

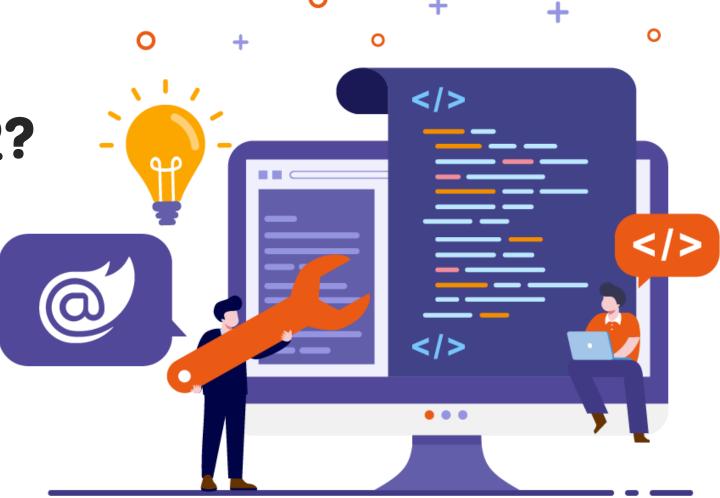




Microsoft Al.NET Day

마우스 클릭 몇 번으로 AI 챗봇을 만들어볼까요?







Speaker

- 김 진석
- (주)케이에이치시스템즈 대표이사
- Microsoft MVP
- @ 한국블레이저개발자모임 / Blazor Korea
- facebook.com/iamjinseok.kr







Blazor, MAUI, WPF을 중심으로 .NET과 관련된 이야기를 하는 정기 밋업 올해부터는 AI도 조금, Cloud도 조금...



Sponsor



- IT 인프라 컨설팅 및 솔루션 제공 업체



- Microsoft에서 진행하는 여성을 위한 IT 교육 프로그램

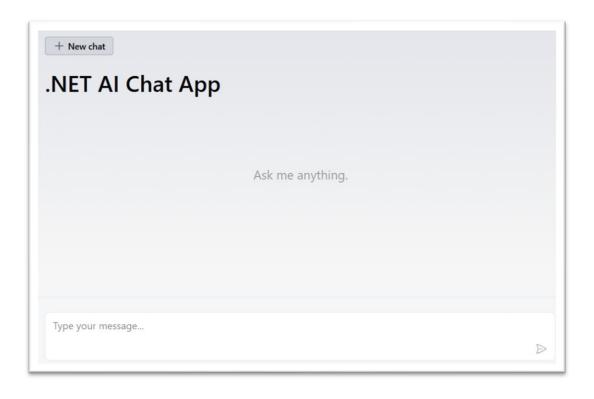


약간의 배경 지식?





오늘의 목표



- .NET AI Template을 알아보자
 - preview
- Microsoft.Extensions.AI가 있다
- AI 관련 개념도 이해하면 좋을 듯
 - RAG



AI 채팅의 UI의 패턴화



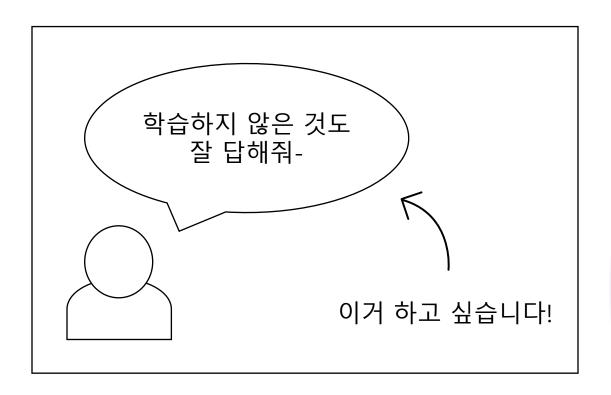




- 채팅창
- 비슷함



RAG?



- Retrieval-Augmented Generation
 - 검색 증강 생성
- LLM이 학습하지 않은 것에 대해서 알려주고 답하게 하는 것
- 조직 내부 데이터 등 유출하고 싶지 않은 데이터를 취급하고 싶음



RAG?

달콤하고 노란 과일이 뭐야?

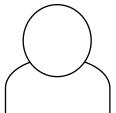
모르지.. 좀 찾아보고

망고 아냐?

그 과일은 망고야

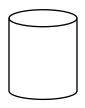
아하!

- 이런 흐름으로 답변
- DB/Vector Store에 정보 저장
- 단순화한 예시!





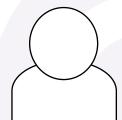
Αl



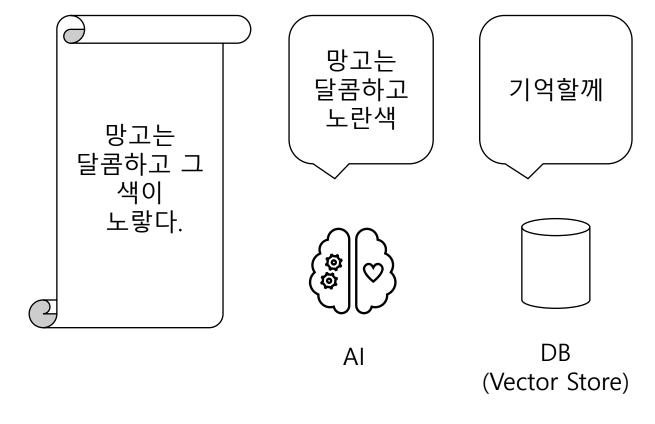
DB (Vector Store)



Αl

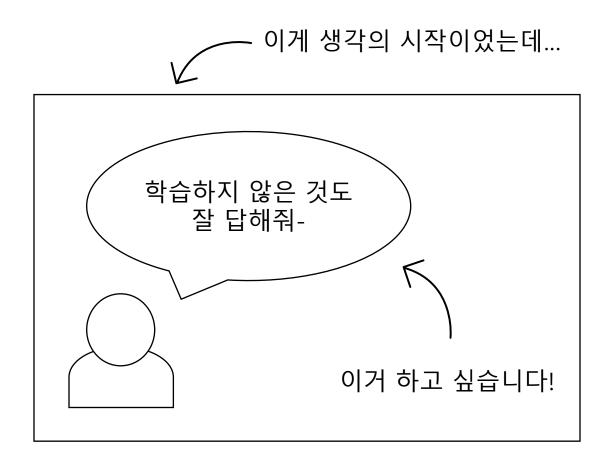


RAG?



- 정보를 기억하는 방법과 다시 찾는 방법이 매우 중요해짐
- Indexing / embedding
- Search

RAG?



- 요즘 모두 이 패턴
- 조금 더 개선되기 위한 추가적인 방법들
 - Advanced RAG
 - 데이터의 전처리와 후처리
 - 조회 방법의 개선
 - ... 등등
- 파인튜닝을 거의 대체하나? (개인의견)

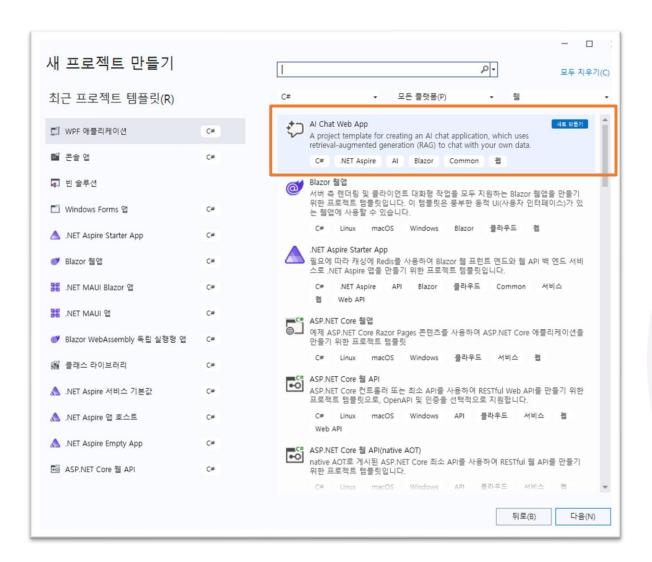
Install Template

- dotnet new install Microsoft.Extension.Al.Templates
- aichatweb
- preview





새 프로젝트 만들기

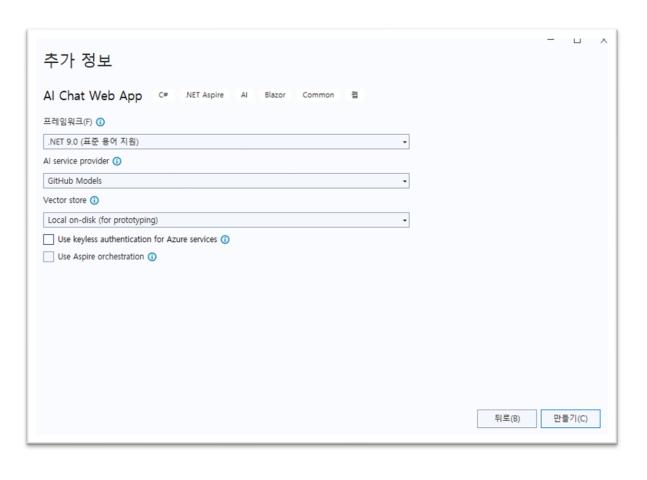


- 새 프로젝트 만들기
- Al Chat Web App





새 프로젝트 만들기



- Al service provider
 - Azure OpenAl
 - GitHub Models
 - Ollama (for local development)
 - OpenAl Platform
- Vector store
 - Local on disk(for prototyping)
 - Azure Al Search
 - Qdrant
- Use keyless authentication for Azure services
- Use Aspire ochestration





환경 설정

```
{
    "AzureOpenAI:Key": "YOUR-AZURE-OPENAI-KEY",
    "AzureOpenAI:Endpoint": "YOUR-AZURE-OPENAI-ENDPOINT"
}
```

- AzureOpenAI: secrets.json
 - Key
 - EndPoint

- Azure Al Foundry : 모델 배포
 - gpt-4o-mini
 - text-embedding-3-small
- Azure Portal : 리소스 삭제



Components/Pages/Chat/Chat.razor

```
private const string SystemPrompt = @"
   You are an assistant who answers questions about information you retrieve.
  Do not answer questions about anything else.
  Use only simple markdown to format your responses.
  Use the search tool to find relevant information. When you do this, end your
   reply with citations in the special XML format:
   <citation filename='string' page_number='number'>exact quote here</citation>
  Always include the citation in your response if there are results.
   The quote must be max 5 words, taken word-for-word from the search result,
   and is the basis for why the citation is relevant.
  Don't refer to the presence of citations;
   just emit these tags right at the end, with no surrounding text.";
```

- Blazor Component

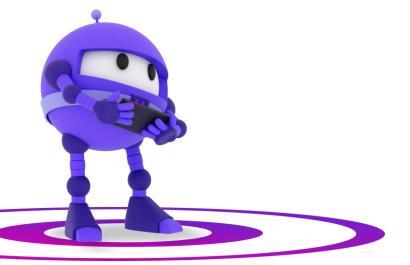
- SystemPrompt





Demo

- new project
- vector store
- my documents

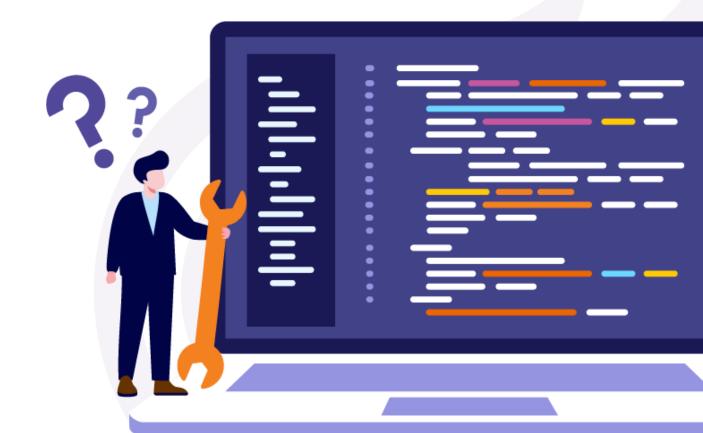






Q&A

- RAG
- Microsoft Extension Al Template





THANK YOU



