

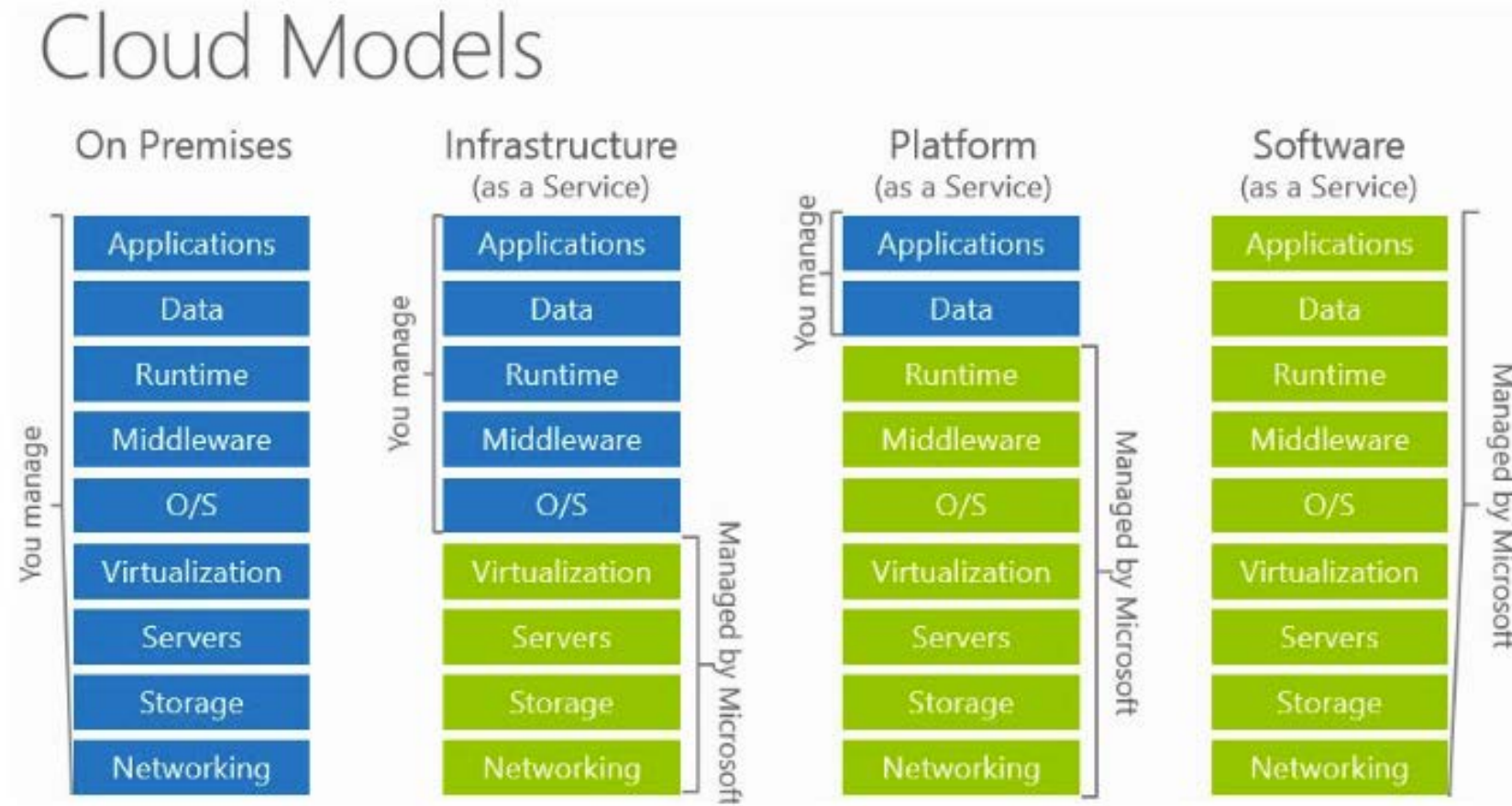
Getting Started with Local Models

JASON HALEY, MICROSOFT AI MVP

MARCH 2025



Remember the Cloud Hosting Spectrum?



AI Model Hosting Spectrum

- **Models hosted or controlled by you** - This is what we are going to focus on

Models as a Service (MaaS)- Azure AI Foundry will provide you many models as a service

- **Models hosted by third parties** - Azure OpenAI is an example of this

External Service/Website/API hosted by the model created - OpenAI's Chat GPT is an example of this

Local Models – Three Options

ONNX Runtime – Open Neural Network Exchange

- <https://learn.microsoft.com/en-us/azure/machine-learning/concept-onnx?view=azureml-api-2>

LM Studio

- <https://lmstudio.ai/>

Ollama

- <https://ollama.com/>

Hands On Lab

<https://github.com/JasonHaley/chainlit-ollama-phi4>