Current Efforts Aren't Enough to Slow the AI Cyber Virus

Despite unprecedented efforts to patch vulnerabilities, AI is helping attackers exploit bugs faster than defenders can fix them. With policy responses lagging, we see the pressures lining up for the cyber epidemic to worsen.

NOVEMBER 13, 2026

HUDSON ATTAR
ARHAM HABIB

© 2026 Bridgewater Associates, LP

BRIDGEWATER

We don't have all of the details related to yesterday's Microsoft outages and series of vishing campaigns across major financial firms, but we do think that episodes like this are one flavor of disruption that is likely to become more common as the cyber-exploit exponential we write about in this piece continues to unfold. We think it is important to pay attention to major outages like this even when we aren't sure that malicious actors are at play because waiting until it is clear that damaging attacks are getting through to key infrastructure will lead to being too late in markets. Bridgewater remains highly vigilant to these evolving risks. We will follow up on this work with another piece covering where we see potential impacts—for now, we want to share how we are tracking and processing cyber risks today.

AI cyber risks, cyber-hacking ability using open-source models has reached dangerous levels, and we are managing money based on a one-in-three chance of an economically destabilizing attack within the next year. AI-enabled cyber warfare is now spreading exponentially with limited controls—like a virus. We see attacks continuing to inflect upward in spite of the race to secure software and the internet, and don't think the potential scale of cyber warfare is well appreciated because of how hard it is to track these attacks systematically. Given that cyber warfare is on an exponential path, reasoning about it well requires staying ahead of the curve, while most decision makers don't even have good visibility into what is happening—that is part of why we are concerned about the risk of a disaster.

Thus far, the plan for managing cyber risks hinges on using the same tools to preemptively identify and patch vulnerabilities before they can be exploited. However, executing that plan is a daunting task because of how much insecure human-written code has been produced over the last several decades and how diffuse ownership of it is. And this approach comes with its own risks in the form of potential loss of containment or misalignment.

In this report, we look at where the AI cyber arms race stands and whether defensive efforts are keeping pace. There is a lot we do not know about how this will unfold, but early signs suggest **current efforts are not doing enough to prevent a continuing surge in damaging attacks**.

- **Defensive efforts are uncovering vulnerabilities at an unprecedented rate, but bugs are still slipping through the cracks.** Programs like Anthropic's Glasswing and OpenAI's Daybreak have driven a record surge in potential exploits. But defenders have to fix every security-critical bug before it is exploited, while attackers only need to find one, and zero-day exploits (those with no lead time to discovery) have roughly doubled since the first truly cyber-capable open-source models were released. Further, not all companies have the resources or incentives to invest in such preemptive protection.
- **AI is allowing attackers to exploit the narrow window between when patches are developed and when they are adopted.** Patching open-source software necessarily exposes the underlying vulnerability, and every downstream user has to adopt the fix before attackers can exploit the bug. AI has made it easier for attackers to win that race. The number of recorded actively exploited vulnerabilities has surged, and the average time between disclosure and exploitation has fallen from a month to 11 days.
- **Policy responses in the hardest-hit economies are directionally helpful but insufficient, and in the West, they remain limited.** Japan and South Korea, which have suffered a wave of damaging attacks on financial institutions, have begun mobilizing resources. Japan has allocated ¥10 billion (about $63 million) to cyber defense, and South Korea has introduced legislation mandating executive-level chief information security officers (CISOs) and fines for repeated breaches. But

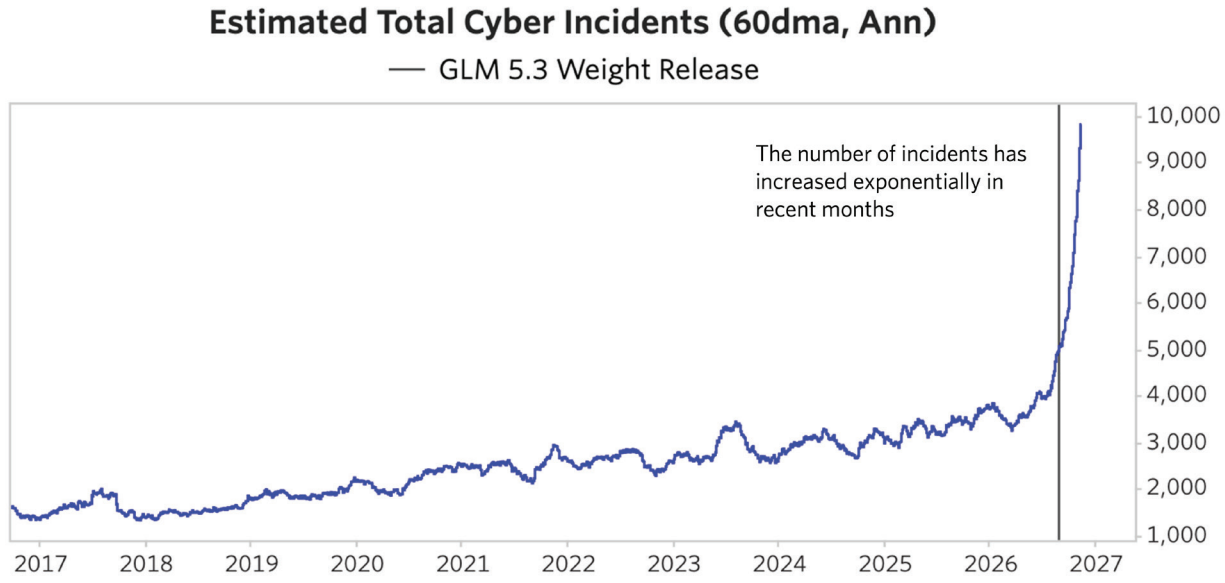
© 2026 Bridgewater Associates, LP

neither addresses the problem that much of the internet runs on open-source software maintained by volunteers and nonprofits with limited resources.

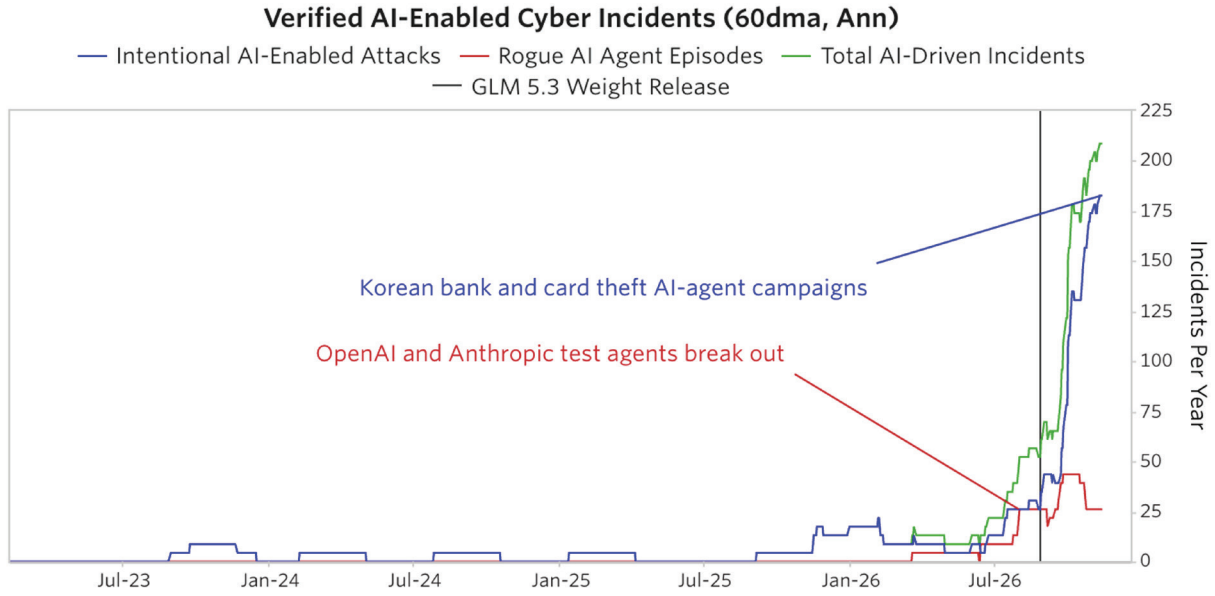
- **The arms race is accelerating as AI expands the attack surface.** While frontier models comb through decades of human-written code, developers are using AI coding tools to add millions of new lines, much of them with their own bugs. Even if AI-written code is marginally safer, the sheer volume means the attack surface is growing faster than defenders can secure it.

Absent an immediate and forceful coordinated policy response, **we see the pressures as lining up for the cyber warfare epidemic to get worse from here**. The "head start" between Glasswing's launch and the release of cyber-capable open-source models this summer was a one-time gift that we're unlikely to see replicated, and its protective benefit will wane over time. Malicious actors are only starting to learn how to use AI to find and exploit vulnerabilities, and we suspect they are earlier on the learning curve than defenders, with more room to improve. Threat actors are not yet broadly deploying fully autonomous attack workflows—a human remains in the loop—but open-source models are not far from the capability level where that is likely to change.

We developed a time series of cyber incidents, a taxonomy for linking incidents to root causes, and a real-time tracking pipeline of exploits in order to monitor what is going on. We scraped public disclosures, media coverage, and corporate filings to assemble a history of past exploits and adjusted them to account for typical reporting lags for different types of players across jurisdictions. We found that the count of cyber incidents and the share of known cyber incidents attributed to AI have both skyrocketed. Following a dramatic wave of major software vulnerabilities, supply chain attacks, and AI-enabled hacking incidents, cybersecurity professionals dubbed the last month of exploits "Black October." The charts below show how the picture since then has been ongoing exponential growth, with little likelihood of abating going forward.



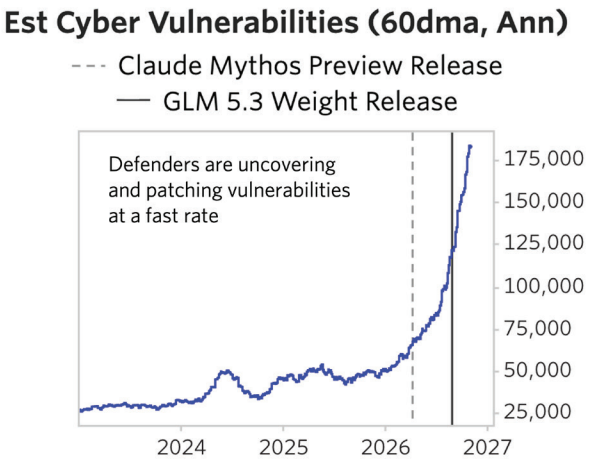
© 2026 Bridgewater Associates, LP

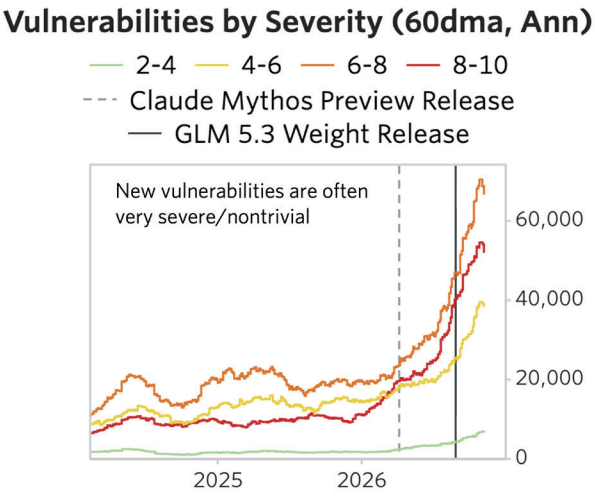


The Arms Race Between Attackers and Defenders Is Barreling Ahead

We are in the middle of an arms race where malicious actors are using agents to find exploitable bugs and defenders are using agents to preemptively find and patch vulnerabilities. While models have significantly accelerated work on both sides of the ledger, the defenders have the much heavier burden of patching all bugs before they can be exploited.

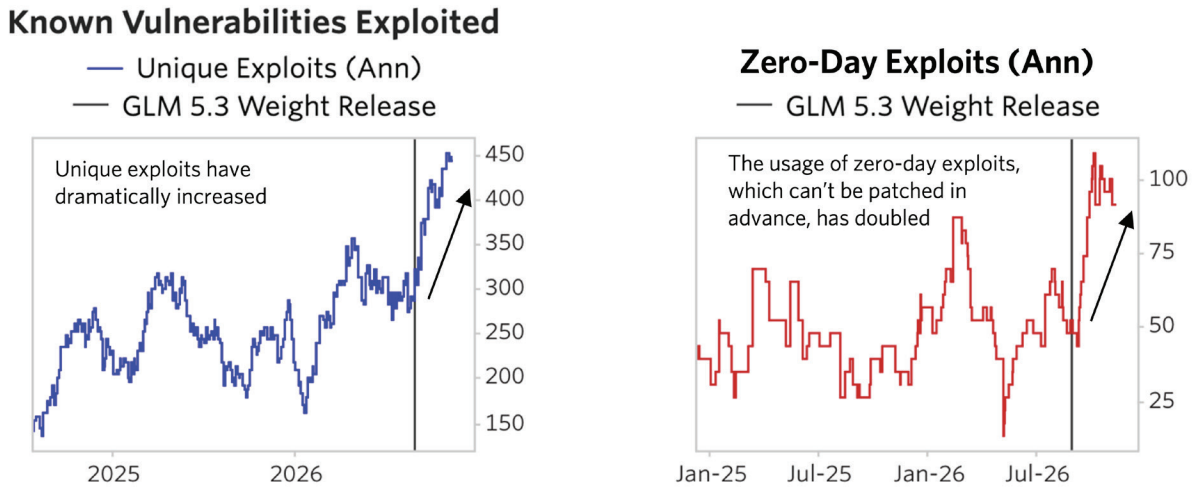
Frontier labs believe that avoiding a catastrophe requires leveraging what's known as the "defender's window," or the period of time in which frontier cyber capabilities are meaningfully better than diffused open-source cyber capabilities, to proof critical software against attacks. They have created programs—Anthropic's Glasswing, OpenAI's Daybreak, and Google's Fairwind—to allow trusted partners to utilize frontier cyber capabilities to identify and patch an unprecedented number of vulnerabilities in widely used software. The number of bugs submitted to the Common Vulnerabilities and Exposures (CVE) program, which is a resource for defenders to name and track software vulnerabilities after patches/mitigations have been developed, has exploded. This is meaningful progress toward a more secure internet.





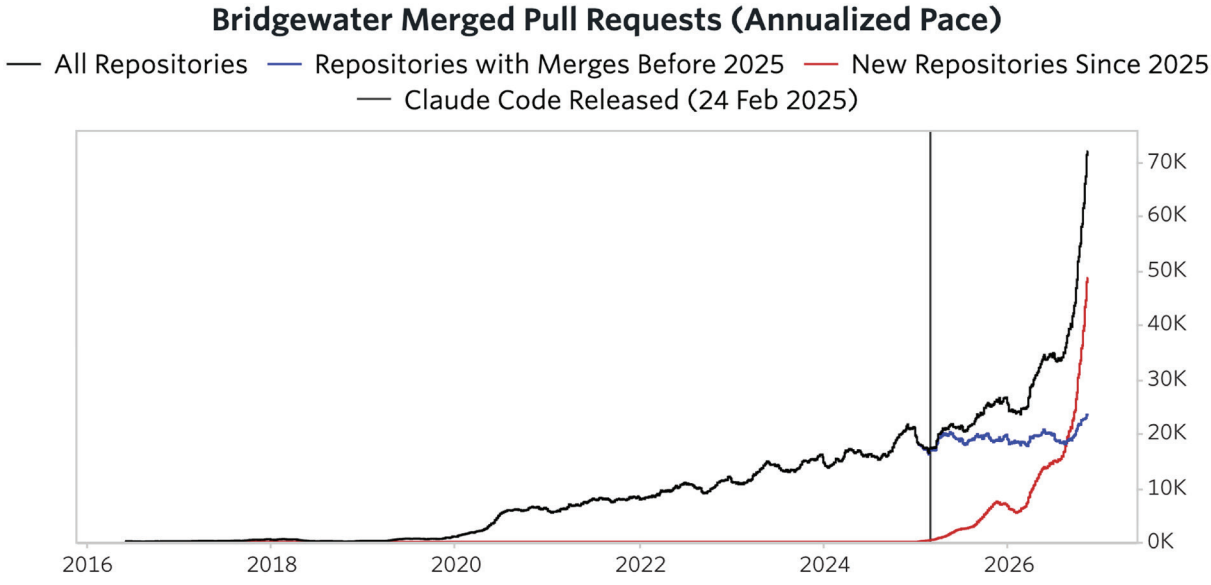
© 2026 Bridgewater Associates, LP

Unfortunately, the task defenders face is difficult: they need to fix all of the bugs, not just most of them, because anything that slips through the cracks can be systematically rooted out and exploited by attackers leveraging AI. The Cybersecurity and Infrastructure Security Agency (CISA) maintains a nomenclature of exploitable and exploited bugs found in popular software projects. The chart below shows the count of exploits added to CISA's Known Exploited Vulnerabilities (KEV) Catalog. One helpful distinction is that between "n-days" and "zero days"; the former had a patch published before a known exploit, while the latter refers to an exploit for which no patch is available when it is found. Zero-day exploits are uniquely troubling because, as the name implies, defenders have no advance notice to diagnose and patch the vulnerability before the vulnerability is exploited. The damage of these can be boundless. The number of such bugs being used in exploits has doubled since the release of the first truly cyber-capable open-source models.



At the same time, AI has also been a force multiplier in producing code, exponentially increasing the amount of software that needs to be secured. While frontier cyber-tuned models race to comb over all the code that's been written by humans in our critical software infrastructure stack over the last 30 years, developers are using coding models to add millions of lines of new code, much of which contains its own bugs. Even if AI-written code is marginally safer than the human-authored equivalent, the sheer volume of code being written means the surface area for exploits is increasing faster than we can identify and secure new vulnerabilities.

The chart below shows the rate at which code production within Bridgewater has exploded in recent years—we think this is indicative of the sea change in code production more broadly thanks to agentic AI.



© 2026 Bridgewater Associates, LP

Current Policy Responses Are Insufficient

Governments in countries with the largest increase in cyberattacks over recent months have started to respond to a degree:

- In October, Japan enabled active cyber-defense powers, which allow authorities to access and neutralize servers used in serious cyberattacks. Under the law, police and the Self-Defense Forces can break into attackers' servers and disable them before damage spreads, with an independent oversight commission set up in April 2026. Weeks later, after dozens of companies reported unauthorized access and possible customer data leaks, Japan launched an emergency package with initial funding of ¥10 billion (about $63 million) creating a "top gun" team at the Ministry of Economy, Trade and Industry that is using AI to counter hackers and find weaknesses, a public/private "cyber rescue team" to handle first responses for attacked companies, and a separate team to help vulnerable companies harden their defenses.
- Also in October, South Korea introduced legislation stipulating a 24-hour incident reporting deadline, turnover-based fines of up to 3% of sales for repeated incidents, and mandatory executive-level CISOs. Separately, the privacy law now raises the maximum penalty for large-scale data leaks to 10% of sales, up from 3%.

In the West, there have been some limited policy initiatives to empower defenders:

- In August, the Trump administration issued a presidential memorandum that directs the creation of a program authorizing vetted private US companies to conduct cyber operations against foreign cyber-enabled criminal groups, under federal control and oversight.
- The UK is strengthening security and reporting requirements for essential services and their suppliers through its Cyber Security and Resilience Bill, including a reporting mandate within 24 hours and a full report within 72 hours of a major cyber incident.

Some proactive efforts are better than none, but the effectiveness of these policies remains unclear and untested. Further, while these efforts are directionally helpful, they don't address the core problems that, without serious policy action, will cause the exponential increase in cyberattacks to continue. First, much of the internet depends on open-source software maintained by volunteers and nonprofit organizations that have limited resources and incentives to find and fix vulnerabilities, making their ability to address problems the bottleneck for creating security. Second, adopting solutions requires vigilance and discipline from all downstream consumers of software, while AI is increasing the cost of a slow response. And third, there is a dependence on the labs themselves to ensure that cyber-defensive efforts utilizing AI are done in a way that appropriately balances the risks of going too slow against the risks of going too fast—loss of containment and misalignment.

© 2026 Bridgewater Associates, LP

Important Disclosures and Other Information

This research paper is prepared by and is the property of Bridgewater Associates, LP and is circulated for informational and educational purposes only. There is no consideration given to the specific investment needs, objectives, or tolerances of any of the recipients. Additionally, Bridgewater's actual investment positions may, and often will, vary from its conclusions discussed herein based on any number of factors, such as client investment restrictions, portfolio rebalancing and transactions costs, among others. Recipients should consult their own advisors, including tax advisors, before making any investment decision. This material is for informational and educational purposes only and is not an offer to sell or the solicitation of an offer to buy the securities or other instruments mentioned. Any such offering will be made pursuant to a definitive offering memorandum. This material does not constitute a personal recommendation or take into account the particular investment objectives, financial situations, or needs of individual investors which are necessary considerations before making any investment decision. Investors should consider whether any advice or recommendation in this research is suitable for their particular circumstances and, where appropriate, seek professional advice, including legal, tax, accounting, investment, or other advice. No discussion with respect to specific companies should be considered a recommendation to purchase or sell any particular investment. The companies discussed should not be taken to represent holdings in any Bridgewater strategy. It should not be assumed that any of the companies discussed were or will be profitable, or that recommendations made in the future will be profitable.

The information provided herein is not intended to provide a sufficient basis on which to make an investment decision and investment decisions should not be based on simulated, hypothetical, or illustrative information that have inherent limitations. Unlike an actual performance record simulated or hypothetical results do not represent actual trading or the actual costs of management and may have under or overcompensated for the impact of certain market risk factors. Bridgewater makes no representation that any account will or is likely to achieve returns similar to those shown. The price and value of the investments referred to in this research and the income therefrom may fluctuate. Every investment involves risk and in volatile or uncertain market conditions, significant variations in the value or return on that investment may occur. Investments in hedge funds are complex, speculative and carry a high degree of risk, including the risk of a complete loss of an investor's entire investment. Past performance is not a guide to future performance, future returns are not guaranteed, and a complete loss of original capital may occur. Certain transactions, including those involving leverage, futures, options, and other derivatives, give rise to substantial risk and are not suitable for all investors. Fluctuations in exchange rates could have material adverse effects on the value or price of, or income derived from, certain investments.

Bridgewater research utilizes data and information from public, private, and internal sources, including data from actual Bridgewater trades. Sources include AERIC INC, BCA, Bloomberg Finance L.P., Candeal, Carbon Arc, CEIC Data Company Ltd., Ceras Analytics, China Bull Research, Citibank, Clarus Financial Technology, CLS Processing Solutions, Consensus Economics Inc., Consumer Edge, CRU Group, DTCC Data Repository, Ecoanalitica, Energy Aspects Corp, Enverus, EPFR Global, Eurasia Group, Evercore ISI, FactSet Research Systems, The Financial Times Limited, Finaeon, Inc., FINRA, GaveKal Research Ltd., GlobalSource Partners, Goldman Sachs, Harvard Business Review, Haver Analytics, Inc., IEA, Institutional Shareholder Services (ISS), The Investment Funds Institute of Canada, ICE Derived Data (UK), Investment Company Institute, Institute of International Finance, JP Morgan, JTSA Advisors, LSEG Data and Analytics, MarketAxess, Metals Focus Ltd, MSCI, Inc., National Bureau of Economic Research, Neudata, Organisation for Economic Cooperation and Development, Pensions & Investments Research Center, Pitchbook, Political Alpha, Renaissance Capital Research, Rhodium Group, RP Data, Rubinson Research, Rystad Energy, S&P Global Market Intelligence, Sentix GmbH, SGH Macro, Shanghai Metals Market, Smart Insider Ltd., Swaps Monitor, Tradeweb, United Nations, US Department of Commerce, Visible Alpha, Wells Bay, Wind Financial Information LLC, With Intelligence, Wood Mackenzie Limited, World Bureau of Metal Statistics, World Economic Forum, and YieldBook.

While we consider information from external sources to be reliable, we do not assume responsibility for its accuracy. Data leveraged from third-party providers, related to financial and non-financial characteristics, may not be accurate or complete. The data and factors that Bridgewater considers within its research process may change over time.

This information is not directed at or intended for distribution to or use by any person or entity located in any jurisdiction where such distribution, publication, availability, or use would be contrary to applicable law or regulation, or which would subject Bridgewater to any registration or licensing requirements within such jurisdiction. No part of this material may be (i) copied, photocopied, or duplicated in any form by any means or (ii) redistributed without the prior written consent of Bridgewater® Associates, LP.

The views expressed herein are solely those of Bridgewater as of the date of this report and are subject to change without notice. Bridgewater may have a significant financial interest in one or more of the positions and/or securities or derivatives discussed. Those responsible for preparing this report receive compensation based upon various factors, including, among other things, the quality of their work and firm revenues.

© 2026 Bridgewater Associates, LP