



CODE Presents:

Prompt Engineering - How to Talk to an AI

Markus Egger

President and Chief Software Architect

CODE
CONSULTING

CODE
STAFFING

CODE
MAGAZINE

CODE
TRAINING



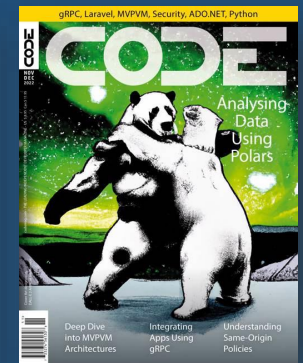
"Helping People Build
Better Software"



About Markus Egger

CODE
PRESENTS

- President and Chief Software Architect
EPS Software Corp. (dba CODE Consulting)
- CEO – CODE Staffing
- Publisher – CODE Magazine
- International Author and Speaker
- Microsoft RD (Regional Director)
- Microsoft MVP 1995-2019
- Email: markus@codemag.com
- Twitter/X: @markusegger



About CODE Consulting



- Custom Software Development
- Artificial Intelligence
- Web, Cloud, Mobile, Desktop, Serverless, Databases,...
- Copilot Development, AI, GPT & Azure OpenAI
- Application Security Testing
- User Interface and Interaction Design
- Project Rescue, App Modernization (VB, VFP, Access, etc.)
- Support & Maintenance for existing applications



AI Consulting Services



CODE CONSULTING Home Magazine Consulting Staffing Training Framework VFP Conversion [Login](#)

Executive Briefing (AI)

Are you wondering how AI can help your business? Do you worry about privacy or regulatory issues stopping you from using AI to its fullest? We have the answers! Our Executive Briefings provide guidance and concrete advice that help decision makers move forward in this rapidly changing Age of Artificial Intelligence and Copilots!

We will send an expert to your office to meet with you. You will receive:

1. An overview presentation of the current state of Artificial Intelligence.
2. How to use AI in your business while ensuring privacy of your and your clients' information.
3. A sample application built on your own HR documents - allowing your employees to query those documents in English and cutting down the number of questions that you and your HR group have to answer.
4. A roadmap for future use of AI tailored to what you do.

A Briefing Combined with Solid Deliverables!

We will send an expert to your office to meet with you. You will receive:

1. An overview presentation of the current state of Artificial Intelligence.
2. How to use AI in your business while ensuring the privacy of your data.
3. A sample application built on your own HR documents - allowing your employees to query those documents in English and cutting down the number of questions that you and your HR group have to answer.
4. A roadmap for future use of AI tailored to what you do.

Executive Briefings are set up specifically for you and usually take several hours and up to a day. It is up to you whether you want to include only C-level Executives or your entire company. The cost is **\$5,000** (which is a discounted special rate, compared to the normal price of \$7,500) plus expenses if we have to travel to your location, or stay over night. This includes the concrete deliverable of a first implementation of AI for your business (see #3 above, although we may also agree that something slightly different is more appropriate for your organization).

[Contact us today to set up your individual Executive Briefing!](#)

Setting Directions for the Future of Your Organization

We are experiencing a time of extreme change in technology. While one might argue that this has always been true, it is clear that current developments have a drastic impact on all organizations at a scale that is orders of magnitude beyond what we have experienced in years past.

This represents a time of extreme opportunity, as well as a time of considerable risks!

Artificial Intelligence (AI) continues to revolutionize industries across the globe, offering immense potential for organizations to enhance operational efficiency, drive innovation, and achieve competitive advantage. Conversely, ignoring these technological trends means that even very recent technology is quickly turning into "legacy" and your competitors now have the opportunity to leapfrog you. Especially the advent of "The Age of Copilots" (powered largely by the onset of Large Language Models - LLMs) promises a productivity improvement of upwards of 50% in some areas year after year, and that isn't the only trend to pay attention to.

What does all of this mean specifically for your organization?

- Check out our new **Executive Briefing** Offer!
- We can help with your AI needs
- What does AI mean for you?
- "Skunk Works" Projects



- www.codemag.com/AI
- www.codemag.com/ExecutiveBriefing

CODE Staffing



- Disrupting the world of staffing!
- Giving our customers the ability to have staff on par with Silicon Valley companies...
- ...and our employees a work environment in a bleeding-edge tech company with the **industry leading benefits!**
- www.codestaffing.com



We Need You!

- We are currently “hiring on all cylinders” across the entire CODE Group Organization!
- Check out www.codestaffing.com/careers



Upcoming CODE Training Classes



- Public AI Training Classes
 - Introduction to Development with Open AI, ChatGPT, Azure Open AI, and more...
 - Building Copilots for your own apps
- Classes are also available as individualized private classes and mentoring sessions



<https://www.codemag.com/training>

Upcoming Events



- State of .NET Tour – Live & In-Person – Fall 2023
- State of .NET Online - .NET 8 – November
- <https://www.codemag.com/StateOfDotNet>



- CODE Presents – Monthly – Mostly AI
- <https://www.codemag.com/CODEPresents>



Free Subscription!

CODE
MAGAZINE

- The leading software development magazine written by expert developers for developers.
- All registered attendees will automatically receive a free digital subscription to CODE Magazine – no need to do anything, it'll happen auto-magically.
- Subscribers get CODE Focus issues free of charge!

- Please share this free subscription link:





<https://tinyurl.com/cp08302023>



Recordings & Slide Decks

- State of .NET Webinar Series
 - codemag.com/StateOfDotNet
- CODE Presents Webinar Series
 - codemag.com/CODEPresents

Recent Event Recordings

State of .NET - The State of .NET App and System Development 2/23/2022	State of .NET - .NET Conf Recap 11/17/2021
	
State of .NET - .NET 6 Preview 10/27/2021	State of .NET - Azure Architecture 9/29/2021
	

Event Survey – Win \$100!

- Complete this very short 12 question survey for a chance at a \$100 Amazon Gift Card!
- Survey must be completed by 11:59pm ET on Friday 9/8/2023 to be eligible!
- THIS SLIDE WILL BE REPEATED AT THE END AND SURVEY LINK REPEATED IN THE CHAT WINDOW!

CODE Presents: Prompt Engineering - How to Talk to an AI Survey

The survey will take approximately 4 minutes to complete.

Thank you for attending! Please complete this brief 12 question survey to be eligible to win a \$100 Amazon Gift Card. Your survey must be completed by 11:59pm ET (UTC-4) on 09/02/2023 to be eligible to win! One entry per person please. Drawing will occur and the individual winner notified by 09/08/2023.

Thank you for attending! Please complete this brief survey. Yes, we still want to hear from you if you were unable to attend but watched the recording instead.

* Required

1. Full Name *

Enter your answer

2. Company Name *

Enter your answer

<https://tinyurl.com/cp08302023survey>

Thank you for attending!



- Welcome new attendees!
- Welcome repeat attendees!
 - We now have quite a number of people who attended 50+ events
- Our AI presentations are the most popular we have ever done
 - ...and we have been at this for 30 years!
- Based on the attendee list, I am guessing we have a mix of developers, decision makers, and users

Agenda

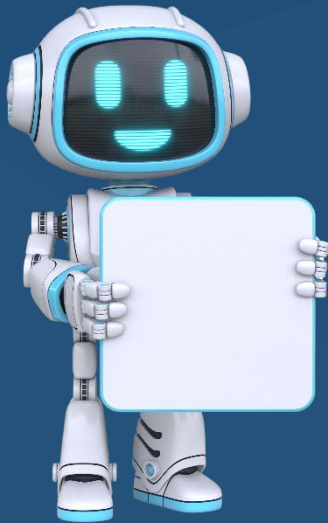
- Some LLM AI Fundamentals
- Prompt Engineering for Users
- Prompt Engineering for Developers
- Advanced Scenarios
- Sharing what we have learned in real-world scenarios so far



LLM AI Fundamentals

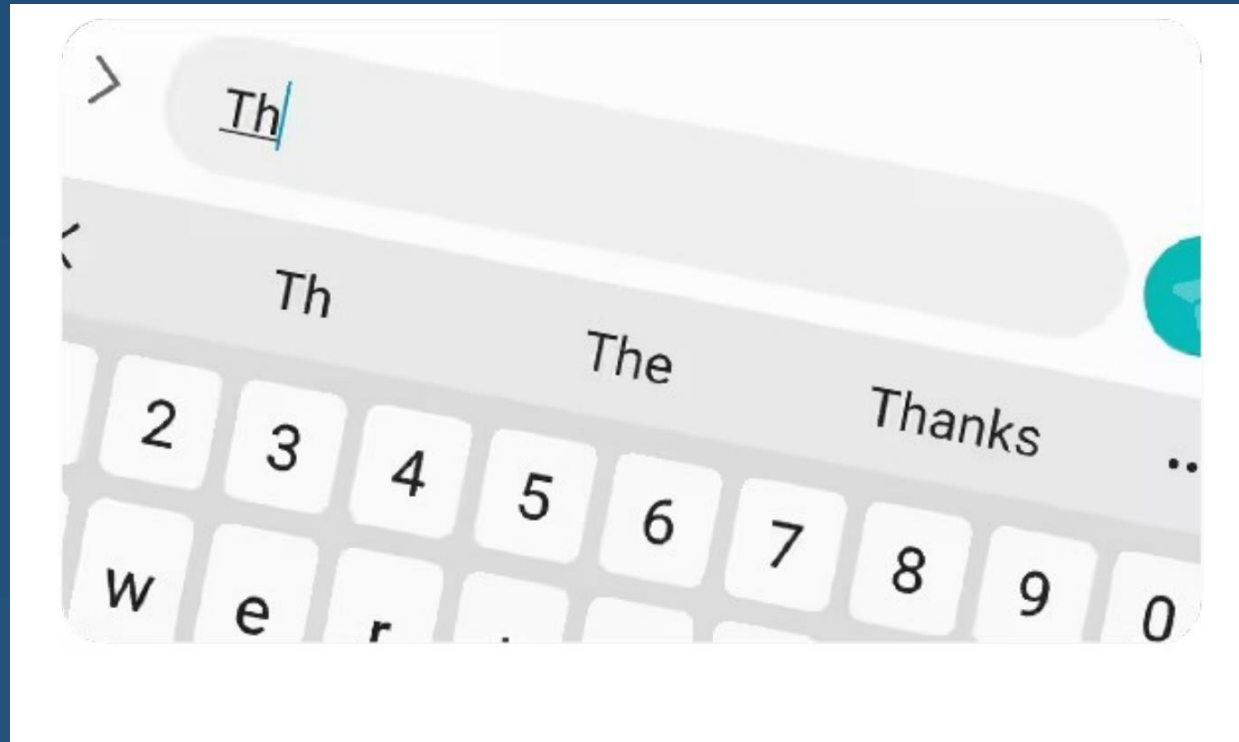
What are LLM AIs?

- There are many different kinds of AIs
- Currently, everyone is talking about LLMs
 - The most prominent example is ChatGPT
- LLMs are what we are focusing on in today's presentation



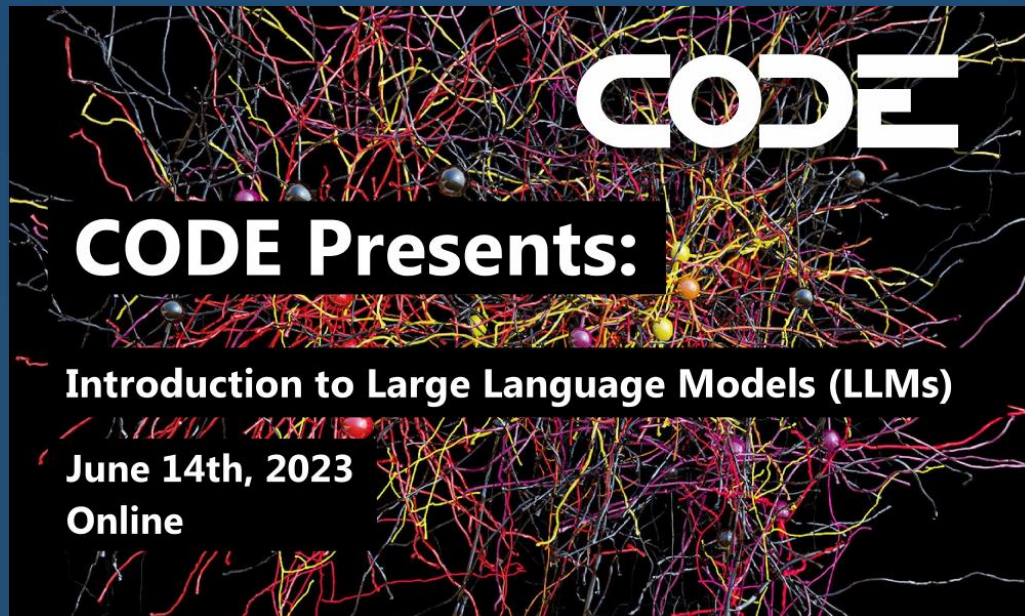
Large Language Models

LLMs based on frequencies



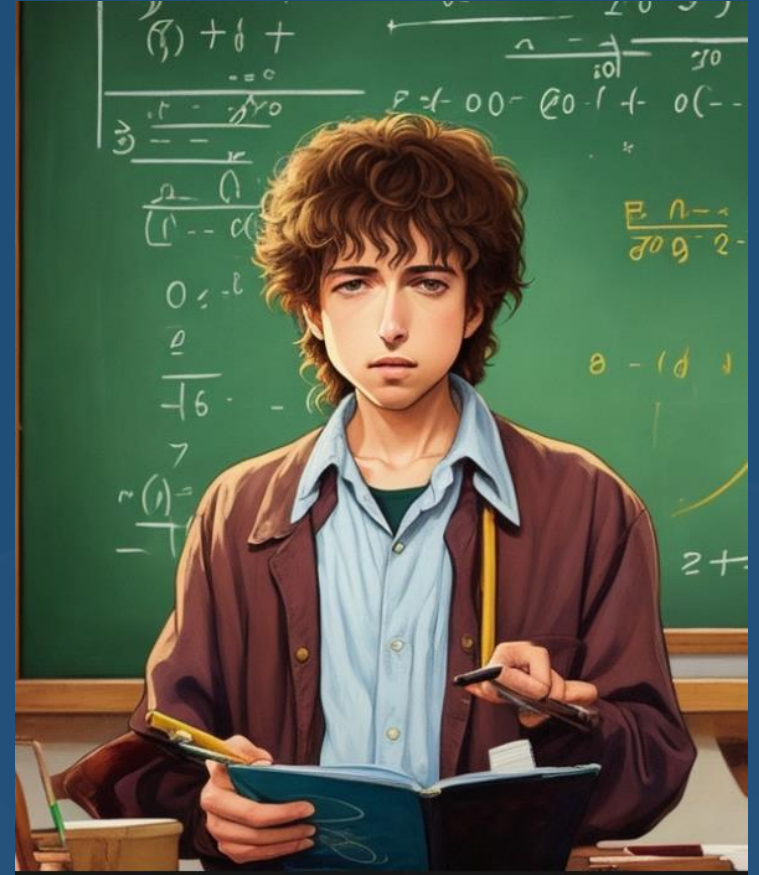
LLMs in Detail

- A past CODE Presents webinar goes into detail about how LLMs work and what they can do (presented by Dr. Otto Dobretsberger)
- A video recording is available (free of charge)

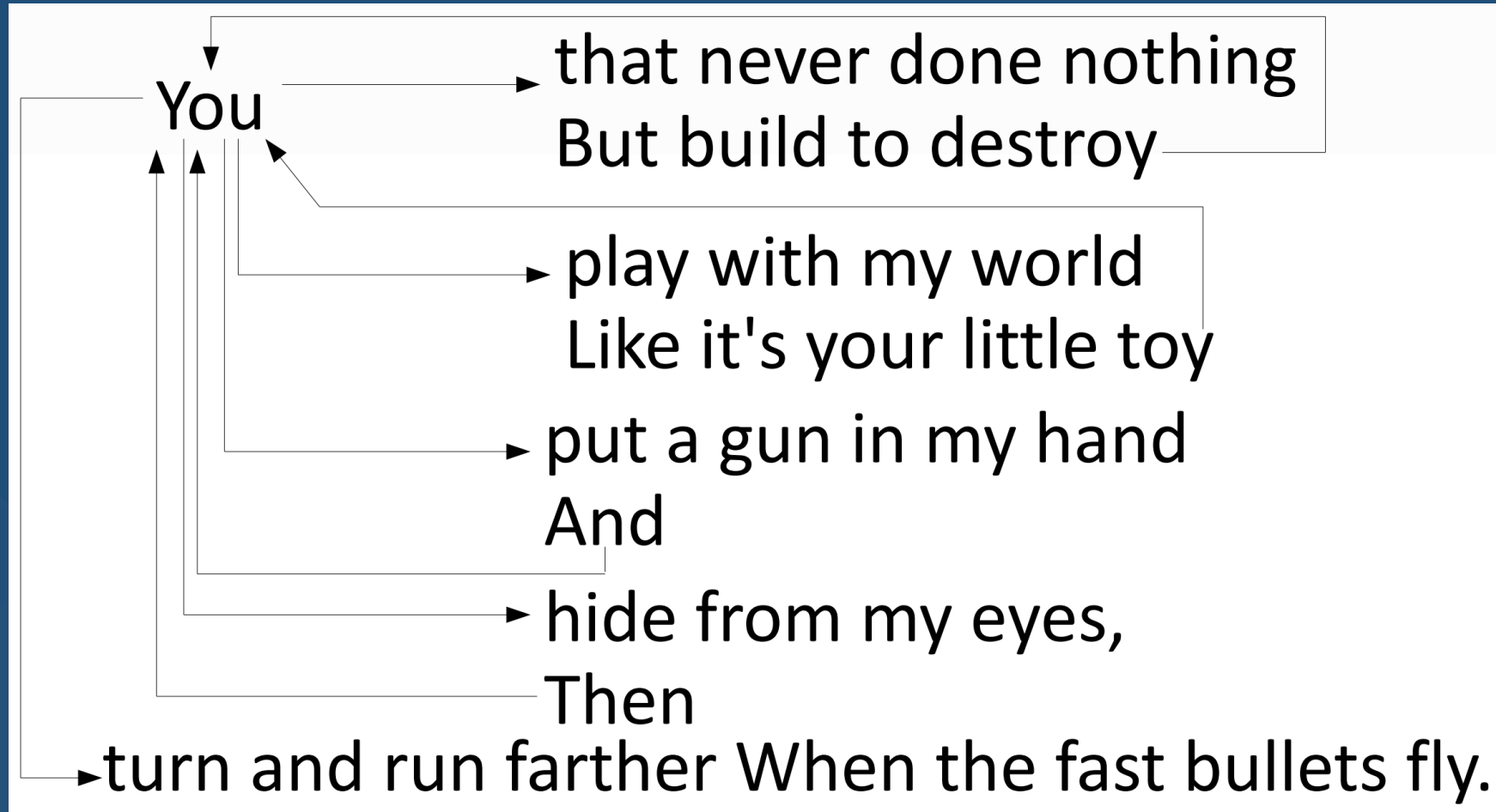


Large Language Models with Bob Dylan

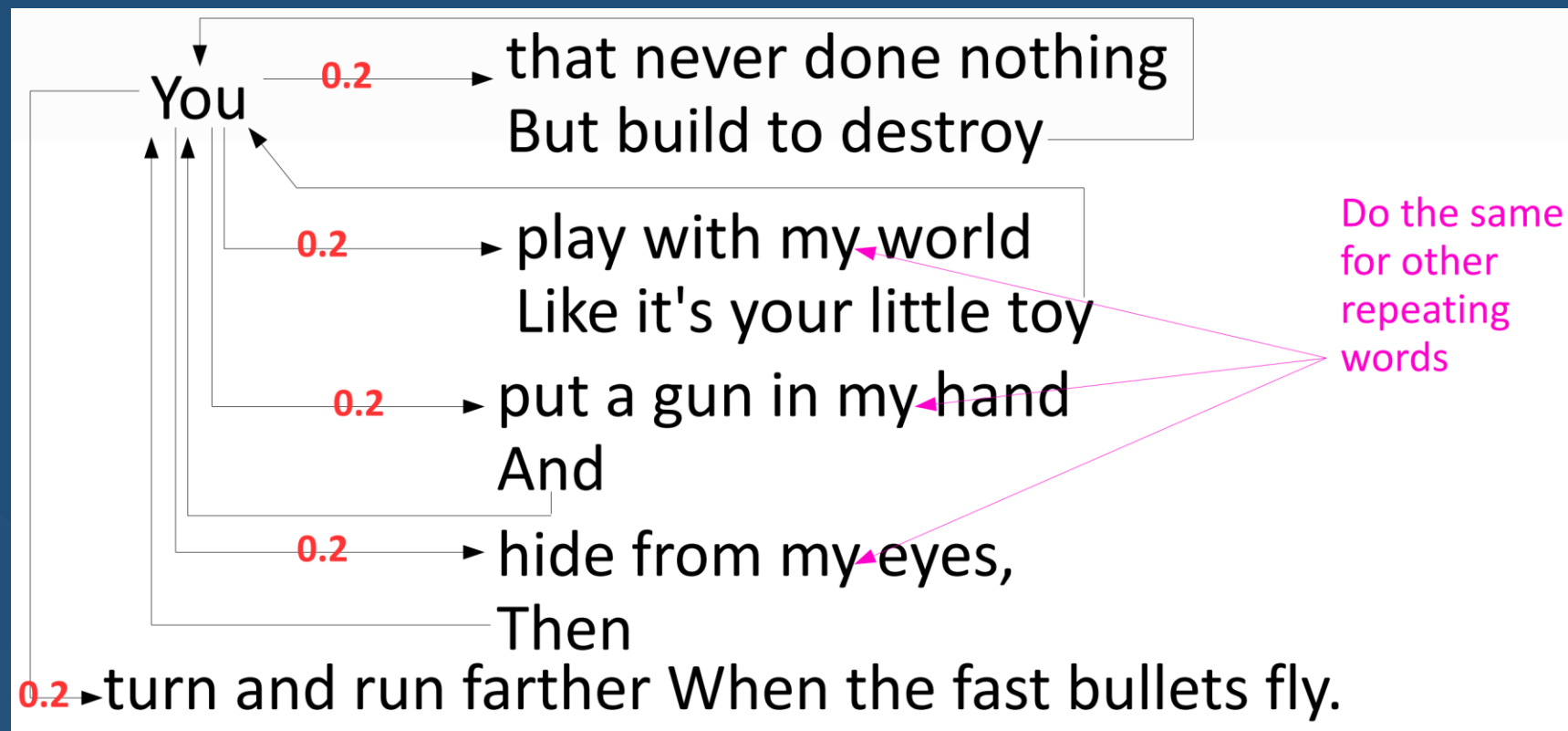
You that never done nothing
But build to destroy,
You play with my world
Like it's your little toy.
You put a gun in my hand
And **you** hide from my eyes,
Then **you** turn and run farther
When the fast bullets fly.



Large Language Models

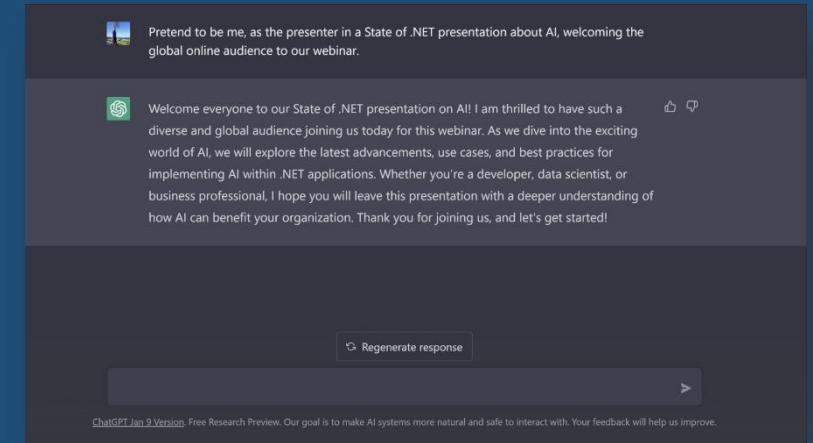


Large Language Models – Edge Probabilities



OpenAI's ChatGPT

- *GPT = Generative Pre-Trained Transformer
- You have probably heard about this 😊
 - Although just very recently
- A technological leap that forms the basis for a lot of new tech
- Note that this is just a "text calculator"... just a very good one 😊
 - The result is that this does a lot more than was anticipated, even by the creators
- Microsoft is very heavily invested



Who is OpenAI?

- Non-Profit OpenAI Incorporated
- For-Profit OpenAI Limited Partnership
- Founded in 2015 in San Francisco
- Microsoft made a \$11 billion investment and reportedly owns 49% of OpenAI LP
- All OpenAI Infrastructure runs on Azure



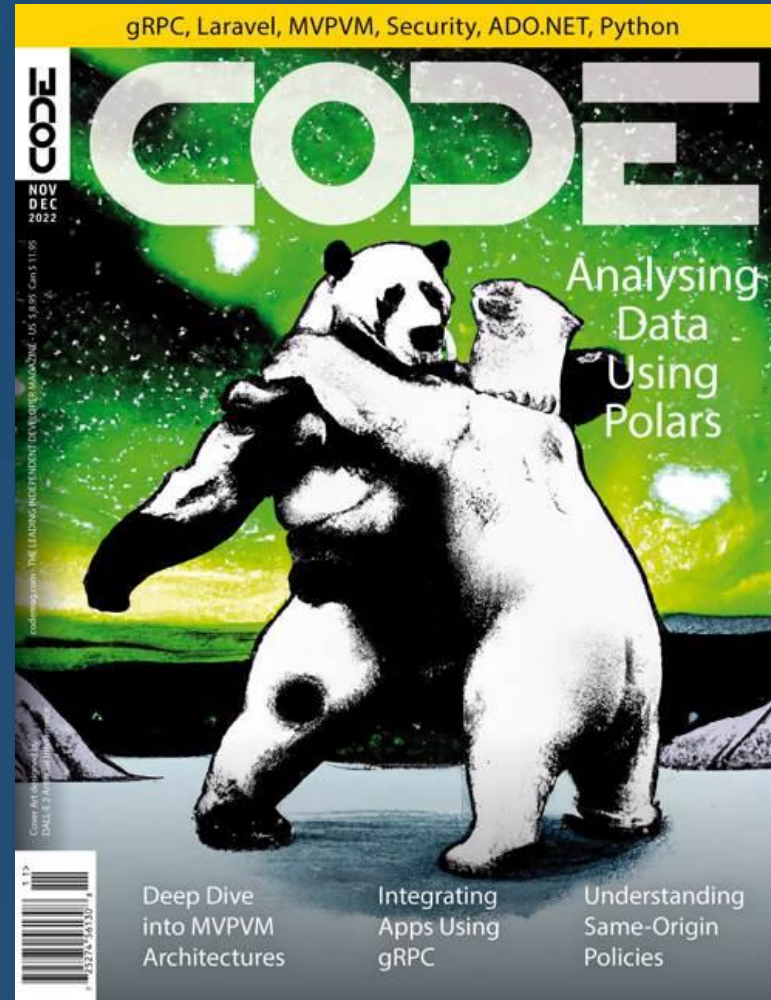
Azure OpenAI for Developers



- OpenAI features are now available to use in our own applications through Azure OpenAI
 - You can currently sign up for approval
- Azure OpenAI provides access to OpenAI features with all the added Azure benefits
 - It is also possible to use OpenAI features directly from OpenAI, but the Azure version already adds a lot, and will likely add more in the future
- This provides an API that allows us to create our own Copilot-style features

OpenAI – DALL-E 2

CODE
PRESENTS



Midjourney

CODE
PRESENTS

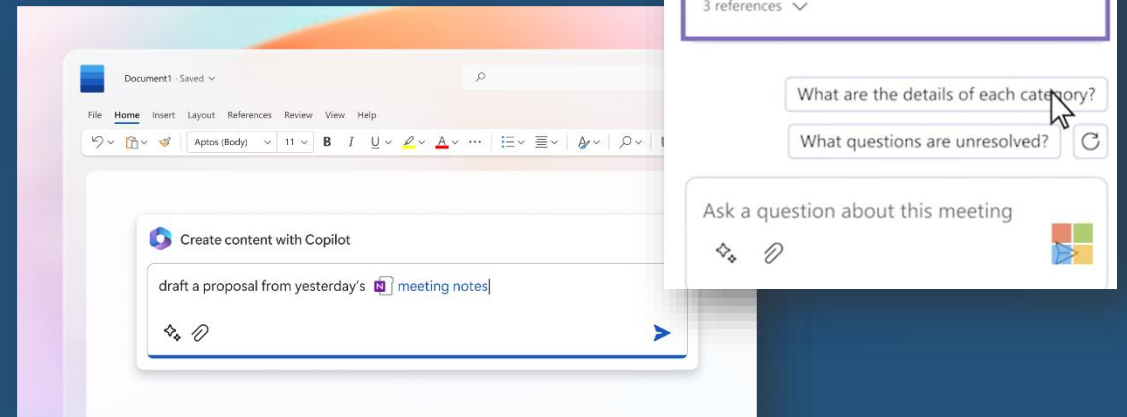


Many Many AIs

- I think the future for most developers will be using a mix of many different AIs that are good (or economic) for specific scenarios

Microsoft Copilot

- Microsoft is now introducing Copilot features in practically all applications
- Copilots are expected to drastically improve productivity
 - They already are today!
- <https://news.microsoft.com/reinventing-productivity/>



GitHub Copilot

CODE
PRESENTS



- GitHub Copilot assists in writing code
 - Now used by over 1 million developers
 - Productivity gains of more than 50%!
-
- Check out the recording from our State of .NET Webinar and on our YouTube channel!

**All apps without
Generative AI features
are now legacy!**

Prompt Engineering

What is Prompt Engineering?

- Prompt Engineering is the Art of Talking to an AI
- ...which means different things to users and developers
- ...although the basic concepts are the same
- The goal is to use the characteristics of a “text calculator” to coax it into producing correct and useful results

The Basics Idea

- When talking to an AI, one sends “prompts” back and forth between the “consumer” and the “AI”
- The “consumer” could be a human user or a program
- The things we are discussing today apply to both
 - ...although the more advanced scenarios require custom programming as there are multiple parameters and longer sequences of “back and forth conversation”

What Exactly are “Prompts”

- Prompts are clear text (in English or any other language) that are sent to, or received from an AI
- There are different types of Prompts:
 - **User Prompt:** What you would type into ChatGPT
 - **System Prompt:** An additional hidden prompt that can be used to “configure” the interaction
 - **Assistant Prompt:** The response from the AI
- There are additional parameters developers can set, but much of the interaction happens in plain text in the prompt

Getting Started

- ChatGPT
- Bing Chat
- OpenAI Playground
- Azure OpenAI Playground
- Programmatic Interaction

Using Appropriate Models

- Different versions of OpenAI models
- Other providers
- Local models (such as Hugging Face Models)
- Be aware of the expense and performance characteristics!!!

What are "Tokens"?

- AIs are generally good with math, but not so good with words
- For this reason, all text is "tokenized", which means it is turned into numbers
 - A simple (although not entirely accurate) way is to think of a "token" as a "syllable"
- Large Language Models all have a specific number of tokens they can deal with on a single request
 - This includes input and output tokens

Human Languages

- The truly large language models (like the OpenAI models) are trained on all available text on the Internet, and are therefore able to deal with many different languages
 - Common languages, like English or Mandarin, have more training data
- It is possible to converse in different languages
- It is also possible to mix and match languages
 - Example: Ask a question in English, but ask for a response in Spanish

Fun Idea: Obscured Answers

- I often have the problem that I cannot show real-world examples due to data privacy problems
- Simple solution: Ask the AI to replace some of the data for privacy reasons!

Complex Reasoning



We have two boxes. John puts a ball into the first box. He then leaves. Bob enters the room and moves the ball from the first to the second box. John returns that evening. Where is he looking for the ball?

Maintaining State

- Sending the conversation back and forth
- Asking the AI to restate the question so not all the state is needed

Summaries and Abstracts

- One can simply ask an AI to create a summary or abstract of text content, such as articles or transcriptions
- The key problem is the length of content, as content often exceeds maximum token counts and has to be split into multiple smaller summaries that are then summarized in combination

Extracting Information

- You can present data/text to an LLM and ask it to extract data in rather sophisticated ways
- This may even be data that does not as such occur in the source information, but can be logically inferred

Special Output Formats

- Define the structure
- True structure (like JSON)

Repeating the Question

- I find that results are improved if the question is stated both at the start and the end of the prompt

Injecting Additional Information

- Once data is retrieved it can be injected into the overall AI interaction
- This is referred to as “Retrieval Augmented Generation” (RAG)
- Note: For specific scenarios, more and more vendors provide out-of-the-box solutions
 - It is still often useful to do this by hand

Asking for Search Terms

- In many scenarios, the AI needs to first search existing data to “ground” the response in
- But how do you turn a longer request into useful search terms?
 - Simple answer: You ask the AI! 😊

Zero-Shot, One-Shot, Few-Shot

- **Zero-Shot:** A simple question
- **One-Shot:** A question and an example Q/A pair for guidance
- **Few-Shot:** A question and a few example Q/A pairs for guidance
- In scenarios where a specific type of output is expected, the Few-Shot approach can achieve considerably better results

Asking for References

- In RAG scenarios, it is useful to include references to the source information
- We find that this greatly improves trust in addition to just being generally useful
- The tricky part is that AI usually doesn't return references in the response, and when it does, it is not consistent enough to be useful
- Solution: Ask it specifically to include it and provide the format you expect! (Often JSON)

Asking for Suggestions

- Like references, you can also ask the AI to provide further suggestions
- We use this a lot to suggest related questions

Streaming vs Non-Streaming

- Streaming responses look good, but it isn't always appropriate
- We find that we often need to wait for the response
 - ...sometimes, we need the response to perform another step
 - ...often, responses are not intended for the user
 - ...structured responses may not be valid until received in full

Chain of Thought Reasoning

- When an AI is asked to provide a short answer to a complex question, it often does not perform as desired
- However, when the AI is asked to explain the process, it often performs much better
- This is known as “chain of thought reasoning”

Asking the AI if it “did it right”

- The most advanced language models are pretty good at knowing whether they performed the task correctly...
- ...but only once the entire response is complete
- This is a side-effect of the “one next word at a time” approach
- Asking an AI to check is a way to cut down on errors when a correct result is crucial
 - But it also means that streaming is not a good option
- *Note that this does not seem to work well with older models, and even with GPT4, the quality goes up and down, so consider this a “forward looking technique” ;-)*

Source Code

- Some LLMs can provide surprisingly interesting information about analyzed source code
- Models can also generate source code in interesting ways
- This is utilized by GitHub Copilot
- We use this extensively to analyze legacy code

Side-Note: Functions

- It is possible to configure some OpenAI models with “function calls”, allowing for the invocation of additional functions as needed
 - This currently works with preview versions of GPT3.5 Turbo and GPT4 models
- Example: The user wants to find hotels at a certain location with a certain price range, which is best done by calling a dedicated function with a specific set of parameters
 - If the model is configured with appropriate functions, it will determine when such a function is to be called

Other Announcements

AI Consulting Services



CODE CONSULTING Home Magazine Consulting Staffing Training Framework VFP Conversion [Login](#)

Executive Briefing (AI)

Are you wondering how AI can help your business? Do you worry about privacy or regulatory issues stopping you from using AI to its fullest? We have the answers! Our Executive Briefings provide guidance and concrete advice that help decision makers move forward in this rapidly changing Age of Artificial Intelligence and Copilots!

We will send an expert to your office to meet with you. You will receive:

1. An overview presentation of the current state of Artificial Intelligence.
2. How to use AI in your business while ensuring privacy of your and your clients' information.
3. A sample application built on your own HR documents - allowing your employees to query those documents in English and cutting down the number of questions that you and your HR group have to answer.
4. A roadmap for future use of AI tailored to what you do.

A Briefing Combined with Solid Deliverables!

We will send an expert to your office to meet with you. You will receive:

1. An overview presentation of the current state of Artificial Intelligence.
2. How to use AI in your business while ensuring the privacy of your data.
3. A sample application built on your own HR documents - allowing your employees to query those documents in English and cutting down the number of questions that you and your HR group have to answer.
4. A roadmap for future use of AI catered to what you do.

Executive Briefings are set up specifically for you and usually take several hours and up to a day. It is up to you whether you want to include only C-level Executives or your entire company. The cost is **\$5,000** (which is a discounted special rate, compared to the normal price of \$7,500) plus expenses if we have to travel to your location, or stay over night. This includes the concrete deliverable of a first implementation of AI for your business (see #3 above, although we may also agree that something slightly different is more appropriate for your organization).

[Contact us today to set up your individual Executive Briefing!](#)

Setting Directions for the Future of Your Organization

We are experiencing a time of extreme change in technology. While one might argue that this has always been true, it is clear that current developments have a drastic impact on all organizations at a scale that is orders of magnitude beyond what we have experienced in years past.

This represents a time of extreme opportunity, as well as a time of considerable risks!

Artificial Intelligence (AI) continues to revolutionize industries across the globe, offering immense potential for organizations to enhance operational efficiency, drive innovation, and achieve competitive advantage. Conversely, ignoring these technological trends means that even very recent technology is quickly turning into "legacy" and your competitors now have the opportunity to leapfrog you. Especially the advent of "The Age of Copilots" (powered largely by the onset of Large Language Models - LLMs) promises a productivity improvement of upwards of 50% in some areas year after year, and that isn't the only trend to pay attention to.

What does all of this mean specifically for your organization?



- Check out our new **Executive Briefing** Offer!
 - We can help with your AI needs
 - What does AI mean for you?
 - "Skunk Works" Projects
-
- www.codemag.com/AI
 - www.codemag.com/ExecutiveBriefing

CODE Staffing



- Disrupting the world of staffing!
- Giving our customers the ability to have staff on par with Silicon Valley companies...
- ...and our employees a work environment in a bleeding-edge tech company with the industry leading benefits!
- www.codestaffing.com



We Need You!

- We are currently “hiring on all cylinders” across the entire CODE Group Organization!
- Check out www.codestaffing.com/careers



Upcoming CODE Training Classes



- Public AI Training Classes
 - Introduction to Development with Open AI, ChatGPT, Azure Open AI, and more...
 - Building Copilots for your own apps
- Classes are also available as individualized private classes and mentoring sessions



<https://www.codemag.com/training>

Upcoming Events



- State of .NET Tour – Live & In-Person – Fall 2023
- State of .NET Online - .NET 8 – November
- <https://www.codemag.com/StateOfDotNet>







- CODE Presents – Monthly – Mostly AI
- <https://www.codemag.com/CODEPresents>



Recordings & Slide Decks

- State of .NET Webinar Series
 - codemag.com/StateOfDotNet
- CODE Presents Webinar Series
 - codemag.com/CODEPresents

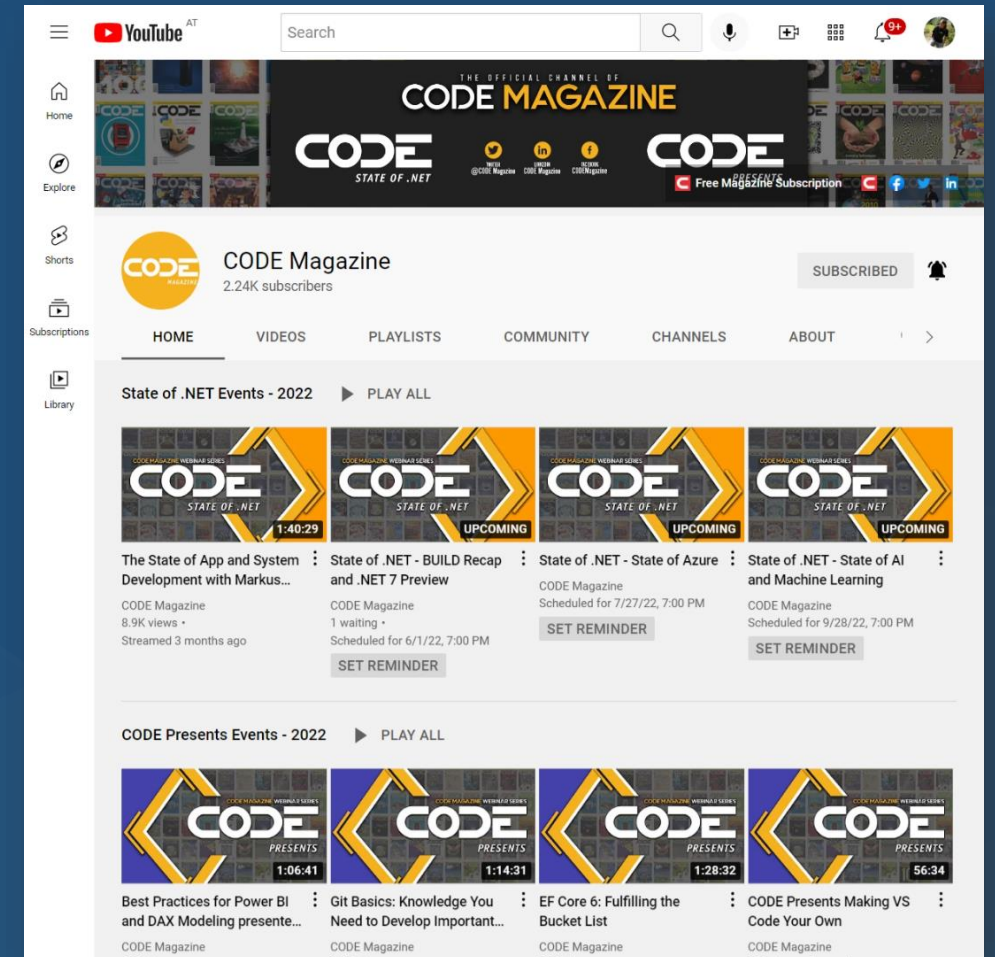
Recent Event Recordings

State of .NET - The State of .NET App and System Development 2/23/2022 	State of .NET - .NET Conf Recap 11/17/2021 
State of .NET - .NET 6 Preview 10/27/2021 	State of .NET - Azure Architecture 9/29/2021 

Subscribe to our YouTube Channel!



- Subscribe and “ring that bell” to never miss any of our content!
- www.youtube.com/Codemag



Event Survey – Win \$100!

- Complete this very short 12 question survey for a chance at a \$100 Amazon Gift Card!
- Survey must be completed by 11:59pm ET on Friday 9/8/2023 to be eligible!
- THIS SLIDE WILL BE REPEATED AT THE END AND SURVEY LINK REPEATED IN THE CHAT WINDOW!



CODE Presents: Prompt Engineering - How to Talk to an AI Survey

The survey will take approximately 4 minutes to complete.

Thank you for attending! Please complete this brief 12 question survey to be eligible to win a \$100 Amazon Gift Card. Your survey must be completed by 11:59pm ET (UTC-4) on 09/02/2023 to be eligible to win! One entry per person please. Drawing will occur and the individual winner notified by 09/08/2023.

Thank you for attending! Please complete this brief survey. Yes, we still want to hear from you if you were unable to attend but watched the recording instead.

* Required

1. Full Name *

Enter your answer

2. Company Name *

Enter your answer

<https://tinyurl.com/cp08302023survey>

Q&A

Contact us with questions!

CODE/EPS Contact:

www.codegroup.io

info@codemag.com

facebook.com/codemag

twitter.com/codemagazine

Presenter Contact:

markus@codemag.com

jduffy@codemag.com