

Pranshu Gupta

pranshug258@outlook.com • +1 (425) 679-2402 • <https://linkedin.com/in/pranshug258/> • <https://pranshug.com>

Professional Experience

Microsoft Azure OpenAI • Senior Software Engineer • Jan 2025 – Present

Lead the development and technical architecture in the data plane of Azure OpenAI service, which handles all the inference requests sent to OpenAI models hosted on Azure, via Microsoft Copilots and customers' AI agents.

- Led and delivered the design and implementation of the rate limiter for inference requests from scratch.
- Contributed to the design and implementation of Azure AI's model router, announced at Build 2025.

Microsoft Azure • Senior Software Engineer • Mar 2024 – Jan 2025

Lead the development and technical architecture in resource provisioning area of Azure resource manager.

- Led and delivered improvements in bulk delete API, reducing P90 latency by 50% and P99 by 65%.
- Designed and led the development of background job quota-based throttling in Azure control plane.

Microsoft Azure • Software Engineer II • Jun 2020 – Feb 2024

Deliver high-quality, robust production code for Azure Resource Manager, the control plane of Azure.

- Implemented a critical security patch and helped more than 200 teams in Azure onboard to it.
- Optimized resource provisioning failure handling, which reduced unnecessary http requests by 90%.
- Implemented callback design for resource actions, which reduced the number of polling requests by 60%.
- Implemented staged retries for dependencies in resource group deletion, which reduced failures by 98%.

Microsoft Core Services • Software Engineer • Jun 2017 – Aug 2019

Design and implement modules for Customer Data Enrichment Service, which enriches customer data from marketing campaigns to help marketers create better sales opportunities for Microsoft.

- Migrated to serverless architecture, reducing operation costs by 90% and improving performance by 3x.

Technical Skills

- Programming Languages : C#, PowerShell, Azure Bicep, Python, C/C++, Kusto QL, HTML, CSS, JS, MD
- Tools and Platforms : Microsoft Azure, .NET Framework, Azure CLI, Azure DevOps

Education

Georgia Institute of Technology, Atlanta • May 2020 • MSCS (ML specialization)

Indian Institute of Technology, Kanpur • May 2017 • B.Tech. CS

Publications

- Saha, K., Gupta, P. et al. Observer Effect in Social Media Use | Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems - <https://dl.acm.org/doi/10.1145/3613904.3642078>
- El Sherief, M., Saha, K., Gupta, P. et al. Impacts of school shooter drills on the psychological well-being of American K-12 school communities: a social media study using multi-modal data analysis. Humanities and Social Sciences Communications 8, 315 (2021) – <https://www.nature.com/articles/s41599-021-00993-6>

Selected Projects

- Image completion with statistics of patch offsets – 44 stars on GitHub.
- Crowd behavior analysis – detect outliers in the trajectories of entities in a moving crowd using the minimum description length principle – 39 stars on GitHub.
- Deep image captioning – generate captions for an image using deep reinforcement learning model with embedding reward – 26 stars on GitHub.