

Mastering LLMs



Large Language Models

ChatGPT, Whisperer, Copilot, Bard, ..

How do LLMs work?

■ Phase 1: Train *knowledge*

- 120 TB of data from the Internet
- Expensive: 2 weeks x 6000 GPU, cost: \$2M, done ~yearly
- Result: condensing it in 120GB of float numbers [0...1]

■ Phase 2: Train to *answer questions*

- Humans provide eg 100.000 questions + ideal answers that AI will imitate
- Cheap: days

■ Phase 3 (optional): User feedback

- Manually correct rejected answers and feed back in Phase 2 ↻

LLMs in Programmer's Life

Coding

- **Write obvious code** / *in the same way*. **Better:** Avoid writing it at all!
- **Build prototypes**, where speed is critical
- **Generate** (draft) **unit tests** on existing legacy code
- **Find bugs** (NPE) and **subtle issues** (performance, vulnerabilities..)
- **Translate code** to other languages/stack
- **Suggest refactoring ideas**, but not execute them
- **Pair-buddy** that keeps you engaged & gives inspiration

Outside code

- **Replacement for Google Search / Stack Overflow** (=garbage/spam recently)
- **Faster Learning: you can *dive in gently* into complex code/specs/topics** 🤗
- **Personal Assistant** writing professional text for you: emails/documentation

AI Problems

- **False information:** Confirm answers! Ask: *Are you sure?*
- **Limited Context:** Copilot: [6K chars](#) in current + near editor tab
- **Garbage IN → Garbage OUT**
 - **Bad context** (legacy code) or **Bad prompts** → **Bad Answers** (aka "Artificial Stupidity")
 - No control on training data set → Bias ("tyranny of majority")
- **Distracting** if too verbose
 - Protect your focus via fast prompts //comments, good names, voice dictation 🎙️
- **No Intellectual Property:** if you commit, you account for it
- **Makes us lazy + critical** & can take away **creativity playground**
 - Addiction prevention: **AI-free day** (eg, Tuesday)

AI vs Intellectual Property

```
const Expanded(
  flex: 5,
  child: Row(
    children: [
      Column(
        children: [
          const Text(
            "© 2022 Intesa Sanpaolo Bank",
            style: TextStyle(
              color: Colors.white,
              fontSize: 12.0,
            ),
          ),
          const Text(
            "Toate drepturile rezervate",
            style: TextStyle(
              color: Colors.white,
              fontSize: 12.0,
            ),
          ),
        ],
      ), // Column
    ], // Row
  ), // Expanded
```

Dear all,

I am trying to use the "R" package "DESeq2" to analyze RNA-seq data. I have a question about the "DESeqDataSetFromMatrix" function. I have a matrix of raw counts, and I would like to use the "DESeqDataSetFromMatrix" function to create a "DESeqDataSet" object. However, I am not sure how to specify the "colData" argument. I have a data frame with the following columns: "Sample", "Condition", "Replicate". I would like to use the "Condition" column as the "condition" argument in the "DESeqDataSetFromMatrix" function. However, I am not sure how to specify the "colData" argument. I have tried the following:

Days before OpenAI



Days after OpenAI



Will AI change Programming?

■ Greater divide between Senior <> Junior

- Seniors get a productivity boost (10x ▶ 100x)
- Juniors get less PR approved% (can't understand/debug AI generated-code)
- Ride your LLMa, don't get carried away: be sure you can write the code w/o

■ Mastered AI = differentiator (faster coder & learner)

➔ Talk to an AI weekly to prepare!

■ Will AI replace programmers?

- Only Code Monkeys 🐵 are in danger
- Programming = **solving problems**, today via **typing** code,
- ... tomorrow via **tiny prompts** that will make AI write it for you
- Domain knowledge will matter more than frameworks/languages - [nvidia CEO](#)

