NEXTDC - AI Executive Playbook

Your infrastructure guide to building an AI-ready future

NEXTDC

where AI lives™

INTRODUCTION

Artificial intelligence

Artificial intelligence (AI) has now created the globally recognised 4th industrial revolution, highlighted by the unprecedented growth of Generative AI as the first ubiquitous killer application of accelerated computer technologies capable of processing huge volumes of data in near real time.

Increasing global investment:

AI spending is projected to reach

US$200 billion in 2025 Goldman Sachs

Rapid market expansion:

Expected to grow from US$757.58 billion in 2025 to approximately

US$3.68 trillion by 2034 Precedence Research

Surging AI-generated data:

By the end of 2025, AI will generate

10% of all global data Markets and Markets

INTRODUCTION

NEXTDC is *where AI lives*™

NEXTDC are investing heavily to play a foundational role in the adoption of AI both in Australian markets and across the Asia-Pacific region. Our nationwide fleet of Uptime Institute-certified Tier III and Tier IV data centres includes AI-ready infrastructure in every mainland capital city.

Certified under the NVIDIA DGX 'AI Factory' data centre reference architecture program, customers can immediately access environments that are engineered from scratch to support the resilience, power, security and interconnection profiles essential to today's and tomorrow's AI success.

NEXTDC customers will experience our industry benchmark 100% uptime guarantee for high density power (up to 600kW/rack), liquid cooling solutions and ultra-low latency interconnection needed to produce AI at scale.

An example of emerging AI Factory architecture standards is the recently announced NVIDIA DGX B200 AI platform, which employs NVIDIA's Blackwell GPU and 208 trillion transistors to deliver this next generation of accelerated compute for AI workloads.

In a technology market growing explosively, (global value to reach $407B US: by 2027 - Forbes), it is evident that most organisations will need to build and connect new digital infrastructure in addition to developing the technology skills required to host and manage their AI platforms.

As digital infrastructure leaders, NEXTDC is ready to facilitate change. We recognise that the business transformational opportunity presented by AI is generational, requiring many players within an inter-connected ecosystem to meet diverse customer requirements.

The innovation focus of NEXTDC's existing ecosystem of clouds, carriers and digital services will ensure that collaborative partnerships and alliances can be forged enabling customers to deploy their choice of AI solutions. NEXTDC will offer the flexibility to customise digital infrastructure so that all leading and emerging platform architectures are supported.

NEXTDC will continue to play a key role as an enabler of resilient, sustainable digital infrastructure and thriving partner ecosystems.

We will support hyperscale, enterprise and government customers by ensuring our premium facilities meet the current and future digital infrastructure requirements of their AI-strategies.

In the chapters that follow, we've outlined "what's NEXT" for data centres in an AI-enabled future – and the role NEXTDC have to play in shaping this future for our partners and customers.

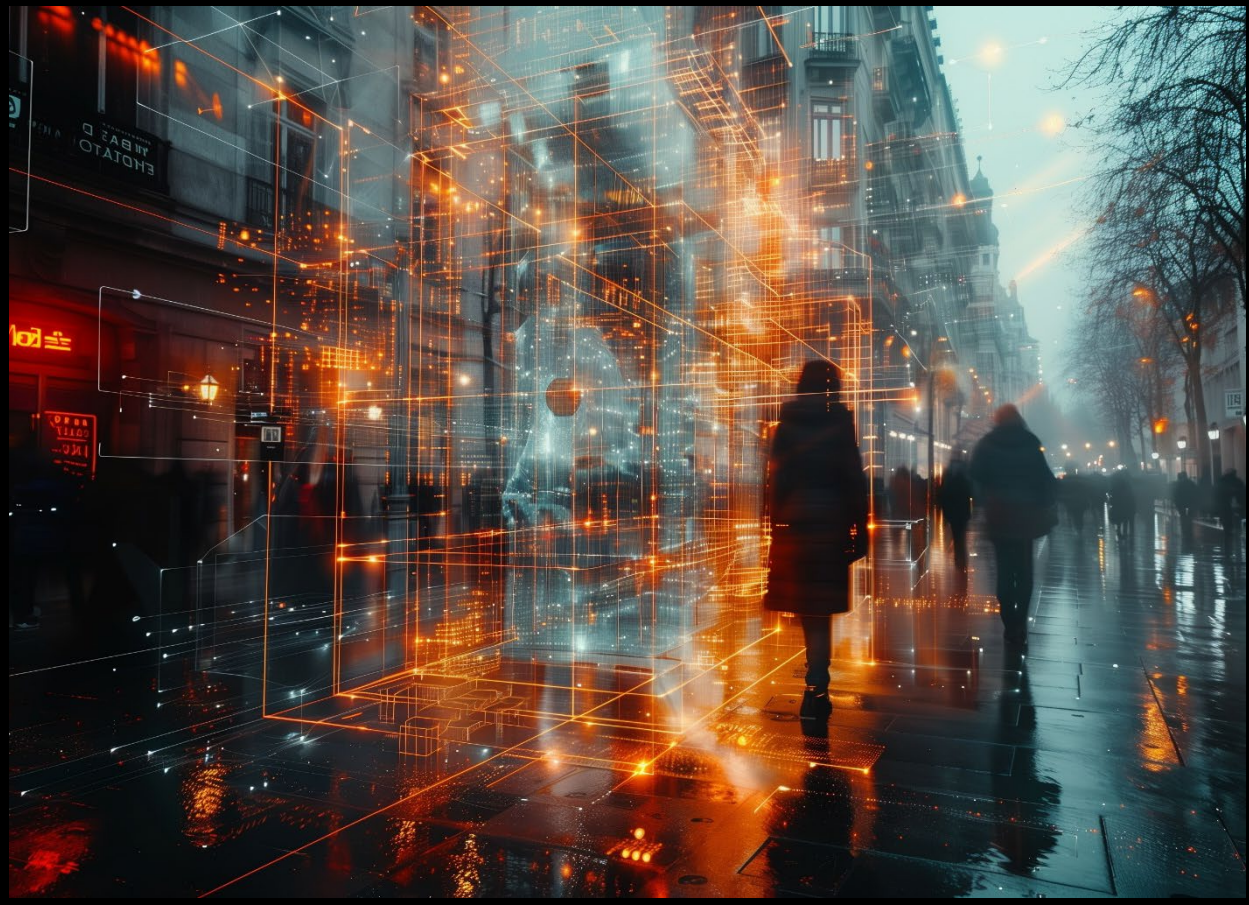
NEXTDC has always been *where the cloud lives*™. NEXTDC are now investing to ensure our facilities are also *where AI lives*™.

"The existing NEXTDC ecosystem will ensure that collaborative partnerships and alliances can be united to enable customers to deploy their choice of AI solutions."

AI is transforming industries

Artificial intelligence (AI) is rapidly transforming the world we live in – as well as the specifications of the fundamental infrastructure required to support it.

It's creating smarter cities, advancing medical and life science breakthroughs, improving productivity, as well as unlocking solutions to some of our most pressing challenges.



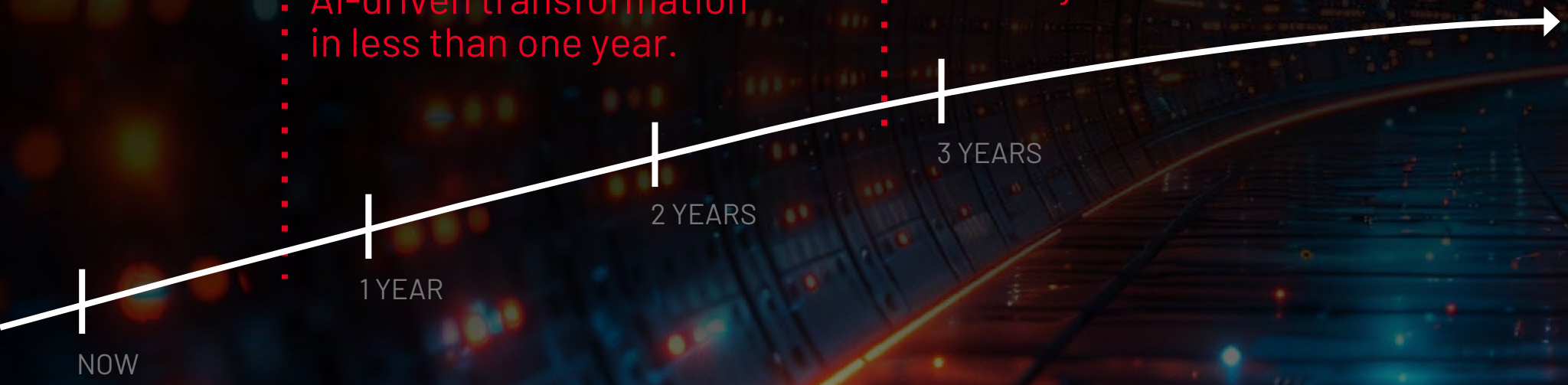
AI offers great potential to create new products, services and business models. We can expect to see exciting developments in this space, particularly those service providers set up to support organisations aiming to be flexible, agile and scalable in their AI ambitions.

AI's accelerated journey into public consciousness since the launch of ChatGPT in November 2022 and the birth of GEN AI in 2023 is yet another signal that a Fourth Industrial Revolution (Industry 4.0) is underway. Industry 4.0 is about the production of intelligence. Ubiquitous and democratised technology adoption continues to drive an explosion of global data, with commerce and lifestyle evolutions taking digital services into exciting new frontiers.

The AI race is well and truly on – and it's happening right now:

31%
of business leaders expect substantial AI-driven transformation in less than one year.

48%
expect it in one to three years.



Source: Deloitte, Now Decides Next, January 2024

The role of digital infrastructure as a critical foundation for AI

The need for specialised digital infrastructure to support AI's demanding computational workloads is already surging.

Not only must this NEXT generation of infrastructure measure up technically – particularly when it comes to power and cooling considerations – but it must also be secure, interconnected, scalable, flexible and sustainable, in order to grow and adapt to this new era of accelerated computing.

We have the established infrastructure platform and the proven track record of agile data centre construction that will ensure we are at the heart of this transition from AI readiness to AI scale.

The emerging concept of AI Factories is evidence that data centres are a strategic pillar for the growth and adoption of AI and will play a key role in guiding its journey.

Meanwhile, it's important to note that infrastructure requirements will be different for diverse applications of AI.

For example, cloud processing may be the best option for some applications (such as AI training models).

Other applications (such as AI inference models) that inform decision-making in closer to real-time, will require complex high-performance interconnection with proximity (low latency) to multiple data sources and infrastructure. They will also feature extremely dense racks (requiring lots of power and cooling) and high availability. For these applications, colocation designed in a specialised AI factory will be an effective infrastructure solution.

Much like cloud technologies, we expect there will be hybridised public and private infrastructure, as well as new products and services that evolve in the months and years to come.

How prepared are most organisations for AI?

60% of technology leaders say that their organisation's tech and digital infrastructure are not fully prepared to optimise AI.

Source: Deloitte, Now Decides Next, January 2024

AI concepts and definitions you need to know

The arrival of AI has created a slew of new technical terms to get our heads around.

AI Factory

Purpose-built, privately managed colocation facilities designed to support the specific demands of AI workloads.

Sovereign AI

Refers to a nation's capabilities to produce AI using its own infrastructure, data, workforce and business networks.

(Source: NVIDIA)

AI training

The process of training AI models to make accurate, humanistic inferences.

Machine learning (ML)

A type of AI that trains computers to learn on their own without direct human intervention. Deep learning is a further subset of ML.

Generative AI (e.g. ChatGPT)

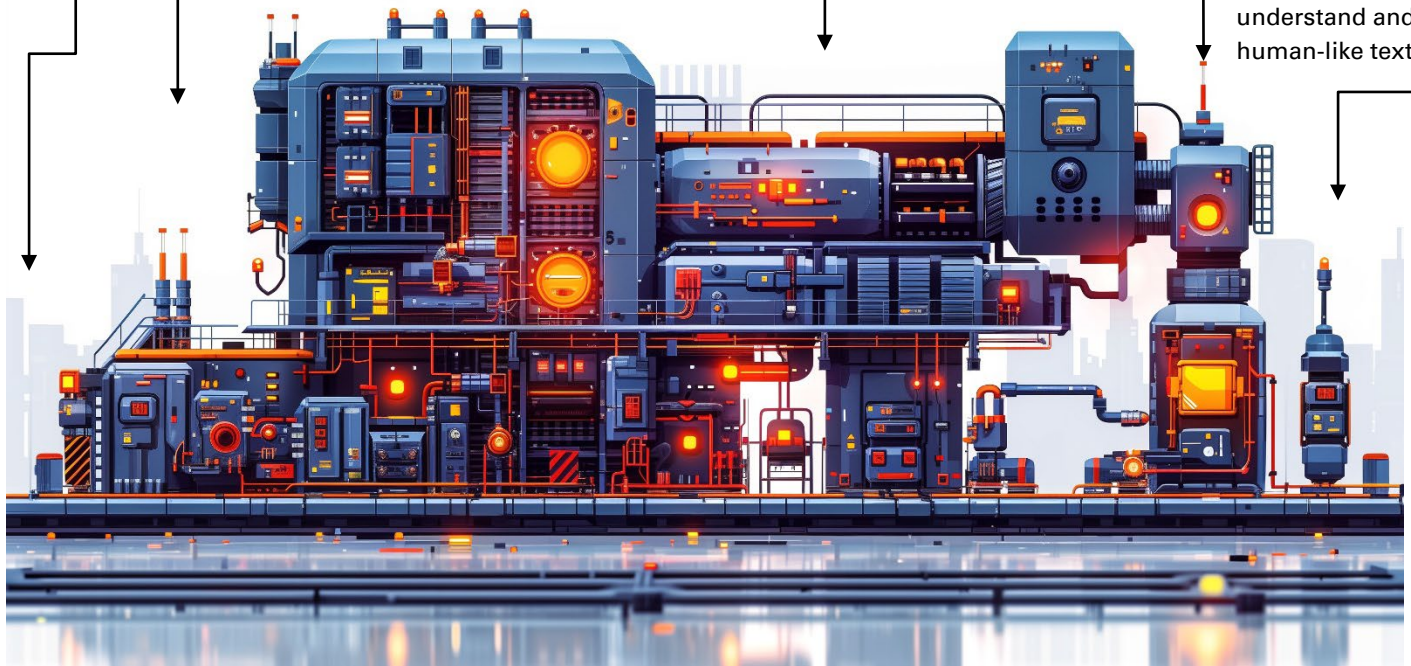
A type of AI that creates content such as text, video, code and images.

Large language model

A subset of generative AI, where the model has been trained (using natural language processing tools) to understand and produce human-like text.

AI

An overarching term to describe any program seeking to emulate the way humans think.



AI inference

The real-time process of an AI model making predictions, drawing conclusions or any other output based on dynamic data sets. Using data to influence and automate business decisions.

Private AI

Refers to methods of building and deploying AI technologies that respect the privacy and control of user and organisational data.

How do we think the AI story will evolve?

No one can say exactly if, when and to what extent AI will completely transform our economies and communities. What we can observe, based on breakthrough technologies from the past few decades, is that there is a general adoption pattern of awareness, hype, disappointment and then sensible growth once those initial phases have passed (see: the Gartner Hype Cycle methodology).

Another observation is that technology adoption and innovation follows a two-speed evolutionary process – slow and fast. So far, AI seems to be following that general pattern, but at a much faster pace than any previous technology. Consider that ChatGPT was only released in November 2022, and how much progress has been made with generative AI models since that time.

We are confident the AI journey over the coming months and years will resemble other innovation journeys and trends such as cloud developments and that the benefits and strategies will evolve over time, as the technology and skills mature and are able to prosecute strategic outcomes.

Ultimately, the agile organisations who excel at testing, failing fast and creating a strategic competitive advantage will be the winners.

The clear challenge for most organisations who want to move from "slow" to "fast" is access to AI talent and skills:

Wave 1
(fast paced adoption)

- Productivity
- Cost efficiencies
- Speed to market

Wave 2
(slower realisation of benefits)

- Data driven and automated problem solving
- Realised innovation
- New business models

56% of business leaders say their organisation's generative AI expertise is low-to-moderate.

Source: Deloitte, Now Decides Next, January 2024

How to be future-ready for AI

It is clear that premium data centres and AI Factories are critical to an AI-enabled future

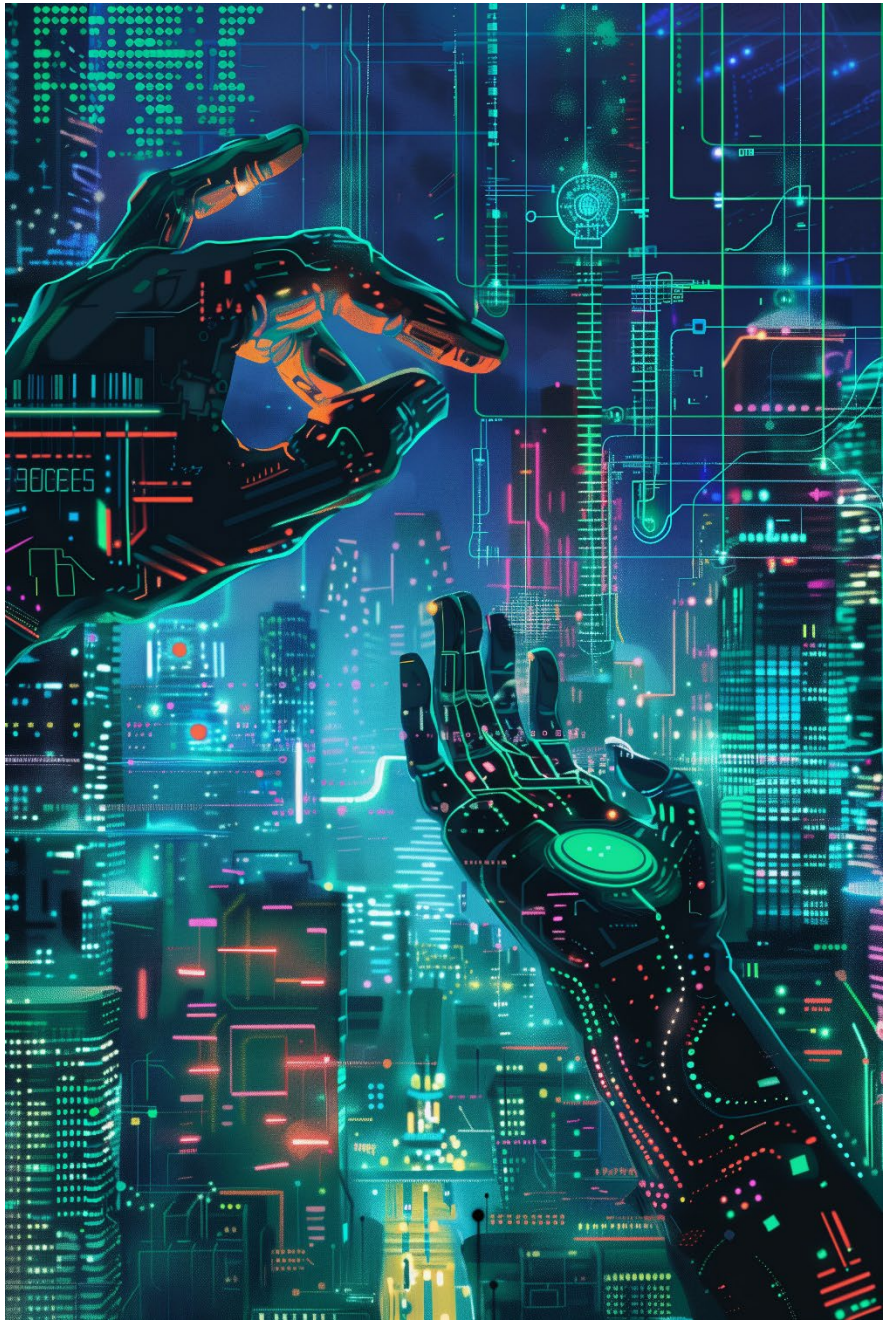
These facilities must be specifically designed and engineered to meet demanding computational workloads, from deep learning algorithms to the vast neural networks that drive today's most innovative AI applications.

Many are already adapting to offer the necessary support, since the vast majority of organisations will not have the technical skills and knowledge to deploy the required infrastructure quickly and cost-effectively, and with the appropriate levels of resilience, security and interconnectivity.

The critical foundations for AI success

Key enablers for digital infrastructure designed as AI-ready

- Support for high-density computing and GPU workloads, enabling the efficient processing of massive data sets. This includes the provision of specialised high-density resources and hardware for AI applications.
- Effective management of the immense amounts of heat generated by high-performance AI workloads, through innovative designs for cutting-edge cooling technologies. Liquid cooling technologies are a cornerstone in driving energy efficiency and sustainability outcomes for this class of infrastructure. At NEXTDC, we're embracing a suite of bespoke and standardised liquid cooling solutions and have deployed environments for customers that include direct-to-chip, rear door heat exchangers and immersion cooling innovations.
- A sharp focus on energy efficiency and sustainability, to keep power costs down and to address broader ESG implications around expanding carbon footprint accompanying ballooning power consumption.



- Multiple locations for customers to deploy in. AI inference models (real time data analysis, informing business decisions) require ultra-low latency and locations close to metro areas – for example, in AI factories. Meanwhile, training models are not so time-sensitive or resilience dependent so can be conducted most cost-effectively in more distributed locations, including traditional data centres or city-fringe locations as hyperscale campuses close to suitable sources of power.
- Providing a robust partner ecosystem to ensure access to data, innovative hubs, talent with AI expertise, and AI partnerships with established solution providers.
- High-bandwidth, low-latency interconnectivity. AI is fundamentally changing the role of connectivity in the data centre, for efficient data transfer and communication between AI systems. For AI to make informed decisions in real-time, large amounts of data will typically be sourced from multiple locations and applications. Resilient, centralised, private and sovereign networks with utility billing models are needed for cost reduction, efficiency, speed and security.
- Consideration of security and data sovereignty issues, including physical and data security, access management, and mitigation of sovereignty risks associated with ensuring data is retained within corporate, state or national boundaries.
- Power availability to deal with the scale of AI power requirements. Experienced, agile and flexible data centre partners will be best positioned to source power and make it available with 100% uptime for critical digital environments.
- Collaboration and partnerships with global and local AI experts, as well as technology and advisory companies. The onus is on data centre providers to host and facilitate these networks to ensure customers are poised to harness the full potential of the emerging AI community.

Popular AI deployment models

As mentioned earlier, there is no “one size fits all” model for deploying AI workloads. Just as with cloud computing — where private, public, hybrid and multi cloud environments are favoured — organisations will need to make strategic decisions about their AI environment, depending on the use cases which can support their strategies.

Four preferred models are already emerging.

Public AI

As the most well-known model, public AI is accessible to a broad audience (usually the general public, sometimes specific user groups). These workloads are designed to be deployed and accessed through public and private clouds, as well as data centres.

Private AI

An architectural approach that aims to balance the business gains from AI with the practical privacy and compliance needs of the organisation. This means data is not made available to develop other external models.
Private AI workloads can be deployed in both public and private clouds, as well as data centres, as long as the privacy and access controls are satisfied.

Sovereign AI

A deployment model where nations can produce AI using their own infrastructure, data, workforce and business networks. This gives them ownership of their own data and the intelligence it produces.

Cloud AI

The amalgamation of cloud computing and AI. The seamless fusion of AI tools, algorithms, and cloud services into day-to-day business operations, processes and teams.

The NVIDIA AI Factory architecture

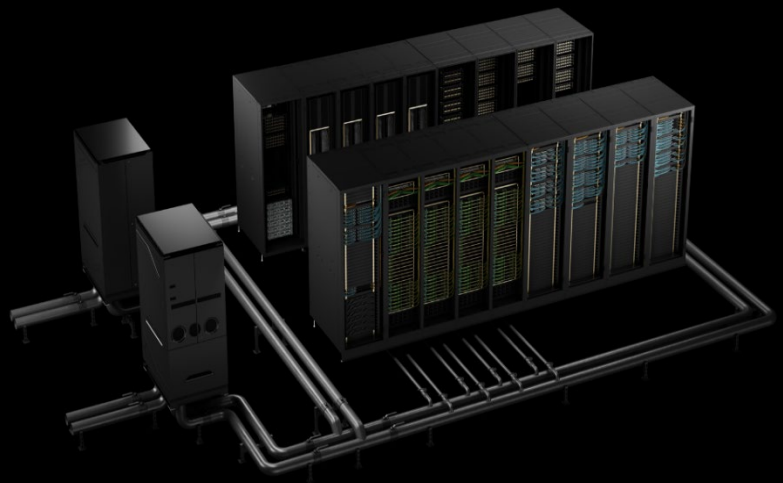
Get to market quickly. Turn data into intelligence NOW. Nation-wide availability with NEXTDC.

Premium data centres or 'AI Factories' are your fast-track to AI-enabled growth and innovation.

Get started now with infrastructure designed and engineered to meet NVIDIA reference architecture.

Specifically designed and operated:

- Agile deployment of NVIDIA Reference Architectures
- Trusted liquid cooling solutions (Direct to Chip & Rear Door Heat Exchangers)
- Resilient power for up to 600kW to individual racks
- Flexible, agile and direct interconnection to everything
- A rich ecosystem of clouds, carriers and digital service providers
- Physical security to the highest global and industry standards



NVIDIA

PREFERRED PARTNER

Understanding the value of AI Factories

NEXTDC has been strategically planning for the democratisation of this step change in technology for the past decade, having researched, designed, built, delivered and successfully operated multiple liquid cooling configurations across multiple facilities in Melbourne and Sydney. In fact, our background as an organisation as an infrastructure leader has affirmed this approach with the successful deployment of cloud enabled data centres, including the growth of Hyperscale facilities.

The high-performance computing future became apparent during NEXTDC's first 80+kW water cooled supercomputer deployments in 2015, where immense computer power was supported in a rack footprint 40% less than would have otherwise been required in a conventional data hall.

Meanwhile, our energy efficiency gains through the adoption of free-air-cooling technologies have also contributed greatly to certifications for Australia's first 5-star NABERs rated data centre.

When combined with our market-leading data centre engineering, operational processes and facilities management expertise, there is a significant opportunity for NEXTDC to build market leading AI Factories, by customising and assigning existing premium grade facilities to the specific needs of AI workloads.

AI Factory features, specs and "must haves"

Liquid cooling is a cornerstone in this infrastructure, offering unparalleled cooling efficiency and sustainability outcomes. Direct-to-chip cooling, rear door heat exchangers, immersion cooling and other innovations will all play a role in the NEXTDC AI Factories of the future.

AI Factories will also support **modular, reference architectures** for our customers to design to. This approach will enable customers to start small and then scale out by adding compute and GPU or storage as required.

While offering **power density** five to 20 times that of the traditional data centre, AI Factories such as NEXTDC's NVIDIA-certified AI Factories nation-wide are likely to be smaller than traditional hyperscale data centre facilities. Supporting inferencing AI models, new AI factories will be located closer to CBDs where customers are headquartered or as new configurations within existing facilities that have sufficient power available.

Managed services for AI will provide expertise in developing, deploying, managing and optimising AI infrastructure.

Customers need to be supported by and have easy access to robust, AI aware and **AI capable interconnectivity**.

Diversity in the power grid will also be a key feature to lessen the distribution impacts on community-based power grids.

AI Factories will offer more opportunities for **customised designs** matched to customer requirements and workloads, including AI training and inference models. Bespoke super pods, operational environments and platforms will be designed for each customer.

Utility billing-based "as-a-Service" (aaS) models will be a critical:

- Infrastructure-as-a-Service (IaaS) for AI will offer pre-configured AI-environments on demand.
- GPUaaS and TPUaaS will become a convenient way to access high-performance computing resources for AI and other data-intensive applications. Customers will be able to access advanced computational capabilities without expensive hardware or privately owned infrastructure.

The value of future-ready infrastructure

At NEXTDC, our insight and expertise ensure we are able to keep our customers aware of future trends and opportunities. The pursuit of excellence and being prepared for "Whatever comes NEXT" is a concept that is embedded in our brand and culture, with a significant focus on helping our customers and partners to understand that being future-ready is critically important to digital transformation success.

We're not only building the ground-breaking infrastructure and sovereign ecosystems where AI can evolve, innovate and thrive; we are also committed to enabling our customers to achieve future-ready AI transformation.

CASE STUDY

S6 Sydney: An AI deployment model to consider

NEXTDC's S6 Sydney data centre – in Artarmon, just across the road from S3 Sydney – is Australia's first purpose-built AI Factory. Located close to the heart of Australia's largest city, S6 provides the necessary high-density power, cooling, security and high-performance ecosystem interconnection required for AI to process and analyse data sustainably and at unprecedented scale. Find out more here.

S6 Sydney is Australia's first standalone, ultra high-density AI Factory, engineered exclusively to support sovereign AI architectures and workloads, ensuring our customer and partner ecosystem has the infrastructure it needs to remain at the forefront of innovation and technology.

Under NVIDIA's DGX-Ready reference architecture program, all NEXTDC data centres are now certified to deliver 'AI Factory' environments provided with all the traditional features of NEXTDC facilities such as customer service, security, high performance and resilient private cloud and ecosystem connectivity and broad compliance with global and industry standards.

S6 all other NEXTDC AI Factories will embrace a suite of liquid cooling solutions for customers' bespoke operating environments including direct-to-chip, rear door heat exchangers and immersion cooling. They will be flexible to each of these technologies thereby helping to ensure optimal HPC performance while minimising environmental impact.

Finally, we have also developed a ground-breaking system that harnesses heat to water exchangers, providing improved total system reliability, segregation of risk and complete separation of client systems.

Up to 600kW of power to each rack.

Your nine point guide to enabling digital infrastructure that empowers AI solutions

1

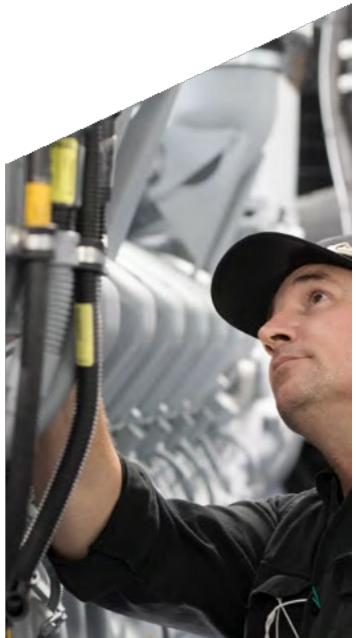
Innovative cooling solutions

Ensure your data centre partner has the flexibility to embrace a suite of liquid cooling solutions for customers including direct-to-chip, rear door heat exchangers, and immersion cooling. Each of these technologies play a pivotal role in dissipating the intense heat generated by AI computations, ensuring optimal performance while minimising environmental impact by allowing for unprecedented density and efficiency.

2

High-density power solutions

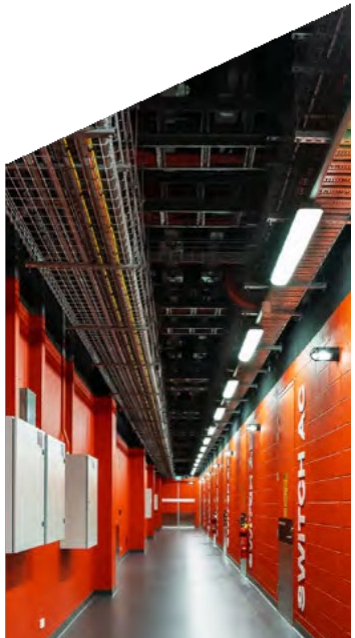
High-density power delivery that's completely fault tolerant is a complex design and deployment challenge in the AI era. It needs to not only be operated for 100% uptime but also flexible to individual rack requirements which will vary based on workloads. Environments are already being designed and deployed in Australia to deliver up to 600kW of power to each rack while Cushman & Wakefield recently forecast that average DC rack power will reach 120kW by 2025.



3

Scalable and reliable infrastructure

Ensure your data centre partner has a growth pathway to build highly certified facilities at scale. Look for established history of agile development and Uptime Institute Tier IV-certified facilities, ready to meet the dynamic resilience, compute and storage demands of AI.



4

Expertise and experience

Developing and recruiting increasingly specialist in-house skills, experience and expertise required to build bespoke and cutting-edge infrastructure environments will be competitive and expensive. Select a data centre partner that has a track record of guiding and supporting customers through the complexities of AI infrastructure.

5

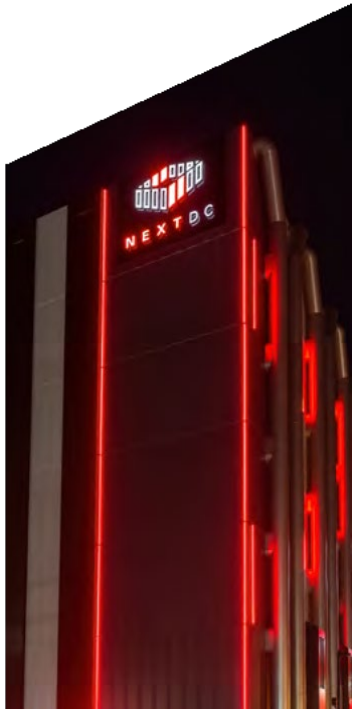
Security, certifications and compliance

In the AI future, data is gold and your infrastructure needs to protected to the highest global, Commonwealth and industry standards for data protection, sustainability and physical security. AI Factory facilities should be certified to be DGX-Ready via the NVIDIA Data Centre program.

6

Sovereign AI infrastructure

For Australian data, select an Australian data centre services provider with a sovereign national footprint of standardised facilities. Your data must stay onshore, within a secure sovereign ecosystem of infrastructure, cloud providers, carriers and digital service providers.



7

Faster time to value

For rapid deployment of value- proven AI applications, the fastest route ROI and the achievement of business benefits will be by tapping into established, pre-configured environments and managed services.

8

Sustainability and energy efficiency

Industry leading data centre sustainability and energy efficiency that does not compromise the delivery of world-class operational excellence is an essential feature of AI ready infrastructure. As net neutral carbon imperatives become ever more critical the minimisation of power utilisation and associated carbon footprint of your AI deployment will be a key success metric.

9

Alliances and partners

Embedding yourself in infrastructure that is in close proximity to a multitude of partner and alliance relationships including the key global AI technology and advisory companies will be critical to success.

Harnessing the experience and expertise of the emerging “AI community” is an essential element for design, build and operational success.

Fast track to AI value: the NEXTDC advantage

When it comes to accelerating your AI program, and unlocking its business benefits, a trusted infrastructure partner is essential.

Customer first

Infrastructure guidance is critical at every step of the journey. Engage deeply to make sure you get the best, most experienced engineering, pre-sales support, sales engagement, security compliance, delivery management and customer experience teams.

Support your team

You need to stay focused on innovation and achieving your targeted business outcomes. Your data centre partner should make it easy to navigate the maze of configuration, management and optimisation of AI environments through flexibility, innovation and agility with all the necessary data centre skills and expertise to supplement your internal technology team.

Geodiversity and redundancy

Your AI environment will be unique. You need flexibility to build your secure, sovereign, private and interconnected AI infrastructure exactly as you require it within a single ecosystem, with multi-site, multi-city capability for back-up and disaster recovery.

Flexible, scalable growth

In most cases AI adoption will start small and scale incrementally. Ensure your infrastructure partner has plans for future growth requirements of AI environments. You want to gradually scale your footprint as AI adoption accelerates without losing the value of proximity to existing services and infrastructure.

Culture of innovation

Do you believe in the “Pursuit of Excellence:”? Hitch a ride on partners who measure and are transparent about innovation and continual improvement. AI technologies and valuable associated service provision will evolve rapidly in coming years. It is important to stay at the cutting edge.

How AI is enabling our operations

NEXTDC'S future-ready infrastructure platform benefits from our use of AI. We capture, analyse and leverage large amounts of data in the real time management of resilience and performance of our services to the benefit of our customers.

In conjunction with La Trobe University, we've been investing in how AI and ML can be applied to improve the performance of our NaaS virtual interconnection network, specifically through proactive monitoring and management to review past patterns and address potential issues before they happen.

We also see a significant opportunity for AI in driving data centre sustainability outcomes. We've previously invested in PhD research matching historical weather data to critical plant and equipment settings under a range of power loads with a view to optimising energy efficiency through tuning variables such as fan speeds, pump speeds and approach water temperatures.

Finally, we are also rolling out our AI alliance program. This is a range of AI ready workshops to progress and mature the AI capabilities of our customers and partners.

Watch this space for more – we're just getting started.

Conclusion

When it comes to AI, every deployment and use case will be unique. That's why it's critical that design, innovation and operational excellence remains fluid to whatever comes next. AI is in its infancy. Every deployment today needs to be flexible, agile and scalable to unknown but imminent disruption and supporting all bespoke environments that meet the exact customer requirements.

Technology infrastructure is a critical consideration for leaders who are currently looking to embrace AI as an essential element in their strategic organisational transformation. Where sovereignty, resilience, agile scalability and deep ecosystem engagement will be pre-requisites for AI success, the partner you choose will be critical consideration in this transformative journey.

All hyperscale, enterprise and government organisations should **contact NEXTDC** to understand more about how the right infrastructure partner will accelerate the opportunity to harness the power and potential of an AI-driven future.

NEXTDC
where AI lives™

Contact us

136 398
sales@nextdc.com

nextdc.com