

LIGHTHOUSE CANTON THEMATIC RESEARCH · ISSUE 01

— POWER SEMICONDUCTORS · THE OTHER AI TRADE

From Torque to *Tokens*.

Power Semiconductors — The Other AI Trade

Why the silicon carbide and gallium nitride supply chain underwrites every megawatt of AI compute — and why the market is still pricing it as a recovering EV cyclical.

POWER SEMI TAM

\$55–57B

2025A → \$77–95B by 2031–35E

SIC DEVICES

20% CAGR

\$3B → \$10.3B by 2030E (Yole)

GAN DEVICES

~42% CAGR

\$1.2B → \$3B by 2030E (Yole)

HYPERSCALER CAPEX

~\$384B

FY2025A, four largest hyperscalers

From Torque to Tokens: Power Semiconductors — The Other AI Trade

May 2026 · Lighthouse Canton Investment Research

The Re-Rating Thesis

There is a peculiar irony at the heart of AI infrastructure. The chips generating the most excitement in financial markets — the GPUs at the centre of every earnings call and capital allocation decision — cannot function until an entirely different set of semi-conductors solves a problem most investors have never considered.

Every watt of power consumed by an AI accelerator must be transformed, regulated, protected, and delivered with extraordinary precision before a single calculation occurs. The companies making this possible are, in aggregate, less discussed than the GPU designers they serve. They are also, in several important cases, more defensibly positioned — because their revenues accrue regardless of which GPU vendor wins the AI compute race. **Every rack pays the same toll to the same supply chain.**

The physics is straightforward. Two equations explain why the entire power semiconductor supply chain must be re-engineered. First: $P = V \times I$ — power equals voltage multiplied by current. As AI racks push from 120 kilowatts toward one megawatt, the industry faces a choice: deliver that power by increasing current, or by increasing voltage. The previous generation chose current — running at 54 volts and pushing thousands of amps through copper. But the second equation determines why that approach has hit a wall: $P_{\text{loss}} = I^2 \times R$ — waste heat in copper scales with the *square* of the current. Double the current, quadruple the loss. At megawatt scale, the copper alone would weigh 200 kilograms per rack. The only viable path forward is to raise the voltage — from 54V to 800V — which cuts the required current by roughly 15×, slashing resistive losses by over 200×. **This voltage transition is not a design preference. It is a law of physics, and it structurally requires silicon carbide and gallium nitride semiconductors that silicon cannot replace.**

The sector was mispriced as an electric vehicle cyclical. When EV demand slowed in 2023, the market treated the entire power semiconductor industry as broken. It was not broken — it was incomplete. The EV recovery remains on track, with next-generation 800V vehicle platforms driving meaningfully higher semiconductor content per car. But two additional demand vectors — AI infrastructure and grid decarbonisation — have since emerged that were absent from every 2022 bull case. **The market is currently pricing AI infrastructure power semiconductors as one recovering cyclical. The report argues that mispricing is the investment case.**

A note on valuation: Several of the names highlighted in this report have already re-rated meaningfully from their 2023–2024 troughs. We acknowledge the initial move. Our thesis, however, is that the earnings inflection ahead — driven by the structural 54V-to-800V transition and the emergence of AI and grid demand — will dwarf the share-price recovery seen so far. More importantly, if this demand proves secular rather than cyclical, these companies deserve a fundamentally different multiple. The absence of the boom-bust pattern that has historically compressed power semiconductor valuations is itself a source of future capital appreciation that the market has not yet priced in.

Market at a Glance

TOTAL POWER SEMI TAM ~\$55–57B 2025A global market (Mordor Intelligence, Global Market Insights, SNS Insider) → \$77–95B by 2031–2035E (~5–6% CAGR, Mordor/GMI 2026)	SIC POWER DEVICES ~\$3B 2024A; correction persisting through 2025–26 (Yole: ~50% upstream utilisation) → ~\$10.3B by 2030E (20% CAGR 2024–30, Yole 2025)	GAN POWER DEVICES ~\$1.2B 2024A est. (broader market incl. consumer; Yole power-device segment: ~\$355M) → ~\$3B by 2030E (~42% CAGR, Yole 2025)	HYPERSCALER CAPEX FY2025A (4 PLAYERS) ~\$384B Amazon \$132B · Alphabet \$91B · Microsoft \$89B · Meta \$72B (FY2025 actuals) Four largest hyperscalers; all five spent ~\$450B+
--	--	--	---

Sources: Mordor Intelligence (2026), Global Market Insights (Mar 2026), SNS Insider (Apr 2026) for TAM; Yole Group Power SiC 2025 and Power GaN 2025 for sub-segment forecasts; company reported financials for hyperscaler capex. GaN figure reflects broader power market including consumer; Yole power-device segment only ~\$355M in 2024. All forecasts subject to revision.

Why the Transition Is Happening Now — Four Forcing Functions

01

The GPU power curve went parabolic

The electricity demand of a single NVIDIA AI server rack has grown roughly 60-fold in five years — from about 10 kilowatts (the A100 generation, 2020) to a projected 600 kilowatts for the next generation arriving in 2027. That is enough power to run several hundred homes from a single equipment rack. At these extremes, routing electricity sideways across a circuit board generates so much heat that the board itself becomes the bottleneck. The solution — moving the final power conversion step directly underneath the chip — is no longer optional engineering. It is a physical necessity.

02

Silicon ran out of road

Standard silicon semiconductors work well up to around 600 volts. Above that, they waste too much energy as heat — a fundamental property of the material that no amount of clever engineering can fix. Silicon carbide (SiC) handles three times the voltage at a fraction of the heat loss. Gallium nitride (GaN) switches on and off 50 to 100 times faster than silicon, enabling far more compact power supplies. Both are essential for the voltage levels that modern AI data centres and electric vehicles now demand.

03

The EV correction mispriced the whole sector

The 2023–2025 SiC inventory cycle compressed multiples across the entire sector — including companies whose AI data centre revenues were growing rapidly. The market treated them as one story. They are not. Critically, AI data centre power demand is non-cyclical in the way EV is — hyperscaler capex commitments are multi-year and largely insensitive to consumer sentiment. The entry opportunity exists because investors remain anchored to the wrong demand driver.

04

Grid decarbonisation adds a third, uncorrelated pillar

European and North American policy-driven build-out of wind, solar, and grid-scale storage all require SiC and GaN at the inverter stage. This demand is uncorrelated with EV cycles and AI capex — it structurally widens the earnings floor and justifies a higher mid-cycle multiple.

The Full Value Chain — Where Value Concentrates and Where It Doesn't

LAYER	DESCRIPTION	KEY PLAYERS	INVESTMENT RATIONALE	RISK / CHOKEPOINT
Raw Materials	The raw crystals and elements from which power chips are made	Wolfspeed, Coherent, ROHM/SiCrystal for SiC; China controls ~80% of gallium supply	SiC supply constrained through 2027. China's 2023 export controls on key materials remain an active risk.	China export controls on key materials. Wolfspeed — the leading independent SiC substrate supplier — has been under significant financial pressure, highlighting how concentrated and fragile parts of this supply chain remain.
Epitaxy Equipment	Specialised crystal-growing reactors (HTCVD for SiC, MOCVD for GaN)	ASMI/LPE (SiC epi); AIXTRON (GaN/SiC MOCVD); Naura (CN)	12–18 month switching costs lock in customers. Current trough = entry point. Recovery expected 2027–2028.	Chinese competition advancing. SiC oversupply extending. GaN ramp timing uncertain.
Back-End Substrates	Ceramic tiles that sit between a power chip and its metal housing, conducting heat away	Maruwa (5344.T) ~15% global share; Kyocera; Rogers; Denka — three-player oligopoly	The chokepoint device-layer analysis misses. Even perfect SiC ramp execution at Infineon and onsemi cannot outrun Si ₃ N ₄ substrate constraints. Maruwa: ~15% global share, ~45% OPM on highest-growth segment. Three-player oligopoly. Under-monitored by most Western analysts.	Expanding capacity requires specialist equipment with multi-year lead times. Most Western investors have never heard of this company.
Power Device (IDM)	The actual silicon carbide and gallium nitride chips that do the voltage conversion	Infineon, onsemi, STMicro, ROHM, Wolfspeed, Navitas (the companies that design and manufacture these chips)	Infineon holds the primary GPU power position. SiC recovery underway from 2025–2026 trough. GaN gaining share in AI power supplies.	Must re-win position with each new GPU generation. Chinese competition compressing margins in standard silicon products.
Protection & Isolation ICs	Safety and control chips — supervise power flow and protect circuits from faults	Analog Devices (ADI), Texas Instruments, Littelfuse	ADI grows regardless of which GPU power supplier wins. TXN present at multiple stages of every AI rack.	Slower growth than chip-level names during peak cycles. China competitive at lower-end products.
System Integration	Complete power delivery units — power supplies, distribution panels, backup systems, next-gen modules	Delta Electronics (2308.TW), Flex Power, Bel Fuse, Murata, ABB	Delta bridges the device layer and hyperscaler infrastructure. Buys power semiconductor components from manufacturers such as Infineon and onsemi and assembles them into the complete power delivery hardware — server power supplies, rack distribution units, and backup power systems — that hyperscalers physically install in their data centres. Infineon–Delta VPD collaboration is the concrete illustration of cross-layer value chain interaction.	Delta has warned of rising input costs — driven by oil and the same AI demand fuelling its revenues. TWD/USD currency exposure adds volatility.

Where Value Actually Concentrates

Highest-quality moats: Equipment incumbency (ASMI, AIXTRON) — 12-18 month switching costs protect revenue across cycle turns, entry at trough. Protection infrastructure (ADI, TXN) — grows with GPU power density, generation-agnostic, no socket reset risk. System integration (Delta) — less volatile than pure device names, direct AI CapEx linkage.

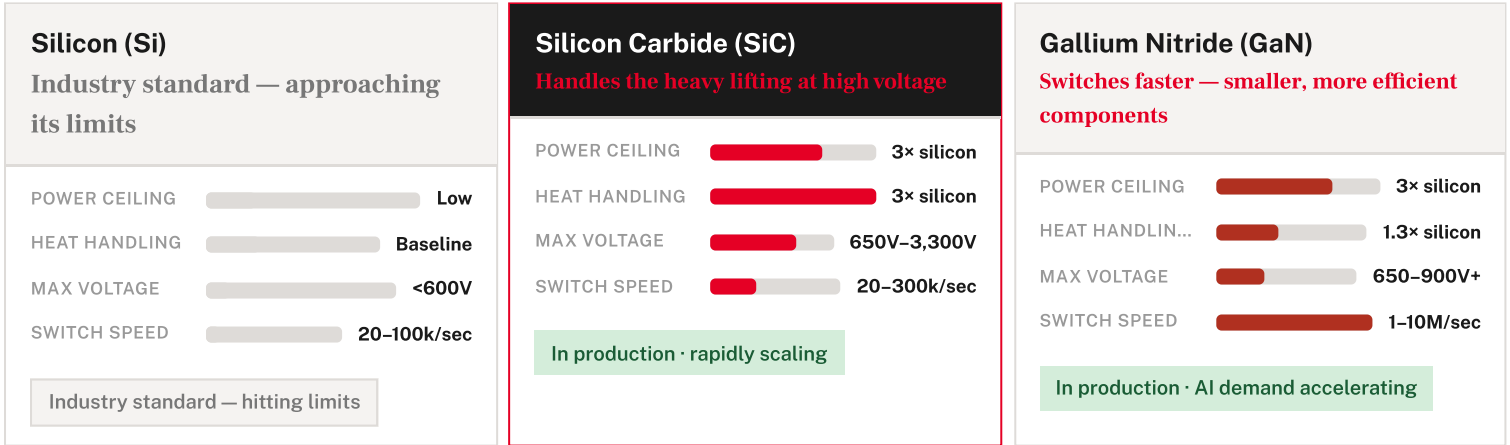
Hidden chokepoint: Maruwa's ceramic substrates — a near-oligopoly upstream of the chip manufacturers that standard analysis ignores. Worth tracking before the next SiC expansion wave makes the constraint visible.

Avoid: Commodity Si MOSFET and standard packaging — Chinese vendors closing the quality gap with state subsidy. Margin compression is structural, not cyclical.

Highest near-term leverage: Chip-level GPU power positions (Infineon, MPWR, Vicor) carry the highest earnings upside from platform wins — and the highest risk of disruption when NVIDIA changes generations. These are event-driven positions, not long-term compounders.

Silicon Hits a Wall — and Wide-Bandgap Materials Step In

Most electronics run on standard silicon chips — the same material in your phone or laptop. Silicon works well for most applications. But it has a hard ceiling: above a certain voltage and temperature, it wastes too much energy as heat and simply stops working efficiently. AI data centres and electric vehicles have now pushed past that ceiling. Two newer materials — silicon carbide (SiC) and gallium nitride (GaN) — handle much higher power levels with far less heat loss. They are not upgrades to silicon; they are replacements built for a different class of problem.



† GaN voltage range reflects current commercial devices (650V lateral GaN widely available; 900V devices shipping). Vertical GaN architectures — including onsemi's GaN-on-GaN platform announced October 2025 — are pushing beyond 1.2kV, directly targeting 800V data centre and EV applications. A fourth material (gallium oxide) exists with even better theoretical properties but is still in laboratory development — not relevant for investment before 2028-2030 at the earliest. Bar lengths are comparable within each row only.

NVIDIA's GPU Roadmap — Power Demand Doubles With Every Generation

GENERATION	RACK POWER	PEAK CURRENT	POWER ARCHITECTURE	SOCKET / STATUS
Blackwell (GB200/GB300)	~120 kW	~1,200A	54V DC bus, horizontal delivery	Infineon primary · Renesas secondary
Vera Rubin (R100/R200)	~240-300 kW	~1,500A+	800V DC transition begins · VPD	Requalification open — MPWR binary Q2-Q3 2026
Rubin Ultra / Kyber NVL576	~600 kW	~2,000A+	VPD mandated — new rack architecture	Vicor FPA best-positioned architecturally
Feynman NVL1152	~1.2 MW	>3,000A	Full VPD + CPO optical scale-up Architecture TBD · 2028+ horizon	

Rack power figures are engineering estimates; projections for Kyber and Feynman generations are forward-looking. The three hyperscaler capex figures are FY2025 reported actuals — money spent, not forecasts. The four listed hyperscalers (Amazon, Microsoft, Alphabet, Meta) spent ~\$385B in FY2025; all five major hyperscalers combined exceeded \$450B.

Why the Next Generation of AI Servers Needs a Fundamentally Different Power Design

A GPU chip runs at less than one volt — roughly the voltage of a AA battery — but a next-generation AI rack consumes up to 600,000 watts. Delivering that much power at such a low voltage means pushing an enormous amount of electrical current through the system. The problem: copper wiring resists current flow, generating heat. Even a short section of circuit board trace wastes 800 watts at these current levels — roughly the heat output of a small electric heater, concentrated in a few centimetres of wire.

The underlying constraint is the relationship between power, voltage, and current: $P = V \times I$. At today's 54-volt rack architecture, delivering 120 kilowatts requires over 2,200 amps. The heat wasted in copper conductors scales with the square of the current (I^2R), so every increase in rack power drives a disproportionate increase in waste heat and copper mass. At megawatt scale, 54V distribution would require up to 200 kilograms of copper busbar per rack — the weight of the rack structure itself. The industry's response is to raise the distribution voltage from 54V to 800V, reducing the required current by approximately 15× and cutting I^2R losses by over 200×. This is the same principle that drives long-distance electricity transmission at hundreds of thousands of volts. It is also why this transition structurally requires wide-bandgap semiconductors — silicon carbide and gallium nitride — which can operate efficiently at these higher voltages where standard silicon physically cannot.

The solution — now being adopted across the industry — is to move the final voltage conversion step from the edge of the circuit board to directly beneath the chip itself, shortening the current path from several centimetres to a few millimetres and eliminating most of that heat loss. This is called **Vertical Power Delivery (VPD)**. It is not a design choice; it is what the physics demands at the power levels the next generation of AI hardware requires.

Six Conversion Stages, Six Toll Booths — GPU-Vendor Agnostic Revenue

Electricity from the grid cannot go straight into an AI chip. It must pass through six separate conversion and regulation stages — each one stepping down the voltage and managing the power more precisely — before reaching the chip at the correct level. Think of it as a series of six toll booths on the road from the electricity grid to the GPU. Each toll booth is a separate commercial opportunity, served by different companies, generating revenue regardless of which chip manufacturer wins the AI race.

#	STAGE	WHAT HAPPENS	WHO BENEFITS	WHY IT MATTERS
1	Grid Entry Grid → data centre	Grid electricity is converted to high-voltage DC power for distribution across the data centre. Current Blackwell-generation systems use 54V DC in-rack distribution. The industry is preparing to upgrade to 800V systems — a transition that begins with the Vera Rubin generation and becomes mandatory at Kyber / Rubin Ultra scale.	Infineon (SiC), STMicro, onsemi, Wolfspeed, Power Integrations (POWI), Navitas Semiconductor	Every AI campus built to the forthcoming 800V standard will require SiC and GaN chips at this stage regardless of GPU vendor. GaN-based rectification (POWI, Navitas) enables the 98%+ efficiency targets that make 800V systems economically viable at megawatt scale — silicon cannot meet this efficiency requirement at these voltages.
2	Bus Protection Safety & protection	Safety components that monitor the high-voltage power line and protect the rest of the system from faults or spikes. Present in every rack, regardless of which GPU or power supplier is used.	Analog Devices (ADI) dominant · Texas Instruments	ADI earns revenue regardless of who wins the GPU power competition. Every converter made by any supplier still needs ADI's safety components.
3	Intermediate Bus Step down for distribution	Steps the high voltage down to a lower level (48V) for distribution across the server. GaN semiconductors are enabling smaller and more efficient converters than the previous generation of silicon-based designs.	Vicor (FPA), Infineon (CoolSiC), Renesas, TI	The entire hyperscaler industry is mid-transition to 48V systems. GaN is the enabling technology. This is where GaN's commercial momentum is most visible today.
4	Point-of-Load VRM Final step to GPU chip	The final step — reducing voltage to the tiny fraction of a volt the GPU actually runs at, while pushing an enormous amount of current. This is the most commercially valuable component position in the chain. Suppliers must compete for it afresh with each new NVIDIA GPU generation.	Infineon (Vera Rubin primary) · Renesas (secondary) · MPWR (requalifying) · Vicor (Kyber VPD)	The key competitive battle. Who wins this position with NVIDIA determines the biggest earnings swings in the sector. See Page 6.
5	Energy Storage Power shock absorber	Acts as a buffer between the power supply and the GPU — absorbing sudden spikes in electricity demand when AI workloads ramp up instantly. Think of it as a shock absorber for electricity.	AOSL, Texas Instruments, Eaton, Delta Electronics	AOSL has exposure to both AI server power delivery and optical networking — two fast-growing AI themes. This combination is not yet fully reflected in its valuation.
6	Upstream Equipment Crystal growth reactors	Specialised reactors that grow the semiconductor crystals from which all power chips are made. Once a chipmaker qualifies a reactor for their production line, switching costs (12–18 months of re-testing) are prohibitively high — creating a very durable competitive position.	ASMI/LPE (SiC crystal reactors) · AIXTRON (GaN crystal reactors)	Near-duopoly with deeply entrenched customer relationships. Current oversupply in SiC creates an entry opportunity ahead of a recovery expected around 2027.

The Layer 4 Socket Race — The Single Most Proximate Catalyst

Q2-Q3 2026 · MPWR VERA RUBIN REQUALIFICATION WINDOW

Monolithic Power held the primary supplier position at NVIDIA's H100 generation, lost significant share on Blackwell after performance concerns, but retained some confirmed orders. By early 2026, the analyst consensus had shifted meaningfully positive on Vera Rubin — with some estimates suggesting MPWR could hold a dominant position on that platform. The key question for investors is no longer binary win/loss but whether current valuations already reflect this recovery.

Infineon — the incumbent (Vera Rubin primary). Direct architecture collaboration with NVIDIA on next-generation 800V DC power delivery makes within-generation displacement very difficult. CoolMOS and CoolSiC platforms benchmark best-in-class for high-current GPU VRM synchronous rectification. Kyber path depends on VPD compatibility of next-generation platform.

Renesas — the qualified second. Co-qualified across Blackwell and Vera Rubin. Automotive VRM heritage (more demanding than server VRM by almost every specification) translates directly to AI demands. Path to primary requires Infineon execution failure or a VPD transition that resets the competitive field.

Vicor — the long-duration VPD option (Kyber H2 2026).

Factorised Power Architecture places the Voltage Transformation Module directly beneath the GPU package substrate — the exact topology Vertical Power Delivery demands. Architecture-matched. Execution unproven at hyperscaler volume. Kyber specification finalisation in H2 2026 is the defining binary event. Size explicitly for the optionality.

Tiber (private) — the wildcard. Substrate-embedded voltage conversion could change the physical form of the Layer 4 socket entirely if adopted at Vera Rubin or Kyber scale. A monitoring item for all VRM positions with a 2027–2028 risk horizon.

Six Risks That Could Break the Thesis

AI CAPEX PAUSE Hyperscaler monetisation shortfall reprices the entire supply chain simultaneously. Leverage is higher than "picks and shovels" framing implies — device names have significant earnings sensitivity to platform timing. Amazon, Google, and Microsoft have committed capital but have redirected CapEx before when return profiles disappointed.	CHINESE VENDOR QUALIFICATION Chinese semiconductor manufacturers — supported by substantial government funding — are rapidly closing the performance gap at the lower end of the power semiconductor market. Reaching the quality level required for premium AI applications is a harder target, but the 2028–2030 timeframe is credible, not speculative. The practical response is to favour companies whose revenues are not exposed to commodity price pressure.	SYSTEM INTEGRATOR MARGIN RISK Delta has explicitly warned of rising costs from oil and AI-demand-driven materials shortages. Gross margin expansion is not linear — dual-sided dynamic warrants explicit position-sizing discipline rather than a momentum approach.
SIC CYCLE EXTENSION Overcapacity extending beyond 2027 delays ASMI/AIXTRON recovery. The Si₃N₄ constraint becomes more visible during the trough as remaining high-utilisation fabs compete for limited Maruwa substrate supply.	TIBER / SUBSTRATE EMBEDDING A private company called Tiber is developing technology that would embed the voltage conversion function directly inside the GPU's own packaging — potentially eliminating the need for a separate external component at the highest-value position in the power chain entirely. A 2027–2028 risk horizon for Infineon, Renesas, MPWR, and Vicor positions.	GEOPOLITICAL / EXPORT CONTROLS China's 2023 gallium/germanium controls established the precedent for materials-level disruption. Escalation to SiC substrates would hit the entire wide-bandgap (SiC and GaN) supply chain simultaneously. Taiwan Strait risk reprices the sector overnight — there is no Plan B at scale.

IMPORTANT INFORMATION

Disclaimer

CONFIDENTIALITY

The contents of this document are confidential and are meant for the intended recipient only. If you are not the intended recipient, please delete all copies of this document and notify the sender immediately.

GENERAL COMMENTARY

This document, provided as a general commentary, is for informational purposes only and is not to be construed as an offer to sell or solicit an offer to buy any financial instruments in any jurisdiction. This does not constitute any form of regulated financial advice, and your independent financial advisor should be consulted prior to taking any investment decision(s). This document is based on information from sources which are reliable but has not been independently verified by Lighthouse Canton Pte. Ltd. and its affiliates ("LC"). LC has taken reasonable steps to verify the contents of this document and accepts no liability for any loss arising from the use of any information contained herein. Please also note that past performances are not indicative of future performance. Information contained herein are those of the author(s) and does not represent the views held by other parties. LC is also under no obligation to update you on any changes made to this document.

REGULATION

This document is prepared by Lighthouse Canton Pte. Ltd. and its affiliate, Lighthouse Canton Capital (DIFC) Pte. Ltd., which are regulated by Monetary Authority of Singapore ("MAS") and Dubai Financial Services Authority ("DFSA") respectively. MAS and DFSA have no responsibility for reviewing, verifying and approving the contents of this document and/or other associated documents. The contents of this document may not be reproduced or referenced, either in part or in full, without prior written permission from LC.

OUR OFFICES

SINGAPORE

Singapore

16 Collyer Quay #11-02
Collyer Quay Centre
Singapore 049318
+65 6713 0570

UAE

Dubai

The Exchange, Gate Village 11
Unit 802, DIFC
PO Box 507026, Dubai
+971 4 586 1500

UNITED KINGDOM

London

24 Hanover Square
London W1S 1JD
+44 1642 843 487

INDIA

Mumbai

Unit 507/508, A Wing, INS Tower, G Block, BKC,
Mumbai - 400051
+91 96504 73961

New Delhi

Unit 104A, Worldmark 2, Aerocity, New Delhi 110037
+91 96504 73961

Bengaluru

1st Floor, WeWork 37, Cunningham Cross Rd, Vasant
Nagar, Bengaluru 560001
+91 99000 96873

Chennai

RK Swamy Centre, Hansa Building, Door No: 3,
Thousand Lights, Chennai 600006
+91 96504 73961

Hyderabad

Suite 502, Building 450, Central Plaza, Genome
Valley, Shameerpet, Hyderabad 500078
+91 96504 73961

Gandhinagar

Unit FF-10, Pragya Accelerator, Block 15T, GIFT CITY,
Gujarat 382355
+91 96504 73961

✉ info@lighthouse-canton.com

in Lighthouse Canton