European Semiconductors

Profiling Private Semiconductor Companies - 2nd Edition

Equities

Global

Semiconductors

Harry Blaiklock, CFA
Analyst
harry.blaiklock@ubs.com
+44-20-7568 5385

Francois-Xavier Bouvignies
Analyst
francois.bouvignies@ubs.com
+44-20-7568 7105

Private Semiconductor Company Profiles in the AI compute space

In 2024 we profiled the five largest private semiconductor players in Europe, one of which, Graphcore, was acquired by SoftBank for $625m in 2024 and another, Nexperia, was at the heart of geopolitical supply disruption in 2025, displaying the relevance of the private space to the wider sector. In this note we profile a further ten private semiconductor companies, with a specific focus on the AI compute market.

The drive for compute alternatives beyond merchant GPUs

All of these companies have positioned themselves as alternatives to merchant GPUs currently used in AI data centers. As AI usage scales and workloads increasingly shift from training to inference, energy consumption, the "memory wall" and supply chain constraints have become critical limits. New architectures aim to deliver better performance per watt, lower cost per token, and more sustainable, scalable compute, particularly for inference and edge deployments where GPUs may be less optimal.

Proof points from the shift to custom ASICs

While GPUs still account for the majority of the installed base of AI compute, custom ASICs from hyperscalers are starting to represent an increasing share. The UBS Global Semis team estimates that the AI cloud ASIC market will grow at a c70% CAGR, expanding from $17bn in 2025 to >$100bn by 2028. ASICs can deliver superior workload specific performance and lower total cost through tailored architectural optimisations and are often deployed alongside GPU platforms within large scale data centres rather than as direct replacements.

Two emerging architectures - in-memory compute and optical compute

In-memory and optical computing are emerging architectures designed to overcome the power and performance limitations of traditional processors. In-memory computing reduces the need to move data between memory and compute by performing calculations within or adjacent to memory, while optical computing uses photons instead of electrons to execute key AI workloads with higher efficiency. These approaches are the focus of several of the profiled private companies.

How can privates impact public?

Private AI compute companies can influence public companies by reshaping expectations around performance, cost, and architecture. They can emerge as acquisition or partnership targets for incumbents looking to secure talent, IP, or architectural options in inference, photonics, memory centric compute, or interconnects, prompting earlier stage M&A or minority investments. Even without commercial scale, their architectures could help validate investment in adjacent technologies from larger public peers.

Private companies highlighted

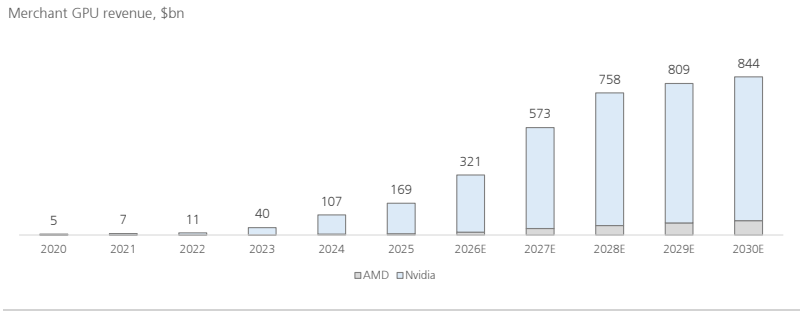
We highlight ten private companies within the AI compute space including Arago, Axelera, Euclyd, Fractile, Lumai, Olix, Optalysys, Q.ANT, SiPearl and VSORA.

This report has been prepared by UBS AG London Branch. **ANALYST CERTIFICATION AND REQUIRED DISCLOSURES, INCLUDING INFORMATION ON THE QUANTITATIVE RESEARCH REVIEW PUBLISHED BY UBS, BEGIN ON PAGE 19.**

AI accelerators - the heart of the AI Boom

Over the past five years graphics processor units or GPUs, have evolved from a niche gaming console component to the hardware underpinning the AI boom. Unlike traditional CPUs, the parallel processing ability of GPUs makes them significantly more efficient for AI training. GPU revenue has grown rapidly over the same period from $5bn in 2020 to $169bn 2025, with consensus expecting $844bn by 2030.

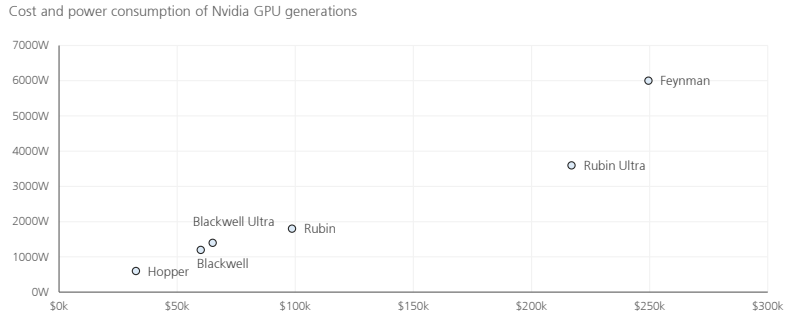
Figure 1: GPUs have been the key driver of the AI boom



Source: VA cons, company data

As the AI infrastructure buildout continues, several factors are driving the rise of alternative solutions. Cost and power consumption of GPUs has continued to climb, as shown below. As AI workloads shift from training to inference, more specialised chips such as custom ASICs are often better suited. Challenges like the memory or networking wall are also driving players to look at alternative architectures to GPUs.

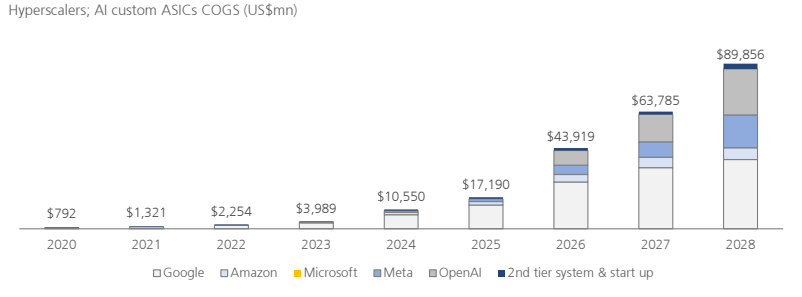
Figure 2: traditional accelerators are becoming increasingly expensive and energy hungry



Source: UBS estimates, company data

These alternatives have largely come from publicly listed players, specifically in the form of custom ASICs from hyperscalers. This has driven the size of the custom ASICs market from $0.8bn in 2020 to $17bn in 2025 and it is set to grow to $90bn by 2028. Increasingly, though, many private companies are also disrupting the space. Cerebras, which builds wafer scale processors, recently listed at a $95bn valuation, and another, Groq, entered a $20bn licensing deal with Nvidia for its inference chips.

Figure 3: the industry is increasingly turning towards custom ASICs



Source: UBS estimates, company data

The drive for alternatives to GPUs

These alternatives to GPUs are being sought because AI workloads are becoming increasingly constrained by power consumption, memory bandwidth, availability and cost. GPUs are the most flexible and widely supported option but that flexibility is often not needed for predictable, high-volume workloads. Purpose-built accelerators or alternative compute methods can optimise data movement, logic and memory architecture for specific training or inference tasks, potentially delivering better performance per watt and lower cost per token.

What is a custom ASIC ?

A custom ASIC is a chip designed for a particular workload, rather than the broad range of tasks handled by a general-purpose CPU or GPU. By focussing only the functions needed, for example, AI tensor operations, it can deliver better speed and energy efficiency, but sacrifices flexibility.

Figure 4: key differences between GPUs and custom ASICs

	GPUs	Custom ASICs
1. Workload flexibility	Highly programmable; the same hardware can support training, inference, graphics and scientific computing.	Optimised around a narrower set of workloads and less able to accommodate major architectural changes.
2. Performance on target workload	Strong across a broad range of parallel workloads, but carries some general-purpose overhead.	Can deliver higher performance when the workload closely matches the chip's purpose-built architecture.
3. Energy efficiency	Generally lower performance per watt because silicon resources support a wider range of functions.	Generally higher performance per watt by removing unnecessary circuitry and optimising data movement and computation.
4. Training versus inference	Usually preferred for rapidly evolving, heterogeneous training workloads; also capable at inference.	Particularly attractive for high-volume, repeatable inference, although some designs support both training and inference.
5. Upfront development cost	No bespoke chip-development expense for the user; hardware can be bought or rented as a standard product.	High non-recurring engineering cost for architecture, verification, tape-out and software development.
6. Cost at scale	Higher flexibility reduces utilisation risk, but acquisition and energy costs can be significant.	Potentially lower cost per inference at high utilisation and volume, once substantial upfront costs are absorbed.
7. Time to deploy or adapt	Faster to deploy and can support new models through software and compiler updates.	Longer design cycle, with silicon changes required if future workloads exceed the original design envelope.
8. Software ecosystem and portability	Mature, broad ecosystem, particularly NVIDIA CUDA, with extensive framework, library and developer support.	Typically a narrower, vendor-specific software stack, increasing optimisation requirements and possible platform lock-in.

Source: UBS

Two of the most prominent custom ASICs come from Google and Amazon, who are focussing on reducing reliance on GPUs and optimising performance, power consumption and cost for their own cloud and AI workloads. Google's Tensor Processing Units (TPUs) are purpose-built for AI training and inference and tightly integrated with its software and data centre infrastructure, while Amazon uses Trainium for model training and Inferentia for inference.

Beyond mainstream custom ASICs

The private companies we profile in this note address multiple levels of the AI value chain and do so in a wide variety of ways. VSORA offers perhaps the closest equivalent to more traditional custom ASICs outlined above, while others are proposing novel architectures that differ drastically from the current mainstream. Optalysys, for example, is developing programmable photonic processing, which will enable compute in transit. We summarise and group these companies at a high level below.

Figure 5: summarising our profiled private companies

	Company	HQ	Comment
Optical compute	Arago	France	Arago positions its hybrid electronic-photonic JEF accelerator as a practical, energy-efficient alternative to GPUs for AI inference, combining optical compute with conventional electronics rather than pursuing an all-optical design.
	Lumai	UK	Optical compute is aimed primarily at the compute-bound prefill phase, silicon scaling limits and the data-centre power, cooling and rack-space ceiling
	Olix	UK	Olix is designing an optical-digital inference processor around SRAM and a new interconnect architecture. Its distinctive claim is that it can avoid HBM and advanced-packaging constraints
	Optalysys	UK	Its "compute-in-transit" architecture puts programmable photonic compute into the optical data path. It explicitly targets data movement, energy/conversion overhead and scaling cost
	Q.ANT	Germany	Q.ANT positions itself as a full-stack photonic-computing company, combining thin-film lithium-niobate chips, a proprietary pilot line and deployable server systems to offer energy-efficient AI and HPC acceleration alongside conventional processors.
In-memory compute	Fractile	UK	Fractile's "memory-compute fusion" approach places or interleaves memory close to compute to reduce weight movement during inference. The pitch is faster token generation, lower energy and cost
	Axelera AI	Netherlands	Digital in-memory compute performs matrix operations inside SRAM, reducing intermediate data movement and power. The emphasis is affordable, compact edge inference
Conventional compute	SiPearl	France	Offers a high-performance, energy-efficient European CPU for HPC, AI and sovereign infrastructure. Its on-package HBM targets bandwidth-limited workloads
	Euclyd	Netherlands	Tackles compute, memory bandwidth and memory capacity together using custom processors, custom ultra-bandwidth memory and advanced packaging.
	VSORA	France	Positioned as breaking the memory wall for large-scale inference, reducing idle compute and power draw while improving cost per query

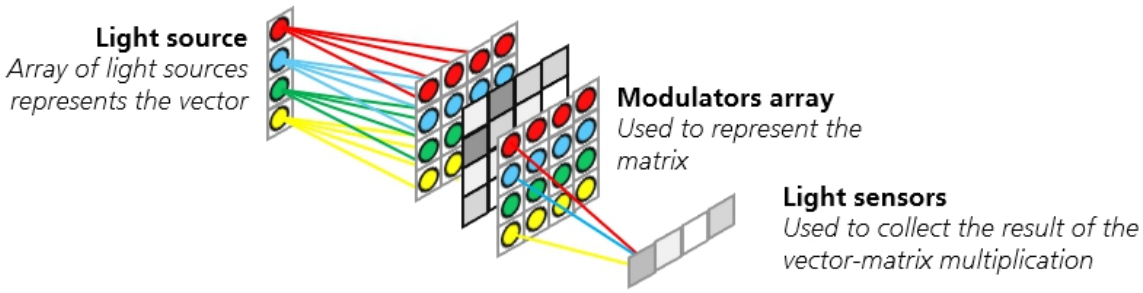
Source: Company data, UBS

Below we explore in greater detail two of the more novel approaches pursued by these private companies, namely optical compute and in-memory compute.

Optical compute

Optical computing uses light, rather than electrical signals, to perform or transmit calculations. In a conventional processor, information is represented by electrical signals moving through transistors and wires. An optical system instead encodes information into properties of light and passes it through optical components that transform those signals. The result is then detected and converted back into an electronic form.

Figure 6: key elements of an optical processor



Source: UBS

The technology is particularly relevant to artificial intelligence because neural networks repeatedly perform vector-matrix multiplications, where large grids of numbers are multiplied together to process model inputs and generate outputs. These operations are highly repetitive and parallel, making them well suited to optical processing. Multiple beams or wavelengths of light can travel through a system simultaneously, allowing many calculations to be performed in parallel rather than sequentially.

Why could it be attractive?

The principal attraction is the potential for high throughput with lower power consumption. Conventional AI processors consume energy not only when performing calculations, but also when moving data between memory and compute units. This generates heat and places demands on cooling and datacentre power infrastructure. Optical systems seek to execute the most computationally intensive operations with less electrical switching and greater parallelism.

Figure 7: comparison of electronic vs. optical compute with key benefits in green

	Electronic compute	Optical compute
Information carrier	Electrons/voltage	Photons/light waves
Basic machinery	Transistors and logic gates	Lasers, modulators, waveguides, detectors
Best at	General-purpose, precise digital logic	Highly parallel, repetitive maths and data movement
Parallelism	Many electronic cores	Multiple beams and wavelengths can operate together
Heat	Switching and resistance generate heat	Can be very efficient, although lasers, conversion and control still consume power
Memory	Mature electronic SRAM/DRAM	Practical optical memory is much less mature
Programmability	Extremely flexible	Usually optimised for a narrower set of operations
Output	Already electronic	Normally detected and converted back to electronic

Source: UBS

However, optical computing should not be thought of as an entirely light-based replacement for a GPU. Current designs are generally hybrid optical-electronic systems. Optical hardware handles the matrix-heavy part of the workload, while conventional electronics continue to provide memory, system control, software execution and operations that are less naturally suited to light. Lumai, for example, describes its Iris architecture as using an optical matrix multiplier for the compute bottleneck while digital electronics handle the rest of the system.

What still needs to be proven?

The challenge is turning impressive component-level performance into a reliable and economical complete system. Data must often be converted from electronic to optical form before a calculation and then converted back afterwards. Those conversions consume power and can reduce the benefit of the optical calculation itself. Ultimately customers will care less about the individual component and more about end-to-end performance, power consumption, latency, cost, software compatibility and reliability of the system as a whole.

In memory compute

In-memory computing is an alternative processor architecture that performs calculations inside, or immediately alongside, the memory storing the data. In a conventional processor, memory and compute are separate, data is retrieved from memory, transferred to the processor, used in a calculation and then written back. In-memory computing reduces these repeated journeys by allowing the calculation to occur in the memory structure itself, or in compute circuitry integrated very closely with it. The approach is particularly suited to the matrix-vector multiplication used extensively in neural networks. Model weights can be stored as a matrix in a memory array, inputs are applied to that array and many multiply-and-accumulate operations are performed in parallel where the weights already reside.

SRAM, or static random-access memory, is a common foundation for in-memory designs. It is fast and can be integrated on the processor die, but it occupies more silicon area than denser off-chip memories. The central economic question is therefore whether reductions in data movement can compensate for the cost and capacity constraints of placing more memory close to the compute. SRAM-based in-memory computing is being studied precisely because its proximity to processing elements can reduce memory-access latency for data-intensive applications such as AI.

Why could it be attractive?

In-memory compute could be attractive because it addresses the energy, latency and cost involved in repeatedly moving data between separate memory and processing units. The architecture can reduce data movement and improve throughput and energy efficiency, particularly for the matrix-heavy inference workloads used by modern neural networks. This could become increasingly valuable as models grow, inference volumes rise and power and cooling become more binding constraints in both datacentres and edge devices. In-memory compute may therefore offer a path to lower cost per inference and better performance per watt than conventional architectures.

What still needs to be proven?

Vendors need to demonstrate reliable performance on current, evolving AI models after accounting for peripheral circuitry, data conversion, control logic and communication overhead. They must also show that the technology can deliver sufficient numerical accuracy, scale to commercially useful workloads, achieve acceptable manufacturing yields and reliability, and integrate with mainstream software frameworks without excessive model redesign. Ultimately, adoption will depend on evidence that these architectures can be produced and deployed at scale while sustaining a meaningful advantage in cost per inference, throughput and power consumption versus rapidly improving GPUs and conventional accelerators.

Figure 8: comparison of profiled private companies offerings vs. mainstream GPUs and custom ASICs

	Company	Processor	Primary workload	Architecture / process	Peak compute per processor or package	Memory	Memory bandwidth	Power / implied efficiency	Product status
GPUs	NVIDIA	Rubin GPU	Training + inference	General-purpose GPU; dual-die; TSMC 3nm-class	50 PFLOPS NVFP4 inference; ~35 PFLOPS NVFP4 training; published third-party figures indicate ~17.5 PFLOPS FP8	288 GB HBM4	22 TB/s	2.3 kW per GPU package	Full production
	AMD	Instinct MI350X	Training + inference	CDNA 4; chiplets; TSMC 3nm/6nm	9.23 PFLOPS MXFP4/MXFP6; 4.61 PFLOPS FP8 with sparsity; 2.31 PFLOPS dense FP8	288 GB HBM3E	8 TB/s	1kW per GPU package	Full production
Established custom ASICs	Google	TPU7x	Training + inference	7th-generation TPU ASIC; 2 TensorCores + 4 SparseCores per chip; dual-chiplet	4.6 PFLOPS FP8; 2.3 PFLOPS BF16	192 GiB HBM3E	7.38 TB/s	980W per TPU	Full production
	AWS	Trainium3	Training + inference	3nm AWS ASIC; 8 NeuronCore-v4 cores	2.5 PFLOPS MXFP8/MXFP4; 0.7 PFLOPS BF16/FP16/TF32; 0.2 PFLOPS FP32	144 GiB HBM3E	4.9 TB/s	500W per chip	Full production
Optical compute	Arago	JEF photonic AI accelerator	Inference	Hybrid optical/electronic "multi-physics" computing core; CARLOTA software stack interfaces with PyTorch	Multi-PetaOp/s claimed	Not disclosed	Not disclosed	Claims 10×–30× lower energy consumption than conventional architectures	Development-stage
	Lumai	Iris Nova / Iris platform	Inference	Hybrid free-space optical matrix multiplier + digital processor; matrix size up to 2,048 × 2,048	Platform target of 1 ExaOPS in 10 kW; INT4/INT8-equivalent precision	Three-tier architecture referenced; absolute capacity not disclosed	Not disclosed	Claims up to 90% lower energy than conventional architectures	Customer sampling
	Olix	Optical Tensor Processing Unit	Inference	Hybrid photonic/electronic processor; model held in SRAM, deliberately avoiding HBM and incumbent advanced-packaging constraints	Not disclosed	SRAM-based	Not disclosed	Not disclosed	Development-stage.
	Optalysys	Compute-in-Transit platform	In-network compute	Programmable silicon photonics embedded in the optical transport/data path	Not disclosed	Not disclosed	Not disclosed	Not disclosed	Development-stage.
	Q.ANT	Native Processing Unit Gen 2	Inference	Analog photonic processor	8 GOPS sustained throughput on nonlinear functions; not directly comparable with GPU tensor FLOPS	Host/server memory	Not disclosed	Claims 30× greater energy efficiency than conventional architectures	Sampling
In-memory compute	Fractile	Inference processor	Inference	Physically interleaved memory + compute; in-memory arithmetic; Andes AX45MPV 64-bit RISC-V vector control processor	Not disclosed - company target is up to 25× faster inference and 1/10 the cost of current systems	On-chip/in-memory architecture	Not disclosed	Not disclosed	Development-stage.
	Axelera AI	Europa AIPU	Edge inference	8 second-generation digital in-memory-compute AIPUs + 16 RISC-V vector cores	629 TOPS INT8; also supports INT4/INT16; auxiliary vector performance quoted at 4,915 GOPS	128 MB on-chip L2 SRAM; external LPDDR5 capacity not stated	200 GB/s LPDDR5	45W per AIPU	Development-stage.
Conventional compute	SiPearl	Rhea1	Inference CPU	80 Arm Neoverse V1 cores; 2 × 256-bit SVE units/core; TSMC N6	Not disclosed	4 integrated HBM2E stacks + 4 DDR5 interfaces, 2 DIMMs/channel	Not disclosed	Not disclosed	Sampling
	Euclyd	CRAFTWERK SiP	Inference	16,384 custom 128-way SIMD processors; large multi-chiplet SiP; advanced 2.5D/3D packaging	32 PFLOPS FP4; 8 PFLOPS FP16	1 TB custom UBM	Claimed 8,000 TB/s	3 kW per SiP	Development-stage.
	VSORA	Jotunn8	Inference	Chiplet architecture; TSMC 5nm, CoWoS-S, 3× reticle-size package	3.2 PFLOPS	288 GB HBM3E	Not disclosed	Claims 50% less power than "current market-leading chips"	Ramping production

Source: UBS, company data

Optical Compute

PRIVATE COMPANY PROFILE: ARAGO

COMPANY NAME	Arago
SECTOR / SUB-SECTOR	Semiconductor, Compute, Photonics
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** France **Presence: US, Israel**
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD **Private:** Lightmatter, Celestial AI, Optalysys, Etched
KEY SUPPLIERS / CUSTOMERS	--
COMPETITIVE POSITIONING	Arago positions its hybrid electronic-photonic JEF accelerator as a practical, energy-efficient alternative to GPUs for AI inference, combining optical compute with conventional electronics rather than pursuing an all-optical design. Its claimed differentiation is compatibility with existing AI software, infrastructure and manufacturing,
YEAR FOUNDED / ESTABLISHED	2024
WEBSITE	https://www.arago.inc/
MANAGEMENT TEAM / FOUNDER(S)	**Nicolas Muller (Co-founder & CEO):** MIT-trained in machine learning; previously worked at deep-tech start-ups, including as growth lead at a quantum/semiconductor company **Ambroise Müller (Co-founder & CSO):** Computational-physics researcher and former AI algorithm engineer at Eviden, with a background from École Polytechnique.
SIZE OF COMPANY	c80 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Seed; $26m; July 2025 **Investors:** Earlybird Venture Capital, Protagonist, Visionaries Tomorrow
LATEST ROUND VALUATION	--
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: LUMAI

COMPANY NAME	Lumai
SECTOR / SUB-SECTOR	Semiconductor, Compute, Photonics
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** UK **Presence:** California, Europe
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Intel **Private:** Lightmatter, Celestial AI, Optalysys
KEY SUPPLIERS / CUSTOMERS	--
COMPETITIVE POSITIONING	Lumai is differentiated by its free-space, 3D optical architecture, which performs large matrix multiplications for LLM inference without relying on conventional GPUs or integrated photonics; the company claims up to 50× GPU performance with 90% lower power consumption. Its competitive advantage is potentially superior inference economics under data-centre power constraints. The first Iris Nova systems are currently offered for customer evaluation rather than volume deployment.
YEAR FOUNDED / ESTABLISHED	2022
WEBSITE	https://lumai.ai/
MANAGEMENT TEAM / FOUNDER(S)	**Tim Weil (CEO, Co-founder):**Deep-tech entrepreneur focused on scaling advanced computing ventures **Xianxin Guo (Co-founder, Head of Research):** Academic expert in optical neural networks. **Founders:** Xianxin Guo, Alex Lovsky, Thomas Barrett, James Spall
SIZE OF COMPANY	c50 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Early-stage VC; $10m; July 2025 **Investors:** Constructor Capital, IP Group, PhotonVentures, Journey Ventures, LIFTT, Qubits Ventures, State Farm Ventures, and TIS Inc.
LATEST ROUND VALUATION	--
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: OLIX

COMPANY NAME	Olix
SECTOR / SUB-SECTOR	Semiconductor, Compute, Photonics
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** UK **Presence:** US, Canada
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD **Private:** Lightmatter, Celestial AI, Etched, Optalysys
KEY SUPPLIERS / CUSTOMERS	--
COMPETITIVE POSITIONING	Olix is positioned as a differentiated AI-inference challenger. Its Optical Tensor Processing Unit combines digital photonics with SRAM to address data-movement bottlenecks while avoiding HBM, advanced packaging, and leading-edge silicon dependencies. Relative to Nvidia and photonic peers such as Lightmatter and Celestial AI, its claimed advantage is higher throughput and interactivity at lower cost and energy use. Planned product shipments are for 2027.
YEAR FOUNDED / ESTABLISHED	2024
WEBSITE	https://olix.com/
MANAGEMENT TEAM / FOUNDER(S)	**James Dacombe (CEO, Founder):** Serial entrepreneur; previously founded neurotech company CoMind **Tom Elvidge (COO):** experienced technology operator who has held senior leadership roles at Uber, WeWork and Arrival
SIZE OF COMPANY	c150 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Series A; $312m; August 2026 **Investors:** Arm, Hummingbird Ventures, Plural, LocalGlobe, Vertex Ventures, and Entrepreneurs First
LATEST ROUND VALUATION	Exceeded $3bn
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: OPTALYSYS

COMPANY NAME	Optalysys
SECTOR / SUB-SECTOR	Semiconductor, Compute, Networking, Photonics,
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** UK **Presence:** UK
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Intel, Marvell (Celestial AI) **Private:** Lightmatter, Lumai, Ayar Labs
KEY SUPPLIERS / CUSTOMERS	--
COMPETITIVE POSITIONING	Optalysys differentiates itself by combining programmable silicon-photonic compute with a particular focus on fully homomorphic encryption, targeting secure AI and cloud workloads where data can be processed while remaining encrypted. Its potential edge is energy-efficient, privacy-preserving computation and compatibility with existing software ecosystems, but it remains pre-revenue and commercially earlier than scaled accelerator vendors.
YEAR FOUNDED / ESTABLISHED	2013
WEBSITE	https://optalysys.com/
MANAGEMENT TEAM / FOUNDER(S)	**Dr Nick New (CEO, Co-founder):** Physicist and serial deep-tech founder in photonics **Robert Todd (CTO, Co-founder):** Optical systems architect specializing in photonic compute.
SIZE OF COMPANY	c30 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Series A; $31m; January 2026 **Investors:** Northern Gritstone, imec.xpand, Lingotto Horizon, UK NSSIF
LATEST ROUND VALUATION	NA
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: Q.ANT

COMPANY NAME	Q.ANT
SECTOR / SUB-SECTOR	Semiconductor, Compute, Photonics
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** Germany **Presence: US**
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Intel **Private:** Lightmatter, Optalysys, Akhetonics and Neurophos
KEY SUPPLIERS / CUSTOMERS	**Suppliers:** IMS CHIPS **Customers:** IONOS
COMPETITIVE POSITIONING	Q.ANT positions itself as a full-stack photonic-computing company, combining thin-film lithium-niobate chips, a proprietary pilot line and deployable server systems to offer energy-efficient AI and HPC acceleration alongside conventional processors. Its differentiation versus other optical-compute start-ups is its emphasis on commercially deployable analog co-processors with relatively high numerical precision.
YEAR FOUNDED / ESTABLISHED	2018
WEBSITE	https://www.qant.com/
MANAGEMENT TEAM / FOUNDER(S)	**Michael Förtsch (Founder & CEO):** Mathematician and physicist with a PhD in quantum information processing. Previously worked in corporate development and as executive assistant to TRUMPF's CTO before founding Q.ANT in 2018 **Bruno Spruth (CTO):** A semiconductor veteran and former IBM Vice President of Power Processor Development, Spruth leads Q.ANT's technology strategy and product development, as well as its North American business. **Tim Stiegler (CFO):** An experienced finance leader who joined Q.ANT after holding senior controlling and finance roles across the TRUMPF group.
SIZE OF COMPANY	c200 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Series A; €62m; July 2025 **Investors:** Cheery Ventures, UVC Partners, imec.xpand, TRUMPF
LATEST ROUND VALUATION	--
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website



In-memory compute

PRIVATE COMPANY PROFILE: AXELERA AI

COMPANY NAME	Axelera AI
SECTOR / SUB-SECTOR	Semiconductor, Compute, Edge AI
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** Netherlands **Presence:** Europe
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Intel **Private:** Hailo, DeepX, EdgeCortix, SiMa.ai, Kneron
KEY SUPPLIERS / CUSTOMERS	**Suppliers:** TSMC, Samsung **Customers:** 500+ across industrial, defense, telecom and retail
COMPETITIVE POSITIONING	Axelera AI is positioned as an edge-first AI-inference player, differentiating through digital in-memory compute and an integrated hardware/software stack that targets high performance within tight power, thermal and cost constraints. Versus general-purpose GPU leaders such as Nvidia and AMD and specialist accelerators such as Hailo it aims to compete less on training breadth and more on efficient local inference, privacy/ data sovereignty and ease of deployment.
YEAR FOUNDED / ESTABLISHED	2021
WEBSITE	https://axelera.ai/
MANAGEMENT TEAM / FOUNDER(S)	**Fabrizio Del Maffeo (CEO, Co-founder):** previously Head of AI at Bitfury Group, a global technology incubator; VP and MD of AEEON Technology Europe, the AI and IoT computing company within ASUS Group **Evangelos Eleftheriou (CTO, Co-founder):** previously worked for IBM Research over a 35 year period reaching the highest technical honour of becoming an IBM Fellow. He is also a Fellow of the IEEE. **Founders:** Evangelos Eleftheriou, Fabrizio Del Maffeo, Bram Verhoef, Ioannis Papistas, Marco Barbato
SIZE OF COMPANY	c285 employees
GROWTH TRAJECTORY	Annual revenue estimates of $10-70m
ANY MAJOR M&A / CORPORATE ACTIONS	NA
INVESTORS / INVESTMENT STAGE	**Investment stage:** Series C; $250m; February 2026 **Investors:** Innovation Industries, SiteGround Capital, BlackRock, Samsung Catalyst Fund
LATEST ROUND VALUATION	Estimated $1-1.5bn
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	NA

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: FRACTILE

COMPANY NAME	Fractile
SECTOR / SUB-SECTOR	Semiconductor, Compute
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** UK **Presence:** UK
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Cerebras **Private:** Groq, SambaNova, Etched, Tenstorrent
KEY SUPPLIERS / CUSTOMERS	**Suppliers:** -- **Customers:** Anthropic
COMPETITIVE POSITIONING	Fractile is building a full-stack inference system that uses in-memory compute, aiming to overcome the data-movement bottleneck while delivering low latency and high throughput for frontier AI models. It claims up to 25× faster inference at one-tenth the cost of existing hardware.
YEAR FOUNDED / ESTABLISHED	2022
WEBSITE	https://www.fractile.ai/
MANAGEMENT TEAM / FOUNDER(S)	**Dr Walter Goodwin (CEO, Founder):** Oxford-trained AI researcher focused on hardware acceleration for large-scale inference **Matt Arnold (CFO):** previously held senior roles at Goldman Sachs, Oxford Science Enterprises and Mind Foundry
SIZE OF COMPANY	c120 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Series B; $220m; May 2026 **Investors:** Accel, Factorial Funds, and Founders Fund
LATEST ROUND VALUATION	--
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

Other Compute

PRIVATE COMPANY PROFILE: EUCLYD

COMPANY NAME	Euclyd
SECTOR / SUB-SECTOR	Semiconductor, Compute
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** Netherlands **Presence:** US
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD **Private:** Groq, Cerebras, Axelera, Fractile
KEY SUPPLIERS / CUSTOMERS	**Suppliers:** Samsung Foundry **Customers:** NA
COMPETITIVE POSITIONING	Euclyd is positioning itself as a next-generation AI inference chip company focused on delivering better power efficiency and lower cost per token than Nvidia GPUs for large-scale AI inference workloads. It differentiates through a compute-near-memory architecture, called Craftwerk, targeting data center operators that need cheaper and more energy-efficient AI deployment.
YEAR FOUNDED / ESTABLISHED	2024
WEBSITE	https://euclyd.ai/
MANAGEMENT TEAM / FOUNDER(S)	**Bernardo Kastrup (CEO, Founder):** Former ASML executive and technology entrepreneur. Leads the company's strategy, fundraising, and product vision. **Founders:** Gerard Egelmeers, Ingolf Held, Harm Peters
SIZE OF COMPANY	c20 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Seed; €10m **Investors:** Peter Wennink ex-ASML
LATEST ROUND VALUATION	--
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: SIPEARL

COMPANY NAME	SiPearl
SECTOR / SUB-SECTOR	Semiconductor, Compute, CPU
SIZE OF TOTAL ADDRESSABLE MARKET	According to UBS estimates, the server CPU market is currently worth ~$20B/year growing to > $100B by 2030E
GEOGRAPHIC PRESENCE	**HQ:** France **Presence:** Europe
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Intel **Private:** Ampere, Tenstorrent
KEY SUPPLIERS / CUSTOMERS	**Suppliers:** TSMC **Customers:** Jupiter exascale supercomputer
COMPETITIVE POSITIONING	SiPearl is positioned as Europe's sovereign alternative in high-performance server CPUs, with its Arm-based Rhea1 combining 80 cores, in-package HBM and energy efficiency for HPC and AI inference while remaining interoperable with third-party accelerators. Its differentiation versus Intel, AMD and Nvidia is less about a broad commercial ecosystem and more about European-controlled design, security, memory bandwidth and strategic adoption in EuroHPC systems such as JUPITER. Its product Rhea1 is scheduled for general availability at the end of 2026.
YEAR FOUNDED / ESTABLISHED	2019
WEBSITE	https://sipearl.com
MANAGEMENT TEAM / FOUNDER(S)	**Philippe Notton (CEO, Founder):** Former Atos/Bull HPC executive driving EU semiconductor sovereignty
SIZE OF COMPANY	c200 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	--
INVESTORS / INVESTMENT STAGE	**Investment stage:** Series A; E130m; July 2025 **Investors:** EIC Fund, France 2030, Cathay Venture, Arm, Atos/Eviden
LATEST ROUND VALUATION	--
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	--

Source: Crunchbase and Company's reports and website

PRIVATE COMPANY PROFILE: VSORA

COMPANY NAME	VSORA
SECTOR / SUB-SECTOR	Semiconductor, Compute
SIZE OF TOTAL ADDRESSABLE MARKET	Global AI inference market of $255bn by 2030
GEOGRAPHIC PRESENCE	**HQ:** France **Presence:** Global
NOTABLE COMPETITORS	**Listed:** Nvidia, AMD, Cerebras, Alphabet, Amazon **Private:** Groq, Etched, SambaNova and Tenstorrent.
KEY SUPPLIERS / CUSTOMERS	**Suppliers:** TSMC, GUC **Customers:** --
COMPETITIVE POSITIONING	VSORA positions Jotunn8 as a purpose-built inference chip that aims to address the memory wall through continuous data flow, high embedded memory and chiplet scalability, targeting high throughput and low latency at lower power and cost per query. Its differentiation versus general-purpose GPUs and training-led architectures is an inference-only design, claimed at 3,200 TFLOPS with 288GB HBM3e and 50% lower power than current market-leading chips.
YEAR FOUNDED / ESTABLISHED	2015
WEBSITE	https://vsora.com/
MANAGEMENT TEAM / FOUNDER(S)	**Khaled Maalej (CEO, Co-founder):** Former Texas Instruments executive with deep DSP/ semiconductor experience
SIZE OF COMPANY	c200 employees
GROWTH TRAJECTORY	--
ANY MAJOR M&A / CORPORATE ACTIONS	NA
INVESTORS / INVESTMENT STAGE	**Investment stage:** Seed; July 2026 **Investors:** Ardian Semiconductor and Otium Capital led round
LATEST ROUND VALUATION	NA
VALUATION APPROACHES / METHODOLOGIES USED IN THE SECTOR	For early stage compute companies, we believe it is appropriate to use P/S or EV/S valuation multiples, as profitability may vary as the company scales. P/E is typically used for listed peers as given a more mature profitability profile.
SECONDARY PRICING DATA	NA

Source: Crunchbase and Company's reports and website

Valuation Method and Risk Statement

The upside risks to the semiconductor sector include stronger end demand from OEMs and tightness of supply due to the financial distress of competitors. Downside risks include macro-economic factors, over-capacity in times of peaking demand and poor yields. The semiconductor sector is highly cyclical and vulnerable to sudden shifts in customer sentiment while many companies also have high cost bases meaning they can go loss-making in the downturn.