Platform Engineer, Scale GenAI Platform

Description

Join us at SGP (Scale Generative Platform), the cutting-edge, full-stack platform tailored for enterprise innovation in AI. At SGP, we empower businesses to design, develop, and deploy AI applications from the ground up within the secure confines of their private clouds, ensuring maximum data privacy and eliminating the risk of data breaches.

At the heart of SGP is our robust testing and evaluation segment. Here, precision meets innovation: We are pioneering the development of automated AI evaluations, designed to accelerate project timelines significantly and combine this with our meticulous human annotation processes to ensure unparalleled AI accuracy.

Our platform is versatile, supporting an array of models from leading AI innovators including OpenAI, Cohere, and Google. This opens up a realm of possibilities for enterprise applications beyond the conventional uses of technologies like ChatGPT.

SGP is at the forefront of AI application, backed by profound insights and extensive visibility gained through collaborations with major model developers and large enterprises. Our rigorous testing and evaluation methods enhance trust and reliability in AI models, setting a new standard in the industry.

Discover how SGP is enabling thousands of businesses worldwide to harness the power of AI and redefine their industries. Join us in shaping the future.

We are seeking a Platform Engineer who is enthusiastic about bringing generative AI applications from concept to production. This vital role involves deploying these sophisticated applications within customer-specific Virtual Private Cloud (VPC) environments and optimizing deployment processes across various cloud platforms.

Join us in our mission to revolutionize how AI is integrated and managed in enterprise settings, ensuring every deployment is a benchmark for industry excellence.

You will:

- Drive the architecture, design, implementation and support of our foundational platforms and systems, working closely with stakeholders and internal customers to understand and refine requirements.
- Collaborate with cross-functional teams to define, design, and deliver new features.
- Proactively identifying opportunities for, and driving improvements to, current programming
- practices, including process enhancements and tool upgrades.
- Present technical information to internal and external teams and stakeholders, providing guidance and insight on development processes and technologies.
- Craft scalable systems that maintain their integrity across different cloud providers.
- Meticulously manage multiple deployment versions, ensuring each aligns with specific customer requirements.

Hiring organization
Candidate-1st

Employment Type
Full-time

Beginning of employment
asap

Job Location
Budapest, Hungary

Working Hours
40

Base Salary
euro EUR 57K - 106K *

Date posted
May 21, 2024

- Navigate the unique challenges of a non-traditional SaaS product environment.

Ideally you'd have:

- **Professional Experience**: Minimum of 5 years of full-time engineering work post-graduation, with a focus on back-end systems.
- **Technical Expertise**: Extensive knowledge in software development with a robust understanding of distributed systems, cloud platforms, and data science.
- **Containerization & Deployment Technologies**: Proficient in using standard technologies such as Kubernetes, Terraform, Docker, and Helm across various cloud environments.
- **Orchestration Platforms**: Experience with orchestration tools, including Temporal and AWS Step Functions.
- **System Integration**: Skilled in building and maintaining scalable systems, including pipeline creation, process enhancements, continuous integration, and continuous deployment (CI/CD) operations (CircleCI).
- **Cloud Services**: Familiarity with multiple cloud providers is advantageous but not required.
- **Communication Skills**: Strong ability to communicate technical concepts and collaborate effectively with multidisciplinary teams.

Nice to have:

- **Nice to haves**

  - Excitement to work with AI technologies.
  - Familiarity with advanced DevOps practices and tools, enhancing automation and monitoring in a production environment.
  - Experience with authentication/authorization systems (Zanzibar, Authz, etc.)
  - Experience scaling products at hyper-growth startups.

About Us:

At Scale, we believe that the transition from traditional software to AI is one of the most important shifts of our time. Our mission is to make that happen faster across every industry, and our team is transforming how organizations build and deploy AI. Our products power the world's most advanced LLMs, generative models, and computer vision models. We are trusted by generative AI companies such as OpenAI, Meta, and Microsoft, government agencies like the U.S. Army and U.S. Air Force, and enterprises including GM and Accenture. We are expanding our team to accelerate the development of AI applications.

We believe that everyone should be able to bring their whole selves to work, which is why we are proud to be an affirmative action employer and inclusive and equal opportunity workplace. We are committed to equal employment opportunity regardless of race, color, ancestry, religion, sex, national origin, sexual orientation, age, citizenship, marital status, disability status, gender identity or Veteran status.

We are committed to working with and providing reasonable accommodations to applicants with physical and mental disabilities. If you need assistance and/or a reasonable accommodation in the application or recruiting process due to a disability, please contact us at accommodations@scale.com. Please see the United States Department of Labor's [Know Your Rights poster](#) for additional information.

We comply with the United States Department of Labor's [Pay Transparency provision](#).

PLEASE NOTE: *We collect, retain and use personal data for our professional business purposes, including notifying you of job opportunities that may be of interest and sharing with our affiliates. We limit the personal data we collect to that which we believe is appropriate and necessary to manage applicants' needs, provide our services, and comply with applicable laws. Any information we collect in connection with your application will be treated in accordance with our internal policies and programs designed to protect personal data.*

How the process will look like

Your teammates will gather all requirements within our organization. Then, once priority has been discussed, you will decide as a team on the best solutions and architecture to meet these needs. In continuous increments and continuous communication between the team and stakeholders, you're part of making data play an even more important (and understood) part withing Brand New Day.

Job Benefits

 EUR 57K – 106K *