```
Please use '!option handwritten true' to enable handwritten
```

OllamaNet Architecture - Technical Details

«person»

User

Uses

OllamaNet Platform
[system]

«container»
API Gateway
[Ocelot/.NET]

Routes requests, handles
cross-cutting concerns

Routes to   Routes to   Routes to   Routes to   Routes to

«container»
Auth Service
[.NET]

Authentication, JWT, RBAC

«container»
Explore Service
[.NET]

Model discovery, search

«container»
Admin Service
[.NET]

Model management, admin
operations

«container»
Inference Service
[Python/FastAPI]

LLM inference, ngrok tunnel

«container»
Conversation Service
[.NET]

Chat, RAG, document
management

Caches   Reads/writes   Caches   Reads/writes   Caches   Reads/writes   Subscribes   Reads/writes (if any)   Publishes URL   Caches   Reads/writes   Subscribes   Queries

«container»
Redis
[Cache]

Distributed caching

«container»
SQL Server
[Database]

Shared relational data

«container»
RabbitMQ
[Message Broker]

Async messaging, service
discovery

«container»
Pinecone
[Vector DB]

Embeddings for RAG