CEN 5016:
Software
Engineering

Fall 2024

University of
Central Florida

Atish Kumar Dipongkor

Week 7- Class 1:
More LLMs & Open Source Software

Administrivia

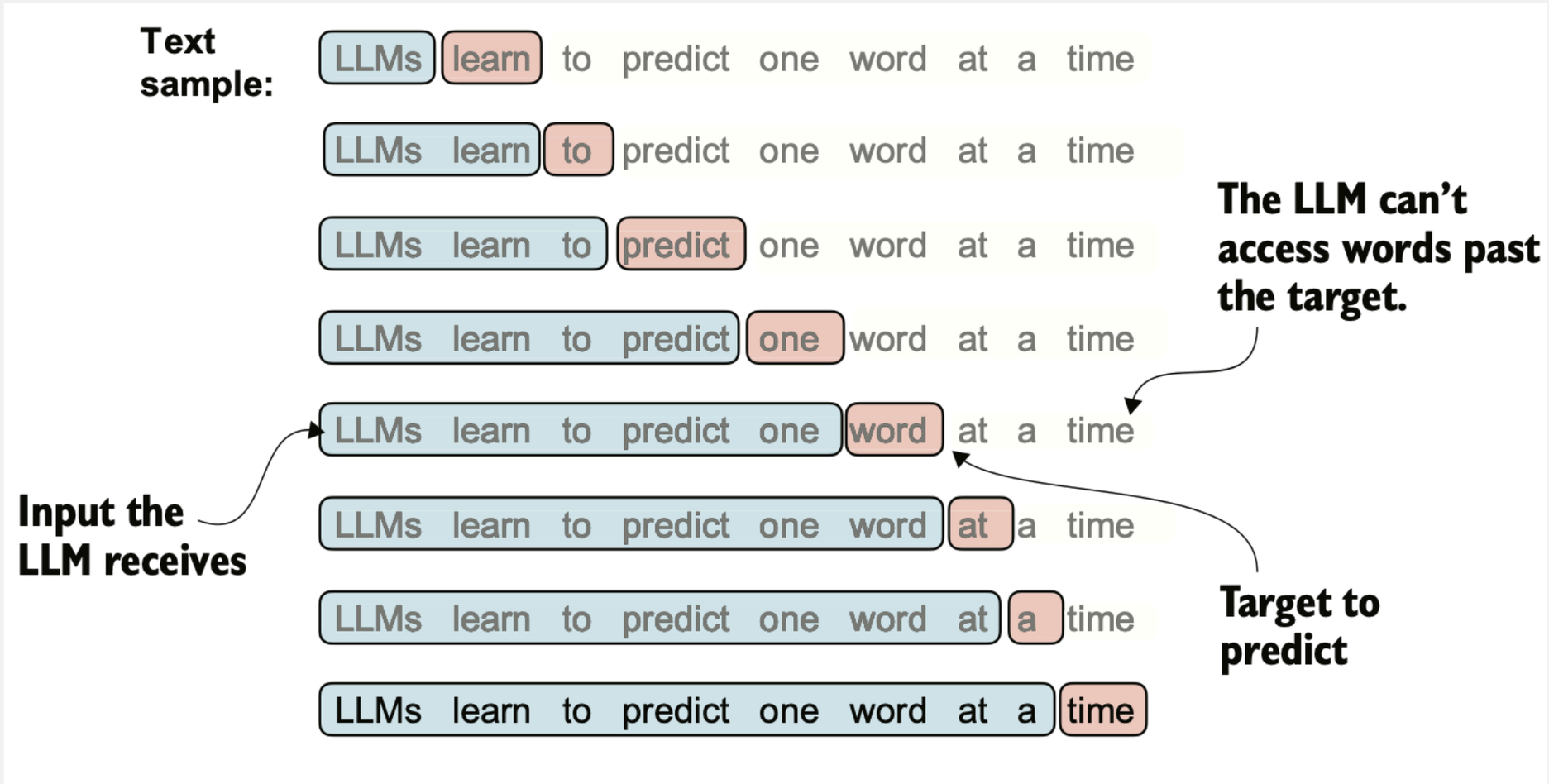
- *Assignment 4*
  - Due Monday, October 7th
  - Exploring Static Analysis Tools and CI with a simple Python app
  - Accept the Assignment on GitHub Classroom
- *SDE Project Part 2*
  - Due Tuesday, October 15th (updated deadline!)
  - Dr. Moran is working on Feedback
  - Two parts:
    - Process & Implementation Snapshot
    - Checkpoint Presentation

A Software Engineer's Guide to LLMs (continued)

Today's Presentation Overview

- *Input: Prompt Engineering*
- *How do LLMs generate Output?*
- *LLM Evaluation*
- *Accessing LLM*
- *Productizing LLM*

How LLMs generate Long Responses?



Building a Large Language Model - Sebastian Raschka

Input to LLM: Prompt Engineering

- Rewording text prompts to achieve desired output. Low-hanging fruit to improve LLM performance!
- Popular prompt styles:
  - Zero-shot: instruction + no examples
  - Few-shot: instruction + examples of desired input-output pairs
  - Chain of Thoughts: Breaking down into multiple prompts

Few-shot Prompting

Task: Classify the following bug report as either a "Bug" or a "Feature Request."

Few-shot Prompt:

Example 1:

- *Bug Report:* "The application crashes every time I try to open the settings menu."
- *Classification:* **Bug**

Example 2:

- *Bug Report:* "It would be great if the app supported dark mode for better visibility at night."
- *Classification:* **Feature Request**

Example 3:

- *Bug Report:* "The app doesn't save my preferences after closing it, and I have to set them again."
- *Classification:* **Bug**

Example 4:

- *Bug Report:* "Can you add a function to export data in CSV format?"
- *Classification:* **Feature Request**

New Input:

- *Bug Report:* "The app freezes when I try to upload a file larger than 10MB."
- *Classification:* **[MODEL OUTPUT: Bug]**

Chain of Thought Prompting

- Few-shot prompting strategy
  - Example responses include reasoning
  - Useful for solving more complex word problems [arXiv]
  - Example:
    Q: A person is traveling at 20 km/hr and reached his destiny in 2.5 hr then find the distance? Answer Choices: (a) 53 km (b) 55 km (c) 52 km (d) 60 km (e) 50 km
    A: The distance that the person traveled would have been 20km/hr * 2.5 hrs = 50km
    The answer is (e).

Chain of Thought Prompting

Problem: A school is organizing a field trip and they need to arrange transportation for 120 students. They can rent vans that can hold 12 students each or buses that can hold 40 students each. If the school rents 2 buses, how many vans are needed to transport the remaining students?

Prompt 1: Restate the problem and break it down.

Response 1: There are 120 students in total, and we are renting 2 buses. Each bus can hold 40 students. We need to find out how many vans, which can hold 12 students each, are required for the remaining students

Chain of Thought Prompting

<u>Prompt 2:</u> Compute the number of students accommodated by the buses.

<u>Response 2:</u> Each bus can hold 40 students, and we are renting 2 buses. So, 2 buses can transport 2 × 40 = 80 students.

<u>Prompt 3:</u> Subtract the number of students already accommodated from the total number of students.

<u>Response 3:</u> We have 120 students, and 80 of them can be transported by the buses. Therefore, there are 120 - 80 = 40 students remaining.

Prompt 4: Calculate how many vans are required for the remaining students.

Response 4: Each van can hold 12 students, so we need to divide the remaining 40 students by 12 to see how many vans are needed. 40 ÷ 12 = 3.33. Since we can't have a fraction of a van, we round up to 4 vans

Fine-Tuning

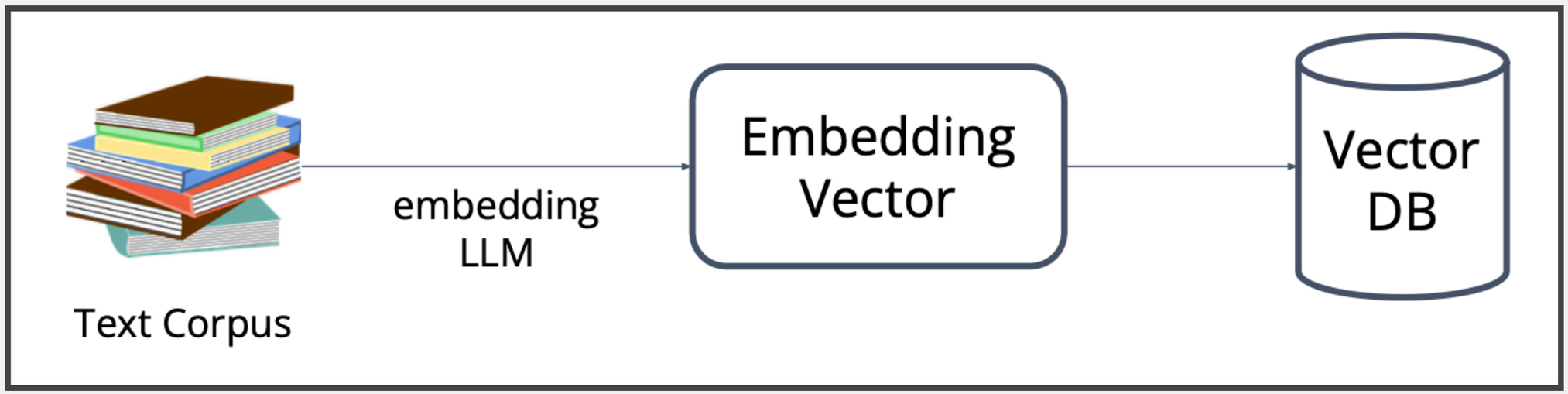
- Retrain part of the LLM with your own data
- Create dataset specific to your task
- Provide input-output examples (>= 100)
- Example:
  - I have a online store that sells books
  - A Human agent replies all customers' query
  - How can I replace the agent with an LLM?

Information Retrieval and RAG

- *RAG: Retrieval-Augmented Generation*
- Used when you want LLMs to interact with a large knowledge base (e.g. codebase, company documents)
  1. Store chunks of knowledge base in Vector DB
  2. Retrieve most “relevant” chunks upon query, add to prompt
- Pros: Only include most relevant context → performance, #tokens
- Cons: Integration, Vector DB costs, diminishing returns

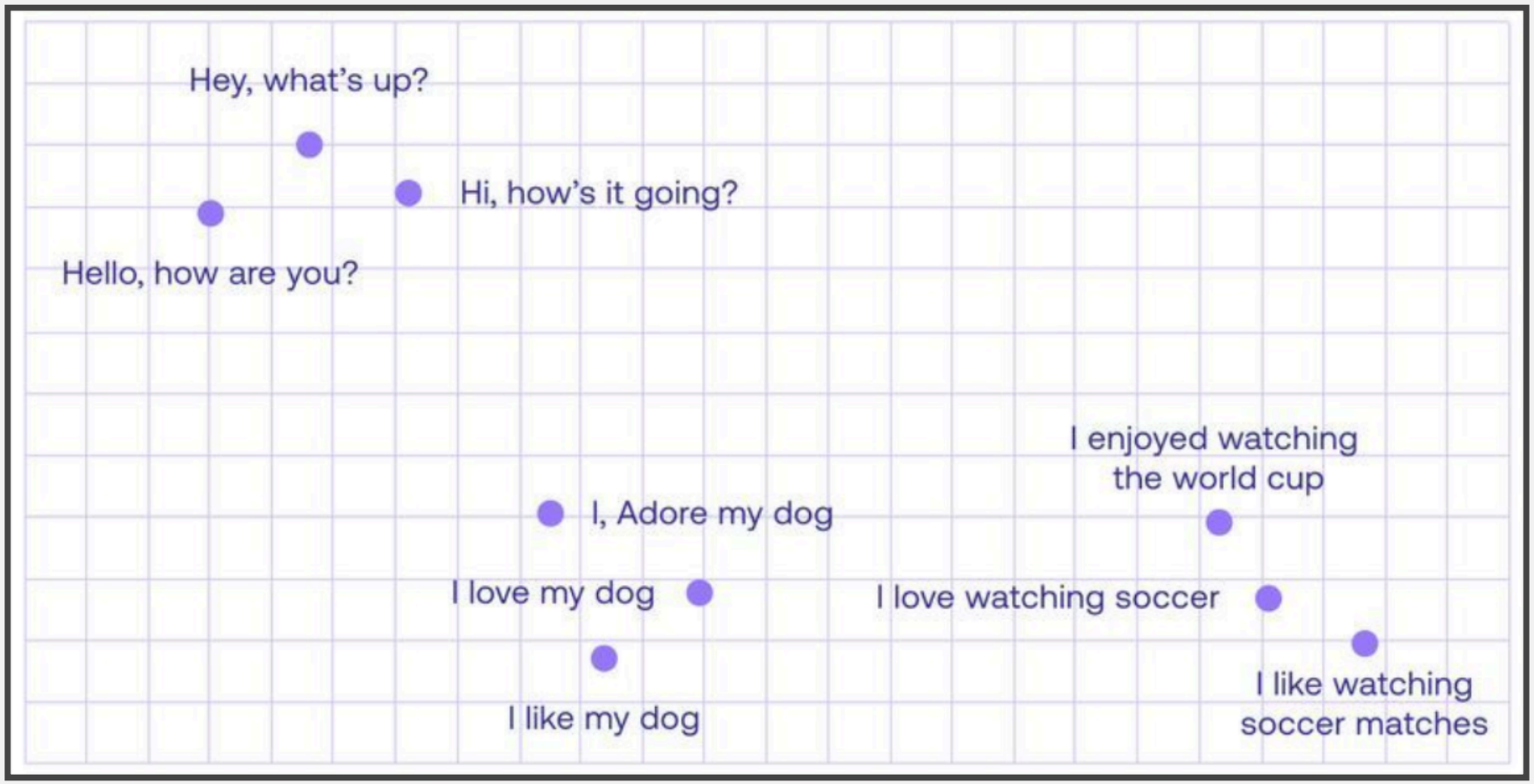
Information Retrieval and RAG

- *1. Store semantic embeddings of documents*



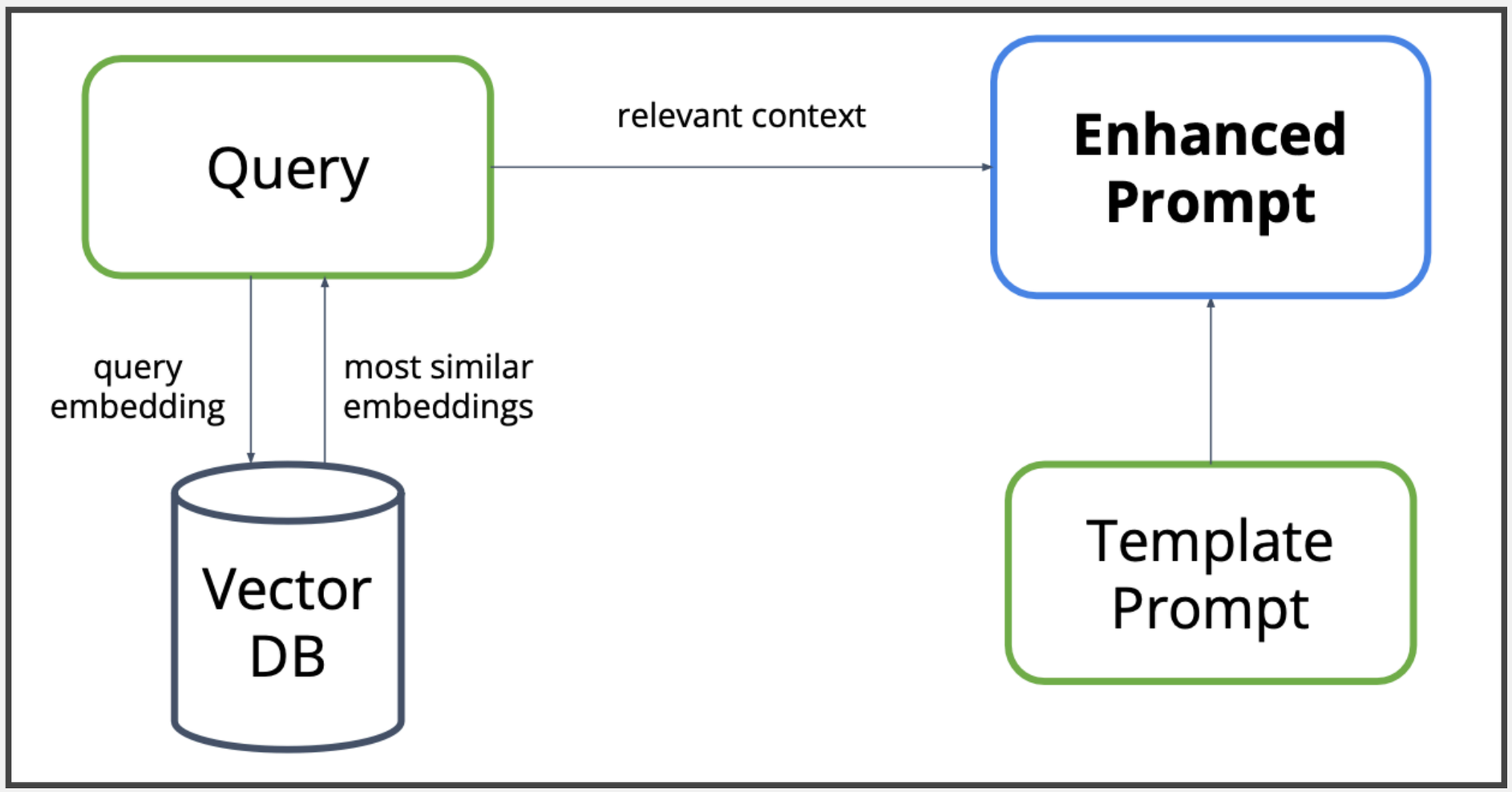
Embeddings Vectors

- Embeddings are a representation of text aiming to capture semantic meaning.



Information Retrieval and RAG

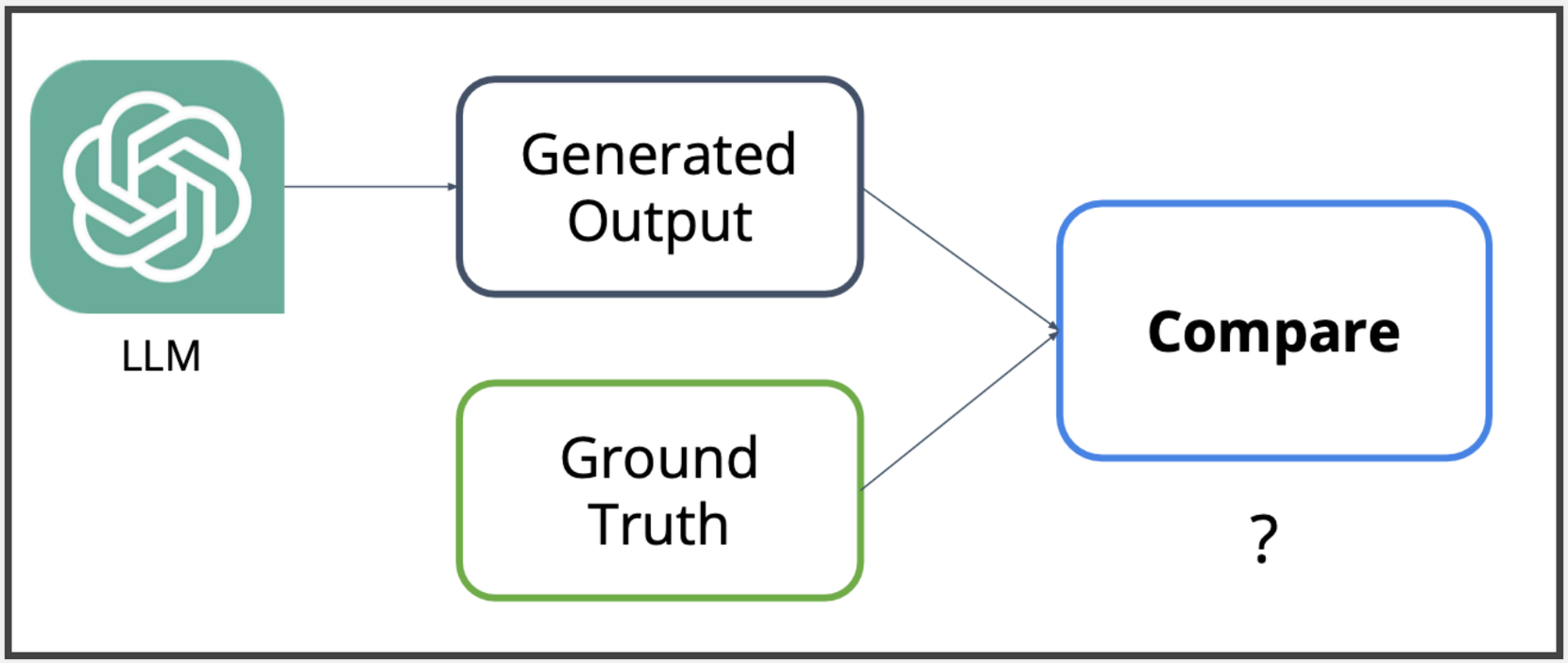
- *2. Retrieve most relevant embeddings, combine with prompt*



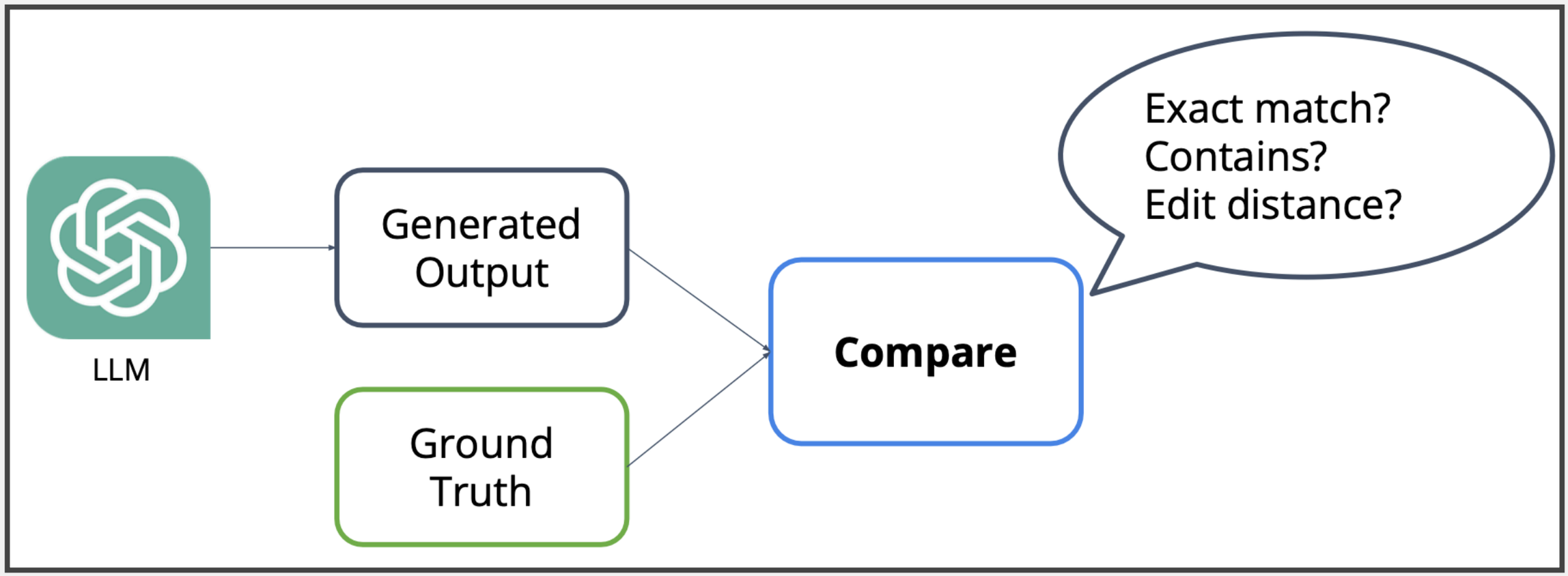
LLM Evaluation

Evaluation: Is the LLM good at our Task?

- First, do we have a labeled dataset?

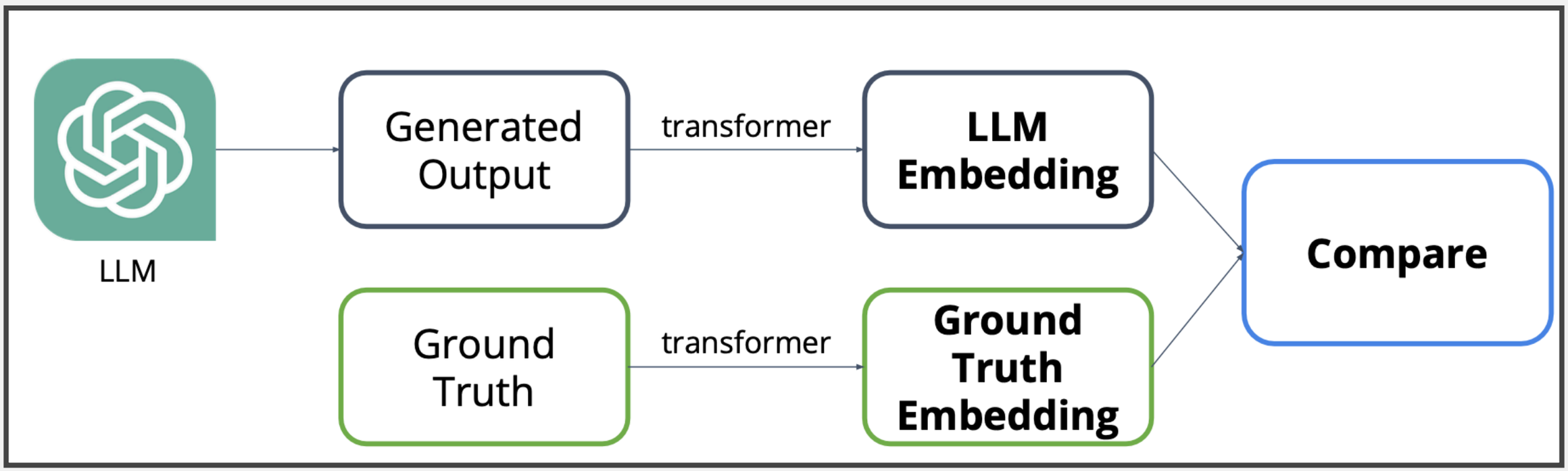


Textual Comparison: Syntactic Checks

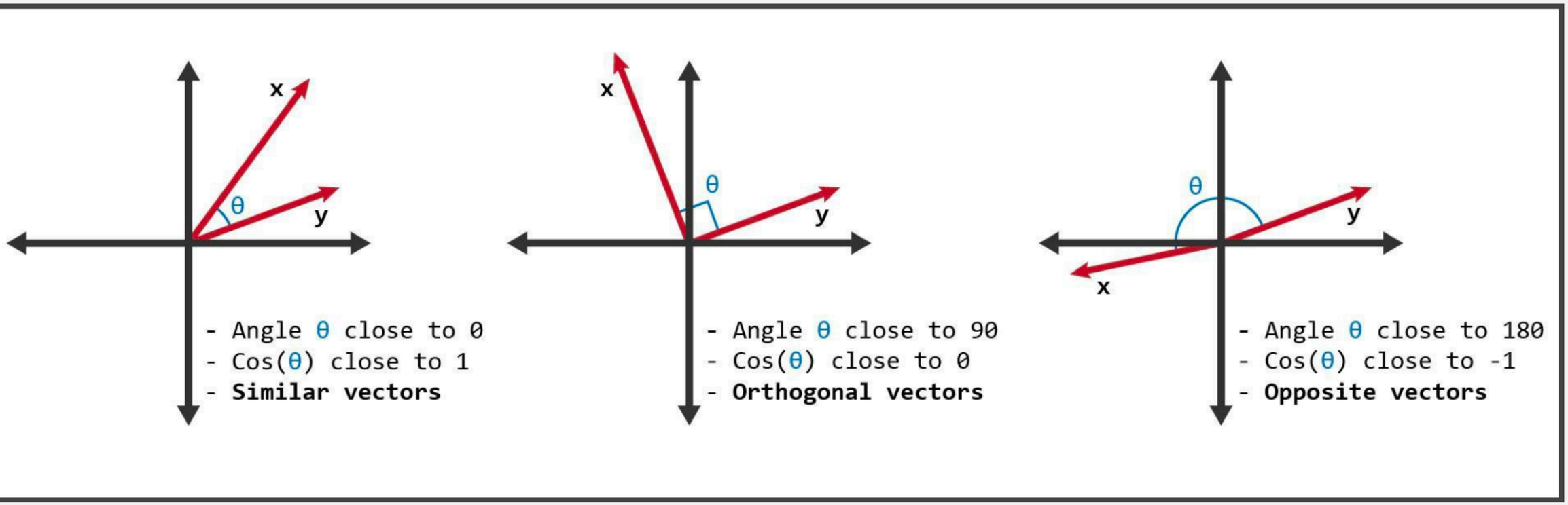


Textual Comparison: Embeddings

- Embeddings are a representation of text aiming to capture semantic meaning.



Textual Comparison: Cosine Similarity



Evaluation

- Suppose we don't have an evaluation dataset.
- What do we care about in our output?
- Example: creative writing
  - Lexical Diversity (unique word counts)
  - Semantic diversity (pairwise similarity)
  - Bias

Evaluation: Test Generation

1. Correctness of Test Cases

- **Quality**: The unit tests should correctly cover the intended functionality of the code, ensuring that the assertions and logic align with the expected behavior.
- **Heuristic**:
  - **Manual Review**: Perform a code review to check if the assertions correctly match the expected behavior for each function.
  - **Automated Pass/Fail**: Run the generated unit tests on a set of known-good code implementations to see if they pass or fail correctly based on expected outcomes.

2. Code Coverage

- **Quality**: The generated tests should cover a significant portion of the codebase, including edge cases and all branches of logic.
- **Heuristic**:
  - **Code Coverage Tool**: Use a tool like **pytest-cov** or **JaCoCo** to measure how much of the code is covered by the generated unit tests. Target a high percentage of branch and statement coverage.

Accessing LLM

- Web Interface
- API

Function Calling

- LLM returns sequence of calls to your function
  - Supported on GPT-3.5, GPT-4
- 1. List all APIs/functions the LLM has access to.
- Additional prompt to figure out which APIs to use

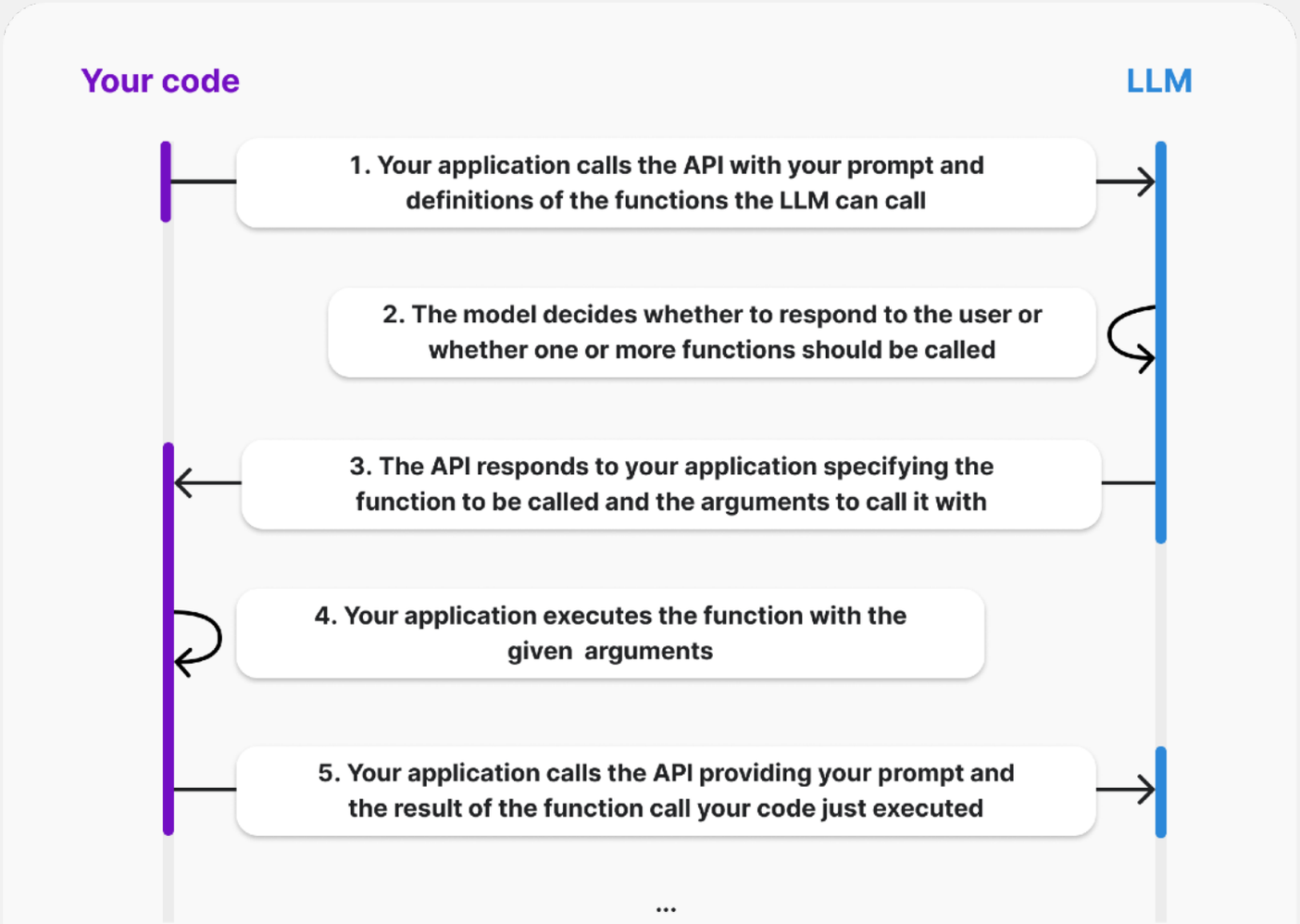
Function Calling

- 1. Specify Available Functions
- Example from OpenAI

```
  "model": "gpt-3.5-turbo-0613",
  "messages": [
    {"role": "user", "content": "What is the weather like in Boston?"}
  ],
  "functions": [
    {
      "name": "get_current_weather",
      "description": "Get the current weather in a given location",
      "parameters": {
        "type": "object",
        "properties": {
          "location": {
            "type": "string",
            "description": "The city and state, e.g. San Francisco, CA"
          },
          "unit": {
            "type": "string",
            "enum": ["celsius", "fahrenheit"]
          }
        },
        "required": ["location"]
      }
    }
  ]
}'
```

Function Calling

- 1. Specify Available Functions
- Example from OpenAI



Function Calling

- 1. Model Response Contains Function Calls
- Example from OpenAI

```
{
  "id": "chatcmpl-123",
  ...
  "choices": [{
    "index": 0,
    "message": {
      "role": "assistant",
      "content": null,
      "function_call": {
        "name": "get_current_weather",
        "arguments": "{ \"location\": \"Boston, MA\"}"
      }
    },
    "finish_reason": "function_call"
  }]
}
```

Function Calling

```
curl https://api.openai.com/v1/chat/completions -u :$OPENAI_API_KEY -H 'Content-Type: application/json' -d '{
  "model": "gpt-3.5-turbo-0613",
  "messages": [
    {"role": "user", "content": "What is the weather like in Boston?"},
    {"role": "assistant", "content": null, "function_call": {"name": "get_current_weather", "arguments": "{ \"location\": \"Boston, MA\"}"}},
    {"role": "function", "name": "get_current_weather", "content": "{\"temperature\": "22", \"unit\": \"celsius\", \"description\": \"Sunny\"}"}
  ],
  "functions": [
    {
      "name": "get_current_weather",
      "description": "Get the current weather in a given location",
      "parameters": {
        "type": "object",
        "properties": {
          "location": {
            "type": "string",
            "description": "The city and state, e.g. San Francisco, CA"
          },
          "unit": {
            "type": "string",
            "enum": ["celsius", "fahrenheit"]
          }
        },
        "required": ["location"]
      }
    }
  ]
}'
```

Productizing an LLM

Estimating Operational Costs

- Most LLMs will charge based on prompt length.
- Use these prices together with assumptions about usage of your application to estimate operating costs.
- Some companies (like OpenAI) quote prices in terms of tokens - chunks of words that the model operates on.
- GCP Vertex AI Pricing
- OpenAI API Pricing
- Anthropic AI Pricing

Optimizing Latency + Speed

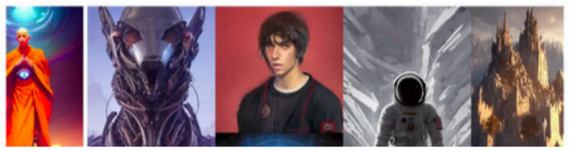
- Making inferences using LLMs can be slow...
- Strategies to improve performance:
- **Caching** - store LLM input/output pairs for future use
- **Streaming responses** - supported by most LLM API providers. Better UX by streaming response line by line.

Open Intellectual Property Concerns

- Was the data used to train these LLMs obtained illegally?
- Who owns the IP associated with LLM outputs?
- Should sensitive information be provided as inputs to LLMs?

ARTIFICIAL INTELLIGENCE / TECH / CREATORS

AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit



/ The suit claims generative AI art tools violate copyright law by scraping artists' work from the web without their consent.

ARTIFICIAL INTELLIGENCE / TECH / LAW

The lawsuit that could rewrite the rules of AI copyright

/ Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI. But the suit could have a huge impact on the wider world of artificial intelligence.

Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT

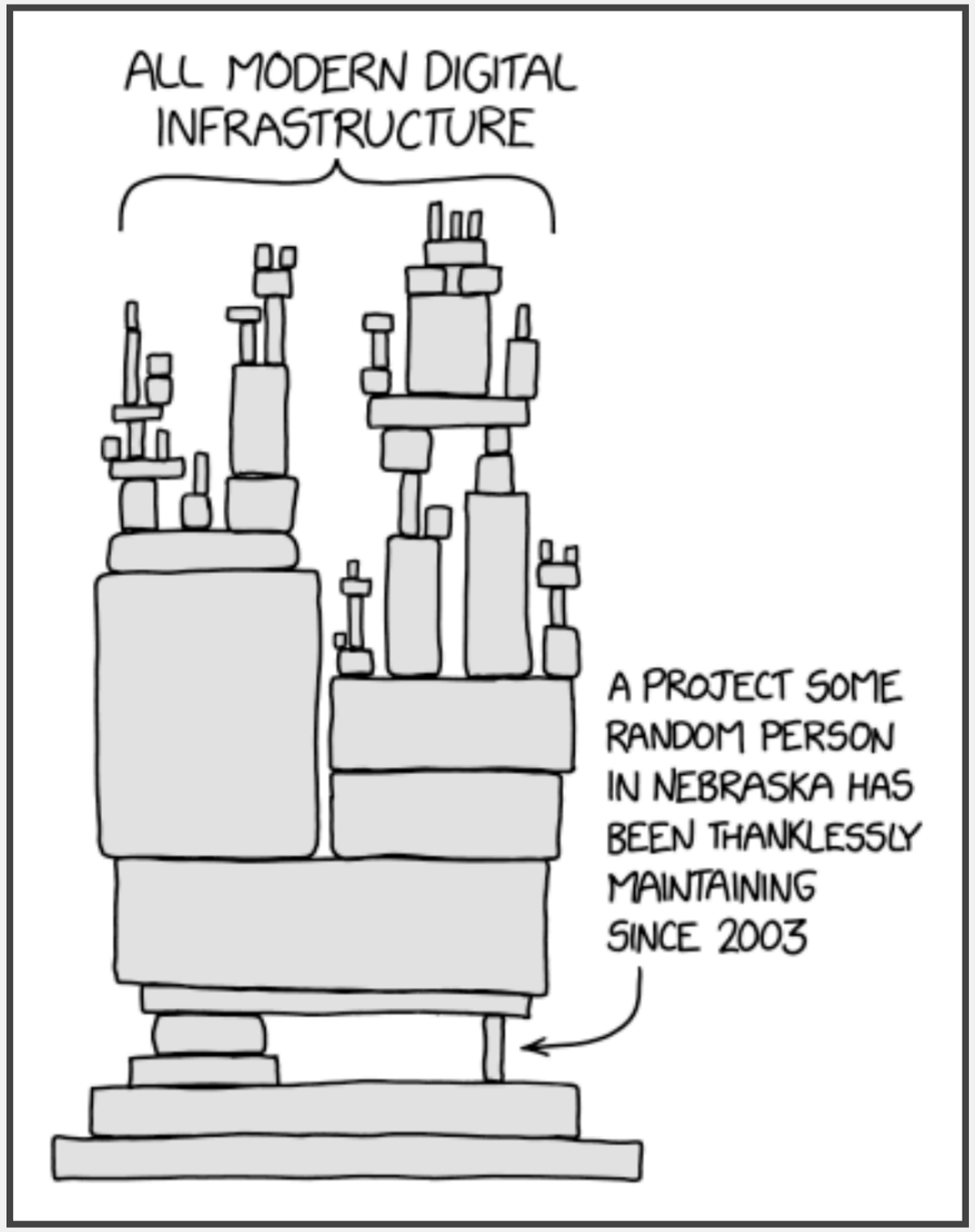
ChatGPT doesn't keep secrets.

Open-Source Software

Learning Goals

- Distinguish between open-source software, free software, and commercial software.
- Identify the common types of software licenses and their implications.
- Distinguish between copyright and intellectual property.
- Express an educated opinion on the philosophical/political debate between open source and proprietary principles.
- Describe how open-source ecosystems work and evolve, in terms of maintainers, community contribution, and commercial backing
- Identify various concerns of commercial entities in leveraging open-source, as well as strategies to mitigate these.

The Importance of Open-Source



What is Open Source Software?

Open-source

Proprietary

Uber

NETFLIX

What is Open Source Software?

- Source code availability
- Right to modify and creative derivative works
- (Often) Right to redistribute derivate works

Common Misconceptions about Open Source

- **Quality**
  - **Myth:** Lower quality than proprietary
  - **Reality:** Often meets or exceed industry standards
- **Support and Maintenance**
  - **Myth:** Lack of Professional support
  - **Reality:** Robust support community
- **Security**
  - **Myth:** Less secure because code is public
  - **Reality:** Transparency allows quicker identification and fixing

Benefits of Open Source Software

- **For Individuals**
  - Learning opportunities
  - Customization
  - Cost Saving
- **For Business**
  - Flexibility
  - Security
  - Community Support

Open Source Licenses

- **Copyleft:** Requires derivatives to maintain the same license (e.g., GPL)
- **Permissive:** Allows proprietary use of modified code (e.g., MIT, Apache)

GNU General Public License (GPL)

Type: Copyleft

Key Terms:

- Any derivative work must be distributed with the same GPL license.
- The source code must be made available to users, enabling them to modify and redistribute it.
- Commercial use is allowed, but any distributed version of the software (including commercial ones) must adhere to the GPL terms.

MIT License

Type: Permissive

Key Terms:

- The software can be used for personal, commercial, or open-source purposes.
- There’s no requirement to release derivative works as open source.
- The original copyright notice and license must be included in all copies or substantial portions of the software.
- No warranties or liability are provided by the authors of the software.

Apache License 2.0

Type: Permissive with additional patent rights

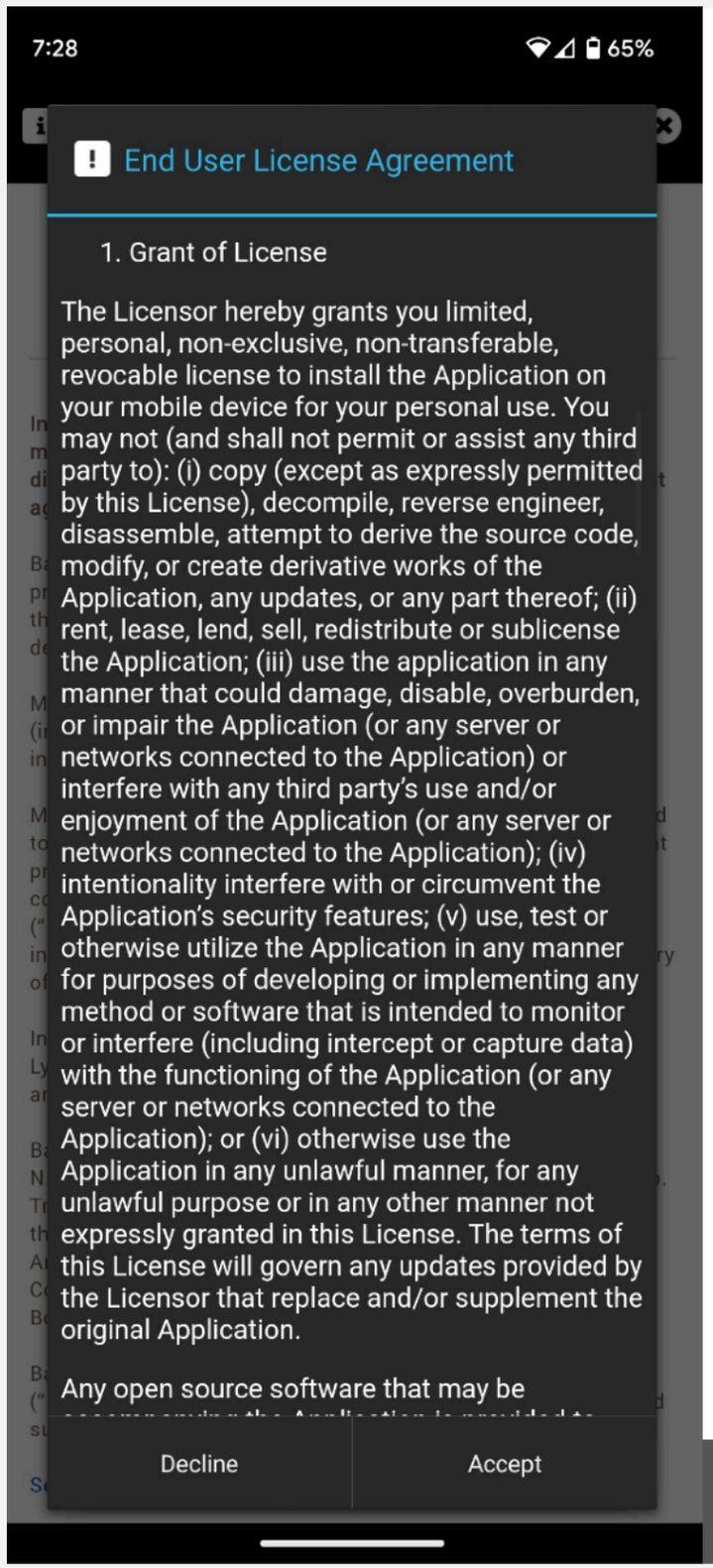
Key Terms:

- Users can use the software for both open-source and proprietary purposes.
- The license includes an express grant of patent rights, ensuring that contributors cannot sue users for patent infringement related to their contributions.
- Modifications to the original software must be clearly marked.
- The original copyright notice and license must be included in any derivative works.

Contrast with Proprietary Software: A Black Box

- Intention is to be used, not examined, inspected, or modified.
- No source code – only download a binary (e.g., an app) or use via the internet (e.g., a web service).
- Often contains an End User License Agreement (EULA) governing rights and liabilities.
- EULAs may specifically prohibit attempts to understand application internals.

Contrast with Proprietary Software: A Black Box



Free Software vs. Open Source

- *Free software origins (70-80s ~Stallman)*
  - ~~Cultish~~ Political goal
  - Software part of free speech
    - free exchange, free modification
    - proprietary software is unethical
    - security, trust
  - GNU project, Linux, GPL license

- *Open source (1998 ~O'Reilly)*
  - Rebranding without political legacy
  - Emphasis on internet and large dev/user involvement
  - Openness toward proprietary software/coexist
  - (Think: Netscape becoming Mozilla)

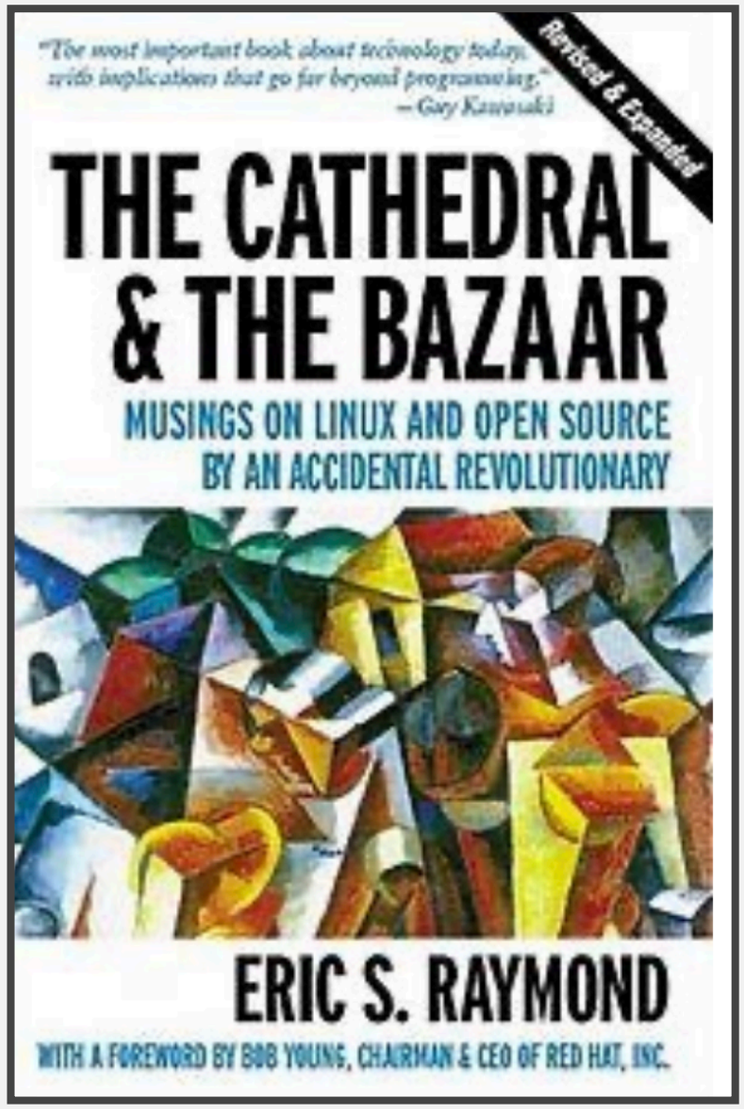
Free Software vs. Open Source

Perception (from some):

- Anarchy
- Demagoguery
- Ideology
- Altruism

Open-Source Ecosystems

The Cathedral and the Bazaar



The Bazaar Won

Cathedral

- Developed centrally by a core group of members
- Available for all once complete (or at releases)
- Examples: GMU Emacs, GCC (back in the 1990s)
- “Sort of” examples today: Chrome Intellij

Bazaar

- Developed openly and organically
- Wide participation (in theory, anyone can contribute)
Examples: Linux

OSS has many stakeholders / contributors

- Core members
  - Often (but not always) includes the original creators
  - Direct push access to main repository
  - May be further split into admin roles and developers
- External contributors
  - File bug reports and report other issues
  - Contribute code and documentation via pull requests
- Other supporters
  - Beta testers (users)
  - Sponsors (financial or platform)
  - Steering committees or public commenters (for standards and RFCs)
- Spin-offs
  - Maintainers of forks of the original repository

Contributing Processes

- Mature OSS projects often have strict contribution guidelines
  - Look for CONTRIBUTING.md or similar
- Common requirements:
  - Coding style (recall: linters) and passing static checks
  - Inclusion of test cases with new code
  - Minimum number of code reviews from core devs
  - Standards for documentation
  - Contributing licensing agreements (more on that later)

Governance

- Some OSS projects are managed by for-profit firms
  • Examples: Chromium (Google), Moby (Docker), Ubuntu (Canonical), TensorFlow (Google), PyTorch (Meta), Java (Oracle)
  - Contributors may be a mix of employees and community volunteers
  - Corporations often fund platforms (websites, test servers, deployments, repository hosting, etc.)
  - Corporations usually control long-term vision and feature roadmap
- Many OSS projects are managed by non-profit foundations or ad- hoc communities
  • Examples: Apache Hadoop/Spark/Hbase/Kafka/Tomcat (ASF), Firefox (Mozilla), Python (PSF), NumPy (community)
  - Foundations fund project infrastructure via charitable donations
  - Long-term vision often developed via a collaborative process (e.g., Apache) or by benevolent dictators (e.g., Python, Linux)
- Corporations still heavily rely on community-owned OSS projects • Many OSS non-profits are funded by Big Tech (e.g., Mozilla by Google)

Example: Apache

WHAT MAKES THE APACHE WAY SO HARD TO DEFINE?

The Apache Way is a living, breathing interpretation of one's experience with our community-led development process. Apac
unique, diverse, and focused on the activities needed at a particular stage of the project's lifetime, including nurturing comm
building awareness. What is important is that they embrace:

- ***Earned Authority:*** all individuals are given the opportunity to participate, but their influence is based on publicly earne
community. Merit lies with the individual, does not expire, is not influenced by employment status or employer, and is r
project cannot be applied to another). More on merit.
- ***Community of Peers:*** individuals participate at the ASF, not organizations. The ASF's flat structure dictates that roles ar
equal weight, and contributions are made on a volunteer basis (even if paid to work on Apache code). The Apache com
with respect in adherence to our Code of Conduct. Domain expertise is appreciated; Benevolent Dictators For Life are c
participation.
- ***Open Communications:*** as a virtual organization, the ASF requires all communications related to code and decision-m
asynchronous collaboration, as necessitated by a globally-distributed community. Project mailing lists are archived, pub
  - dev@ (primary project development)
  - user@ (user community discussion and peer support)
  - commits@ (automated source change notifications)
  - occasionally supporting roles such as marketing@ (project visibility)

...as well as restricted, day-to-day operational lists for Project Management Committees. Private decisions on code, policies, or
discourse and transactions must be brought on-list. More on communications and the use of mailing lists.

- ***Consensus Decision Making:*** Apache Projects are overseen by a self-selected team of active volunteers who are contrib
Projects are auto-governing with a heavy slant towards driving consensus to maintain momentum and productivity. W
establish at all times, holding a vote or other coordination may be required to help remove any blocks with binding dec
More on decision making and voting.
- ***Responsible Oversight:*** The ASF governance model is based on trust and delegated oversight. Rather than detailed ru
governance is principles-based, with self-governing projects providing reports directly to the Board. Apache Committer
reviewed commits, employing mandatory security measures, ensuring license compliance, and protecting the Apache
abuse. More on responsibility.

THE APACHE SOFTWARE FOUNDATION

OUR SPONSORS

The Apache Software Foundation could not exist without the continued generous support from the community. We would like to take this opportunity to thank our sponsors. If you are interested in sponsoring the ASF, please read our sponsorship page.

FOUNDATION SPONSORS

Platinum Sponsors:

FACEBOOK yahoo!

Facebook Yahoo!

Pineapple Fund

Pineapple Fund

HUAWEI

Huawei

aws

Amazon Web Services

Microsoft

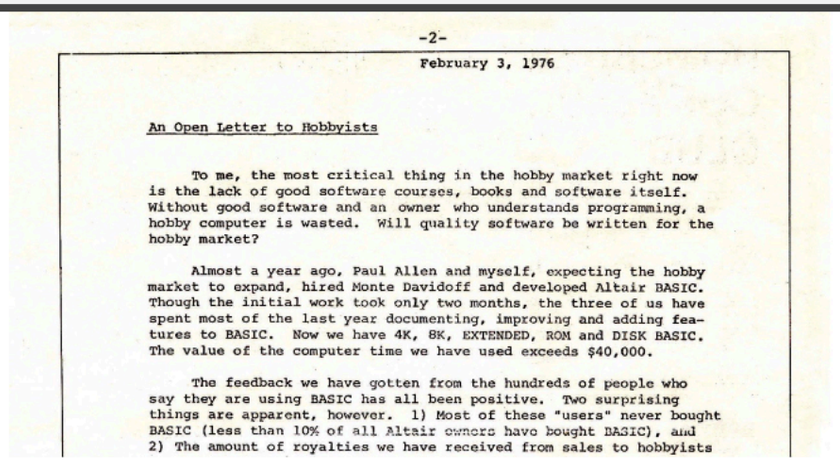
Microsoft

Apple

Google

Google

Corporate Outlook Towards Open-source



-2-

February 3, 1976

An Open Letter to Hobbyists

To me, the most critical thing in the hobby market right now is the lack of good software courses, books and software itself. Without good software and an owner who understands programming, a hobby computer is wasted. Will quality software be written for the hobby market?

Almost a year ago, Paul Allen and myself, expecting the hobby market to expand, hired Monte Davidoff and developed Altair BASIC. Though the initial work took only two months, the three of us have spent most of the last year documenting, improving and adding features to BASIC. Now we have 4K, 8K, EXTENDED, ROM and DISK BASIC. The value of the computer time we have used exceeds $40,000.

The feedback we have gotten from the hundreds of people who say they are using BASIC has all been positive. Two surprising things are apparent, however. 1) Most of these "users" never bought BASIC (less than 10% of all Altair owners have bought BASIC), and 2) The amount of royalties we have received from sales to hobbyists

"...most of you steal your software..."

Is this fair? One thing you don't do by stealing software is get back at MITS for some problem you may have had. MITS doesn't make money selling software. The royalty paid to us, the manual, the tape and the overhead make it a break-even operation. One thing you do do is prevent good software from being written. Who can afford to do professional work for nothing? What hobbyist can put 3-man years into programming, finding all bugs, documenting his product and distribute for free? The fact is, no one besides us has invested a lot of money in hobby software. We have written 6800 BASIC, and are writing 8080 APL and 6800 APL, but there is very little incentive to make this software available to hobbyists. Most directly, the thing you do is theft.

What about the guys who re-sell Altair BASIC, aren't they making money on hobby software? Yes, but those who have been reported to us may lose in the end. They are the ones who give hobbyists a bad name, and should be kicked out of any club meeting they show up at.

I would appreciate letters from any one who wants to pay up, or has a suggestion or comment. Just write me at 1180 Alvarado SE, #114, Albuquerque, New Mexico, 87108. Nothing would please me more than being able to hire ten programmers and deluge the hobby market with good software.

Bill Gates

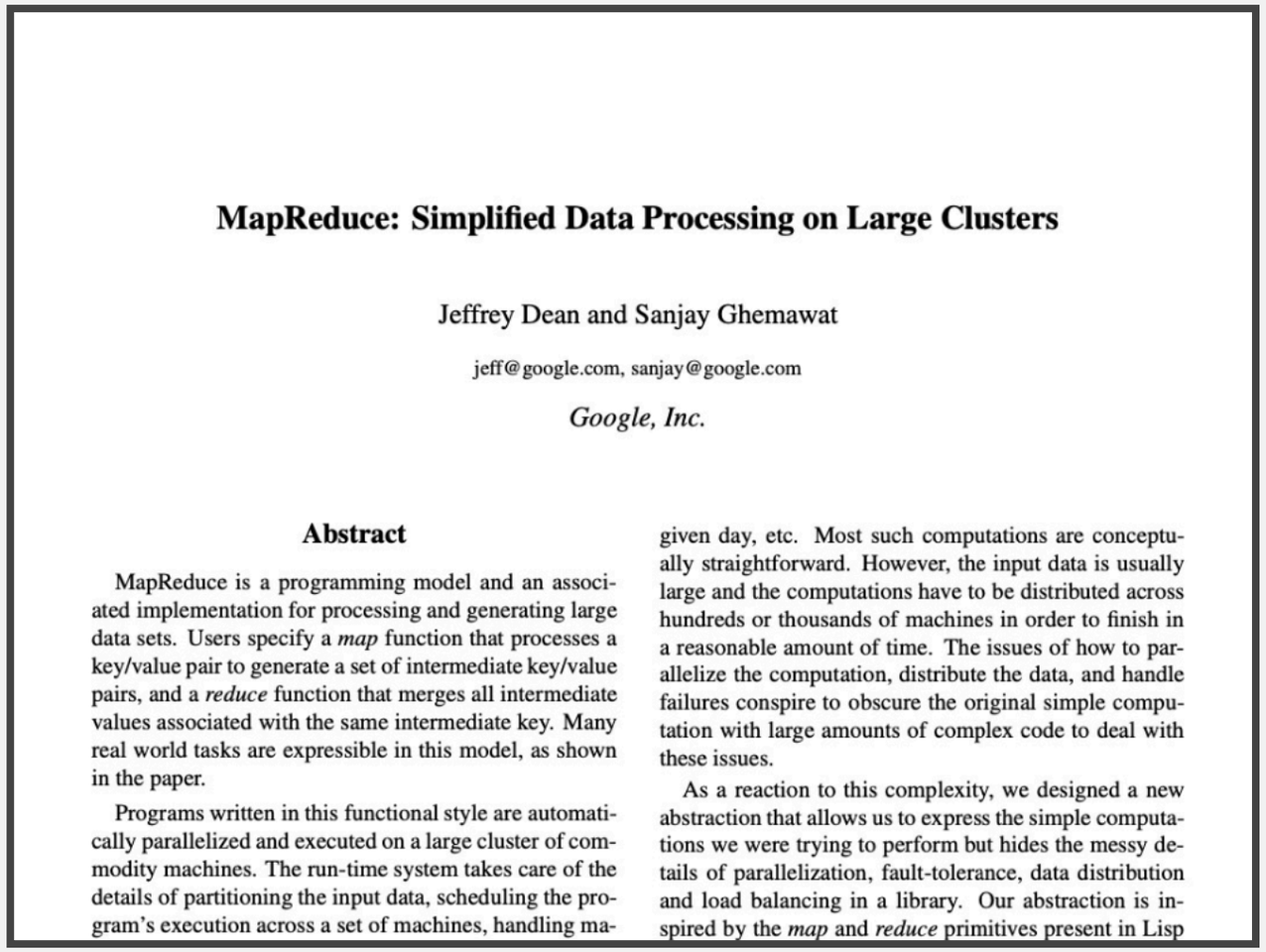
Bill Gates
General Partner, Micro-Soft

Redmond top man Satya Nadella: 'Microsoft LOVES Linux'

Open-source 'love' fairly runneth over at cloud event

20 Oct 2014 at 23:45, Neil McAllister

Risks in ***not*** Open-Sourcing?



MapReduce: Simplified Data Processing on Large Clusters

Jeffrey Dean and Sanjay Ghemawat

jeff@google.com, sanjay@google.com

Google, Inc.

Abstract

MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a *map* function that processes a key/value pair to generate a set of intermediate key/value pairs, and a *reduce* function that merges all intermediate values associated with the same intermediate key. Many real world tasks are expressible in this model, as shown in the paper.

Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system.

Our implementation of MapReduce runs on a large cluster of commodity machines and is highly scalable: a typical MapReduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of MapReduce programs have been implemented and upwards of one thousand MapReduce jobs are executed on Google's clusters every day.

1 Introduction

Over the past five years, the authors and many others at Google have implemented hundreds of special-purpose computations that process large amounts of raw data, such as crawled documents, web request logs, etc., to compute various kinds of derived data, such as inverted indices, various representations of the graph structure of web documents, summaries of the number of pages crawled per host, the set of most frequent queries in a given day, etc. Most such computations are conceptually straightforward. However, the input data is usually large and the computations have to be distributed across hundreds or thousands of machines in order to finish in a reasonable amount of time. The issues of how to parallelize the computation, distribute the data, and handle failures conspire to obscure the original simple computation with large amounts of complex code to deal with these issues.

As a reaction to this complexity, we designed a new abstraction that allows us to express the simple computations we were trying to perform but hides the messy details of parallelization, fault-tolerance, data distribution and load balancing in a library. Our abstraction is inspired by the *map* and *reduce* primitives present in Lisp

Use of Open-Source Software in Companies

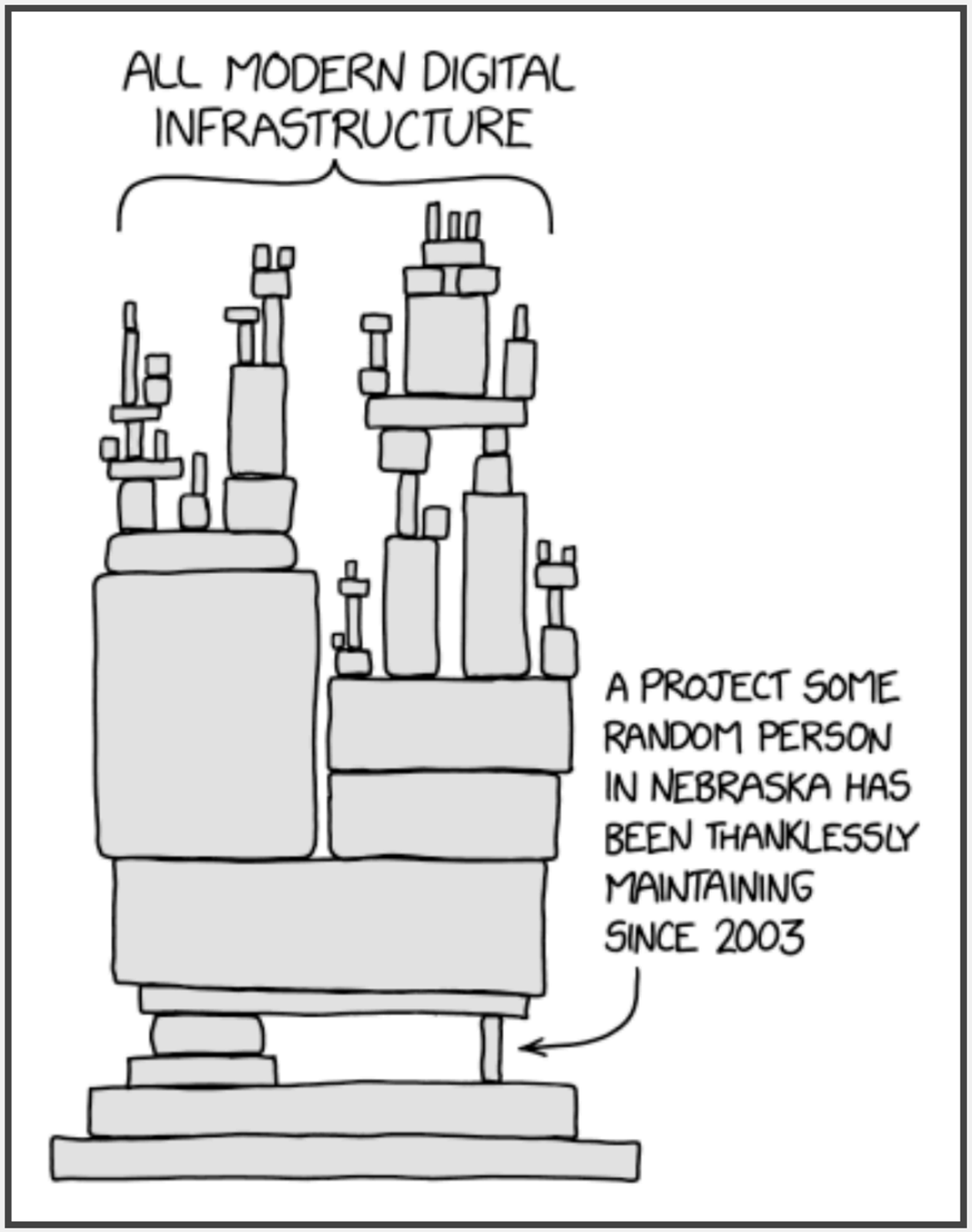
- Is the license compatible with our intended use?
  - More on this later

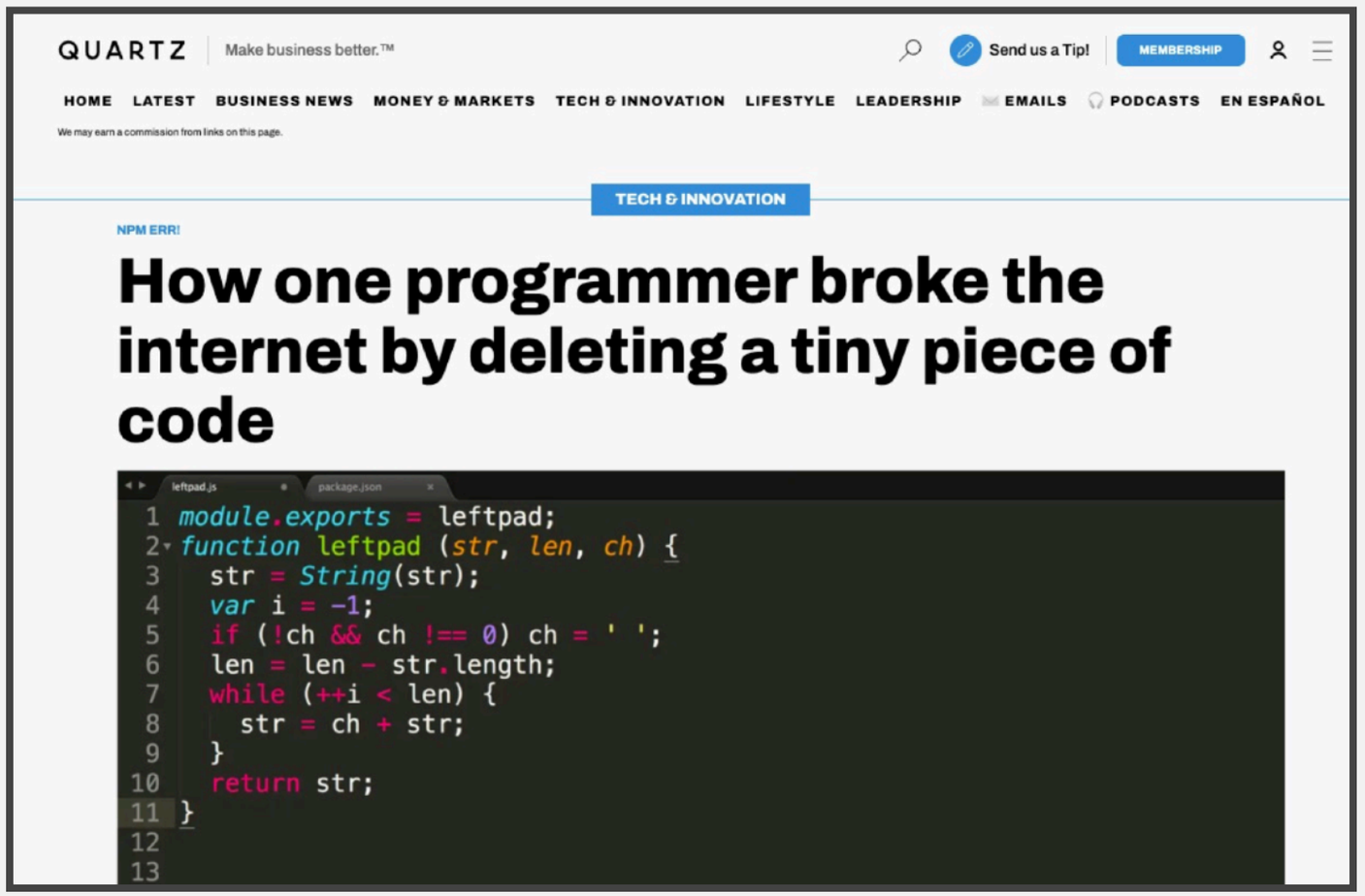
- How will we handle versioning and updates?
  - Does every internal project declare its own versioned dependency or do we all agree on using one fixed (e.g., latest) version?
  - Sometimes resolved by assigning internal “owners” of a third-party dependency, who are responsible for testing updates and declaring allowable versions.

- How to handle customization of the OSS software?
  - Internal forks are useful but hard to sync with upstream changes.
  - One option: Assign an internal owner who keeps internal fork up-to-date with upstream.
  - Another option: Contribute all customizations back to upstream to maintain clean dependencies.

- Security risks? Supply chain attacks on the rise.

Use of Open-Source Software in Companies

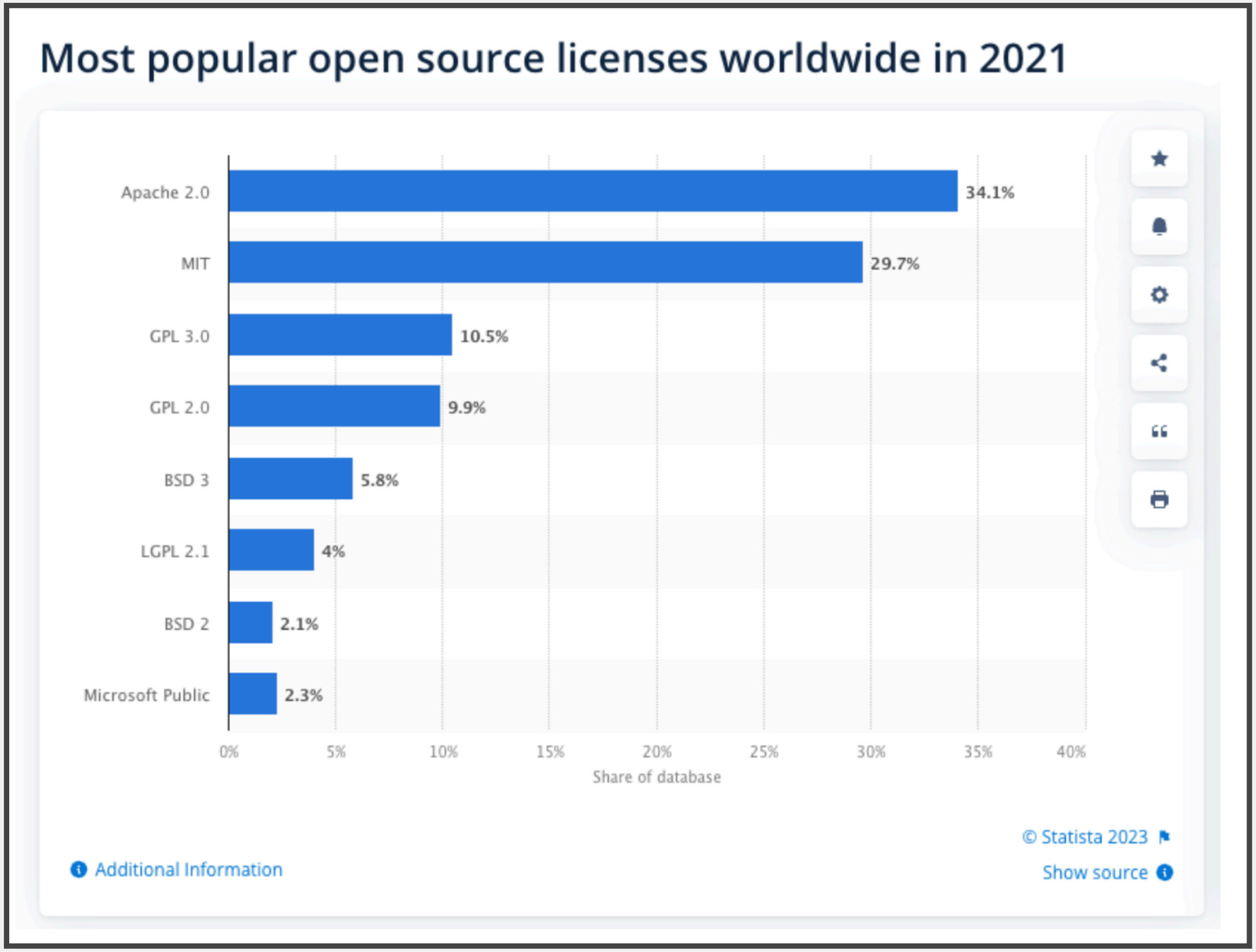




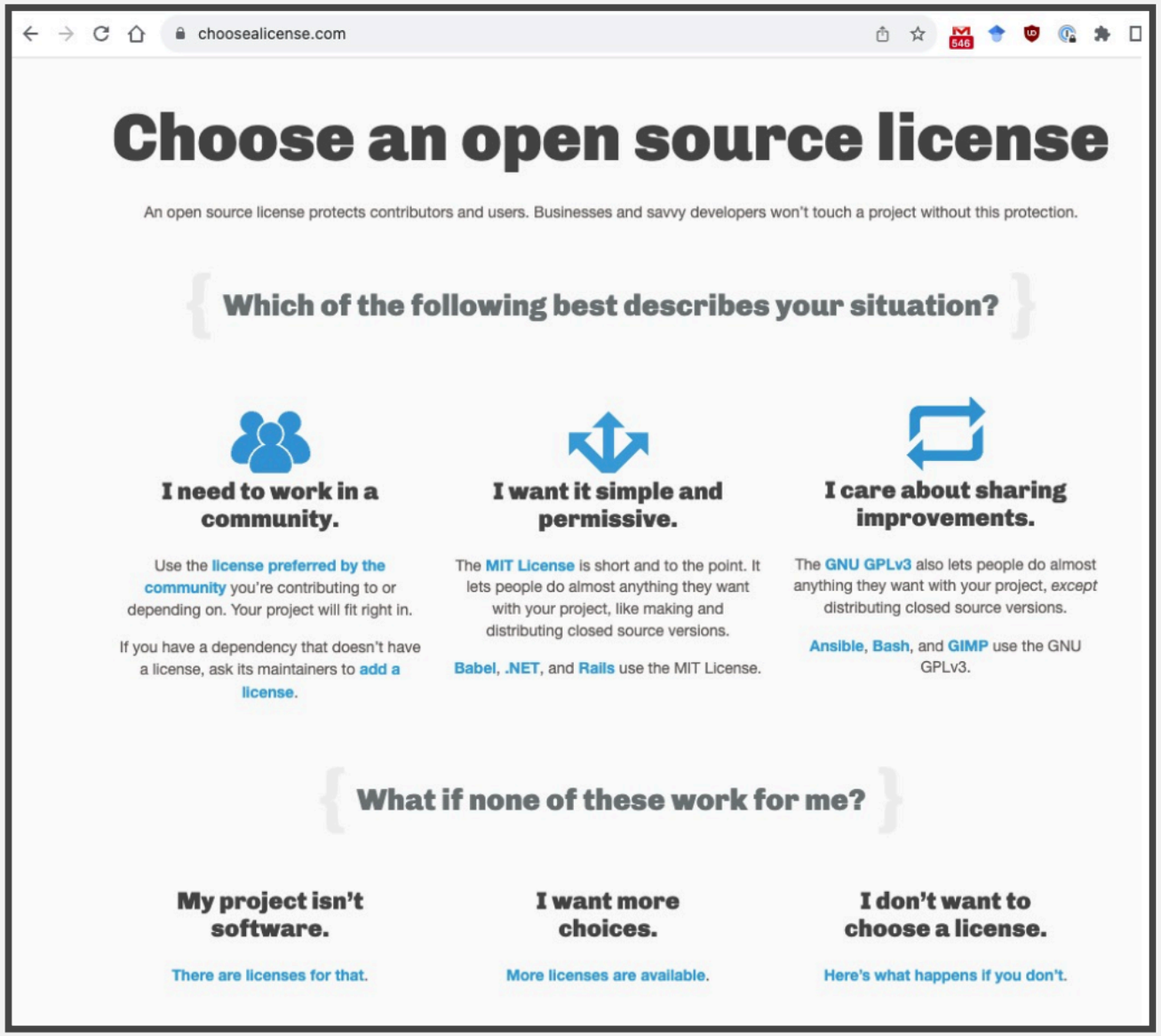
Software Licenses

Note: I am not a lawyer (this is not legal advice)

Most popular Software Licenses



Which License to Choose?



GNU General Public License: the Copyleft License

- Nobody should be restricted by the software they use. There are four freedoms that every user should have:
  - the freedom to use the software for any purpose,
  - the freedom to change the software to suit your needs,
  - the freedom to share the software with your friends and neighbors, and
  - the freedom to share the changes you make.
- Code must be made available
- Any modifications must be relicensed under the same license (copyleft)

Risks of "Copyleft" Licenses

- Nobody should be restricted by the software they use. There are four freedoms that every user should have:
  - the freedom to use the software for any purpose,
  - the freedom to change the software to suit your needs,
  - the freedom to share the software with your friends and neighbors, and
  - the freedom to share the changes you make.
- Code must be made available
- Any modifications must be relicensed under the same license (copyleft)

Lesser GNU Public License (LGPL)

- Software must be a library

- Similar to GPL but does not consider dynamic binding as “derivative work”

- So, proprietary code can depend on LGPL libraries as long as they are not being modified

- See also: GPL with classpath exception (e.g., Oracle JDK)

MIT License

- Simple, commercial-friendly license
- Must retain copyright credit
- Software is provided as is
- Authors are not liable for software
- No other restrictions

Risk: Incompatible Licenses

- Sun open-sourced OpenOffice, but when Sun was acquired by Oracle, Oracle temporarily stopped the project.
- Many of the community contributors banded together and created LibreOffice
- Oracle eventually released OpenOffice to Apache
- LibreOffice changed the project license so LibreOffice can copy changes from OpenOffice but OpenOffice cannot do the same due to license conflicts

Copyright vs. Intellectual Property

- IP and Patents cover an idea for solving a problem
  - Examples: Machine designs, pharma processes to manufacture certain drugs, (controversially) algorithms
  - Have expiry dates. IP can be licensed or sold/ transferred for $$$.

- Copyrights cover particular expressions of some work
  - Examples: Books, music, art, source code
  - Automatic copyright assignment to all new work unless a license authorizes alternative uses.

- Exceptions for trivial works and ideas.

Contributor License Agreements (CLA)

- Often a requirement to sign these before you can contribute to OSS projects
- Scoped only to that project
- Assigns the maintainers specific rights over code that you contribute
- Without this, you own the copyright and IP for even small bug fixes and that can cause them legal headaches in the future

Summary

- Open-source software harnesses the collective power of stakeholders not directly associated with main developers
- Open-source ecosystems thrive in many application domains where reuse is common (e.g., platforms, frameworks, libraries)
- Corporations rely on open-source even if they develop proprietary software or services.
- Open-source licenses must be chosen carefully to align with intended use case.
- You will all contribute to OSS in this class!