NOVARTIS

Associate Director - AI Infrastructure Architect (AI Factory)

Job ID
REQ-10088123
Окт. 30, 2026
Чехия

Сводка

Novartis architecture principles call for AI adoption guided by business value: AI is evaluated for new solutions and enhancements wherever it delivers a clear benefit, and it must be delivered on infrastructure that is secure, compliant and economically sustainable. Within Cloud & Hosting Services, AI & HPC Engineering is the single product management and delivery channel for AI infrastructure and AI use-case onboarding. The AI Infrastructure Architect sets the architecture those platforms are built to.

The role defines the reference architecture and guardrails for AI workloads across Azure, AWS and private AI environments: training and inference platforms, GPU capacity and scheduling, data pipelines and data access, model operations, logging, monitoring, cost control and lifecycle governance. It establishes the standard intake pattern by which an AI use case moves from idea to a hosted, monitored, auditable service, so that scientists and business teams get speed without the organisation inheriting uncontrolled risk.

This is an individual contributor architecture role. Its authority comes from design quality, evidence and adoption rather than from line management.

About the Role

Key Responsibilities:

- Own the AI infrastructure reference architecture and the standards that govern training, fine-tuning, inference and retrieval-augmented patterns across public cloud and private AI hosting.
- Define the standard AI workload intake pattern, including architecture review criteria, data classification handling, environment selection and the conditions under which a use case may go to production.
- Set guardrails for sensitive data, model access, prompt and inference logging, monitoring, resilience, retention and cost, in partnership with Security, Data Privacy and Quality.
- Design GPU and accelerator capacity patterns - shared pools, reservations, scheduling, quotas and burst to cloud - and support the associated capacity and cost models.
- Support qualification of AI platforms, regions and services for enterprise consumption together with AI & HPC Engineering and the public cloud architects.
- Define the MLOps and LLMOps reference chain: model registry, versioning, evaluation, deployment, drift and performance monitoring, and decommissioning.
- Maintain architecture decision records for AI platform choices, linking each to business value, risk posture and total cost of ownership.
- Advise business, research and Infrastructure Solution Delivery stakeholders on feasible AI infrastructure options, and translate scientific and commercial requirements into technical designs.
- Track the AI infrastructure market, evaluate emerging accelerators, model-serving stacks and platform services, and bring credible innovation into the roadmap.
- Coach engineers and architects on AI infrastructure practice and act as a role model for the Novartis Values and Behaviours.

Essential Requirement :

- 8+ years' experience in infrastructure, cloud, or platform architecture, with a strong track record of designing and supporting enterprise-scale technology solutions.
- Demonstrated expertise in building, deploying, and operating AI and machine learning infrastructure within production environments.
- Experience enabling AI solutions to transition from proof-of-concept stages into scalable, governed, and business-critical services.
- Strong knowledge of GPU and accelerator technologies, Kubernetes, container orchestration, and cloud-based AI platforms such as Azure and AWS.
- Hands-on experience with model serving, inference optimization, MLOps/LLMOps practices, and modern data and vector pipeline architectures.
- Solid understanding of identity and access management, observability, cost optimization (FinOps), and secure AI platform operations.
- Familiarity with Responsible AI principles, including explainability, auditability, logging, human oversight, and risk management controls.
- Excellent stakeholder management and communication skills, with the ability to explain complex AI infrastructure concepts to technical and non-technical audiences while driving collaboration across engineering, security, privacy, and business teams.

Desirable Requirement:

- Fluent English, written and spoken. Additional languages are an asset.

Expected Annual Base Salary Range for role:1,120,210.00 - 2,080,390.00CZK

The base salary offered is determined based on gender-neutral objectives, such as relevant skills, competencies and experience in accordance with the Novartis pay

setting policy and upon joining Novartis will be reviewed periodically.

In addition to your base salary, you may be eligible for a performance-based bonus depending on certain performance parameters.

The rewards of being part of our team go far beyond base pay and incentives. We also offer a variety of competitive benefits in kind to help you thrive personally and professionally, such as insurance plans, retirement plans, wellbeing resources and global recognition programs. In addition, we provide flexible and hybrid working options, where possible, and minimum 14 weeks paid parental leave.

Pay equity is a fundamental principle of our employment policy and reflects our commitment to create a diverse, equitable and inclusive environment that treats all employees with dignity and respect, as outlined in our Code of Ethics.

Read our brochure to learn more about our global total rewards offering:https://www.novartis.com/sites/novartis_com/files/novartis-life-handbook.pdf

Note: Benefits and compensation may vary by country and are subject to local legal requirements, including provisions of collective bargaining agreements where applicable. A full overview of your compensation package, including any relevant collective bargaining agreement details applicable to your role based on your employment location and Novartis employer entity, will be communicated separately to you during the application process.

Commitment to Diversity & Inclusion:

Novartis is committed to building an outstanding, inclusive work environment and diverse teams representative of the patients and communities we serve.

Accessibility and accommodation:

Novartis is committed to working with and providing reasonable accommodation to all individuals. If, because of a medical condition or disability, you need a reasonable accommodation for any part of the recruitment process, or in order to receive more detailed information about the essential functions of a position, please send an email to di.cz@novartis.com and let us know the nature of your request and your contact information. Please include the job requisition number in your message.

Why Novartis: Helping people with disease and their families takes more than innovative science. It takes a community of smart, passionate people like you. Collaborating, supporting and inspiring each other. Combining to achieve breakthroughs that change patients' lives. Ready to create a brighter future together? https://www.novartis.com/about/strategy/people-and-culture

Benefits and Rewards: Learn about all the ways we'll help you thrive personally and professionally.
Read our handbook (PDF 30 MB)

Primary location salary range
Kč1,120,210.00 - Kč2,080,390.00
Дивизион
Operations
Business Unit
Information Technology
Место
Чехия
Сайт
Prague
Company / Legal Entity
CZ02 (FCRS = CZ002) Novartis s.r.o.
Functional Area
Информационные технологии
Job Type
Full time
Employment Type
Regular
Shift Work
No

Accessibility and accommodation

Novartis is committed to working with and providing reasonable accommodation to all individuals. If, because of a medical condition or disability, you need a reasonable accommodation for any part of the recruitment process, or in order to receive more detailed information about the essential functions of a position, please send an e-mail to di.cz@novartis.com and let us know the nature of your request and your contact information. Please include the job requisition number in your message.

Job ID
REQ-10088123

Associate Director - AI Infrastructure Architect (AI Factory)

Apply to Job
Job ID
REQ-10088123

Associate Director - AI Infrastructure Architect (AI Factory)

Apply to Job

Source URL: *https://novartis.ru/careers/career-search/job/details/req-10088123-associate-director-ai-infrastructure-architect-ai-factory*

List of links present in page

1. https://www.novartis.com/sites/novartis_com/files/novartis-life-handbook.pdf
2. mailto:di.cz@novartis.com
3. https://www.novartis.com/about/strategy/people-and-culture
4. https://www.novartis.com/sites/novartis_com/files/novartis-life-handbook.pdf
5. https://novartis.wd3.myworkdayjobs.com/en-US/Novartis_Careers/job/Prague/Associate-Director---AI-Infrastructure-Architect--AI-Factory-_REQ-10088123
6. https://novartis.wd3.myworkdayjobs.com/en-US/Novartis_Careers/job/Prague/Associate-Director---AI-Infrastructure-Architect--AI-Factory-_REQ-10088123