FGCU

U.A. WHITAKER COLLEGE OF ENGINEERING

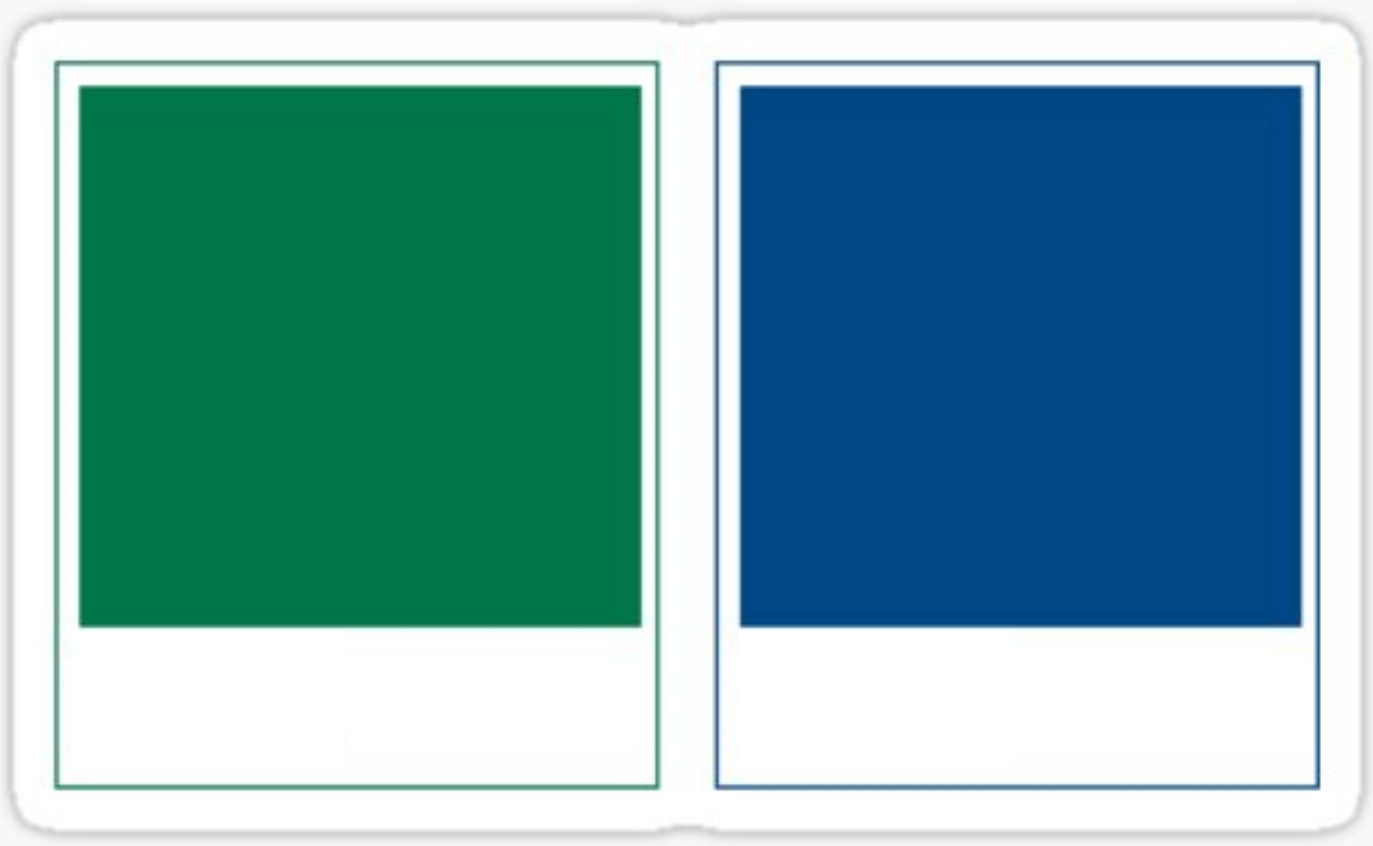
Dendritic: A Human-Centered Artificial Intelligence and Data Science Institute

Prompt Engineering

John Peller

May 16, 2025

What are FGCU's Colors?

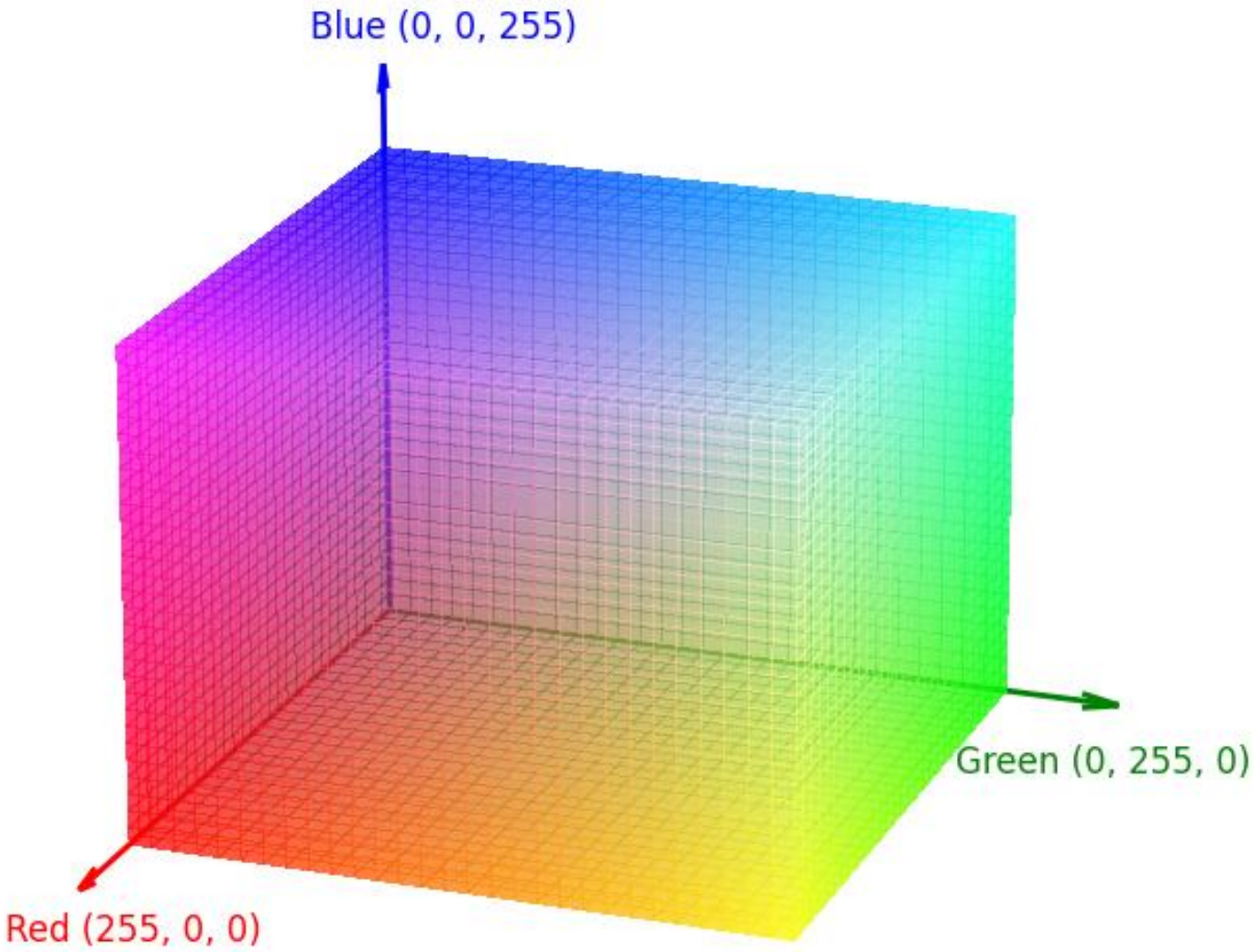


What are FGCU's Colors?

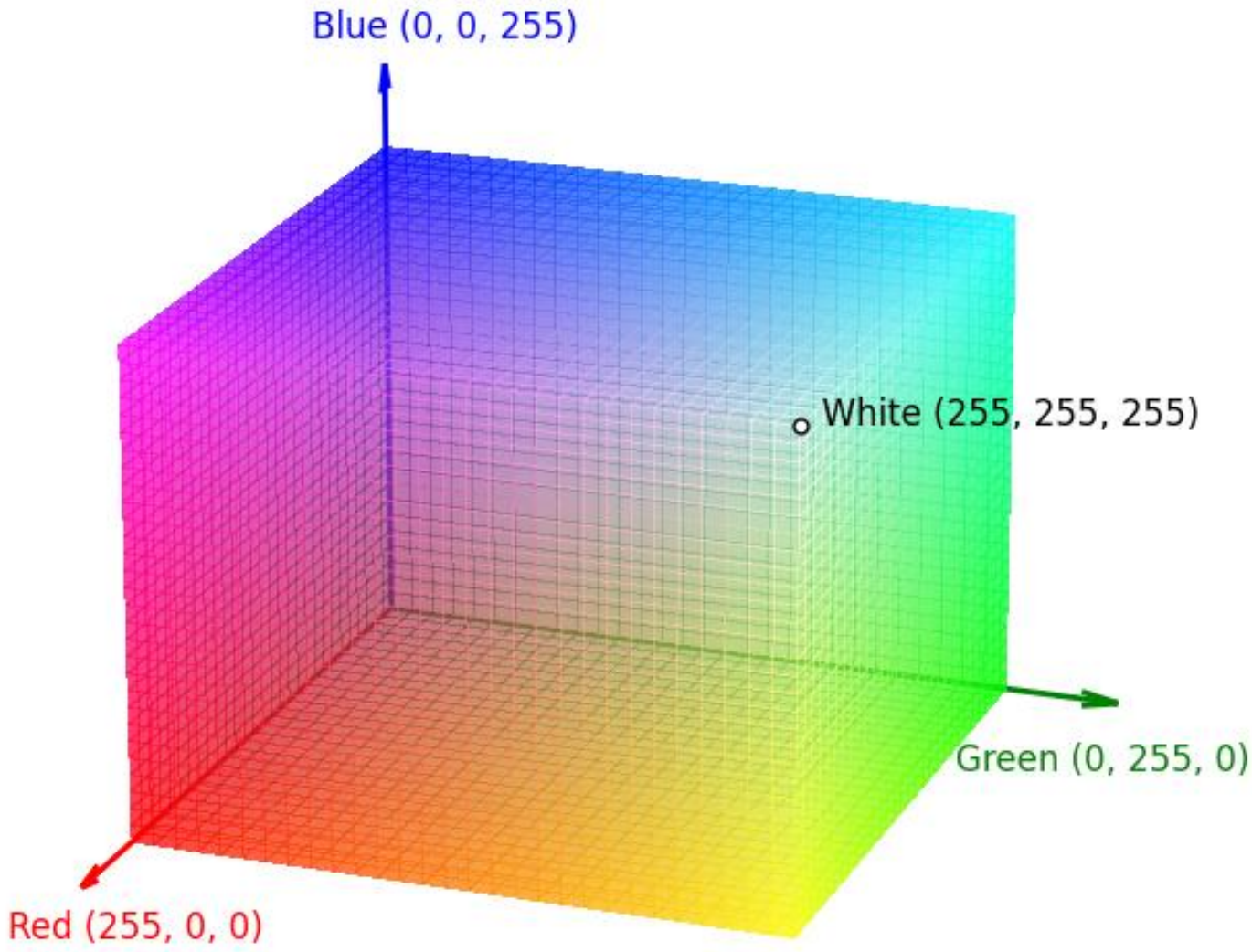
On Color

Additive color mixing

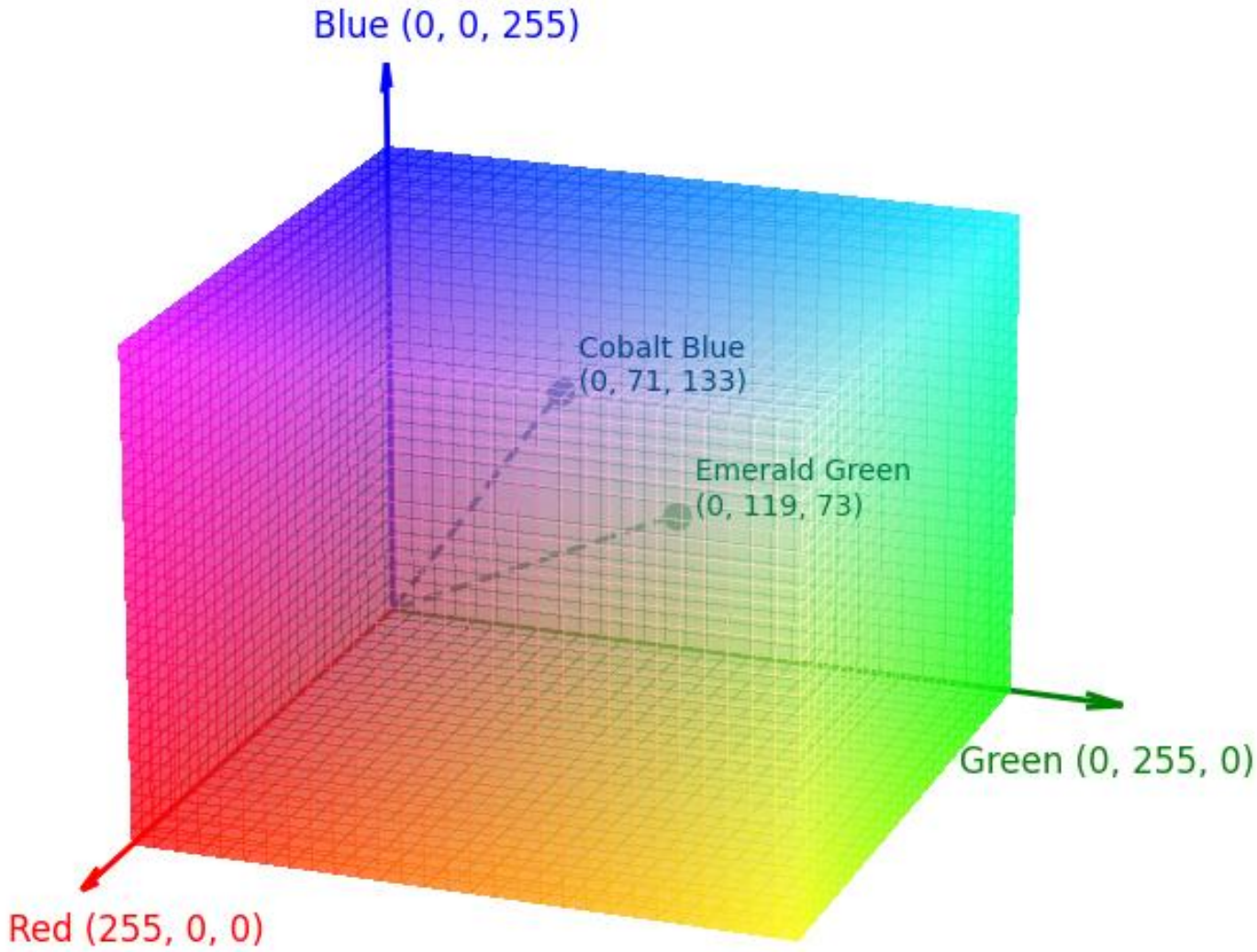
On Color



On Color



On Color

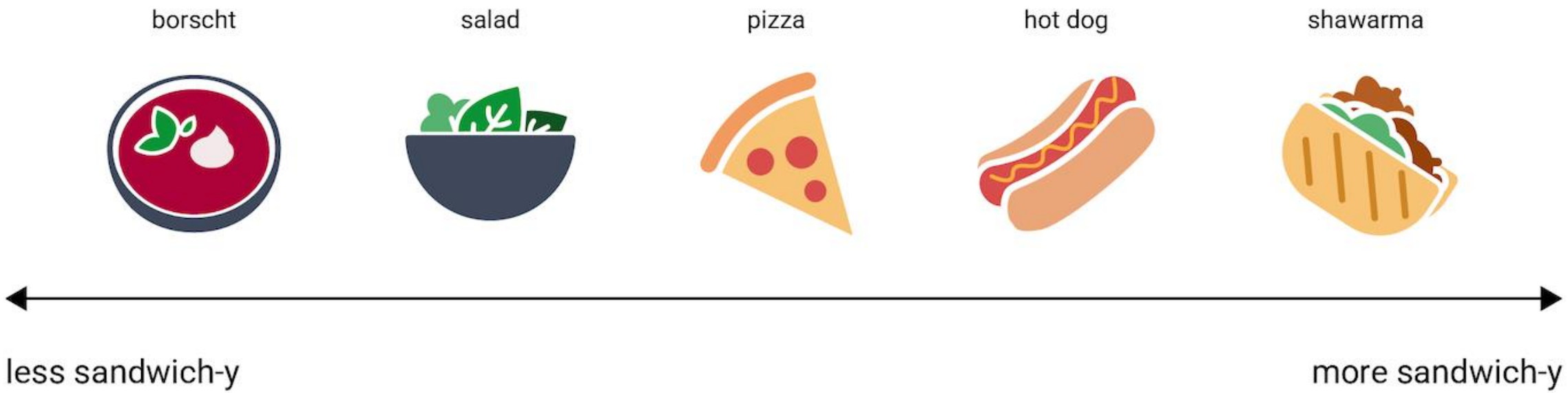


EAGLES
3415 C
EMERALD GREEN

EAGLES
288 C
COBALT BLUE

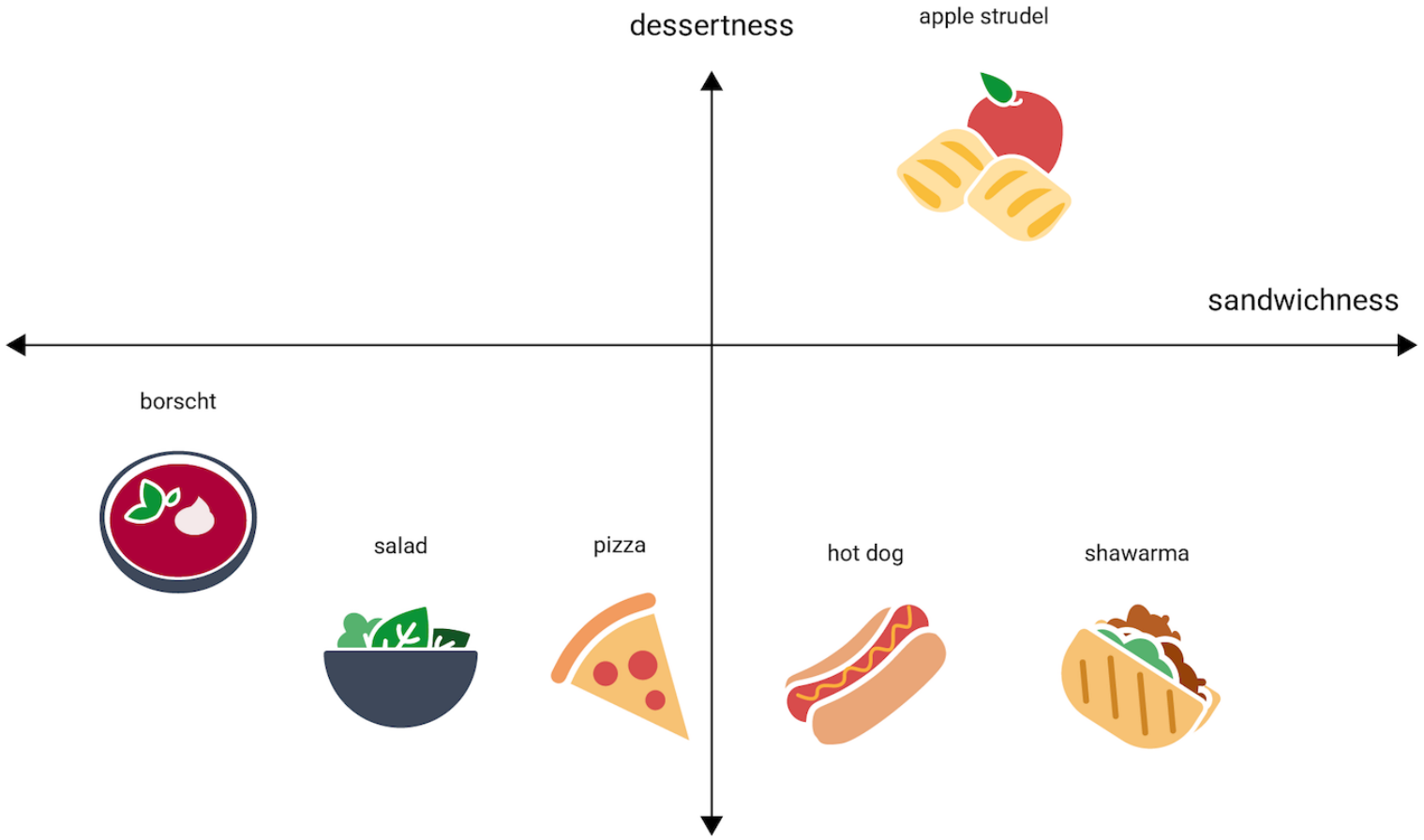
FGCU

On Vectors



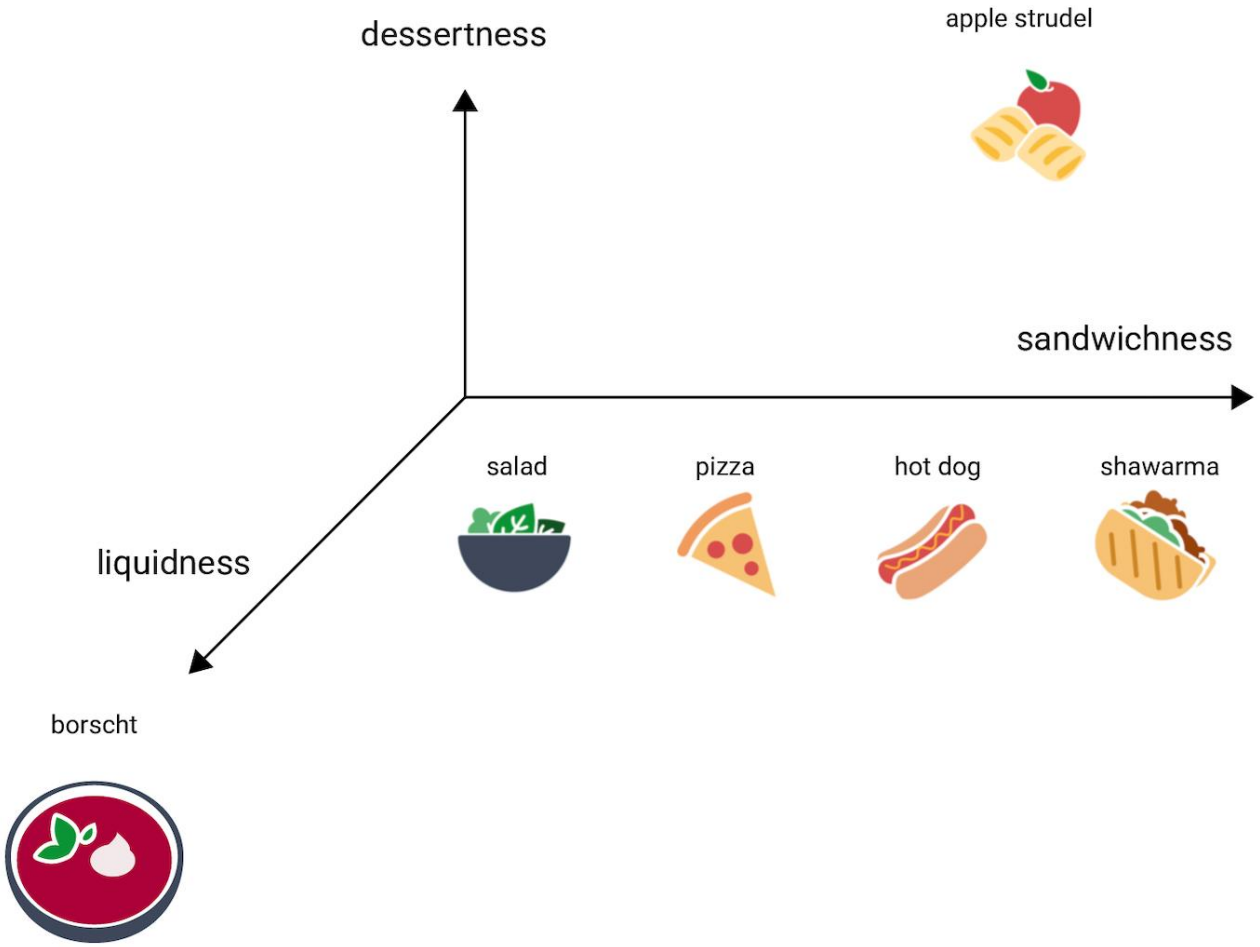
https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

On Vectors



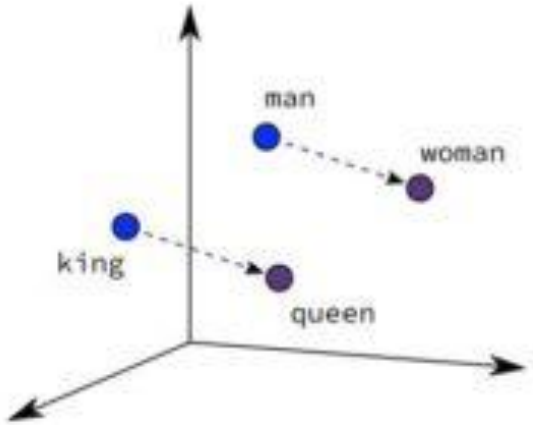
https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

On Vectors

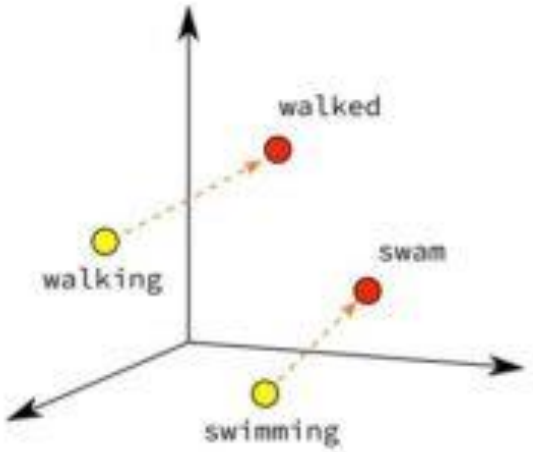


https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

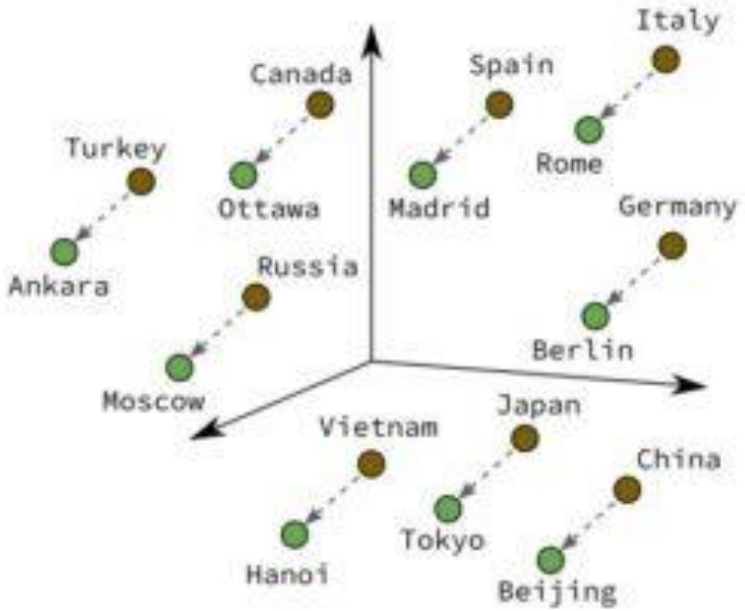
On Words



Male-Female



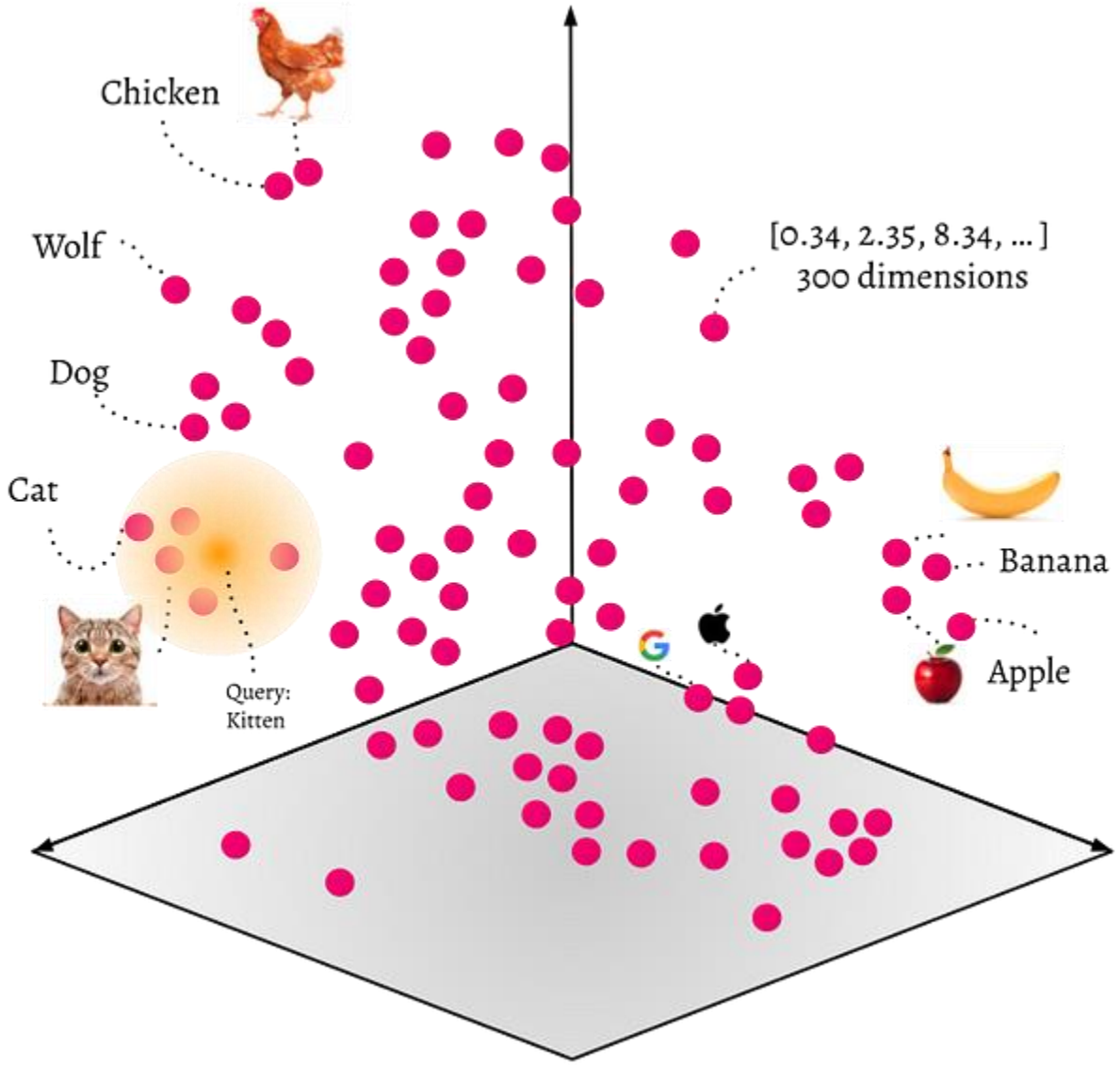
Verb Tense



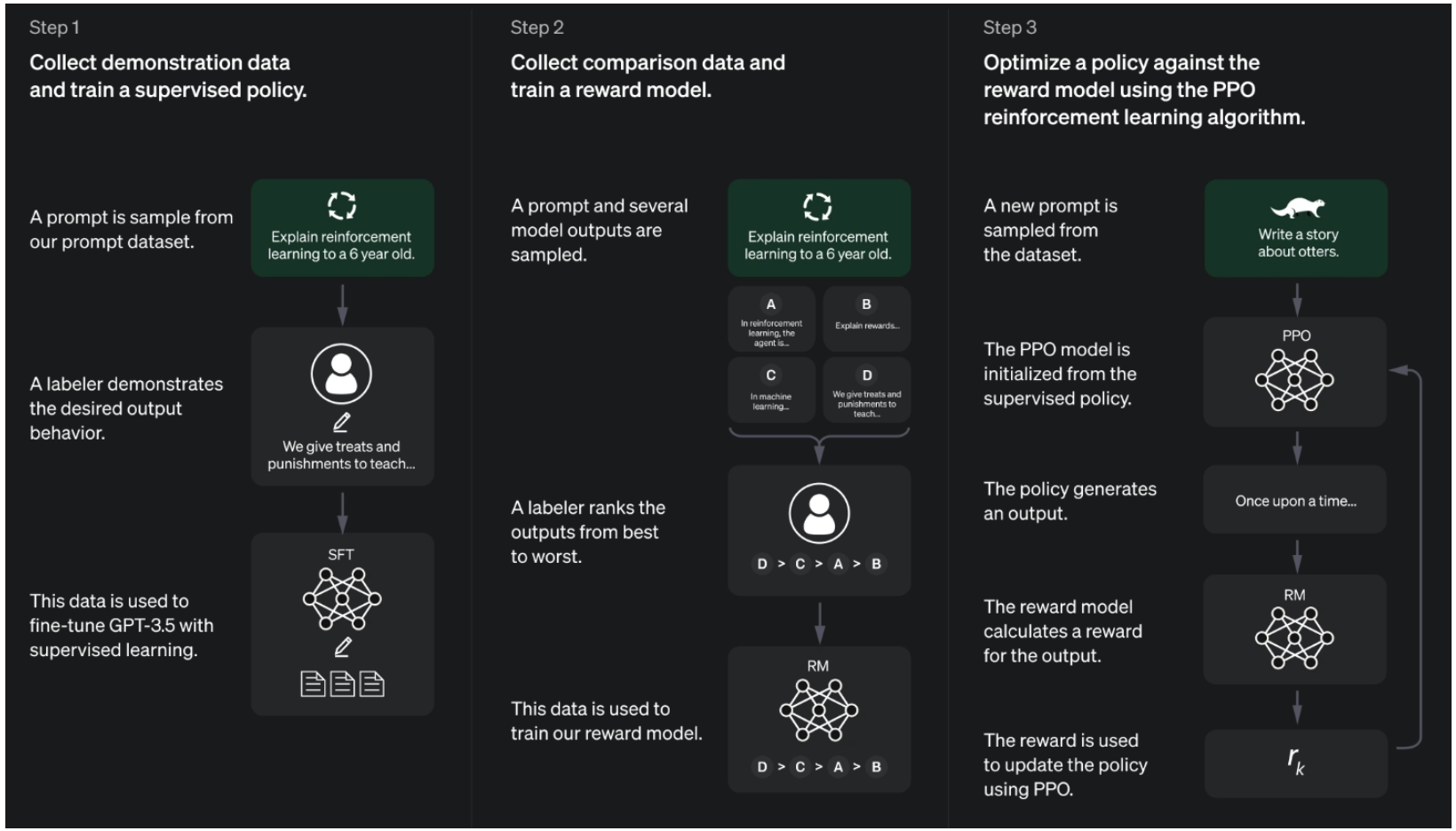
Country-Capital

https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

How an LLM ‘thinks’



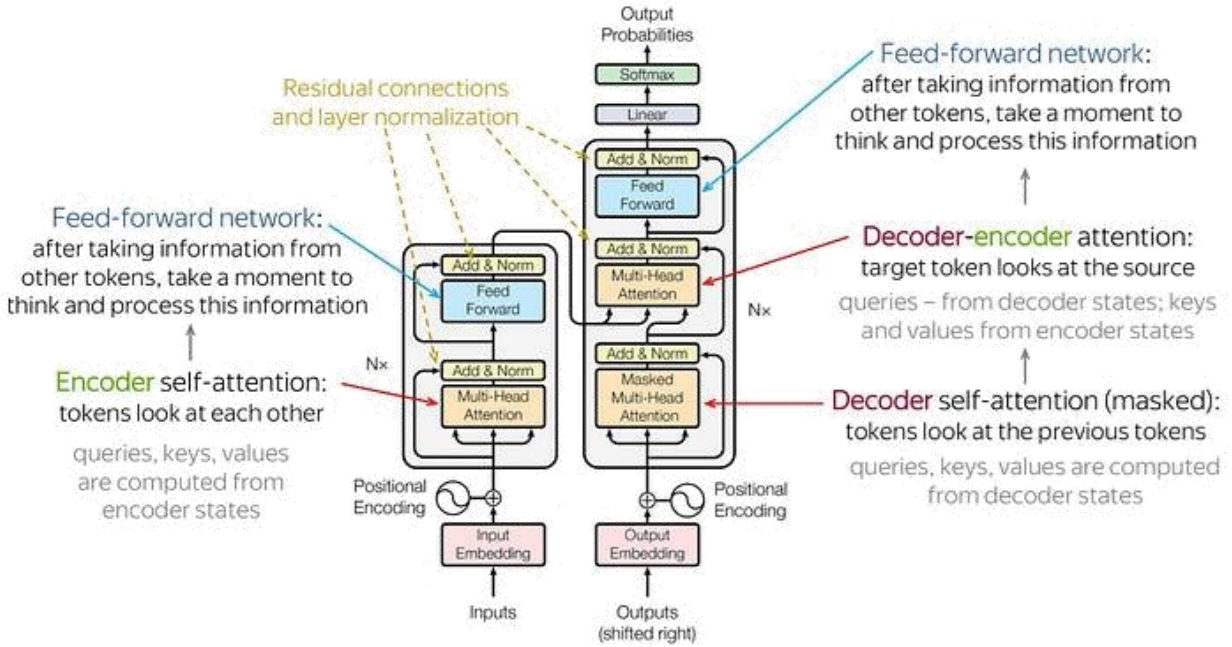
How an LLM ‘thinks’



https://openai.com/index/chatgpt/

Transformers

How Transformers Work: A Step-by-Step Breakdown



Input Sequence

A sequence of words or tokens that the model will process.

https://arxiv.org/abs/1706.03762

Tokenization

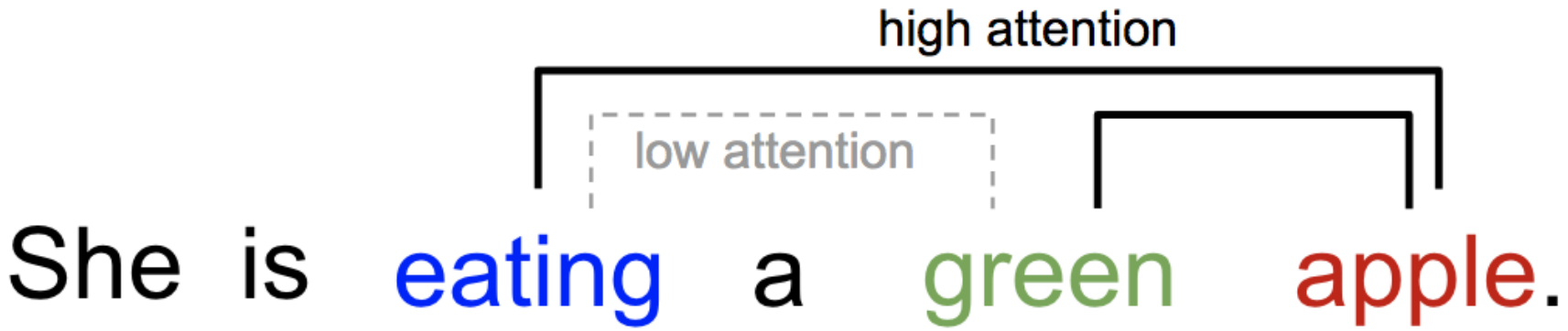
"It’s not the size of the dog in the fight—it's the size of the fight in the dog."

"It’s not the size of the dog in the fight—it's the size of the fight in the dog."

[68093, 802, 625, 290, 2349, 328, 290, 6446, 306, 290, 9848, 2322, 64190, 290, 2349, 328, 290, 9848, 306, 290, 6446, 3692]

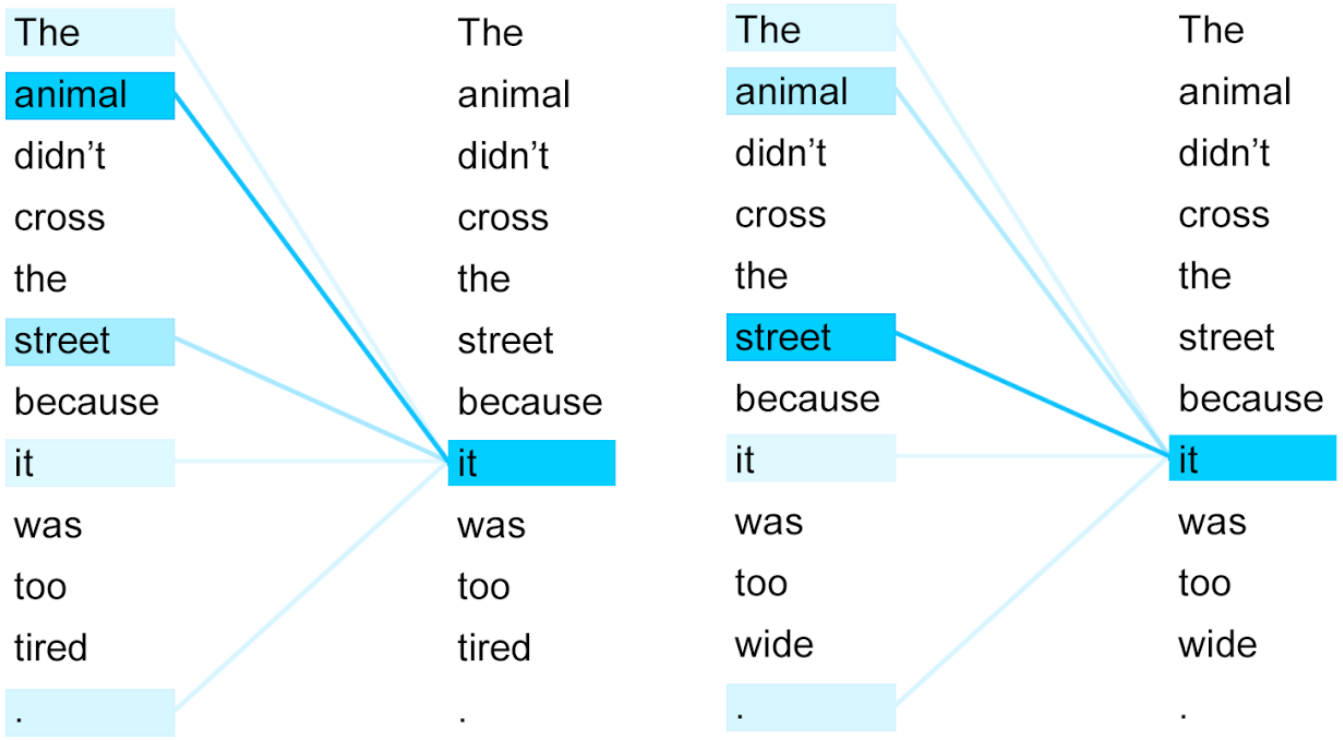
https://platform.openai.com/tokenizer

Self-Attention



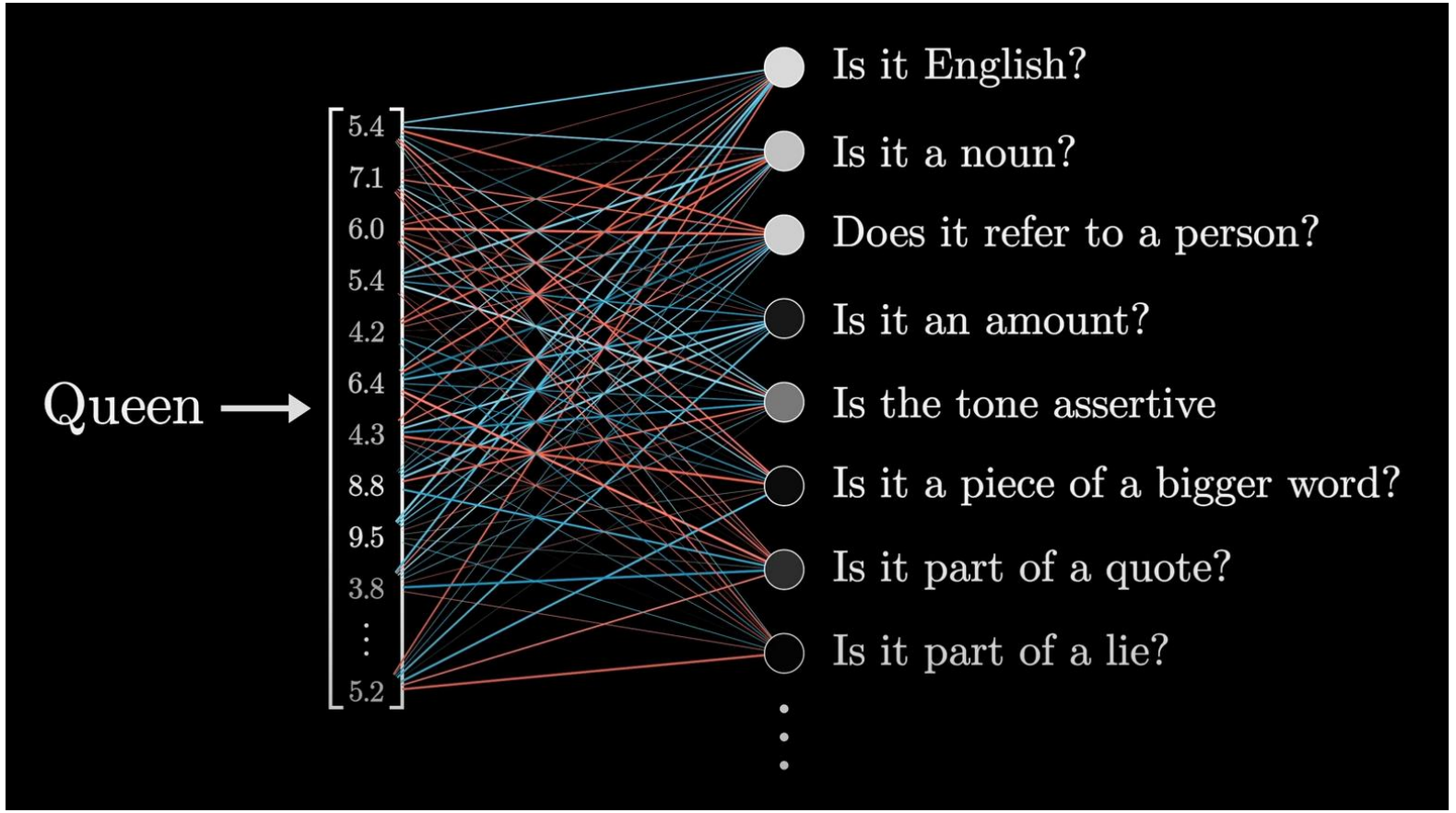
https://devopedia.org/transformer-neural-network-architecture

Self-Attention



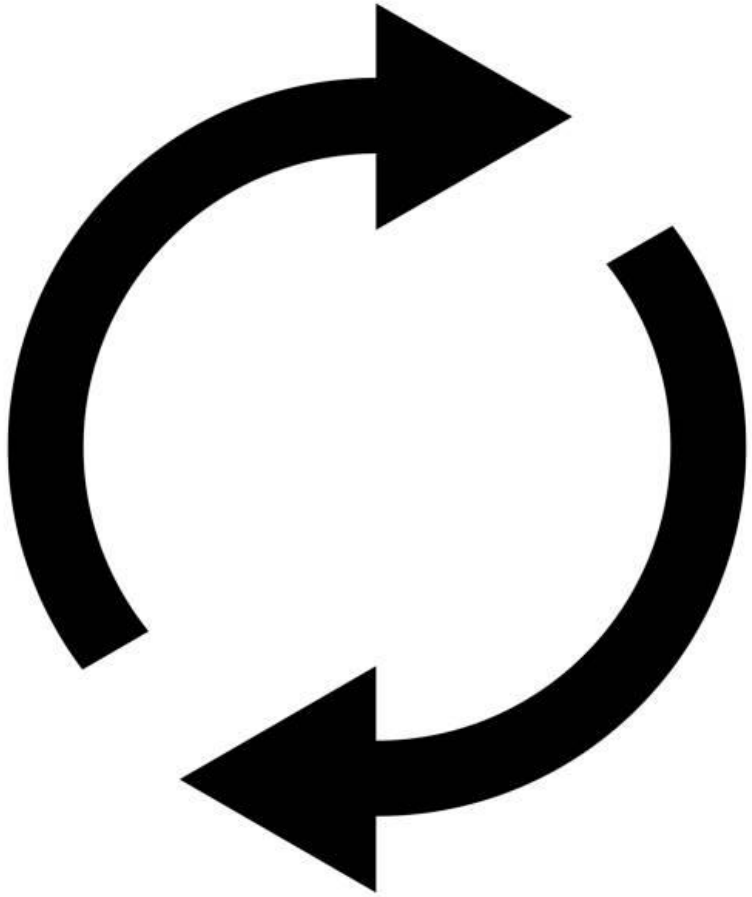
https://github.com/jessevig/bertviz

Feed-Forward



https://www.youtube.com/watch?v=wjZofJX0v4M

Layer Stacking



Unembedding

- (0, 119, 73) → **Emerald Green**
- (0, 71, 133) → **Cobalt Blue**
- (2, 670, 76, 2159, 551, 24590, … 1072) → ‘mat’

What It Means for Us

What It Means for Us

- Position Sensitivity

What It Means for Us

- Position Sensitivity
- Length, Structure, and Phrasing change Attention

What It Means for Us

- Position Sensitivity
  - Front-load your key instructions
- Length, Structure, and Phrasing change Attention
  - Use summaries and headings
  - Delimit and isolate
  - Be clear and consistent

Recap

Few-Shot

- Input-output example pairs
- https://doi.org/10.48550/arXiv.2005.14165

Chain-of-Thought

- Articulate intermediate reasoning
- https://doi.org/10.48550/arXiv.2201.11903

Generate Knowledge

- Background information first
- https://doi.org/10.48550/arXiv.2110.08387

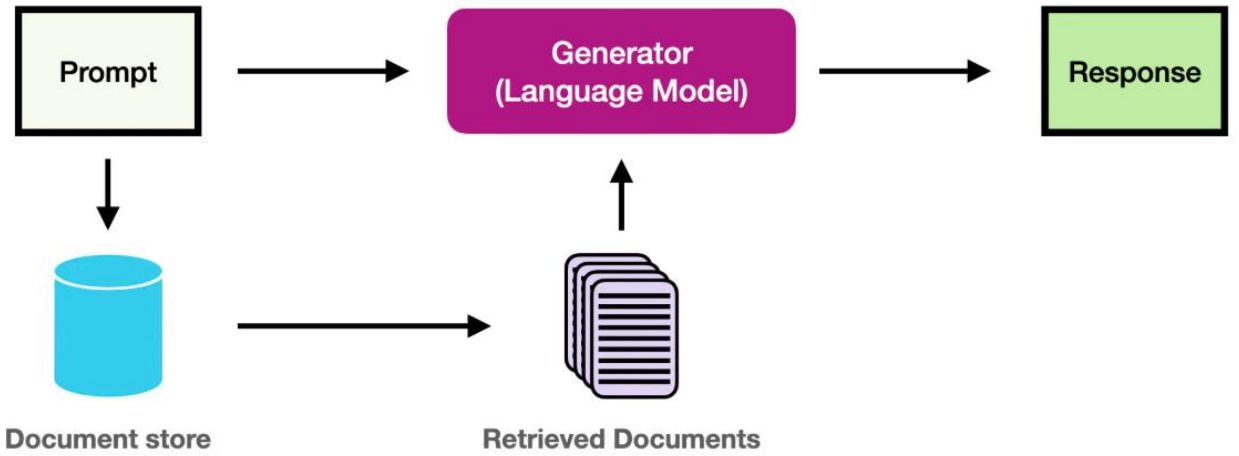
Role Prompting

- Adopt a persona or viewpoint
- https://doi.org/10.48550/arXiv.2308.07702

FGCU

Retrieval-Augmented Generation (RAG)

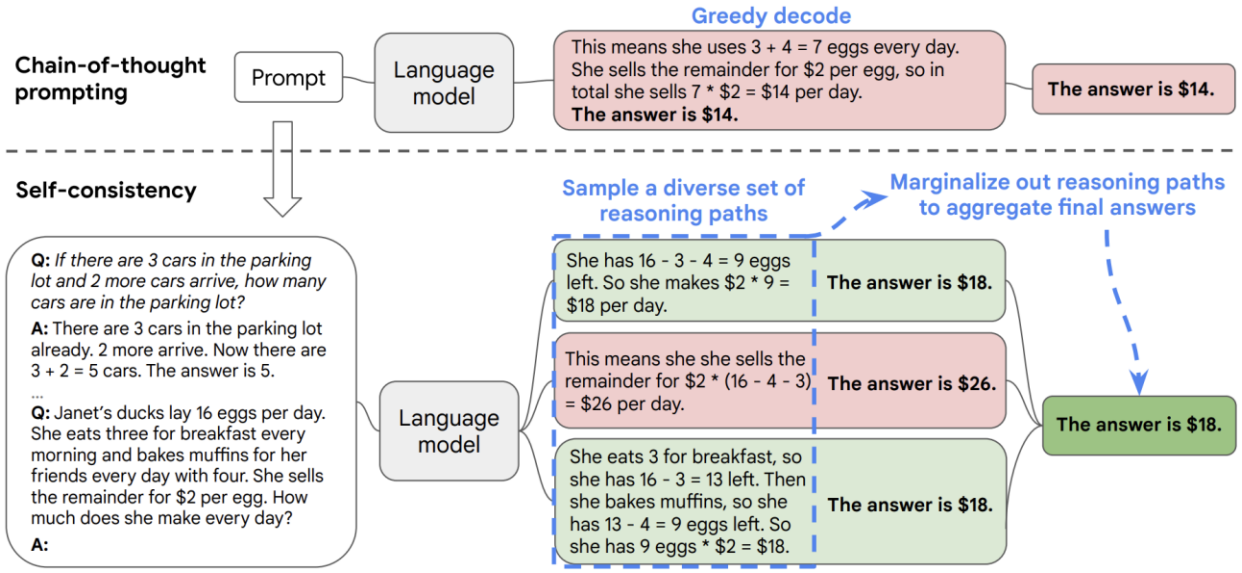
- Fetch relevant external documents into the prompt before generating response
- Grounds responses in domain-specific information, reducing hallucinations and increasing relevancy and accuracy



https://gettectonic.com/retrieval-augmented-generation-techniques/

Self-Consistency

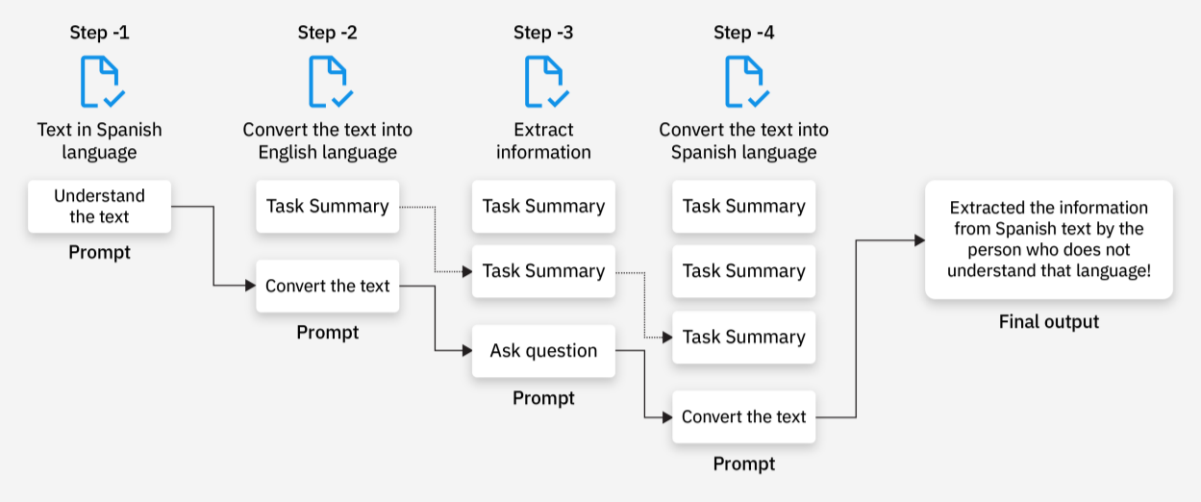
- Generate multiple independent chain-of-thoughts for the same query and choose or aggregate
- Reduces random errors in individual reasoning paths, improving reliability in ambiguous situations



https://www.prompthub.us/blog/self-consistency-and-universal-self-consistency-prompting

Prompt Chaining

- Break complex task into a series of focused consecutive prompts
- Modularizes and preserves context across stages, yielding more available attention



https://www.ibm.com/think/topics/prompt-chaining

Prompt Engineering Mindset

- Position, Length, Structure, and Phrasing
- Know your environment
  - Data Description & Preparation – set the role and scope
  - Descriptive Data Analysis – quantify and confirm
  - Data Visualization – balance, take recommendations, or choose
  - Data Storytelling (Narrative) – declare the audience and output
  - Dashboard Design – methodology for tangible items

FGCU

Works Cited

- Anthropic. “Chain Complex Prompts for Stronger Performance.” Build with Claude Prompt Engineering, https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/chain-prompts.
- Lewis, Patrick, et all. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” arXiv Preprint arXiv:2005.11401v4, 12 Apr. 2021, https://doi.org/10.48550/arXiv.2005.11401.
- Saravia, Elvis. “Prompt Engineering Guide.” GitHub, Dec. 2022, https://github.com/dair-ai/Prompt-Engineering-Guide.
- Sun, Shichao, et all. “Prompt Chaining or Stepwise Prompt? Refinement in Text Summarization.” arXiv Preprint arXiv:2406.00507v1, 1 Jun. 2024, https://doi.org/10.48550/arXiv.2406.00507.
- Vaswani, Ashish, et all. “Attention Is All You Need.” arXiv Preprint arXiv:1706.03762v7, 2 Aug. 2023, https://doi.org/10.48550/arXiv.1706.03762.
- Wang, Xuezhi, et all. “Self-Consistency Improves Chain of Thought Reasoning in Language Models.” arXiv Preprint arXiv:2203.11171v4, 7 Mar. 2023, https://doi.org/10.48550/arXiv.2203.11171.

FGCU

FGCU

U.A. WHITAKER COLLEGE OF ENGINEERING

Dendritic: A Human-Centered Artificial Intelligence and Data Science Institute

Thank You!