

Beyond Copy-Paste Regulation: Innovation-Weighted Risk Classification for AI Governance in Emerging Economies

Egemen Bora Tunçarslan^{1,2}, Ece Melisa Şerik^{1,3}, Samet Çayır^{1,2}

¹TOBB Gençol Turkish AI Law and Governance Workshop, Türkiye

²Maltepe University, Istanbul, Türkiye

³Beykoz University, Istanbul, Türkiye

egementuncarslan@gmail.com, ecemserik@gmail.com, samedbatur02@gmail.com

Abstract

The EU AI Act will begin full enforcement for high-risk applications in August 2026. For emerging economies, this presents a tough choice: adopt a regulatory framework mainly aimed at large firms with strong compliance resources, or delay action, leaving citizens vulnerable to unregulated algorithmic decision-making. Our team decided to explore whether a third option might exist. Over ten months, researchers from computer engineering, software engineering, law, medicine, and international trade examined how AI governance works in seven sectors: Fintech, healthcare, education, energy, media, logistics, and AI literacy and compared their findings with six national regulatory approaches: the EU, US, UK, Singapore, South Korea, and Japan.

We found a structural mismatch in how risk categories are applied. The EU AI Act places the oversight of an autonomous nuclear plant on the same level as predicting energy production in a wind farm, despite the very different consequences of failure. This means that a small energy startup working on weather prediction models faces the same compliance requirements as a vendor of nuclear safety systems. In the six jurisdictions we reviewed, we did not find a way to adjust regulatory obligations based on the actual innovation value of a technology or its domestic origin.

The US model seemed more promising at first because it operates on a sector-by-sector basis. It is not. Its dependence on voluntary commitments allows companies to mostly avoid compliance, which defeats the purpose for economies wanting enforceable baseline protections.

We took a different route by creating a weighted risk scoring system. In this model, innovation potential and domestic development capacity serve as negative weight parameters, lessening the regulatory burden for applications that provide clear value. At the same time, we established a strict safety floor. Any system that scores at maximum impact severity will receive the most stringent oversight, regardless of other factors. The guiding principle was clear: no matter how innovative an AI system is, if it controls a nuclear reactor, it must be regulated accordingly.

For the financial sector, we introduced what we call a “Too big to fail AI” parameter. The idea is simple. A single algorithmic system’s failure can ripple through interconnected institutions. To our knowledge, no current framework tries to measure this.

Instead of using a single sandbox model across all sectors, we created sector-specific mechanisms. One size does not fit. Healthcare follows a three-stage process: simulation, controlled clinical deployment, and limited market release. Energy systems must complete 10,000 hours of digital twin validation before linking to the live grid. Fintech employs a graduated scaling approach from micro-pilots to nationwide roll-out. Each design came from the risk characteristics specific to its sector.

One unexpected finding was the extent to which policymakers believed the EU framework could be adopted with minimal adjustments. Most did. Few assessed whether their domestic market structures, often led by SMEs with limited compliance capacity, could realistically manage the costs involved. Our graduated obligation model was designed specifically to tackle this issue, differentiating requirements across enterprise tiers based on organizational capacity.

Introduction

The August 2026 enforcement deadline for the EU AI Act has turned a theoretical debate into a binding question for jurisdictions outside Europe. Our own country feels this acutely. We sit inside a customs union with the EU, which means CE-marked goods, including AI-enabled industrial equipment, already need to comply with European standards. But our domestic AI ecosystem looks nothing like the European one. The financial sector alone contains over 200 fintech firms, most of them with fewer than 50 staff. Asking a fifteen-person company to run the same conformity assessments designed for a multinational bank seemed clearly unworkable, but ignoring the regulatory question altogether was not a real option either.

The compliance cost picture is what made this concrete for us. The CEPS impact study estimates baseline quality management system setup costs of €193,000–€330,000 per high-risk system, with €71,400 in annual maintenance (Renda et al. 2021). The Center for Data Innovation’s analysis of the EU’s own impact assessment puts total compliance costs for an SME deploying a single high-risk product at up to €400,000, sufficient to erode roughly 40% of operating profits for a typical European SME (Mueller 2021). Independent industry studies covering the post-AI Act period place SME compliance ranges at €50,000–€500,000 depending on system complexity (Intellera Consulting 2024). For a domestic fintech with a fifty-person headcount and a technol-

ogy budget that often does not exceed a few hundred thousand dollars, these numbers do not describe a manageable compliance line item. They describe an exit from the market.

A ten-month workshop was convened to explore what could be done about this gap. It was a big group. The participants came from computer engineering, software engineering, law, medicine, and international trade. We examined seven sectors and six national approaches to AI governance: the EU, the US, the UK, Singapore, South Korea, and Japan. We came in expecting to find a model worth adapting. What we found instead was that none of the existing models maps cleanly onto economies with our structural features. The problem is not that the frameworks are poorly designed; they were designed for a different kind of market.

Out of that recognition came a scoring model that differs in three respects from anything we encountered in the survey. Most frameworks classify AI risk by the domain in which the system is deployed: this system is healthcare, that one is finance, the rules follow accordingly. Our model classifies the system itself across a broader set of questions. How severe is the harm if it fails, and how many people does it touch? How likely is failure given the system's maturity and operating conditions? Is the damage reversible? But also: how much innovation value does this system actually deliver, and can our regulators look inside it? That last question is one that wealthy jurisdictions tend to overlook because they rarely import the systems they regulate. We do, and the question matters.

A second problem surfaced in the financial sector that none of the six frameworks engages with directly. We came to call it the “too-big-to-fail” problem in AI. When a single algorithmic system handles a sufficient share of national equity volume, its failure does not stay contained inside the operating firm. It cascades through every institution wired to that system, much the way the 2008 financial crisis revealed cascading failure through concentrated points of dependence. The same logic now applies to AI systems embedded deeply enough in financial infrastructure to matter. We tried to formalise it.

This paper makes four contributions. First, it compares six national approaches across five dimensions and locates the specific places where each one fails to serve emerging economies. Second, it proposes a risk-scoring model in which innovation value and regulator auditability act as negative weights against a strict safety floor. Third, it formalises the systemic risk problem in financial AI through a sixth parameter, P_6 , with three operational dimensions and concrete thresholds. Fourth, it argues that sandbox regulation has to be sector-specific because the failure modes of healthcare, energy, and finance are not points on the same scale; they are categorically different problems demanding categorically different testing architectures.

Related Work

Risk Classification: What Exists

Of the six frameworks we read closely, the EU's is the most developed. The AI Act (European Parliament and Council

2024) sorts systems into four tiers, mostly by domain, as listed in Annex III. Veale and Borgesius (2021) flagged a problem with this scheme early, and it was a problem we kept running into throughout the workshop. Because classification follows the domain and not the system, very different applications inside the same sector receive the same regulatory treatment. An AI that schedules hospital appointments and one that reads mammograms can both fall under the same high-risk label, since both touch healthcare. The label is the same; the systems are not.

The US picture is almost the inverse. There is no unified AI regulation. FDA covers medical AI, SEC covers financial AI, DOT does autonomous vehicles. The NIST AI Risk Management Framework was supposed to introduce some coherence, but it is voluntary (NIST 2023). Calo (2017) put the issue with voluntary frameworks bluntly. The firms that adopt them are mostly the firms that would behave responsibly anyway. The firms you actually want to oversee can just ignore the framework. We think this is the central design flaw of the US approach, and not a fixable one without legislation.

The UK in its 2023 white paper declared itself “pro-innovation” and chose not to set up a central AI regulator at all (UK DSIT 2023). Singapore has, in our view, the most impressive sandbox ecosystem of the six, particularly through the Monetary Authority, although at the broader level it remains guidance-based (Singapore IMDA 2020). South Korea passed an AI Basic Act in early 2025, with a three-tier classification, but implementation is still early and we are watching to see how it develops. Japan has taken a principle-based path through its Society 5.0 initiative, focused on human-AI coexistence rather than binding rules (Smuha 2021). Six countries, six philosophies. None converged.

The thing that struck us when we put all six side by side was something common to all of them. Not one factors innovation value into the risk assessment itself. Not one. Not one. Countries that want to promote AI do so through separate programs: grants, tax incentives, research funding. The risk assessment stays focused on harm, and only on harm. That separation may be perfectly workable in the EU or US, where the institutional infrastructure for both tracks already exists. For economies trying to do promotion and protection at the same time, under tighter constraints, the separation creates problems we describe in the next subsection.

The Gap for Emerging Economies

Jobin, Ienca, and Vayena (2019) surveyed 84 AI ethics documents from around the world and found that most of them came from North America and Europe. This is not just an equity issue. The governance models we have today reflect the assumptions of the economies that produced them, and three of those assumptions break down in our context.

The first assumption is capacity. Regulated firms are assumed to be able to absorb complex compliance. But when the dominant fintech in a market has fewer than 50 employees, expecting it to run conformity assessments designed for a multinational bank is not realistic, and the cost figures from

the introduction make that point concrete rather than rhetorical.

The second assumption is institutional infrastructure. European countries built their conformity assessment bodies, national standards organisations, and accumulated regulatory expertise over decades. A country whose AI regulatory apparatus is still being assembled cannot just transplant the institutional design. The comparative law literature calls this the “transplant effect”: Berkowitz, Pistor, and Richard (2003) analysed 49 countries and found that the way law was initially transplanted is a stronger predictor of institutional quality than which legal family the law came from. Countries that received foreign law without adapting it or without a population already familiar with its principles developed weaker institutions. Watson (1974) made the foundational observation that legal borrowing is the most common source of legal change, but the outcome depends on whether the receiving context can support what is transplanted.

The third one is the one we found most consequential, and the one we kept coming back to during the workshop. Wealthy countries can afford separate bureaucracies, one for safety regulation and another for innovation support. Smaller economies cannot. Full stop. They have to combine both inside a more consolidated institutional setup. We started calling this the dual mandate problem somewhere around the fourth or fifth plenary session. The name stuck.

There is also a separate line of work we want to mention, although our paper does not engage with it at the philosophical level. Birhane (2020) argues that the import of Western AI systems and norms into Global South contexts often takes place with very little local scrutiny, and that this import displaces domestic alternatives in the process. Our concern is not exactly the same, but it shares a root. A framework written in Brussels carries assumptions about market structure, institutional capacity, and the asymmetry between regulator and regulated entity that simply do not hold where we work.

Bradford (2020) has documented how EU regulations propagate globally through what she calls the “Brussels Effect.” For countries with direct trade ties to the EU, this isn’t abstract. When you share a customs union, the question is not whether to engage with European standards but on what terms. The worry we kept returning to is that engagement turns into passive adoption without the institutional capacity to implement what was adopted. Floridi et al. (2018) has called for governance approaches that account for different national contexts, but the practical question (how to actually build such approaches) has not been worked out in any systematic way. That gap is what our work set out to address.

Sandboxes and Sectoral Differences

Regulatory sandboxes started in UK fintech and have since spread to more than 50 countries (Bromberg, Godwin, and Ramsay 2017). Zetzsche et al. (2017) traced this diffusion and argued that sandbox design needs to evolve from simple derogation from existing rules toward what they call “smart regulation” that embeds learning mechanisms. The EU AI Act has its own sandbox mechanism, in Articles 57 through 63. What kept bothering us is that these provisions do not distinguish between sectors. The same testing framework is

supposed to apply whether you are validating a diagnostic imaging algorithm or a high-frequency trading bot. On the surface that sounds fair enough. We think in practice it misses something.

The reason a healthcare AI needs testing is not the same as the reason a financial AI needs testing. Not even close. In healthcare, the worry is that a wrong diagnosis could permanently harm a patient. In finance, the worry is that one system’s malfunction can cascade through dozens of interconnected institutions. Different failure modes. Different timescales. Different victims. These are not different intensities of the same problem. They are different problems, and they need different testing setups. We were honestly surprised by how little published research compares sandbox design requirements across sectors. Most of it focuses on fintech specifically, or treats the sandbox as a general-purpose tool.

Methodology

Design Science as a Frame

The research was structured around the Design Science Research methodology of Hevner et al. (2004) and the refinements added by Peffers et al. (2007). DSR fit because we were not only trying to describe a problem. We were trying to build something. In DSR terms: problem identification came out of the six-jurisdiction comparison, which surfaced three structural mismatches. The solution objectives centred on two things: pulling innovation incentives into the risk assessment itself, and capturing systemic risk in financial AI. The design and development phase produced the weighted scoring model, the P_6 parameter, and three sector-specific sandbox designs. The demonstration is the worked example with three distinct AI systems that appears later. The evaluation, in the stricter DSR sense, is incomplete. We are upfront about this. The model has been proposed formally and submitted to a government ministry; a pilot has not yet taken place. This paper plays the role of the communication phase, and we come back to what the incomplete evaluation means in the limitations.

Comparing Jurisdictions

We needed a structured way to compare jurisdictions that would go beyond narrative description. After some back-and-forth, we settled on five dimensions. D1 captures how fine-grained the risk classification is. D2 asks whether innovation incentives are integrated into the regulatory structure or kept separate from it. D3 covers the sandbox approach. D4 covers accommodation for smaller firms. D5 covers the enforcement mechanism.

Country selection was driven by two things: geographic spread and regulatory diversity. The EU, US, and UK gave us three Western philosophies. Singapore, South Korea, and Japan gave us three Asian ones. Together they cover mandatory regulation, voluntary frameworks, principle-based approaches, and co-regulatory hybrids. We thought this would be enough range for what we were trying to do. It probably is, although we acknowledge that excluding Latin America and Africa from the comparison is a real gap that limits the

generalisability claim we can make. Following Yin (2018), we claim analytical rather than statistical generalisability: the three structural mismatches we identify are derived from the design assumptions of the frameworks themselves, not from features unique to our country. Any emerging economy that imports a significant fraction of its AI technology stack, has an SME-dominated AI sector, and operates under EU trade alignment pressures should encounter the same structural tensions.

Workshop Process

The work was done through a workshop hosted by a national business association. Seven sectoral working groups, about 40 participants in total — roughly 18 from industry (including 6 who run AI-focused companies), 13 from academia, and 9 from legal practice — drawn from computer engineering, software engineering, law, medicine, and international trade. Not all groups were equally staffed. We tried to make sure every sectoral group had at least one industry practitioner, one academic with relevant research background, and one legal expert with regulatory experience. The triangulation was deliberate, and it shows up in the disagreements documented in our internal records.

For convergence we used a structured process. Each working group produced an initial draft from its own sectoral expertise, then circulated the draft to the other six groups for cross-sectoral review during ten monthly plenary sessions. Where disagreements appeared on cross-cutting questions (definitions of high risk, treatment of imported systems, scope of human oversight requirements, which together accounted for the majority of recorded cross-group objections), we resolved them through structured discussion rather than majority vote, with resolution requiring that no working group register a formal objection. The document went through two complete revision cycles over ten months. We should say openly that this process leans toward consensus rather than adversarial confrontation, and we come back to what that means for the credibility of our findings in the limitations.

The output was a 501-page master policy document covering fintech, healthcare, energy, education, media, logistics, and AI literacy, formally submitted to the national ministry responsible for industrial policy. This paper focuses on three of those seven sectors: fintech, healthcare, energy. Why these three? Their risk profiles are different in kind, not just in degree. Not just severity. Kind. A diagnostic error in healthcare harms an individual patient, possibly irreversibly. A grid management error in energy can cascade across interconnected infrastructure in minutes. A trading algorithm malfunction in finance can propagate through networked institutions in real time. Testing for these three kinds of risk needs three different sandbox architectures, and that contrast is the heart of what we want to argue.

Comparative Analysis

The Six-Jurisdiction Comparison

Table 1 shows the results of the five-dimension analysis. The clearest pattern, and the one we keep returning to, is

that none of the six puts innovation into the risk formula itself. Zero out of six. Singapore and the UK come the closest by building innovation thinking into their overall regulatory philosophy, but even there, when it comes to deciding whether a specific system is high risk or not, innovation value doesn't enter the calculation. A second pattern: there is an inverse relationship between regulatory strength and flexibility. The EU has the strongest enforcement and the least flexibility. The US has maximum flexibility and no binding effect. None of the six has figured out how to be both binding and adaptive at the same time.

Three Problems with Transplantation

Risk gets treated as uniform within a domain. Annex III of the EU AI Act (European Parliament and Council 2024) lists “critical infrastructure” as a high-risk domain. Under that one label, an AI system that performs autonomous load balancing for a nuclear power plant and an AI that predicts solar panel output for a ten-person startup end up in the same classification. We went through all six frameworks looking for a mechanism that distinguishes consequence levels inside the same sector, and we did not find one. The practical effect of this is that a weather prediction startup faces the same conformity assessment as a nuclear safety vendor, which seems wrong to us on its face.

Small firms cannot absorb the compliance cost. The domestic financial sector includes 67 banks, more than 80 brokerages, and over 200 fintechs, according to the national banking regulator's 2024 year-end data. The fintechs are overwhelmingly small. Most are under 50 employees. The EU AI Act gives smaller firms longer compliance timelines, but the requirements themselves are unchanged. Asking a fifteen-person company to do what a multinational bank does, just nine months later, postpones the capacity problem; it does not solve it. The cost figures from the introduction make this concrete: a €200,000+ QMS setup for one high-risk product is not something you can defer when your entire technology budget is in that range. The US sidesteps the issue by making compliance voluntary, which means there is no consumer protection floor at all.

Innovation is penalised by default. Every framework we examined keeps innovation promotion separate from regulation. This works when both tracks have their full institutional infrastructure running in parallel. When they do not, what happens is that regulation arrives first and innovation support arrives later, or not at all. We saw this in our own workshop. The regulatory drafts always arrived with a much sharper sense of urgency than the corresponding innovation incentives, partly because the EU enforcement deadline created external pressure that no domestic innovation program matched in tempo. For a country trying to build a domestic AI industry while also protecting its citizens, the sequencing is a real disadvantage.

A thread runs through all six frameworks: the assumption that regulated entities are mid-to-large firms with existing compliance infrastructure. That holds for most of Europe and the US. It does not hold where we work. Taking the EU model imposes costs the firms cannot bear. Taking the US model means nobody has to comply with anything. The

Dimension	EU AI Act	US (NIST)	UK (DSIT)	Singapore	S. Korea	Japan
D1: Risk classification	Four tiers, domain-based (Annex III)	Sector by sector, no unified system	Principles rather than formal tiers	FEAT framework, context-dependent	Three tiers, still maturing	Society 5.0 principles, no formal tiers
D2: Innovation incentive	Horizon Europe, IPCEI, separate from regulation	SBIR and NSF, separate	Built into pro-innovation regulatory mandate	Linked to sandbox access	K-AI Strategy, separate	METI programs, separate
D3: Sandbox	Articles 57–63, uniform across sectors	No federal-level sandbox	FCA origin, some sector variation	MAS sandbox, sector-focused	Small pilot programs only	Small pilot programs only
D4: SME accommodation	Deferred timelines, obligations unchanged	Largely voluntary, SMEs effectively exempt	Flexible guidance	Grants and priority sandbox access	Minimal so far	Support programs exist
D5: Enforcement	Mandatory, fines up to 7% of global turnover	Voluntary commitments	Principle-based, non-binding	Mix of co-regulation and guidance	Phased move toward mandatory	Voluntary plus industry codes

Table 1. Six national approaches to AI governance across five dimensions.

remaining four have useful elements, but none adds up to a transferable package on its own.

The Scoring Model

Structure

Instead of classifying systems by domain, we score them on parameters. Three push the risk score upward; two pull it downward. P_1 captures impact severity: how serious is the harm if the system fails, and how many people does it touch? P_2 estimates the probability of failure given the system’s maturity and the conditions in which it operates. P_3 asks whether the damage is reversible.

The two parameters that move the other way are what make our approach different. They cut risk, not raise it. P_4 reflects the innovation value the system delivers to the domestic economy. P_5 measures how transparent the system is to national regulators.

We want to be explicit about what P_5 is and is not, because we know some readers will see it as protectionism in disguise. It is not. P_5 is an auditability proxy, and it is nationality-agnostic in principle. Any vendor, whether domestic or foreign, that provides verifiable source access, training-data documentation, and a locally responsive technical liaison gets the same P_5 score. A foreign firm that opens a local technical office, releases inspectable code, and supports regulator queries receives exactly the same auditability rating as a comparably positioned domestic firm. The general logic appears in fintech sandbox literature (Bromberg, Godwin, and Ramsay 2017), where vendors get equal access on condition of substantive local presence and information sharing. What P_5 penalises is opacity, not foreignness. When regulators cannot look inside a system, test it, or question the people who built it, oversight is structurally weaker, and that weakness maps directly to higher effective risk. An imported system running on overseas servers

behind proprietary licences and non-disclosure agreements creates real enforcement difficulties. The same is true of a domestic system that does these things.

A hard constraint is built into the model. We call it the safety floor. If P_1 reaches its maximum value, the system goes straight into the highest oversight tier. No exceptions. No offsets. It does not matter how innovative the system is, or how transparent the code is. If an AI controls a nuclear reactor, it receives maximum regulation, period.

Operationalising the Innovation Parameter

The natural objection to P_4 is that “innovation value” looks softer than the other parameters. Fair point. A regulator could be captured by firms claiming to be innovative when they are not. We take this seriously. The response is built into how P_4 is assigned: it is not self-assessed by the developer. An independent sectoral technical committee assigns it, using four operational criteria on a graded scale.

The first criterion looks at whether the system reaches populations or use cases that incumbent systems demonstrably fail to serve. The underserved population should be evidenced through data, not through marketing claims. Next: does the system perform a task that no existing deployed system in the domestic market matches at comparable accuracy or cost, with verification against a published benchmark or an audited internal one? Then there is the question of ecosystem contribution: does the development process generate transferable assets (open datasets, open models, published methods) that other domestic developers can build on? Finally, does the system reduce dependence on a single foreign supplier in a strategically sensitive function? P_4 is the sum of these sub-scores, capped and normalised. Sectoral committees publish scoring decisions with reasoning, and the decisions are open to administrative appeal within the standard public-law timeframe.

The committee composition matters. Each sectoral com-

mittee includes at least one academic, one industry representative, and one civil society member, with three-year terms and formal conflict-of-interest disclosure. Anyone holding equity or board positions in firms under review must recuse. We are not going to pretend this fully eliminates capture risk. It does not. But it raises the cost of capture substantially and creates a public record that subsequent decisions can be measured against.

Two more design choices deserve note. P_4 cannot offset the safety floor: a system that reaches maximum impact severity stays in the top tier regardless of its P_4 score. And P_4 expires. A system that earned access-expansion credit at deployment loses that credit once the underserved population has been served by multiple providers. The innovation contribution has ended, and the score should reflect that. Without this expiry mechanism, legacy systems would keep collecting innovation credit they no longer deserve.

Sectoral Weights

We use the same five parameters across all sectors but shift the weights to reflect the dominant risk character of each domain. The weights matter. In healthcare, P_1 takes the largest share, at 35%, because patient safety dominates. In energy, P_1 and P_3 are equally weighted at 25% each because grid failures combine severity with cascade potential. Fintech is different. A sixth parameter becomes necessary.

The Too-Big-to-Fail Problem in AI

This is the section we think is the most original part of the paper. The observation that motivated it is simple. Financial systems are interconnected in ways other sectors are not. When we looked at the 2010 Flash Crash (US equity markets lost roughly a trillion dollars of value in 36 minutes through interactions between algorithmic trading systems, Kirilenko et al. 2017), we noticed a pattern that existing AI governance frameworks do not engage with. The Knight Capital incident in 2012 reinforced the point: a single software-deployment error in a market-making algorithm caused 440 million dollars in losses in under an hour (Securities and Exchange Commission 2013). Both events showed that in financial markets, one system's failure does not stay contained. It propagates.

The Financial Stability Board raised this concern back in 2017, noting that AI concentration in financial services creates systemic vulnerability (Financial Stability Board 2017). The concern has sharpened since: the FSB's 2024 update warned specifically that AI-driven herding behaviour and concentration in model provision create stability risks that microprudential tools were not designed to detect (Financial Stability Board 2024). The European Systemic Risk Board reached a similar conclusion independently: Cecchetti et al. (2025) argued that AI systems can amplify procyclical dynamics and that existing regulatory frameworks lack the tools to measure algorithmic systemic risk. Recent legal scholarship has started engaging with the question explicitly (Schwarcz 2024). But raising a concern and operationalising it into a regulatory classification are two different things. We checked all six jurisdictions. The EU AI Act assesses each

system individually without considering its position in the broader financial network. The NIST framework addresses organisational risk, not cross-institutional cascading. None of them had turned systemic risk into a classification parameter.

The theoretical foundation is established. Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015) showed that in financial networks, diversified interbank linkages act as shock absorbers when shocks are small, but the same linkages become contagion channels once losses exceed a critical threshold; past that point, more connectivity means more damage, not less. Allen and Gale (2000) demonstrated a parallel result: overlapping portfolio exposures create contagion paths that are invisible until a crisis triggers them. P_6 operationalises this threshold logic for AI governance: the three dimensions capture exactly the network properties (concentration, connectivity, lack of alternatives) that determine whether an algorithmic failure stays contained or cascades.

P_6 adds a sixth parameter, specifically for finance, assessed along three dimensions. Market concentration asks what share of total market activity flows through one AI system. If a single trading algorithm handles more than 15% of national equity transactions, that is a red flag. Interconnectedness counts how many institutions depend on the system. If more than 50 banks, funds, or brokerages are connected to one AI platform, failure becomes everyone's problem. Substitutability asks what happens if the system goes down. If fewer than three genuine alternatives are available, the market has a single-point-of-failure problem.

Table 2 summarises the three dimensions and the threshold values we propose. The three dimensions mirror the Basel Committee's G-SIB assessment methodology, which scores banks on size, interconnectedness, substitutability, cross-jurisdictional activity, and complexity to determine systemic importance (Basel Committee on Banking Supervision 2013). We adapted these categories from banking entities to AI systems: where the G-SIB framework asks "how large is this bank relative to the global banking system," P_6 asks "how much market activity flows through this algorithm relative to the national market." The 15% concentration threshold reflects domestic market structure: in our national equity market, algorithmic trading accounts for a growing but still modest share of volume, and a single platform reaching 15% would represent a dominant position warranting heightened oversight. The 50-institution interconnectedness threshold and the 3-alternative substitutability threshold follow the same logic of calibration to domestic scale rather than global markets, and should be understood as starting values subject to empirical adjustment during pilot deployment.

P_6 carries the same weight as P_1 in the fintech model: 25%. Equal weight. Systemic financial risk can affect millions of people who never directly interacted with the AI system in question. That is why we treat it with the same weight as direct impact severity. P_6 inherits the safety-floor treatment as well. If any of its three dimensions reaches the threshold, the system enters the highest oversight tier regardless of its innovation score.

Dimension		What it measures	Top-tier threshold
Market concentration		Share of transactions one AI system handles	More than 15% of market volume
Interconnectedness		Number of institutions dependent on the system	More than 50 dependent institutions
Substitutability		Available alternatives if it fails	Fewer than 3 viable alternatives

Table 2. The three dimensions of P_6 , the systemic risk parameter for financial AI.

Worked Example

Table 3 shows the contrast directly. Three concrete AI systems, classified under the EU AI Act and under our model. The second row is the example that drove the design of the whole framework. A ten-person startup writing weather models should not face the same compliance burden as a nuclear safety vendor. Under the EU Act it does, because both fall under Annex III critical infrastructure. Under our model it does not, because the parameter values diverge sharply. And the safety floor still guarantees that if the startup's system changed in nature (if it began directly controlling load rather than producing forecasts that a human dispatcher consumes), classification would escalate automatically.

Three Sectors, Three Sandbox Designs

Why a Single Sandbox Model Does Not Work

The EU's sandbox provisions in Articles 57 through 63 apply a uniform structure regardless of sector. One framework for all. We think this is the wrong design choice. The reason is not that the provisions themselves are weak. The reason is that the reasons we test AI differ fundamentally across sectors.

Three concrete cases. In healthcare, a misread scan can permanently harm a patient. The testing process has to accumulate clinical evidence over time. There is no short-cut. In energy, a miscalculation of grid frequency can trigger a cascading outage that propagates across infrastructure in minutes. The testing process has to simulate dangerous conditions that nobody can responsibly reproduce on a live network. In finance, an algorithmic malfunction can trigger selling cascades across connected institutions, and systemic risk is invisible at small scale. The testing process has to watch the system as it scales. These are not different intensities of one problem. They are different problems, with different timing structures, and they call for different testing architectures.

The Three Designs

Table 4 puts the three designs side by side. The healthcare track is deliberately the slowest. Clinical evidence takes time to accumulate, and there is no responsible way around that.

Rushing risks the kind of irreversible harm that justifies the extended timeline in the first place. The energy track relies heavily on simulated adversarial testing, because cascading grid failures cannot be deliberately induced on a live network. The fintech track focuses on observed behaviour during staged scaling, because the whole point of P_6 is that systemic risk is a property of scale, not something a small test can reveal.

This paper focuses on the structural design of sector-specific sandboxes rather than their full operational specifications. The details matter, but they are country-specific. Stage durations, exit criteria, and the supervisory mechanisms attached to each will need to be adapted to each country's existing regulatory infrastructure. That is why we describe architectures rather than prescribe identical procedures. The central claim is that the architectures themselves should differ, not that any one instantiation should become a standard.

Graduated Obligations Without Compromising the Floor

In every sector studied during the workshop, the same tension came up. Always. Regulation designed for large firms ends up applied to firms with fifteen or twenty employees. We address this with a graduated obligation model differentiated by enterprise size. The smallest firms (under ten employees) get eighteen additional months to reach compliance, plus free technical assistance. Those with ten to forty-nine employees get twelve additional months and subsidised access to third-party auditing. Mid-sized firms, fifty to 249 employees, follow the standard timeline with sector-specific guidance documents. Large firms, 250 or more, carry the full obligations, including annual independent audits.

The graduated model differentiates procedures (timelines, support structures, audit access) and not substantive standards. The safety floor does not scale with firm size. A three-person startup that builds a trading algorithm handling 20% of national equity volume enters Tier 1 oversight immediately, the same way a multinational bank would. What the graduated model offers is reduced procedural burden during routine compliance work for smaller firms operating below the safety floor. It does not lower the floor itself.

A harder case helps clarify the intent here. Suppose a five-person startup wants to deploy a system that triggers the safety floor (controls a critical infrastructure function, or crosses a P_6 threshold). The graduated model does not protect that startup from Tier 1 obligations. If the firm cannot meet the substantive standards required for safety-floor systems, the firm should not be operating that system. We made this choice deliberately. Reducing substantive standards on the basis of firm size would turn the safety floor into a fiction, and that defeats the whole point. The graduated model offers procedural relief precisely so the substantive floor can hold.

The system	Under the EU AI Act	Under our model	Why the difference matters
Algorithmic trading platform handling 18% of national equity volume, used by 30+ financial institutions	High-risk under Annex III financial services; identical label to a small consumer lending chatbot	Tier 1 (highest). P_6 fires on market concentration (18%) and interconnectedness (30+ institutions)	The EU framework has no mechanism to see the systemic dimension. A lending chatbot and a platform wired into the entire financial system receive identical treatment
Weather prediction tool built by a 10-person domestic startup using open-source code, used for solar energy forecasting	High-risk under Annex III critical infrastructure; same category as nuclear control systems	Tier 3 (lowest regulated level). P_1 is low because forecast error means suboptimal scheduling, not blackout. P_4 and P_5 are high because the team is local and the code is open	Clearest illustration of the parity problem. EU rules impose nuclear-grade compliance on a transparent low-risk tool; our model regulates proportionally
Imported radiology AI with no available source code, developed abroad, used for chest X-ray reading	High-risk under Annex III medical devices	Tier 2. P_1 is high because diagnostic errors can harm patients. P_4 and P_5 are low because the system is imported and opaque. Sandbox entry required before clinical deployment	Tier is similar but the reasoning is visible through parameter scores; sandbox requirement is concrete

Table 3. Three AI systems classified under the EU AI Act and under the proposed model.

Feature	Healthcare sandbox	Energy sandbox	Fintech sandbox
Model name	Clinical pathway	Simulation-validated deployment	Graduated scaling
Stage 1	Synthetic patient data only; no real patients	Open-loop testing: AI recommends, humans decide	Micro-pilot capped at 100 users with limited transaction values
Stage 2	Deployment in a small set of selected hospitals with intensive monitoring	Closed-loop operation under normal conditions; AI acts, humans observe	Regional expansion up to 10,000 users; pattern detection at scale begins
Stage 3	Conditional market release with continuing reporting	Adversarial stress testing plus inter-system interaction testing	National rollout with continuous systemic risk monitoring
Total duration	Approximately 2–3 years	12–14 months across all stages	12–18 months depending on scaling pace
Exit criterion	Performance at or above current human standard of care	Performance at or above non-AI baseline plus safe degradation evidence	Successful scaling with no P_6 dimension crossing threshold

Table 4. Three sandbox designs matched to three sector risk profiles.

Discussion

The Copy-Paste Reflex

The most surprising finding of the workshop process, at least for those of us coordinating it, was the strength of what we ended up calling the “copy-paste reflex.” When the seven sectoral working groups produced their initial drafts, most of them — six of the seven groups — produced texts that adopted the majority of their provisions directly from translated EU AI Act articles, almost clause by clause. Nobody had asked the prior question first: does this make sense for our market?

The shift happened during the cross-sectoral review sessions, when participants from different working groups were reading each other’s drafts for the first time. That changed things. Quickly. Members of the fintech group, several of whom actually run small AI companies themselves, sounded

the alarm first. One of them, a co-founder of a fintech with roughly 20 employees, calculated that the compliance costs in the draft would consume approximately 85% of his company’s annual technology budget. That kind of concrete, first-hand evidence was what broke the copy-paste assumption. Abstract argument that EU rules might not fit was not enough. Specific cost evidence presented by specific affected people was what shifted the discussion.

We think this observation generalises beyond our country. The tendency to adopt established frameworks reflects a real belief, mistaken in our view, that a framework designed for the EU should apply broadly with minor tweaks. The comparative law literature would recognise this as a classic transplant dynamic (Berkowitz, Pistor, and Richard 2003), and parallel concerns have been documented in the Brazilian AI governance context (Belli 2023). Displacing the copy-paste belief requires domain-specific evidence from the peo-

ple who would have to live with the consequences.

What Could Other Countries Use?

The five-parameter scoring model is the most portable component. The parameters are not country-specific; any government could adopt the structure, set its own weights, and obtain a working classification system. The safety floor is universal by design. P_6 was built for finance, but the logic applies wherever institutional interconnection creates systemic vulnerability. The sandbox designs are transferable as principles: healthcare needs slow clinical evidence; energy needs simulated adversarial testing; finance needs progressive scaling. Specific durations and institutional assignments will need adaptation.

Limitations

This work is based on one country. We believe the three mismatches we identified are not unique to our context, but we have not proven that through cross-national comparison, and doing so is the obvious next step. The model has been formally submitted to a government ministry but it has not been piloted. We do not yet know what problems will surface during actual deployment. In DSR terms, we have done design and demonstration; the evaluation stage depends on a pilot that has not happened yet.

The six-jurisdiction comparison is based on published texts and policy documents. How these frameworks operate in practice may differ from how they read on paper, and we did not study implementation. Our scoring in Table 1 involves subjective judgment. A different team using the same five dimensions might assign different scores to the same cells. We have tried to be transparent about the reasoning behind each cell, but the subjectivity is real.

The workshop process itself has limitations. The consensus-leaning approach may have suppressed minority positions that did not crystallise into formal objections. Participants self-selected, biasing the pool toward people already inclined to engage with regulatory questions. The fintech group was more empirically grounded than others because several participants ran the firms being regulated; some sectoral groups had thinner industry representation.

Conclusion

We went looking for a middle path between adopting the EU AI Act wholesale and doing nothing. The weighted scoring model, the systemic risk parameter, and the sector-specific sandbox designs together make up a viable starting point, although each one needs testing in practice. That matters.

What this work showed us is that the problems with regulatory transplantation are not at the level of implementation details. They are structural. The risk parity problem, the SME capacity gap, and the innovation penalty are baked into the design assumptions of existing frameworks. Adjustments won't fix this. It calls for a different way of conceptualising and measuring risk.

The P_6 parameter for systemic financial risk is, in our view, the most time-sensitive contribution. The Flash Crash

was sixteen years ago. Knight Capital was fourteen. AI systems have only become more embedded in financial infrastructure since then. Every year that passes without a mechanism for measuring algorithmic systemic risk is a year in which the gap between regulatory capacity and actual market risk widens.

We want to close with two invitations. Both practical. The first is to researchers working in other emerging economies: test whether the three mismatches we describe hold in your context, because we cannot validate a single-country claim alone. Recent UNCTAD work on technology transfer governance in developing countries (UNCTAD 2024) suggests the tension between imported regulatory frameworks and domestic capacity is not limited to AI, and we suspect the pattern is broader than our data can show. The second is to sectoral regulators in finance: P_6 does not need to wait for a full AI act. The three dimensions (concentration, interconnectedness, substitutability) can be adopted as supervisory metrics on existing algorithmic trading platforms today, and the empirical data this would generate would be valuable on its own terms.

Ethical Statement

This work came out of a participatory process across five disciplinary backgrounds. That breadth was intentional. We designed the graduated obligation model specifically to keep regulation from becoming a barrier that only large corporations can clear. If it does, the domestic AI ecosystem gets shut out of precisely the high-value applications where it could matter most. The P_6 parameter is grounded in consumer protection: ordinary depositors and retail investors carry the heaviest costs when interconnected algorithmic systems fail. The healthcare sandbox centres patient safety by requiring clinical equivalence evidence before any market release. We should acknowledge that the workshop process was bounded by national participation, and the model's effects on populations outside our country (including data subjects of imported systems whose data may be processed under our framework) have not been examined here.

A few second-order effects are worth flagging. P_4 (innovation value) creates an incentive for firms to overclaim novelty; our committee-based process raises the cost of doing so but does not eliminate the risk. P_5 (auditability) may inadvertently disadvantage small foreign developers who cannot afford a local technical liaison, even though we treat it as nationality-agnostic in principle. The graduated obligation model creates a compliance-cost arbitrage incentive: firms approaching SME thresholds may artificially restrict headcount to retain procedural benefits, a known failure mode of size-based tiering in tax and labour regulation.

We are researchers working in an emerging economy with a customs union with the EU. This position shapes the work in ways we want to acknowledge openly. We live with the tension between EU regulatory expectations and local market realities every day. Some of the authors are themselves practitioners building AI products in small firms; some have direct experience with patent processes and conformity assessments. That experience drives our scepticism toward

copy-paste regulation and our insistence that innovation capacity belongs in the risk formula. It also creates bias risks we should be transparent about. We have a personal stake in regulatory regimes that do not impose disproportionate burdens on small firms, and that stake is reflected in how we framed the SME capacity gap. Our perspective is also bounded. This is one country's experience, and whether the patterns we identified replicate elsewhere is something we cannot yet say with confidence.

References

- Acemoglu, D.; Ozdaglar, A.; and Tahbaz-Salehi, A. 2015. Systemic Risk and Stability in Financial Networks. *American Economic Review*, 105(2): 564–608.
- Allen, F.; and Gale, D. 2000. Financial Contagion. *Journal of Political Economy*, 108(1): 1–33.
- Basel Committee on Banking Supervision. 2013. Global Systemically Important Banks: Updated Assessment Methodology and the Higher Loss Absorbency Requirement. Technical report, Bank for International Settlements.
- Belli, L. 2023. Platform Regulation in Brazil: Towards a Brazilian Digital Constitutionalism. In *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*. Emerald.
- Berkowitz, D.; Pistor, K.; and Richard, J.-F. 2003. Economic Development, Legality, and the Transplant Effect. *European Economic Review*, 47: 165–195.
- Birhane, A. 2020. Algorithmic Colonization of Africa. *SCRIPTed*, 17(2): 389–409.
- Bradford, A. 2020. *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.
- Bromberg, L.; Godwin, A.; and Ramsay, I. 2017. Fintech Sandboxes: Achieving a Balance between Regulation and Innovation. *Journal of Banking and Finance Law and Practice*, 28(4): 314–336.
- Calo, R. 2017. Artificial Intelligence Policy: A Primer and Roadmap. *UC Davis Law Review*, 51: 399–435.
- Cecchetti, S.; Lumsdaine, R. L.; Peltonen, T.; and Sánchez Serrano, A. 2025. Artificial Intelligence and Systemic Risk. Technical Report Report No. 16, European Systemic Risk Board Advisory Scientific Committee.
- European Parliament and Council. 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). *Official Journal of the European Union*.
- Financial Stability Board. 2017. Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications. Technical report, FSB.
- Financial Stability Board. 2024. The Financial Stability Implications of Artificial Intelligence. Technical report, FSB.
- Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; and Vayena, E. 2018. AI4People: An Ethical Framework for a Good AI Society. *Minds and Machines*, 28: 689–707.
- Hevner, A. R.; March, S. T.; Park, J.; and Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly*, 28(1): 75–105.
- Intellera Consulting. 2024. An Analysis of the Cost of Compliance with the AI Act for SMEs. Technical report, Intellera Consulting.
- Jobin, A.; Ienca, M.; and Vayena, E. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, 1: 389–399.
- Kirilenko, A. A.; Kyle, A. S.; Samadi, M.; and Tuzun, T. 2017. The Flash Crash: High-Frequency Trading in an Electronic Market. *The Journal of Finance*, 72(3): 967–998.
- Mueller, B. 2021. How Much Will the Artificial Intelligence Act Cost Europe? Technical report, Center for Data Innovation.
- NIST. 2023. AI Risk Management Framework (AI RMF 1.0). Technical report, National Institute of Standards and Technology.
- Peffer, K.; Tuunanen, T.; Rothenberger, M. A.; and Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3): 45–77.
- Renda, A.; Arroyo, J.; Fanni, R.; Laurer, M.; Sipiczki, A.; Yeung, T.; Maridis, G.; Fernandes, M.; Endrodi, G.; and Milio, S. 2021. Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe. Technical report, CEPS for the European Commission.
- Schwarcz, S. L. 2024. Regulating Systemic Risk in Artificial Intelligence. *University of Illinois Law Review*, 2024(3): 905–948.
- Securities and Exchange Commission. 2013. Administrative Proceeding File No. 3-15570: In the Matter of Knight Capital Americas LLC. US SEC.
- Singapore IMDA. 2020. Model AI Governance Framework (2nd Edition). Technical report, Infocomm Media Development Authority.
- Smuha, N. A. 2021. From a 'Race to AI' to a 'Race to AI Regulation'. *Law, Innovation and Technology*, 13(1): 57–84.
- UK DSIT. 2023. A Pro-Innovation Approach to AI Regulation. White paper, Department for Science, Innovation and Technology.
- UNCTAD. 2024. Technology and Innovation Report 2024. Technical report, United Nations Conference on Trade and Development.
- Veale, M.; and Borgesius, F. Z. 2021. Demystifying the Draft EU AI Act. *Computer Law Review International*, 22(4): 97–112.
- Watson, A. 1974. *Legal Transplants: An Approach to Comparative Law*. University of Georgia Press.
- Yin, R. K. 2018. *Case Study Research and Applications: Design and Methods*. Sage, 6th edition.
- Zetzsche, D. A.; Buckley, R. P.; Arner, D. W.; and Barberis, J. N. 2017. Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation. *Fordham Journal of Corporate and Financial Law*, 23(1): 31–103.