LONGRIVER PARTNERS FUND
2ND QUARTER 2025

9 July, 2025

Dear Partners,

The Longriver Partners Fund's net return this quarter was 11.7%, bringing our year-to-date return to 11.4%. Our benchmark, the MSCI AC World USD Net Index, returned 11.5% for the quarter and 10.0% year-to-date. Since inception, the Fund has gained 55.6%, compared to 58.0% for the benchmark.

Our investment strategy is to ride the long-term value created by our companies, so I encourage you to take a similarly long-term view of your investment in the Fund. Consider it as you would an investment in real estate or a private company – assets for which you wouldn't receive a volatile quarterly valuation.

WHY WE INVEST IN BIG TECH

My favourite investments are essentially monopolies with growth. They are more predictable and have natural places to reinvest profits, which lowers the risk of poor capital allocation.

That's what led me over the years to Big Tech and its surrounding ecosystem. Thanks to network effects, scale, bundling and switching costs, companies like Alphabet, Amazon, Apple, Meta and Microsoft have sustained extraordinary profitability and growth. These are not contrarian ideas, but that doesn't mean they're overvalued. Markets often struggle with the idea that something obvious can still be mispriced.

Today, our holdings in this space include Meta, Tencent, TSMC and Nvidia, which I added this quarter. Meta and Tencent aggregate attention and monetise it at scale. TSMC and Nvidia design and manufacture the advanced semiconductors that power the entire stack. Together, these four represent 35% of our portfolio.

ARTIFICIAL INTELLIGENCE ("AI") IS MOVING FROM HYPE TO HABIT

A year ago, I wrote about the cyclical resurgence in TSMC's earnings. AI had triggered one of the largest investment cycles in tech history, and demand for leading-edge chips surged. At the time, AI tools were novel but impractical. It wasn't clear what would drive mass adoption or how Big Tech would earn a return on its staggering investments.

Enough has changed in twelve months to warrant an update. AI is moving from hype to habit. Usage is no longer speculative or marginal. It's reshaping how people create, think, and work.

Longer context windows were the first big unlock. Claude 3 Sonnet can keep about 200,000 tokens in active memory. GPT-4o can hold roughly 128,000. That's enough to load book-length documents in one go and ask the model to reason across them. This is real knowledge work, not prompt juggling.

On top of that, "reasoning-tuned" variants like Claude 3 Opus, GPT-o3 and DeepSeek's R1 have gone through extra rounds of supervised fine-tuning and Reinforcement Learning from Human Feedback (RLHF). These teach the model to follow multi-step logic, infer intent and adapt when users change direction. OpenAI's Deep Research mode builds on this, sustaining focus across complex questions.

I confess that I've become an AI-aholic. It's now central to my work and life. Where Google Search offers links, AI gives

answers. I use it for idea generation, research, writing, and translation. I ask it to pick holes in my investment theses and point out my blind spots. As someone who works solo, it's been transformative. AI is a candid and unflagging partner with an unparalleled breadth and depth of knowledge. It helps me play the game at a higher level.

(I also ask it for advice on making amends to my wife Lizzie when I've put my foot in my mouth and it's pretty good there too!)

The data says I'm not alone. On Microsoft's last earnings call, CEO Satya Nadella shared, "We processed over 100 trillion tokens this quarter, up 5x year-over-year — including a record 50 trillion tokens last month alone." That was March. It's likely higher now. AI isn't theoretical or a gimmick anymore. Usage is real and growing exponentially.

OPENAI'S CHATGPT HAS BECOME A HOUSEHOLD VERB

Twelve months on, the shape of the AI race is clearer. The contest to build the most advanced model continues, but the real battle is in deployment. What matters now isn't the breakthrough, it's the product and the distribution.

OpenAI is the clear frontrunner. In a world saturated with models and chatbots, ChatGPT has become a household verb. Weekly active users have doubled in recent months to nearly 800 million. It processes over a billion queries a day and drove 4.5 billion website visits in March 2025. Like Chrome for browsing, iOS for mobile and WeChat for messaging (in China), ChatGPT has become AI's default interface.

This is because OpenAI has turned ChatGPT into much more than just a model. It's a tightly integrated product that combines reasoning, tools, memory and planning within a fast and clean UI. OpenAI owns the full interaction loop, which lets it learn from usage, iterate quickly and compound its lead. The more you use it, the better it gets. Competing products feel clunky and impersonal by comparison.

That product strength is fuelling platform ambition. OpenAI wants to be a consumer app and an enterprise provider and a hardware company and build Artificial General Intelligence. Each path demands a different business model, go-to-market, and - frankly - culture. But the strategy seems to be: build the best product, drive usage, and let usage compound into dominance.

The catch? OpenAI hasn't quite figured out yet how to monetise usage. GPT Enterprise has seen early traction, but many customers remain price-sensitive and are exploring fine-tuned open models as cheaper alternatives. Free consumer tiers may get ads or affiliate links, but these would have to be done carefully to avoid eroding trust.

Meanwhile, inference costs are steep and have not been coming down fast enough. OpenAI remains deeply unprofitable, even as its appetite for compute continues to grow. As a capped-profit entity born from a nonprofit, its ability to raise capital is constrained. Microsoft, its key partner, is reportedly reluctant to fund further training. So OpenAI is exploring other routes, including a proposed $500 billion decade-long compute partnership with SoftBank and Oracle.

So for now, OpenAI leads on product and usage. But its challenge is to convert habit into cash flow.

META'S SUCCESS AND STRATEGIC CONUNDRUM

Meta, in contrast, is using AI to make money today. Behind the scenes, it has built some of the most advanced infrastructure in the world. This has quietly driven stronger-than-expected revenue growth and margin expansion.

Meta's ad platform is becoming increasingly automated. It already has tools that optimise campaigns for a target return on spend. Now it's building AI systems that can generate creative assets, test variations, and automate entire campaign flows. If it works, the ad loop tightens: better performance, higher margins, and more advertiser lock-in. One expert estimated Meta could improve campaign results by 50 to 100 per cent if CPMs hold steady. That would translate into substantially more spend.

But none of this is visible to the average user on Facebook or Instagram. Meta hasn't shipped anything like ChatGPT. Llama 4 was solid, but the product layer was missing. As one expert put it, "OpenAI shipped tools, memory, orchestration. Meta shipped a model and a blog post."

That is the conundrum. Meta risks becoming the world's most powerful AI-powered ad platform but without a dominant AI interface of its own. That would make it valuable, but fragile - just as it's dependent today on Apple. Until AI is meaningfully integrated into Facebook, Instagram, and WhatsApp, Meta's distribution edge remains potential, not reality. Until then, nearly a billion people are building habits around ChatGPT. In consumer tech, habits are the moat.

Part of the challenge is organisational. Meta's research lab, FAIR, focuses on long-term academic work. Its GENAI team builds models. Product teams are left to figure out how to turn these into usable software. The feedback loop is slow. It is the opposite of OpenAI, where everything feeds the product.

Founder Mark Zuckerberg has gone into 'beast mode' to force a reset. In June, Meta made a minority investment in Scale AI, reportedly worth $15 billion. Scale's CEO, Alexandr Wang, will join Meta to lead AI efforts. The company is also acquiring a stake in NFDG to bring product specialists Nat Friedman and Daniel Gross in-house. These may be the most expensive acqui-hires in history. As I write this, Meta has also poached top talent from OpenAI, DeepMind, and Apple for eye-watering salaries and sign-on bonuses. Zuckerberg clearly wants trusted hands to unify the organisation and start shipping.

The investment case hinges on execution. The components are in place: infrastructure, models, data, scale. What's missing is a product layer that puts it all in users' hands. If Meta can close that gap, the upside is enormous. If not, it may help build an AI future - while someone else owns the customer.

DISTRIBUTION, DEPENDENCY AND THE SHUFFLING OF CLOUD LEADERSHIP

AI has reshuffled the market for cloud computing. Thanks to its early and exclusive partnership with OpenAI, Microsoft has become the default host for AI-native workloads. Every Copilot prompt runs on Azure, and so does every OpenAI API call, even when users never touch a Microsoft front-end.

But the partnership is under strain. OpenAI is pushing its own APIs and enterprise stack. Microsoft wants to be the infrastructure layer for everyone. "They're almost competing now," said one expert, "Microsoft resells GPT. OpenAI is steering people to its own enterprise stack." For now, most usage still benefits Azure, but the lines are blurring.

AWS remains the largest cloud by revenue, but it missed the AI moment. It was late to LLMs, late to inference, and slow to partner with Nvidia. "Amazon has the weakest AI story of the big three," said one expert. "Bedrock is still a framework. Customers want solutions." SageMaker remains underutilised. Trainium and Inferentia have yet to gain real traction, blunting the cost advantages AWS is banking on.

Google has moved faster. Gemini is improving, and TPUs offer performance and cost benefits in the right workloads, but the stack is less open, and developers don't always get what they need out of the box.

The real reshuffle is happening below the Big Three. Oracle leaned into AI infrastructure early and locked in deep ties with Nvidia. CoreWeave, founded in 2017, built a GPU-dense, latency-optimised platform from scratch. Because neither had a legacy business to protect, they could move quickly to seize the moment.

Behind all of this is Nvidia. Supply is tight and it is playing kingmaker, trying to commoditise its largest customers while cultivating a new class of vertically integrated partners. Nvidia is aligned with Oracle and holds equity in CoreWeave. It wants its chips deployed visibly and at scale. So far, it is getting its way.

CHINA'S OPEN FRONTIER

DeepSeek's breakthrough in late 2024 made it clear that U.S. semiconductor sanctions had failed to stop China from building a world-class AI. Its open-source models matched or beat GPT-4 on key benchmarks, while running faster, cheaper, and locally. Its architecture, training efficiency, and speed of deployment are also proof of Chinese firms' ability to innovate at the foundational level.

Meanwhile, Chinese firms are approaching deployment with a sharper commercial lens than many of their American peers. As one expert put it: "They're not chasing AGI. They're building vertical agents that serve a local language, legal system, or customer service script. That's where the money is."

In 2025, we've seen a blitz of real-world adoption across e-commerce, finance, education, industry, and government. Baidu's ERNIE powers maps and search. Alibaba's Tongyi runs marketing and ops. iFLYTEK dominates education. Startups like Zhipu and SenseTime lead in bilingual and visual models.

Tencent, once behind, is catching up fast. Its Hunyuan model now runs across WeChat and Tencent Meeting—powering translation, summarisation, content creation, and search. For specialist tasks, it taps DeepSeek in a modular, mix-and-match approach.

Like Meta, Tencent prioritises scarce compute for internal use, not resale. AI is already lifting Tencent's revenues: ads are more targeted, multiplayer online games are using AI to onboard and retain users, and digital assistants are being deployed across enterprise services. The next big step is a native agent inside WeChat - not a general-purpose chatbot, but a vertically integrated assistant tied to Mini Programs, payments, and the social graph. With so much already embedded in WeChat, that will be a hard product to copy.

CAPEX KEEPS CLIMBING

One thing hasn't changed in the last twelve months: capex keeps rising. But the mix is shifting, from spiky, training-heavy bursts to always-on, product-facing inference infrastructure.

Training spend isn't falling. If anything, the frontier race is accelerating. OpenAI, Google, Meta, Microsoft, xAI, and others are pouring tens of billions into next-gen clusters. That only makes sense if they believe the payoff will be decisive - that the best model wins usage, breaks distribution, and locks in scale.

Inference, meanwhile, is becoming the main driver of incremental infrastructure investment. It is continuous, latency-sensitive, and directly tied to usage. This is reshaping the stack: smarter model placement, lower power draw, edge deployment, and real-time observability. Capex is following suit, into inference appliances, model routers, and regionally distributed systems.

Outside of China, power has become a bottleneck. Compute footprints are massive and still growing, with clusters drawing hundreds of megawatts, enough to power a small city. In the U.S., grid limits, permitting delays, and energy costs now dominate the planning process.

As one expert put it, “We’re not talking about dollars anymore. We’re talking about megawatts. You can’t build a new cluster without a substation.” Amazon has signed multi-gigawatt power deals. Microsoft is exploring nuclear. Meta is re-architecting its data centres. Big Tech is reshaping America’s energy mix.

Looking ahead, training still sets the pace, but inference now shapes the architecture and the budget. Capex is bifurcating. Hyperscalers are building frontier clusters and signing private power deals. Everyone else is focused on delivering faster responses, tighter telemetry, and better token economics for inference workloads. The boom continues, just on different paths.

NVIDIA, ASICS AND AMD

This shift in capex priorities has reopened the debate over compute architectures. Custom silicon promises a way out of Nvidia’s pricing power, especially for inference. But the path is not straightforward.

For hyperscalers, the logic of custom silicon is clear. Nvidia’s pricing power is real and inference costs are spiralling. ASICs offer lower cost, better integration, power efficiency, and more control. But while this sounds ideal on paper, it is hard to deliver in the real world.

Custom chips work best when workloads are stable and scale is extreme. AI is neither. Models have evolved quickly, shifting, for example, from transformers to diffusion, and from instruction-tuned to multimodal. Fixed-function chips, by design, are not built to adapt. If the model shifts, their value evaporates. As one expert put it, “If you spend all this money building something and then you find out the workload changes underneath you, you’re basically stuck.”

GPUs hedge that risk. They are flexible, programmable, and updatable. Moreover, Nvidia’s integration across chips, software, and networking, plus its ability to deliver everything from servers to full data centres, delivers a coherence and efficiency few can match. Its annual product cadence also means that the cost of switching keeps rising.

Nvidia’s NVLink, its high-bandwidth interconnect, underpins training at scale, where GPUs must coordinate across racks. NVLink Fusion, announced this year, may extend that advantage by letting custom chips plug into Nvidia’s system rather than replace it. However, many inference tasks can be handled independently, one GPU at a time. That lowers the importance of networking, and with it, Nvidia’s edge in tightly integrated systems.

This has given AMD a window to become more than a second source. Its MI300X is now deployed at Microsoft, Meta, Oracle, and Dell. In some inference workloads, it beats Nvidia’s H100. As one expert put it, “ROCm used to be a science project. Now we’re finally seeing it run real workloads.” AMD plans to ship full-rack MI400 systems next year. It still trails in training, but inference gives it a real wedge into the market.

AMD is also leaning into openness. ROCm is open source, its interconnects run over Ethernet, not proprietary links, and it is sticking with x86 CPUs. That may appeal to buyers wary of lock-in or reluctant to cross-compile for ARM.

Demand for alternatives is real. But for now, Nvidia still sets the pace. Its software remains unmatched. As one expert put it, “CUDA is Nvidia’s real moat... Even if someone came along with a better chip, you'd need years to match their [ecosytem]”.

It also continues to deliver system-wide. “What you’re really buying is the whole platform, software, networking, orchestration. It’s not just a GPU anymore.” As another expert put it, “What customers want is predictability. Nvidia delivers systems that just show up, plug in, and scale. Nobody else can say that.”

One final advantage is Nvidia’s lock on leading-edge capacity at TSMC, a key enabler of power efficiency and performance. AMD and even well-funded ASIC efforts often find themselves waiting in line.

That said, the risks are real. Nvidia’s growth is tied to hyperscaler capex, which can freeze or shift quickly. ASICs pose a threat in stable, high-volume use cases like video, recommendations, and edge inference. And while Nvidia dominates training, demand may plateau as models get smaller and more efficient. Much of its strength also reflects the fragility of its partners. If OpenAI slows, CoreWeave stumbles, or pricing pressure builds, Nvidia’s leverage could erode.

ALL ROADS STILL LEAD TO TSMC

Underpinning all of this is one constant. TSMC remains the lynchpin of the AI compute stack. Every major chip, whether merchant or custom, still runs through its fabs. No other foundry comes close on yield, throughput, or consistency.

If anything, TSMC’s importance has deepened further, thanks to its own initiative and its competitors’ fumbles. In mid-2024, it unveiled its “Foundry 2.0” strategy, redefining its role beyond traditional manufacturing to include advanced packaging, testing, mask-making, and other services vital to advancing Moore’s Law as transistor density gains slow. By integrating all these aspects of chip design, TSMC can deliver better results for its customers and become even more indispensable.

Nowhere was that shift more visible than in 2024, when TSMC’s 2.5D CoWoS packaging became a chokepoint. Nvidia’s Blackwell hit thermal limits not because of bad silicon, but because packaging couldn’t keep up. TSMC is doubling CoWoS capacity, but the episode showed how sensitive the stack has become.

Of course, TSMC continues to push the limits of manufacturing. Its new N2 process marks a major shift beyond the long-dominant FinFET design to Gate-All-Around, a new structure that offers better control over how electricity flows through each transistor. That means less leakage, less heat, and more room for clever design. These gains don’t come from adding more EUV steps, which have held steady. Instead, they come from smarter layouts, better materials, and structural innovation. The result is higher performance using less power, which is exactly what matters in AI.

TSMC’s lead here is real. Intel’s 18A isn’t in meaningful production. Samsung’s SF2 is still proving itself. Both have roadmaps, but TSMC has the track record - and reputation for customer service. It protects IP, delivers yield, and hits tape-outs. As one customer put it, “They’re boring and they never screw you. That’s worth more than 3 per cent better performance.”

In response to geopolitical risks, TSMC committed to overseas expansion several years ago. This is now bearing fruit, with commercial production underway in Arizona and Kumamoto. These fabs won’t match Taiwan on cost, but customers are willing to pay a premium for the security of supply. Arizona is expected to produce N2 and A16 chips by late 2025, helping anchor future demand.

The risk is complacency, from investors assuming this lead is permanent, or from TSMC assuming no one can catch up. CoWoS is not the only game in town. Intel and Samsung are investing heavily in packaging. If AI shifts toward smaller, cheaper, more distributed models, bleeding-edge demand could ease. But that has not happened yet. Today, TSMC remains the best mix of power, integration, and execution at scale. No matter what kind of chip or who’s designing it, all roads still lead to TSMC.

RE-INVENTING OURSELVES, AGAIN

AI is turning the fixed cost of knowledge work into a variable one. Small businesses like mine can now run leaner. I don't need to hire analysts or consultants when I can subscribe to models that suit my needs. Code, design, research, and support are all available on demand. That changes the cost structure of entrepreneurship. Inevitably, it will do the same for large companies. If they don't adapt, someone else will.

Big Tech already sees it. One expert put it plainly: "We built a team of 40 to do X. Now a single engineer with a good prompt can do it better in an afternoon." Amazon CEO Andy Jassy was blunter: "As we roll out more Generative AI and agents, it should change the way our work is done… We expect that this will reduce our total corporate workforce."

The shift from headcount to hardware will reshape Big Tech's economics: more capex, fewer employees, higher margins but lower free cash flow. Exactly how this plays out will depend on competition, how fast the tech matures and whether productivity gains offset rising depreciation and power costs.

Other sectors will face their own reckoning. History suggests the long-run productivity gains will be real but unevenly distributed. By collapsing a decentralised labour market into a centralised AI one, we are introducing new fragilities and new gatekeepers. Which brings us back to the core question: why are we invested in Meta, Tencent, TSMC, and Nvidia? It's not for AI, per se, but because they remain 'monopolies with growth'.

Growth is the easier case to make, even if there are execution risks and cyclicality. The harder question is whether any edge is defensible in a world of open-source models, falling chip prices, and fast followers. I believe it is. Distribution still beats invention, which favours Meta and Tencent. The bar to compete with TSMC only rises as manufacturing becomes more complex and capital-intensive. And the opportunity cost of not using Nvidia's stack remains steep, if you can even catch up before it ships something better.

AI has already created enormous consumer surplus. But we are still early. Microsoft CTO Kevin Scott recently spoke of a "capability overhang," where today's products underutilise today's models. Even without new breakthroughs, there is still vast room for improvement.

So yes, I joke about being an AI-aholic. But I think the only course is to lean in. AI can make the best of us better. Much else will be commoditised.

I'll leave the last word to Nvidia's CEO, Jensen Huang: "You're not going to lose your job to an AI. But you're going to lose your job to somebody who uses AI."

The same holds true for companies. It's a sobering thought.

As always, thank you for your patience and support.

With my best wishes,

Graham F. Rhodes
Founder & Portfolio Manager
Longriver Investment Partners Limited

LONGRIVER PARTNERS FUND

2ND QUARTER 2025

ABOUT THE FUND:

- Business Owner's mindset to understand and manage investment risk
- Investing globally within our few, narrow areas of expertise, drawing on insights from our experience in Asia
- Prefer long-term investments in companies with predictable business models, ambitious management and good governance, which can re-invest on our behalf at high rates of return
- Concentrated portfolio of 10-20 securities to make the most of scarce ideas, diversified across duration, industry and geography

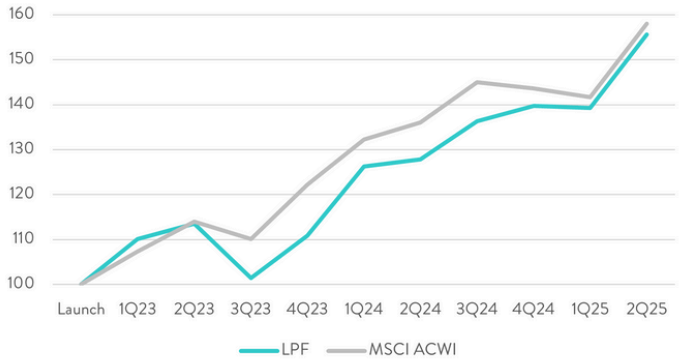
ABOUT THE MANAGER:

- Boutique Hong Kong-based manager with deep roots in Asia
- Major investors in the fund, investing for clients as we do for ourselves
- Spirit of partnership with fair fees, expenses and disclosure
- Employee ownership permits independent judgement
- Focus is always on margin of safety
- Long-term mindset allows for contrarian thinking and accumulation of competitive advantage

NET RETURNS:

Since Inception	Fund	Benchmark	Delta
Cumulative Return	55.6%	58.0%	(2.4%)
Per Annum Return	19.4%	20.2%	(0.8%)

	Q1	Q2	Q3	Q4	YTD	BMK	DELTA
2025	(0.3%)	11.7%	-	-	11.4%	10.1%	1.3%
2024	13.9%	1.3%	6.6%	2.5%	26.0%	17.5%	8.5%
2023	10.1%	3.1%	(10.7%)	9.3%	10.8%	22.2%	(11.4%)



TOP HOLDINGS & PORTFOLIO CHARACTERISTICS:

FUTU | GAMES WORKSHOP | HIKARI TSUSHIN | META | PLOVER BAY TECHNOLOGIES | TENCENT | TOPICUS.COM | TSMC | VITEC SOFTWARE | WISE PLC

Portfolio:		Duration:	
Top 10	89%	Hypergrowth:	-%
25E PE:	24x	Growth:	67%
25E ROE:	34%	Income:	-%
25E EPSg:	22%	Cyclical:	23%
25E Yield:	1%	Cigar Butt:	10%
LTM D/E:	(31%)	Fixed Income:	-%
		Cash:	-%

Industry:		Geography:	
Digital:	38%	Global:	59%
Consumer:	33%	Asia:	27%
Financial Services:	24%	Europe:	14%
Industrial:	5%	Americas:	-%

Notes: Benchmark is the MSCI AC World USD Index. Longriver net performance figures are unaudited. Portfolio characteristics are as at 30 June 2025. Duration, Industry and Geography characteristics are as defined by the Manager. Duration represents the certainty and timing of returns, with cash at one end of the spectrum and companies which re-invest all or more of their earnings into hypergrowth at the other. Portfolio, Industry and Geography metrics exclude cash. Portfolio LTM D/E excludes banks.

KEY INVESTMENT TERMS:

Manager:	Longriver Investment Partners Limited
Mandate:	Long-only, global, primarily equities
Launch:	January 3rd, 2023
Minimum:	USD100,000
Fees:	1.00% p.a. + 10% over 6% hurdle + High Water Mark
Liquidity:	Monthly subscriptions; quarterly redemptions with sixty days written notice

OPERATIONS:

Structure:	Hong Kong Open-ended Fund Company
Regulator:	Hong Kong Securities & Futures Commission
Admin.:	Sinopac Services & Solutions Limited
Custodian:	Interactive Brokers Hong Kong Limited
Auditor:	East Asia Sentinel Limited
Legal:	Charltons
Contact:	grahamfrhodes@longriverinv.com