center investment acutely sensitive to institutional friction. Under conditions of escalating capital commitments and compressed delivery timelines, predictability becomes a binding constraint. As a result, expansion follows a logic of friction minimization rather than pure cost optimization, prioritizing jurisdictions where approvals, power, and connectivity can be secured with temporal and regulatory certainty.

4. From Lock-In to Spatial Continuity: The Structural Belt Model

Building on this lock-in perspective, the paper proposes a structural belt model of data center expansion. Rather than selecting isolated sites that optimize a single factor, hyperscale facilities cluster along contiguous regions where institutional resistance remains consistently low across jurisdictions.

Low-friction infrastructure zones are not defined by any single advantage such as low energy prices or generous incentives. They emerge from the alignment of governance processes, utility expansion capacity, and network accessibility in ways that minimize delay, uncertainty, and coordination costs. What distinguishes such zones is not maximal efficiency, but predictable implementability at scale.

These zones take the form of belts rather than points because the infrastructures on which data centers depend electric grids, long-haul fiber backbones, and logistics corridors are themselves spatially continuous (Graham & Marvin 2001). Once large-scale deployment becomes feasible in one jurisdiction, adjacent areas with comparable institutional arrangements become natural extensions, producing corridor-based agglomeration even in the absence of strong local demand.

Notably, the structural belt is not permanent. As infrastructure accumulates, governance pressures intensify, often prompting institutional recalibration through zoning reform, utility regulation, or public contestation. Rising friction can shift the belt outward, making the model inherently dynamic and linking spatial expansion directly to evolving governance responses.

5. Operationalizing Feasibility

To operationalize the structural belt model, the paper introduces a three-axis analytical framework capturing the institutional and infrastructural conditions that shape large-scale data center siting. Rather than privileging single-factor explanations such as energy prices, tax incentives, or proximity to demand centers, the framework emphasizes institutional capacity under conditions of scale, time pressure, and infrastructural irreversibility.

5.1 The Three-Axis Feasibility Framework

(a) Axis A: Local Institutional Friction

Axis A captures local institutional friction associated with zoning regimes, permitting procedures, litigation risk, and the politicization of land-use decisions. Crucially, institutional friction reflects uncertainty rather than regulatory stringency per se. What matters is the predictability of approval timelines, the likelihood of procedural escalation through discretionary review or public hearings, and the risk of post hoc rule changes.

Low-friction environments are characterized by by-right development pathways, administrative approvals, limited litigation exposure, and stable interpretations of land-use rules. High-friction environments involve multi-stage discretionary review, mandatory public participation with veto potential, and heightened political contestation. For capital-intensive and time-sensitive infrastructure such as hyperscale data centers, institutional friction translates directly into execution risk.

(b) Axis B: Utility Buildability

Axis B measures utility buildability: the capacity of electric utilities to deliver large increments of power on predictable timelines. This dimension extends beyond nominal energy abundance or average electricity prices to encompass grid capacity, interconnection procedures, transmission and substation expansion, and the institutional willingness of utilities and regulators to treat data center load as a primary planning variable.

High buildability characterizes environments in which utilities proactively expand generation and grid infrastructure under regulatory frameworks that support long-term planning and cost recovery. By contrast, regions with abundant energy resources but constrained interconnection queues, prolonged approval processes, or regulatory hesitation score poorly on this axis regardless of headline energy costs.

(c) Axis C: Network Interconnection

Axis C captures the quality and scalability of network interconnection, including access to long-haul fiber backbones, carrier diversity, internet exchange points, and mature interconnection ecosystems. While hyperscale facilities no longer require proximity to coastal landing points or technology clusters, they depend on robust, redundant, and expandable network integration.

Importantly, network interconnection is not a binary condition. What matters is replicability: the ability to extend comparable network quality across multiple sites within a region. Regions that support scalable replication rather than isolated connectivity peaks are therefore structurally advantaged for corridor-based hyperscale expansion.

5.2 Dynamic Coefficients and Phase-Sensitive Weighting of Feasibility

The baseline formulations of the **Growth Attractiveness Index (GAI)** and the **Infrastructure–Governance Asymmetry Pressure Index (IGAP)** employ fixed weights to represent average structural conditions across U.S. states. In practice, however, data center expansion is not a stationary process. The relative importance of institutional friction, utility buildability, and network interconnection shifts systematically across development phases, levels of saturation, and cycles of regulatory response.

These shifts do not alter the substantive meaning of the three axes defined in Section 5.1. Instead, they change the relative salience of each axis as constraints evolve. To capture this temporal variation while preserving the underlying structural logic, the framework introduces dynamic coefficients that reweight the axes endogenously as infrastructural conditions and governance responses change.

This phase-sensitive weighting allows the framework to move beyond static feasibility comparisons and to represent how growth attractiveness and governance pressure co-evolve over time, setting the foundation for the dynamic formulations that follow.

5.2.1 Growth Attractiveness Index

(a) Baseline (Static) Form:

To capture the baseline attractiveness of a jurisdiction for data center growth, we define the Growth Attractiveness Index (GAI) as a weighted composite of utility capacity, network connectivity, and inverse institutional friction:

$$\boldsymbol{GAI = \frac{0.45 \times B + 0.35 \times C + 0.20 \times (5 - A)}{5} \times 100}$$

(5.2.1.1)

where A∈[0,5] denotes local institutional friction (higher values indicate greater resistance), and (5−A) therefore represents institutional permissiveness.

(b) Dynamic Form:

To account for phase-specific constraint dominance, the static weights are generalized into time-varying structural coefficients:

$$\boldsymbol{D\text{—}GAI_t = \frac{\alpha_t \times B_t + \beta_t \times C_t + \gamma_t \times (5 - A_t)}{5} \times 100}$$

(5.2.1.2)

with:

$$\boldsymbol{\alpha_t + \beta_t + \gamma_t = 1}$$

This formulation preserves interpretability while allowing the relative importance of infrastructure and governance dimensions to shift across development phases.

Coefficient Interpretation

- $\boldsymbol{A_t}$: Local Institutional Friction at time *t*, including zoning rigidity, permitting discretion, litigation risk, and political contestation.
- $\boldsymbol{B_t}$: Utility Buildability at time *t*, reflecting available power capacity, substation access, and grid expansion feasibility.
- $\boldsymbol{C_t}$: Network Interconnection at time *t*, capturing carrier density, IX access, and latency-sensitive connectivity.
- $\boldsymbol{\alpha_t}$: Utility Dominance Coefficient. Increases when:
    - Regional load growth exceeds historical baselines;
    - Utilities publicly identify data centers as primary load drivers;
    - Interconnection queues or transmission bottlenecks dominate project timelines.

  Structural meaning: Power delivery becomes the binding constraint; other advantages are secondary.

- β_t: Network Saturation Coefficient. Increases when:
    - Latency sensitivity rises (e.g., AI inference clustering);
    - Carrier density or IX capacity becomes differentiating;
    - Regions approach the frontier of interconnection performance.

  Structural meaning: Expansion shifts from "can we connect" to "can we replicate high-quality connectivity."
- γ_t: Friction Sensitivity Coefficient. Increases when:
    - Zoning reforms, moratoria, or litigation events emerge;
    - Public hearings gain veto power;
    - Regulatory uncertainty lengthens approval timelines.

  Structural meaning: Institutional resistance becomes a first-order risk factor.

Table 1. Illustrative Phase-Specific Weighting Regimes

Development Phase	α_t	β_t	γ_t	Dominant Constraint
Early Expansion	0.50	0.30	0.20	Power availability
Rapid Scaling	0.45	0.35	0.20	Power + network
Mature Saturation	0.35	0.30	0.35	Governance friction
Post-Backlash	0.30	0.25	0.45	Institutional control

5.2.2 Infrastructure–Governance Asymmetry Pressure Index

While GAI captures growth attractiveness, it does not measure governance strain. To address this, we introduce the Infrastructure–Governance Asymmetry Pressure Index (IGAP), which measures the degree to which infrastructure expansion outpaces institutional mediation capacity.

(a) Baseline (Static) Form:

$$IGAP = 100 \times [0.50 \times (5 - A)^{-1} + 0.30 \times B + 0.20 \times C] \tag{5.2.2.1}$$

(b) Dynamic Form:

$$D - IGAP_t = 100 \times [\delta_t \times (5 - A_t)^{-1} + \epsilon_t \times B_t + \zeta_t \times C_t] \tag{5.2.2.2}$$

with:

$$\delta_t + \epsilon_t + \zeta_t = 1$$

Working Paper

Coefficient Interpretation

- δ_t: Governance Lag Amplifier. Increases when:
    - Expansion speed outpaces regulatory adaptation;
    - Infrastructure is approved under legacy zoning regimes;
    - Political response is reactive rather than anticipatory.

  Meaning: Low friction no longer represents efficiency, but governance exposure.
- ϵ_t: Scale Acceleration Coefficient, capturing how power capacity amplifies lock-in effects by accelerating irreversible infrastructure commitments.
- ζ_t: Network Lock-In Coefficient, reflecting sunk interconnection investments that constrain exit options and spatial flexibility.

(c) Interpretive Rule:

IGA pressure rises fastest not where infrastructure is largest, but where it is fastest and least institutionally mediated.

This explains why jurisdictions such as Loudoun County have responded to rising IGA pressure by intentionally increasing Axis A—raising friction through discretionary approvals and public hearings to restore governance balance.

5.2.3 Structural Dynamics vs. Empirical Tuning

It is essential to emphasize that these dynamic coefficients do not function as regression-derived weights or post hoc calibration parameters. Instead, they operate as:

- Phase-sensitive structural modifiers;
- Triggered by observable institutional transitions;
- Interpretable in governance terms;
- Comparable across jurisdictions and over time.

The framework therefore supports scenario analysis rather than point prediction, capturing not only where data centers concentrate, but when and why the dominant drivers of concentration shift (Ansell et al. 2017).

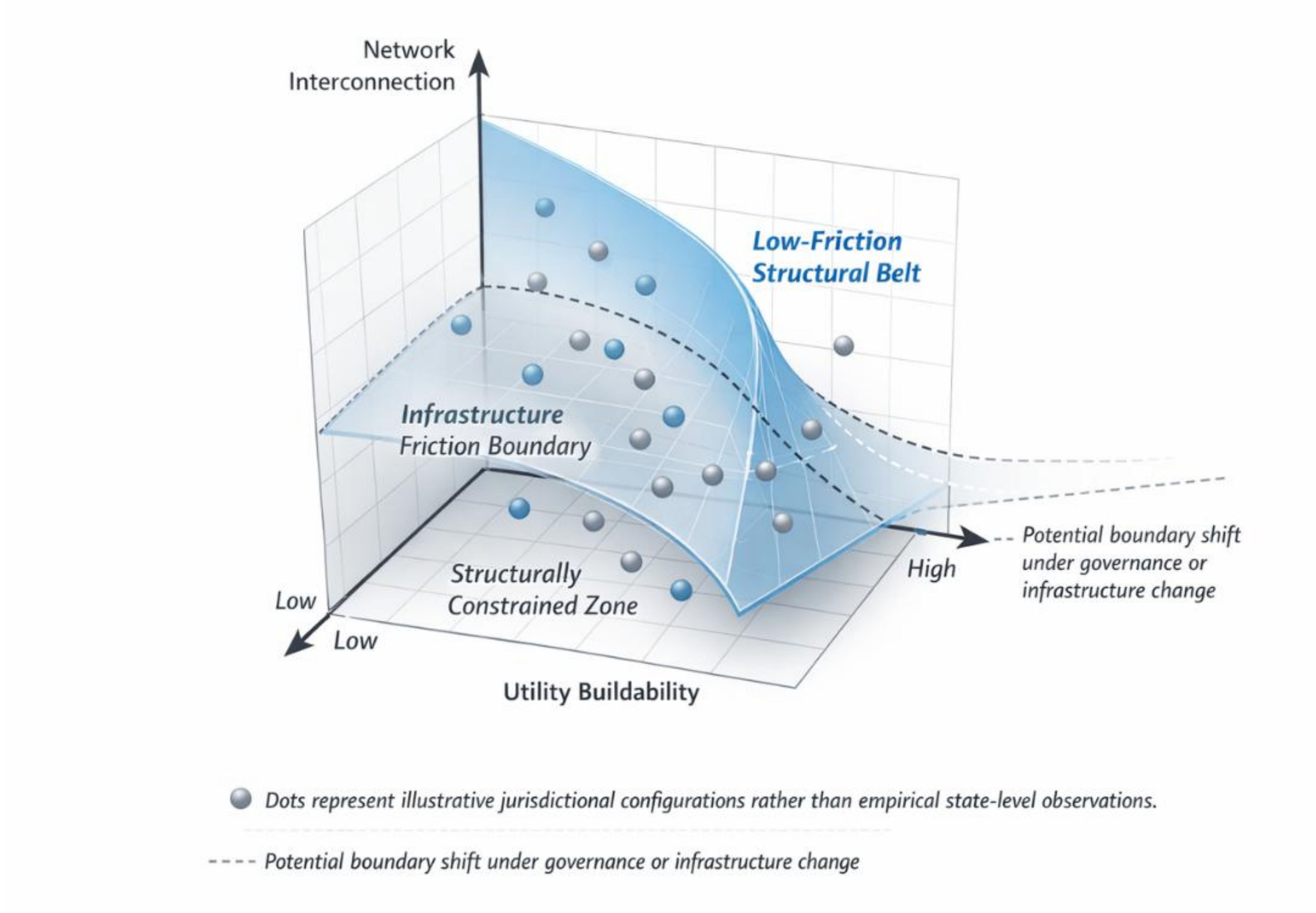

Figure 1. Three-Axis Infrastructure Friction Boundary

Conceptual illustration of the boundary separating low-friction structural belts from structurally constrained zones based on institutional friction (A), utility buildability (B), and network interconnection (C). Points indicate illustrative jurisdictional positions rather than empirical state-level observations. The boundary is dynamic and may shift as governance and infrastructure conditions evolve.

5.3 The Infrastructure Friction Boundary

To synthesize the three-axis framework into a spatially interpretable construct, this paper introduces the **Infrastructure Friction Boundary (IFB)**. The IFB represents a composite feasibility threshold separating jurisdictions capable of rapid hyperscale data center deployment from those structurally constrained by institutional, utility, or network resistance.

Importantly, the IFB is not a geographic latitude line nor a historically fixed regional boundary. It is a **phase boundary in institutional–infrastructural space**. Its apparent north–south manifestation in the United States reflects the spatial projection of underlying governance and infrastructure gradients rather than cultural or climatic divisions.

Formally, each jurisdiction *i* is represented as a point in three-axis space:

$$\boldsymbol{S_i} = (\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{C_i})$$

(5.3.1)

where:

- $\boldsymbol{A_i}$ denotes local institutional friction;
- $\boldsymbol{B_i}$ denotes utility buildability;
- $\boldsymbol{C_i}$ denotes network interconnection capacity.

The Infrastructure Friction Boundary is defined as the locus of points satisfying a composite feasibility condition:

$$\boldsymbol{\alpha_t \times B_i + \beta_t \times C_i + \gamma_t \times (5 - A_i) = \Theta_t}$$

(5.3.2)

where Θ_t represents the minimum feasibility threshold for scalable hyperscale deployment under prevailing institutional and infrastructural conditions, and α_t, β_t and γ_t are the same phase-sensitive coefficients defined in Section 5.2.

Jurisdictions above this threshold constitute the low-friction structural belt, while those below remain expansion-constrained.

Crucially, the IFB is dynamic. As agglomeration intensifies, institutional friction rises endogenously, shifting the boundary outward and inducing spatial diffusion toward adjacent low-friction zones.

The IFB is analytically descriptive rather than prescriptive and does not imply that jurisdictions should seek to position themselves above or below the boundary.

5.4 The Structural Relationship between GAI, IGAP, and IFB

GAI, IGAP, IFB are analytically distinct yet structurally interdependent components of the feasibility framework. At a conceptual level, GAI represents feasibility pull, IGAP captures the accumulation of governance stress, and the IFB functions as a regime boundary separating expansion-dominant from governance-dominant states. Although all three are derived from the same three-axis space—local institutional friction, utility buildability, and network interconnection—they operate at different analytical levels and serve distinct explanatory purposes.

GAI captures the directional logic of infrastructure expansion. It measures the extent to which a jurisdiction's institutional and infrastructural configuration enables rapid, predictable, and scalable deployment of hyperscale capacity. High GAI values indicate environments in which capital and compute loads are likely to concentrate due to low execution friction and delivery uncertainty. As a growth-oriented index, GAI explains where expansion is feasible and attractive, not whether it remains institutionally sustainable over time.

IGAP, by contrast, measures exposure rather than feasibility. It captures the accumulation of governance stress that arises when infrastructure lock-in advances faster than the adaptive capacity of regulatory, fiscal, and political institutions. IGAP is therefore not inversely related to GAI. During early and rapid expansion phases, feasibility pull and governance pressure often intensify simultaneously, reflecting a structural divergence between deployment speed and institutional mediation.

The IFB formalizes the interaction between these dynamics. It is not an index threshold or a fixed geographic line, but a phase boundary in institutional–infrastructural space. Below the boundary, low institutional friction primarily enables rapid deployment. Beyond it, rising friction increasingly reflects endogenous governance responses, such as discretionary permitting, regulatory tightening, or political intervention, aimed at restoring institutional control and temporal alignment.

In this sense, the IFB should be understood as an **emergent structural condition rather than a policy-defined cutoff**. It delineates a transition between two regime states: one in which feasibility advantages dominate spatial outcomes, and another in which governance constraints reassert themselves as first-order determinants. Figures 1 and 2 illustrate this relationship in abstract feasibility space and in its stylized projection onto the United States, respectively.

Taken together, GAI, IGAP, and the IFB allow the framework to explain not only where infrastructure expands, but **why expansion exhibits spatial continuity, phase-specific acceleration, and eventual governance backlash**. The boundary does not halt growth; it redirects it, inducing outward diffusion toward adjacent jurisdictions where feasibility conditions realign under lower accumulated governance pressure.

6. Empirical Illustration: U.S. States

Rather than relying on large-N econometric analysis, this section illustrates the three-axis feasibility framework through comparative state-level cases. The objective is structural validation rather than statistical generalization: to show how different configurations of institutional friction (Axis A), utility buildability (Axis B), and network interconnection (Axis C) generate systematic spatial outcomes within a shared national market for hyperscale infrastructure.

The illustrations are intentionally stylized. Figure 2 does not represent a regression-derived boundary or estimated frontier. Instead, the Infrastructure Friction Boundary (IFB) is projected onto U.S. space using the phase logic developed in Sections 5.2 and 5.3. Jurisdictions experiencing rising institutional friction at comparable levels of utility and network feasibility are positioned on one side of the boundary, while adjacent regions in which the three axes realign under lower accumulated governance pressure are positioned on the other. The figure is interpretive rather than inferential, designed to visualize structural differentiation rather than estimate causal magnitude.

6.1 Analytical Logic: From Core Saturation to Belt Expansion

The cases form a stylized sequence of infrastructure evolution rather than independent observations. The sequence traces a progression from mature cores experiencing rising institutional friction, to adjacent low-friction extension zones, and finally to large multi-node expansion regions. This ordering reflects a core mechanism of the framework: as infrastructure concentrates, institutional friction rises endogenously, reweighting Axis A and inducing outward diffusion toward jurisdictions where feasibility conditions realign.

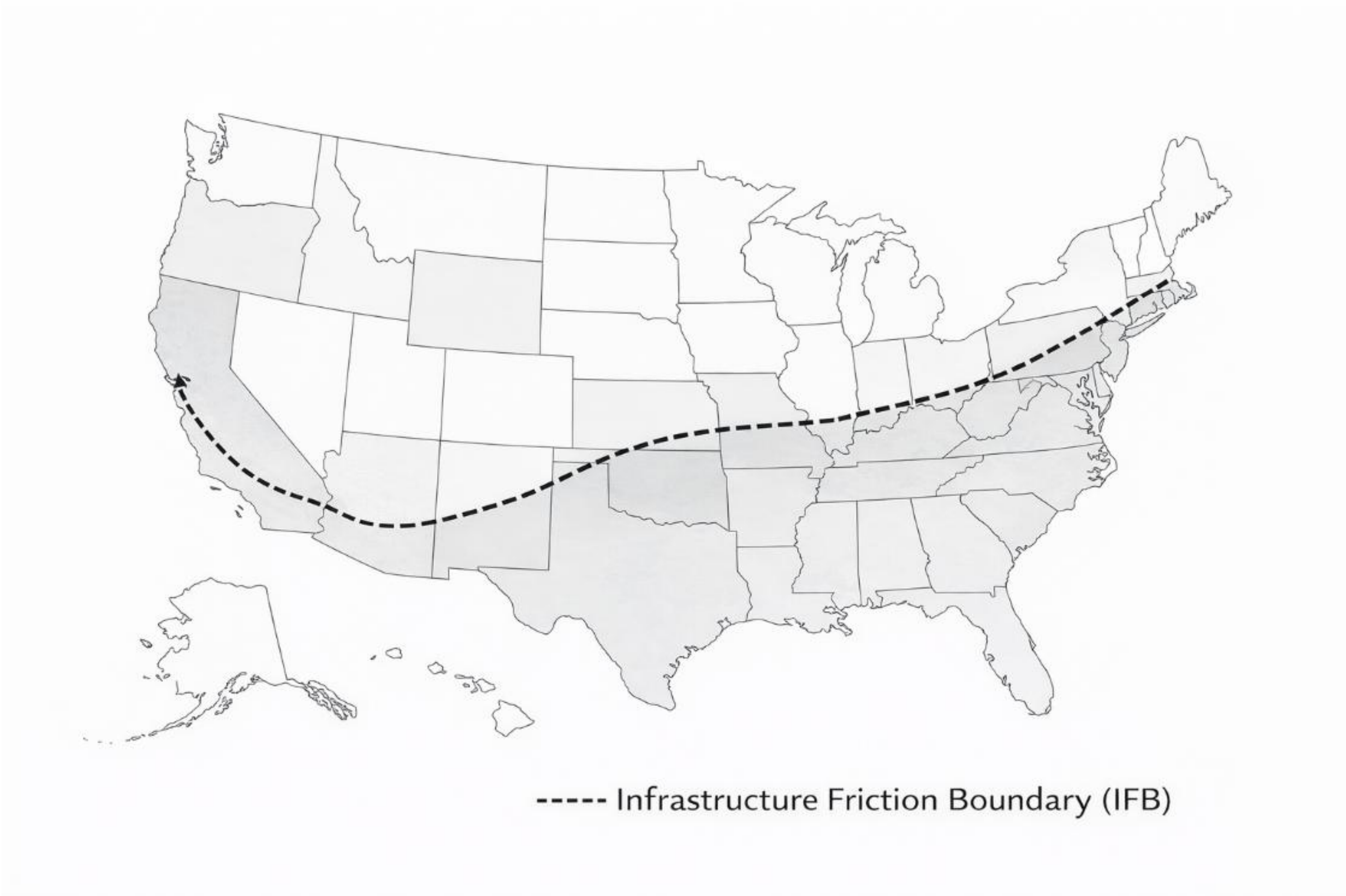

Figure 2. Projecting the Infrastructure Friction Boundary onto the United States

Stylized projection of IFB onto U.S. geographic space. The apparent north–south orientation reflects the spatial manifestation of underlying institutional, utility, and network feasibility gradients, rather than a fixed geographic, political, or historical dividing line.

6.2 Mature Core under Rising Friction: Northern Virginia

Northern Virginia, centered on Loudoun County, represents the most mature data center core in the United States. The region combines exceptional network interconnection (Axis C) with historically high utility buildability (Axis B) supported by long-standing grid investment. However, sustained concentration has been accompanied by rising institutional friction (Axis A). Approval regimes have shifted from largely administrative, by-right processes toward more discretionary pathways involving public hearings, special exceptions, and heightened political scrutiny.

This case illustrates a central dynamic of the framework: agglomeration generates friction. As scale increases, political salience and governance sensitivity rise, constraining marginal expansion despite continued strength along Axes B and C.

6.3 Low-Friction Extension Zones: Georgia (Atlanta)

Georgia, particularly the Atlanta metropolitan region, exemplifies a low-friction extension zone adjacent to a saturated core. The region offers robust and scalable network interconnection (Axis C) while avoiding the congestion and politicization characteristic of legacy hubs. Utility buildability (Axis B) has been comparatively strong, with utilities and regulators treating data center demand as a core planning variable and expanding generation, transmission, and substations accordingly.

Local institutional friction (Axis A) remains moderate and predictable. Approval processes are largely administrative, timelines are legible, and litigation risk is limited. As a result, Georgia scores strongly across all three axes, positioning it as a central node within the southern structural belt.

6.4 Multi-Node Expansion and Coordination Limits: Texas

Texas represents a distinct variant within the structural belt: a large, multi-node expansion region rather than a single metropolitan cluster. In principle, the state exhibits high utility buildability (Axis B), supported by substantial generation capacity and a history of accommodating industrial-scale loads. However, the relative isolation of the ERCOT grid and recent stress events reveal coordination limits associated with rapid load growth.

While local institutional friction (Axis A) often remains low, system-level governance challenges increasingly constrain expansion in certain subregions. This case underscores that utility buildability is dynamic rather than fixed and depends on institutional capacity to integrate scale over time, reinforcing the relevance of the phase-sensitive coefficients introduced in Section 5.2.

6.5 Demand Centers without Expansion Capacity: California

California provides a contrasting case in which demand, capital, and technological leadership remain dominant, but large-scale physical expansion has slowed. The state offers strong network access (Axis C) and proximity to major technology firms, yet performs poorly on the combined alignment of the three axes. Institutional friction (Axis A) is high, driven by complex permitting regimes, politicized land-use decisions, and frequent litigation, while utility buildability (Axis B) is constrained by grid congestion, interconnection delays, and regulatory caution toward new large loads.

As a result, California remains a command and innovation center rather than a primary site of new hyperscale deployment. The case highlights a core distinction of the framework: being a demand center is not equivalent to being an infrastructure expansion zone.

6.6 Why the Structural Belt Tilts South and Southeast

The structural belt identified in this study is not concentrated in the American South because the region dominates along any single dimension. Rather, the South and Southeast are among the few areas in which all three feasibility axes align simultaneously. Throughout this paper, “the South” is used strictly as a structural shorthand denoting a contiguous feasibility regime, not a cultural or historical region.

Across much of this zone, utility buildability (Axis B) is relatively strong, institutional friction (Axis A) remains lower and more predictable than in mature cores, and network interconnection (Axis C) is sufficiently robust and replicable across multiple sites. Together, these conditions generate a functional spatial division consistent with the Infrastructure Friction Boundary described in Section 5.3 and visualized in Figures 1 and 2. Coastal regions continue to serve as centers of demand and strategic control, while the South and Southeast operate as infrastructure-bearing zones. The observed pattern is therefore best understood as a composite feasibility outcome under multiple institutional constraints rather than the result of any single optimizing factor.

7. Governance Implications

Low-friction zones enable rapid deployment of large-scale data center infrastructure, but the same feasibility advantages generate distinct governance challenges. As illustrated in Figure 1, regions located below the IFB, where institutional friction is low and utility and network conditions are favorable, experience accelerated infrastructure lock-in. In such environments, investment and construction timelines routinely outpace the adaptive capacity of regulatory, fiscal, and environmental oversight mechanisms.

This dynamic produces an infrastructure–governance asymmetry. Jurisdictions that host the physical footprint, energy load, and environmental externalities of hyperscale data centers may exercise limited effective control over their cumulative scale or long-term trajectory. Importantly, this asymmetry does not reflect weak governance institutions, but a temporal mismatch between fast-moving infrastructure deployment and slower democratic, regulatory, and planning processes.

The IFB clarifies that this condition is structural rather than incidental. The boundary separating low-friction belts from constrained zones represents a feasibility threshold, not a normative optimum. Crossing this threshold accelerates delivery speed while simultaneously increasing governance exposure by compressing opportunities for iterative review, cumulative impact assessment, and political recalibration.

When projected onto U.S. space (Figure 2), this mechanism acquires a distinct spatial expression. The concentration of recent data center expansion in the South and Southeast reflects the geographic projection of institutional feasibility gradients rather than a fixed regional divide. These regions function as infrastructure landing zones because their institutional arrangements permit rapid scale-up. Over time, however, the same permissiveness intensifies pressure on zoning regimes, utility regulation, and public accountability frameworks.

From this perspective, zoning reforms, expanded public participation, and enhanced utility oversight should not be interpreted solely as obstructions. Instead, they operate as tools of temporal alignment—mechanisms that deliberately introduce governance friction to synchronize infrastructure expansion with democratic oversight, fiscal planning, and environmental management (Mazzucato 2018; Ansell et al. 2017). The endogenous rise in institutional friction observed in mature hubs thus represents a governance response to sustained pressure below the feasibility boundary.

The implications for AI compute governance are substantial. Effective control over the physical substrate of AI systems increasingly resides in subnational jurisdictions where infrastructure is built, permitted, and operated, rather than exclusively in national centers of technological design or policy formulation. Governance capacity therefore follows infrastructure geography. In the AI era, effective compute governance depends not only on national technology policy, but on subnational capacity to manage scale, speed, and lock-in at the sites where compute physically lands.

8. Conclusion

This paper examined the spatial reorganization of data center infrastructure in the United States through the lens of institutional feasibility rather than conventional location advantages. Addressing why recent large-scale expansion has disproportionately concentrated in the American South and Southeast, the analysis showed that this pattern cannot be explained by coastal agglomeration, energy abundance, or fiscal incentives alone.

The paper's primary contribution is conceptual. By introducing a three-axis framework—local institutional friction, utility buildability, and network interconnection—it reframes data center siting as a structurally constrained process in which feasibility emerges only when multiple institutional and infrastructural conditions align simultaneously. The concept of an Infrastructure Friction Boundary further formalizes this insight by identifying a dynamic threshold separating regions capable of rapid hyperscale deployment from those constrained by governance or delivery limits. Crucially, this boundary is not geographically fixed, but shifts endogenously as infrastructure accumulates and institutional responses evolve.

From a governance perspective, the analysis highlights a systematic tension between deployment speed and regulatory adaptation. Regions below the Infrastructure Friction Boundary facilitate rapid infrastructure lock-in, but are also more likely to experience infrastructure–governance asymmetry, in which cumulative impacts outpace mechanisms of democratic oversight and institutional recalibration. This finding suggests that governance friction should not be understood solely as inefficiency, but as a potential instrument for temporal alignment between infrastructure expansion and public accountability.

The findings also underscore that the governance of AI systems cannot be separated from the governance of the physical infrastructure on which they rely. As hyperscale compute increasingly underpins advanced AI capabilities, effective oversight depends not only on national technology policy, but on subnational institutional arrangements governing land use, power delivery, and infrastructure siting. In this sense, AI governance is increasingly shaped by the geography of feasibility rather than the geography of innovation (Crawford 2021; Armond & Manning 2023).

Overall, the paper suggests that contemporary infrastructure geography reflects not the optimization of isolated factors, but the management of compound institutional constraints. Understanding where infrastructure expands therefore requires attention to where governance can keep pace with scale.

References

Ansell, C., Trondal, J., & Øgård, M. (2017). Turbulence and new institutionalism: Governance in times of turbulence. *Governance*, *30*(1), 3–22. https://doi.org/10.1111/gove.12217

Armond, C., & Manning, S. (2023). Governing hyperscale data centers: Energy, land, and regulatory politics. *Energy Research & Social Science*, *98*, 102987. https://doi.org/10.1016/j.erss.2023.102987

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Graham, S., & Marvin, S. (2001). *Splintering urbanism: Networked infrastructures, technological mobilities and the urban condition*. Routledge.

Henderson, J., Dicken, P., Hess, M., Coe, N., & Yeung, H. W.-C. (2020). Global production networks and the globalisation of data centres. *Geoforum*, *108*, 1–13. https://doi.org/10.1016/j.geoforum.2019.10.012

International Energy Agency. (2023). *Data centres and data transmission networks*. IEA. https://www.iea.org/reports/data-centres-and-data-transmission-networks

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, *94*(2), 251–267. https://doi.org/10.2307/2586011

Unruh, G. C. (2000). Understanding carbon lock-in. *Energy Policy*, *28*(12), 817–830. https://doi.org/10.1016/S0301-4215(00)00070-7