Print Edition

Wednesday 26 August 2026

EDITION NO. 34 · 2 PIECES · 9 PAGES

IN THIS EDITION

We Don't Need to Wait for an AI Disaster to Estimate Its Costs

BY AI FRONTIERS · AI FRONTIERS · 24 AUGUST 2026

Daniel Carpenter, Professor of Government at Harvard University, **Feodora Douplitzky-Lunati**, and **Arjun Purohit** — August 24, 2026

In April, US Treasury Secretary Scott Bessent and then–Federal Reserve Chair Jerome Powell gathered leading banks to discuss the threats of Anthropic's new AI model, Claude Mythos. It was a tacit admission: no one, in the public or private sector, is ready for what is coming. The Mythos announcement led top AI firms to attach tighter restrictions to their own models and prompted even relatively laissez-faire voices in the federal government and think tanks to begin considering preapproval regimes for frontier AI models. Wherever those debates go now, Mythos has changed the game, bringing into sharp focus the vast potential scope and scale of catastrophic risks posed by AI—such as an AI-triggered financial meltdown, vast cybersecurity or infrastructure collapse, or enhanced terrorism risk. As a result, companies, the government, engineers, scholars, and all of society are beginning to look for a different set of institutions, whether in regulation or insurance, to manage the risks.

However, there is already an incredibly valuable tool that could be adapted for measuring and managing AI risk. In 2002, in the wake of the September 11 attacks, the US government quietly passed the Terrorism Risk Insurance Act (TRIA) and, with it, the Terrorism Risk Insurance Program (TRIP). TRIP does many things, but its most important lesson for AI policy is the TRIP data call, which combines a kind of war-gaming with insurance underwriting.

Once a year, TRIP requires insurers across the industry to think hard and quantitatively, not about events they have seen but about scenarios they have *never* seen and might not even have imagined. A car bomb packed with the radioactive substance cesium 137 goes off in Atlanta, in the middle of a workday. A massive cyberattack takes down multiple cloud-computing platforms, incapacitating the millions of systems that rely on them, from hospital records to financial transactions to transportation flow. A bomb placed in a shipping container destroys much of an industrial port, setting off chain reactions that sharply curtail world trade. TRIP then asks insurers what total losses for the scenario would amount to, harnessing the preexisting in-house capacity of insurance companies to attach aggregate cost estimates to events.

This essay proposes that the US government establish a similar exercise for catastrophic AI risk, describing how such a program would generate a wealth of data on as-yet-unforeseen risks and the scale of their damages, while building expertise and capacity for thinking more clearly about AI risks.

Analyzing and Insuring Against Novel Threats

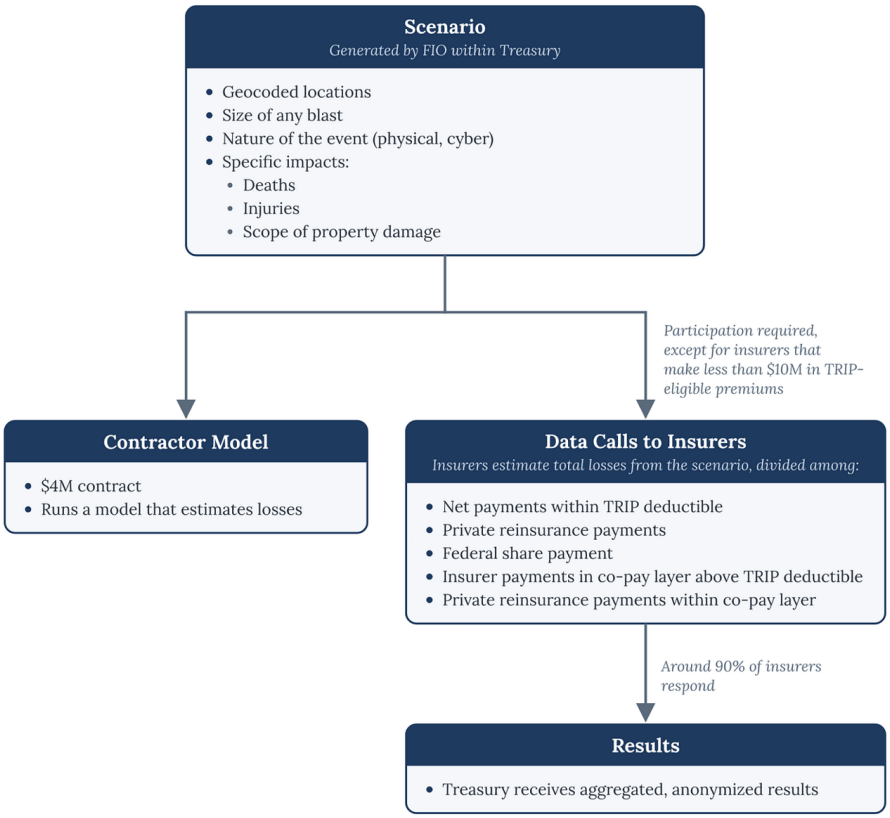
Whether insurance can be applied to frontier AI is a hotly debated question. Some writers point out how inchoate cyber insurance is, even after two decades of development, and question whether insurance can ever price the highly complex, multidimensional, and unforeseen risks posed by frontier AI. Other writers, including some in this publication, argue that AI risks could be insured through financial products already in use to address climate risk (such as catastrophe bonds, which offload extreme risk to capital markets as a kind of backstop to insurance operations).

The core challenge of insuring AI risks is a lack of historical data. For insurance or bond markets to work, they will likely need data that is very hard to come by in the AI world. Insurers and institutional investors estimate risk based on data from the past: namely, observed damages from events similar to those they are trying to insure or invest in. This works well when what might happen in the future looks a lot like what has happened before; it also works when past events are sufficiently similar to compose a "data set" from which one can generalize and predict. But such conditions do not always hold. In the early 2000s, for instance, federal policymakers realized that terrorism posed risks that defied these properties. Nonetheless, they sought a way to estimate and at least partially insure against the threats of terrorism. Enter TRIP.

TRIP enables the government and private insurers to share extreme terrorism risk. TRIP both provides information to the private insurance market about potential losses during terrorist attacks and supports specific insurers in covering such losses. To set TRIP in proper context, we must briefly describe the supportive structure. The program requires certain private insurers to offer terrorism coverage to commercial policyholders and, in the event of a terrorist attack, pay deductibles based on a percentage of their premiums to cover initial claims. However, in cases of extreme damage above a certain threshold, the federal government pays the majority of the remaining losses.

By promising to share losses in case of a large terrorist event, the federal government mitigates extreme risk for insurers, thus stabilizing their profits and avoiding improvised bailouts. Under the program's mandatory-recoupment provisions, the Treasury Department must recover much or all of its expenditures through surcharges on future commercial insurance premiums whenever aggregate insured losses remain below thresholds set by law. In other words, TRIP backstops the private insurance market, but only to a point; it also tries to ensure that insurers have appropriate "skin in the game," incentivizing them to price risk correctly.

Through annual data calls, TRIP gathers information about losses in terrorist attacks. The key innovation in TRIP—a "data call"—did not emerge until 2016. Since then, though, it has been institutionalized. The Treasury Department's Federal Insurance Office (FIO) outlines one scenario per year, imagining in detail an event that might involve terrorist attacks on physical infrastructure as well as cyberattacks on cloud providers and data centers. The data call then unfolds with a publication on the Treasury Department website and in the Federal Register, and one key part asks insurers to estimate the losses that the given scenario would incur. Since 2017, insurer participation in data calls has been mandatory, except for those earning less than $10 million from lines of coverage with TRIP-eligible premiums. The response rate has been 90% or above among relevant insurers. Data is collected through a third-party insurance statistical aggregator, and results are provided to the Treasury Department in an aggregated, anonymized format. A simplified flowchart of the annual TRIP data call appears in the figure below.



Outline of the TRIP data call process.

The 2026 data call imagines a twofold scenario. First, a terrorist organization hijacks a cargo plane laden with fuel and explodes it over several data centers in Virginia, destroying a number of them. Second, at roughly the same time, terrorists unleash a malware-assisted cyberattack that is directed at a large cloud provider. The data call includes precise geographic coordinates of the crash, the kinds of structures affected, and a statement of maximal impact for aggregate losses from the cloud network attack, relying upon assumptions as to how quickly cloud service is restored (within one week, or 168 hours). Participating insurers are asked to consider economic impacts, property damage, and even disability claims resulting from the attack.

Insurers estimate the amounts that they and the government would need to pay. The key work that participating insurers do is to estimate total losses from the scenario, which the insurers then divide into six categories. These categories require insurers to answer a set of questions. For example, would the losses incurred in the scenario exceed the insurer's deductible? If so, how much would they and other insurers have to pay through coinsurance (the percentage of claims above the deductible that they must pay, up to an "out-of-pocket" maximum)? If the risk were much larger and the federal government had to backstop losses, how much would the federal government likely have to pay? In one modeled case of an attack on nuclear power plants with significant release of radioactive material, the total losses exceeded $240 billion (see pages 88–90 of the linked report). As evidence of how the private sector can assist in this

work and how stable these processes have become, consider that the risk analytics firm Moody's is now developing and commercializing a tool that assigns losses to categories.

Next, we consider three important societal benefits of the TRIP data call process.

How TRIP Data Calls Benefit the Public

The annual exercise of the TRIP data calls assists policymakers with managing terrorism risks, in several ways.

Policymakers can extract lessons and mitigation measures from the detailed scenarios. When it comes to the most severe potential catastrophes (whether from terrorism, cyberattacks, or frontier AI), society often lacks data on what kinds of things can go wrong and what the costs would look like. This lack is partly due to a longtime assumption that such events are uninsurable, removing one of the main incentives for thinking about such risks and estimating the associated damage. TRIP data calls help to fill that gap by generating two kinds of data: (1) *narrative data* about a terrorist scenario and (2) *statistical data* about losses incurred in that scenario. Even the narrative data alone is very useful.

The narrative data generated each year by the Treasury's FIO projects the end state of an attack. In some ways, this resembles war-gaming—the generation of as-yet-unrealized scenarios by military and security planners. Using war-gaming to measure AI risks and prepare for AI catastrophes is not new. Drawing upon its extensive experience in the security sector, RAND has conducted war-gaming exercises based on model loss-of-control (LOC) scenarios, which have already produced useful (indeed, sobering) lessons. That same report, for instance, concluded: "Governments and other stakeholders lack a common framework to analyse and respond to LOC risks."

Even if TRIP did not feed the scenarios to insurers for statistical and actuarial analysis, *these catastrophe narratives would be useful in and of themselves*. Suppose that society wishes to regulate frontier AI, whether by collective industry self-regulation, government-imposed constraints, or some combination of the two. Policymakers would want to identify the likely catastrophic risks from frontier AI models and use that data to allow regulators to target the source of the most dangerous weaknesses. Beyond estimation, such "dark speculation" would also allow all of us to think more clearly—both qualitatively and quantitatively—about *mitigation*. As some of us have argued in a recent theoretical paper, the benefits of war-gaming would be massive, even if estimates of damages were noisy. This is in part because the institutionalization of thinking about catastrophe scenarios *prompts consideration of countermeasures, some of which can be implemented now*.

Insurers provide expertise and institutional capacities for quantitative loss estimates. While war-gaming in the frontier AI space already exists, the regular use of expert statistical and actuarial analysis to assess the end state of these war games does not. Put differently, existing war games do not produce general and systematic estimates of the collect-

ive losses from the catastrophes that occur in frontier AI scenarios. Loss estimation requires a set of skills—really, institutional capacities—that traditional war-gaming exercises do not contain. These additional requirements include access to computational capacity and expert statisticians and actuaries. They also include the developed knowledge of organizations that have been both "in the game" and "in the business." "In the game" means they have institutional and organizational procedures and habits (including predictive analytics) for doing the complex analysis that insurance underwriters do all the time, and "in the business" means that underwriters engage in these practices habitually, with money on the line.

One might wonder whether the private sector will supply these benefits on its own—for example, Lloyd's has studied realistic disaster scenarios for decades. But several pieces of evidence suggest that TRIP's efforts go beyond what these private exercises accomplish on their own. First, TRIP leverages the federal government's relationship with all fifty state insurance commissioners, giving it institutional capacity and a statistical check that private companies do not have. TRIP checks its catastrophic estimates yearly against similar estimates supplied by the states. Second, the Congressional Research Service concluded in 2019 that TRIP has generated a more robust private terrorism insurance sector. A 2024 Treasury Department report echoed these conclusions, documenting a more viable insurance market for terrorist events and even cyber insurance. Finally, TRIP operates at a scale that is difficult to replicate in the market. Lloyd's scenario exercises, which span climate risk, terrorism, and cyber, relied on the participation of 57 underwriting firms in 2023. TRIP focuses on terrorist risk alone, and while there is no published data on the exact number of companies participating every year, estimates from the Federal Register suggest more than 700 in 2017. Federal regulators have the power to force insurers to comply and provide their underwriting skills, something insurers would be unlikely to do without government pressure. This aggregation of nearly all insurers in the market gives more reliable underwriting results.

The repeated exercise builds infrastructure and facilitates public-private cooperation. TRIP data calls do not happen in a vacuum. Congress and the Treasury Department have invested in real infrastructure to assist the process. The FIO has now conducted 11 data calls, and since at least 2020 each has involved a new terrorist-attack scenario. The data calls have, moreover, become increasingly sophisticated, with greater use of geo-coded information on attacks and, most recently, the incorporation of cyberterrorism. The Treasury Department has now partnered with the National Science Foundation to establish the Industry-University Cooperative Research Center (IUCRC), which will help insurers estimate risk with modeling and underwriting tools, contribute to expansion of insurance, and develop tools to better inform analysis, management, and treatment of risk in government programs.

Neither underwriting nor scenario-generation is unique to TRIP's data calls. TRIP's innovation is in the *structured, aggregated, and repeated* combination of the two. By conducting the data call exercises year after year,

the Treasury Department builds a set of relationships with both insurers and third-party contractors that supply methodology, software, or scenarios. Across various scenarios, insurers learn how to think about unforeseen (and perhaps unforeseeable) risks, and the Treasury Department learns from the program's own history about what works and what does not. *Structure and repetition require institutionalized organization: a mix of "bureaucracy" and contracted expertise*.

Can the TRIP Model Transfer and Scale?

When it comes to applying the TRIP model to AI, the greatest challenge is perhaps the sheer unpredictability of catastrophic AI risk. TRIP data calls are limited by their specificity: the 2026 call, for instance, which involved the destruction of a data center complex in Virginia, asked insurers only a limited set of questions about cloud providers being incapacitated. AI-assisted or AI-induced catastrophes might be much vaster, even global in character. One adverse event might raise the risk of others, setting off a cascade of catastrophes that could grow to much more extreme levels of damage than most terrorist attacks. Critics might rightly wonder whether data calls for frontier AI catastrophes would need to be scaled up so much that the resulting program would look little like today's TRIP.

There are good reasons to experiment with TRIP for AI, despite feasibility questions. The worry about feasibility is certainly warranted. Yet for three reasons it should not deter us from considering and experimenting with TRIP-like programs. First, information on specific kinds of catastrophes can be used for other, similar projections about aggregate losses from like events. Whether AI risk involves the construction of a bioweapon or a massive cyberattack, the potential for human and economic damages will involve the kinds of events that insurers have to think about in other scenarios where AI does not play a role.

Second, we cannot know about plausibility until we test scenarios. In the autumn of 2001, insuring against terrorism risk was considered a lost cause. Now, a thriving industry extends to all regions of the United States.

Third, just about any new policy initiative (or combination of initiatives) to measure and mitigate the risks of frontier AI will require scaling, in part because frontier AI models themselves are globally scaled and growing rapidly in size, breadth, and power. In this sense, it is worth keeping in mind that TRIP's data call process is remarkably cheap for now. Outlays for TRIP run between $4 million and $7 million per year—a small price given the potential benefits that a similar exercise applied to AI risk could offer.

The discipline of underwriting would prevent, not encourage, catastrophizing about AI. Many observers and scholars worry that society is overreacting to the threat of frontier AI. Would TRIP-like data calls merely institutionalize and amplify this worry, potentially suffocating innovation? We think not. TRIP data calls combine catastrophic scenario-generation with the discipline of underwriting, de-emphasizing the least plausible catastrophes. In fact, a further important benefit of a TRIP-like

program would be to assuage fears about frontier AI that are generated by overly anxious catastrophic thinking.

"TRIP for AI" could have many uses and manifold benefits. In the US, our society, our government, and even our AI sector have not yet agreed on how to manage the risks from frontier AI. However, pretty much any approach would benefit from the kinds of speculative knowledge generated by a TRIP-like program. Considered more imaginatively and on a larger scale, such a program can begin to do for frontier AI what TRIP has done, albeit imperfectly and slowly, for terrorism: help public and private actors get a factual handle on the scenarios that are most troubling and difficult to imagine, encouraging the private and public sectors to cooperate on addressing those possibilities. The breadth and scale of potential AI catastrophes are indeed vast, but a tested method can help us chart this unknown territory.

See things differently? *AI Frontiers welcomes expert insights, thoughtful critiques, and fresh perspectives. Send us your pitch.*

Daniel Carpenter is the Allie S. Freed Professor of Government and Chair of the Department of Government at Harvard University. He works on the political economy of regulation, especially in pharmaceutical regulation, financial regulation and the regulation of AI. His recent research on AI regulation includes analyses of the adaptability of FDA-like approval regulation to AI governance and mathematical models of wargaming-informed underwriting for catastrophic AI risk. A Guggenheim Fellow and an elected fellow of the National Academy of Public Administration, he was recently named a Harvard College Professor for his excellence and innovation in teaching. He received his undergraduate degree in government from Georgetown University and his Ph.D. in political science from the University of Chicago.

Feodora Douplitzky-Lunati is a junior at Harvard University studying Economics and Slavic Languages & Literatures, with a secondary concentration in Government. Her research interests include political economics, public policy, and international relations. She has previously worked on active labor market program design at the World Bank and the politics of trade in the Middle Corridor at the Georgian Institute of Politics.

Arjun Purohit is a recent graduate of Harvard University, where he studied History with a secondary in Economics. His research interests include American grand strategy, international security with a focus on South Asia and Europe, and emerging technologies. He has worked in foreign policy and national security at the American Enterprise Institute, the Bertelsmann Foundation, and the Hudson Institute. He will pursue an M.Phil in Modern European History at the University of Cambridge.

Monthly meetup at Doheny & Nesbitt's

BY PROGRESS IRELAND · PROGRESS IRELAND · 21 AUGUST 2026

We'll be hosting our monthly meetup on Wednesday 26th August from 5.30pm-7.30pm. It's a chance to talk policy, chat with the Progress Ireland team, and hopefully make some new friends.

 Add to your calendar: **Google** · **Outlook**

The details:

We'll meet upstairs in Doheny & Nesbitt's, Baggot Street Dublin. Look out for chalkboards which will point you in the right direction.

To kick things off, Seán Keyes will talk about technocracy and the state of the Western soul: how small policy decisions can have huge consequences for society. He'll cover how this plays out in transport, where we plan for the population we have today and hamstring future generations.

Any queries should be directed to Sarah Scales on X (@sarahjscales) or LinkedIn (https://www.linkedin.com/in/sarahscales/).

We hope to see you on Wednesday.

Seán, Seán, Sam and Sarah.