



Jakarta, Indonesia  
30 November 2024

## Scaling Self-Hosted LLMs on Kubernetes with Fluid

Solo Presentation

**Arief  
Rahmansyah**

Machine Learning Engineer



