

# Big Data Systems

A course for modern data modeling, data management and big data applications

Dr. Rubi Boim

# WHAT IS BIG DATA

Let's try DALL-E 2

**“a 3d render of big data”**

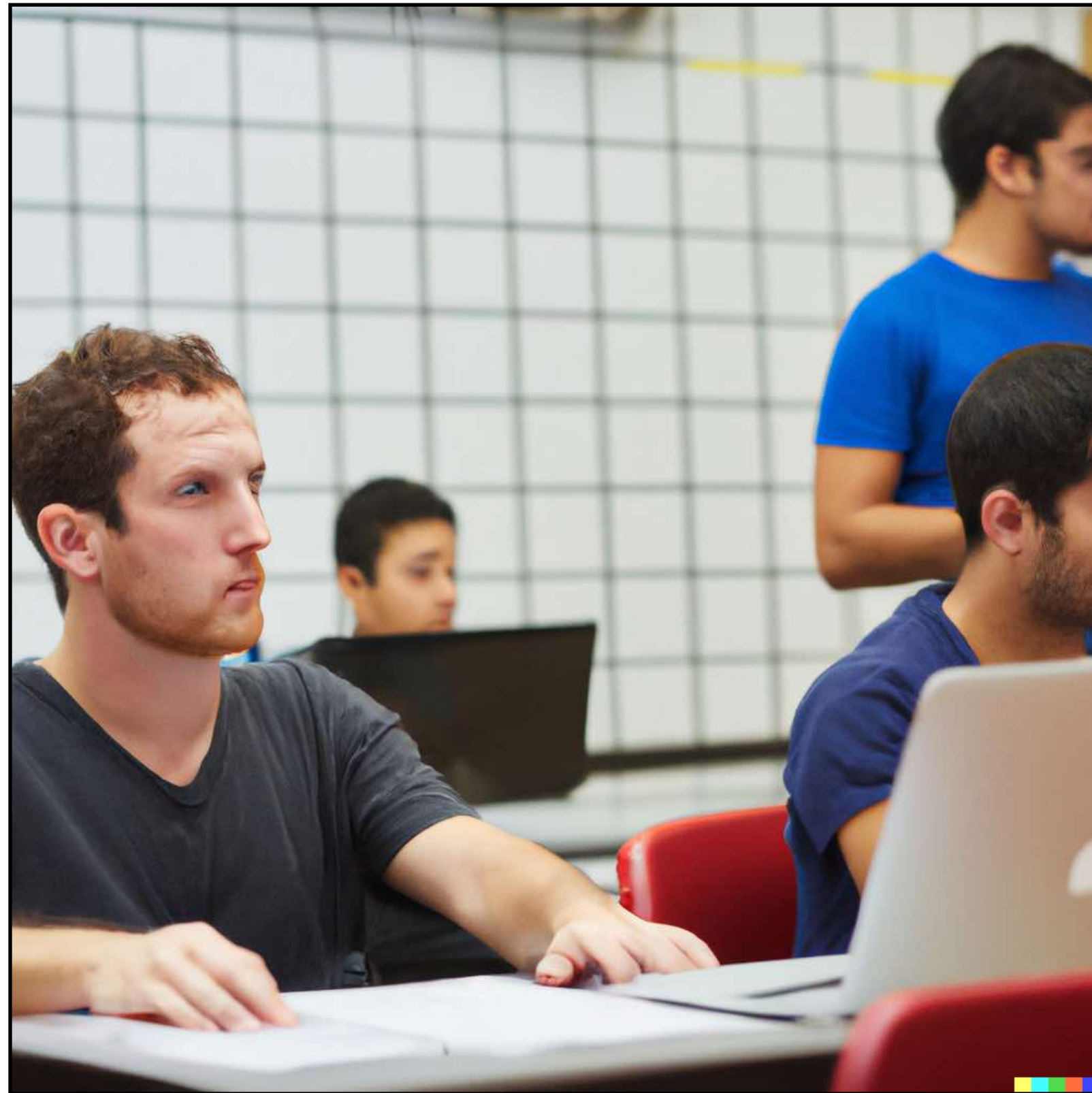


“a 3d render of big data”

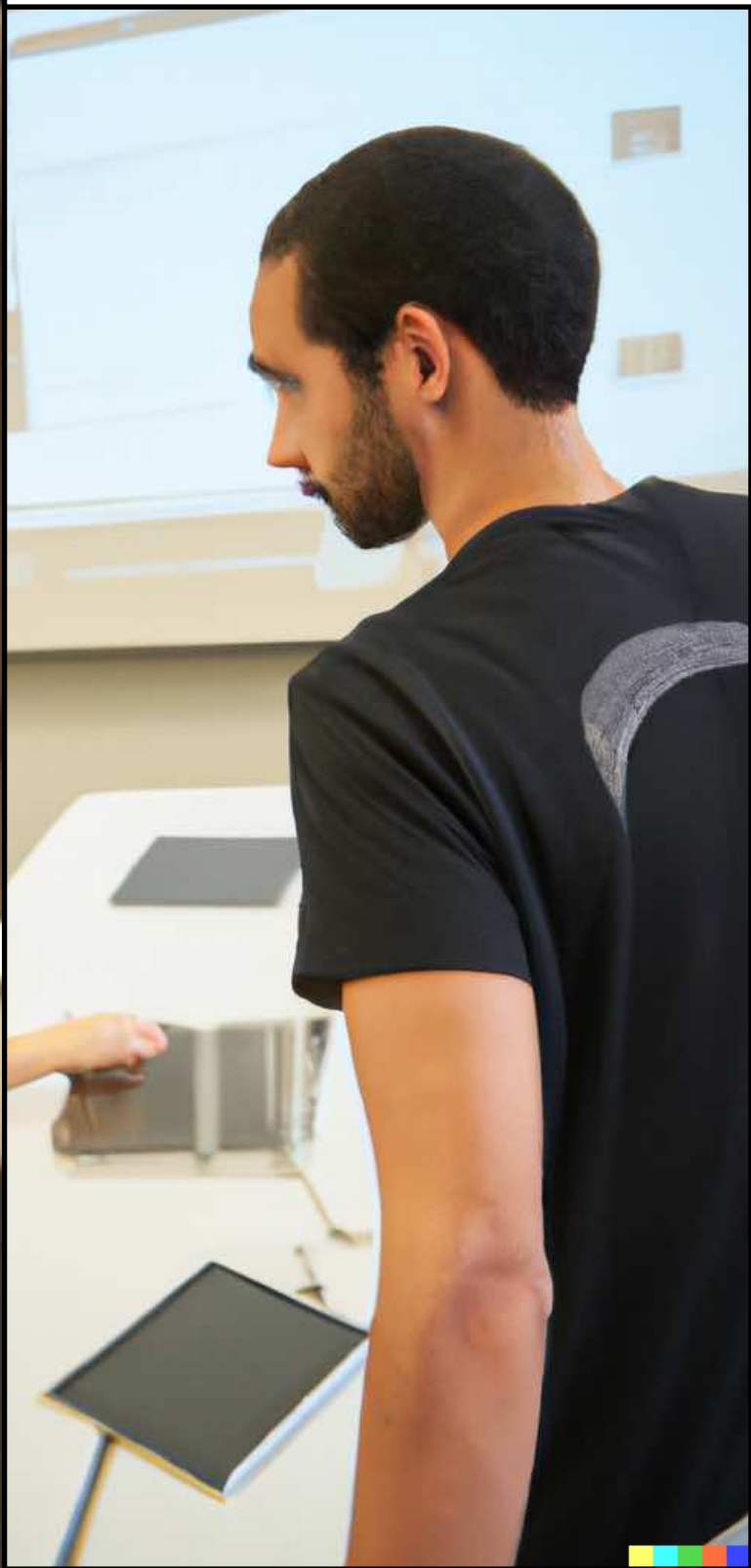


side note - do you think DALL-E 2 likes Metallica?

**“students at bigdata course at tel aviv university”**



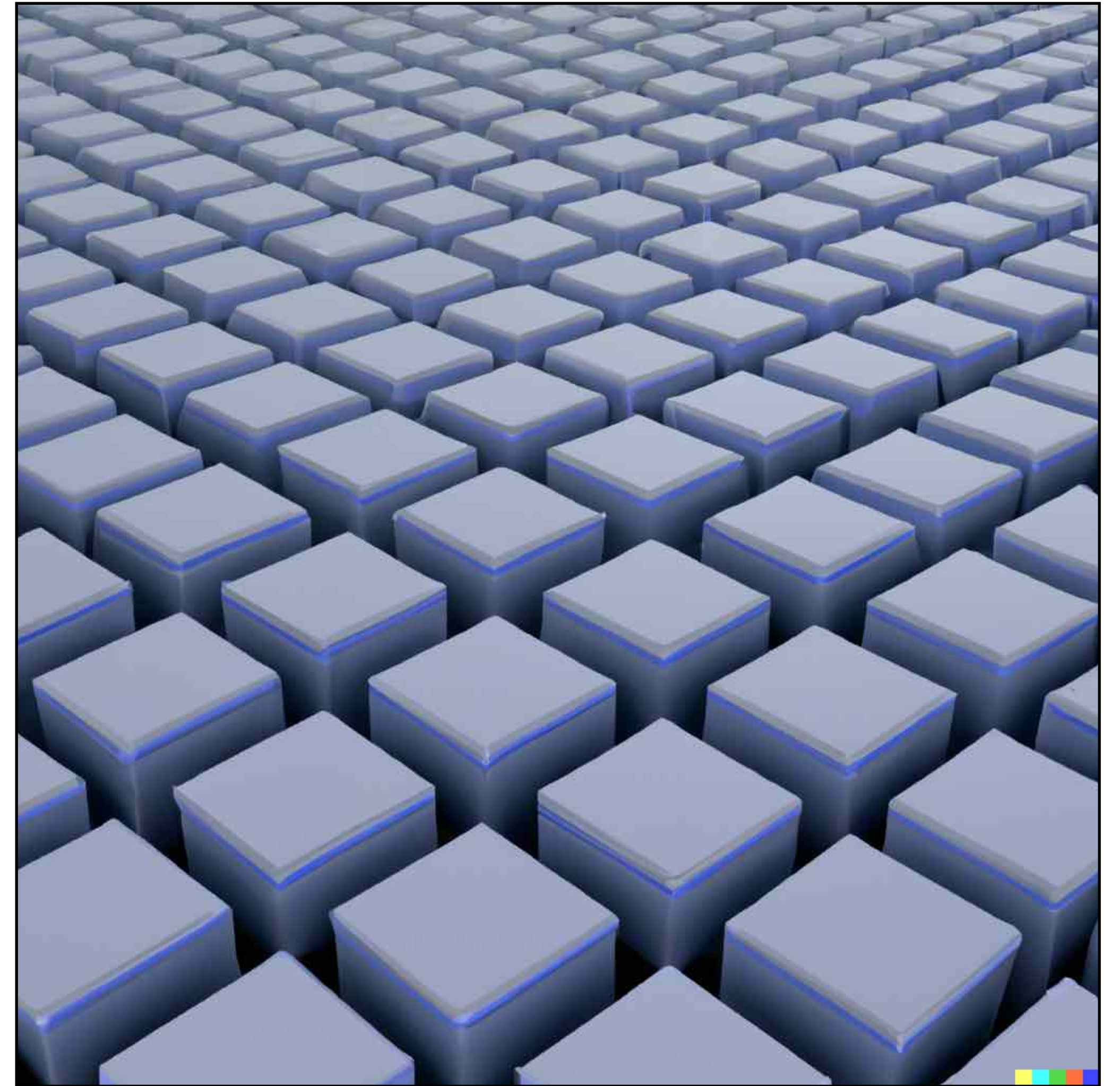
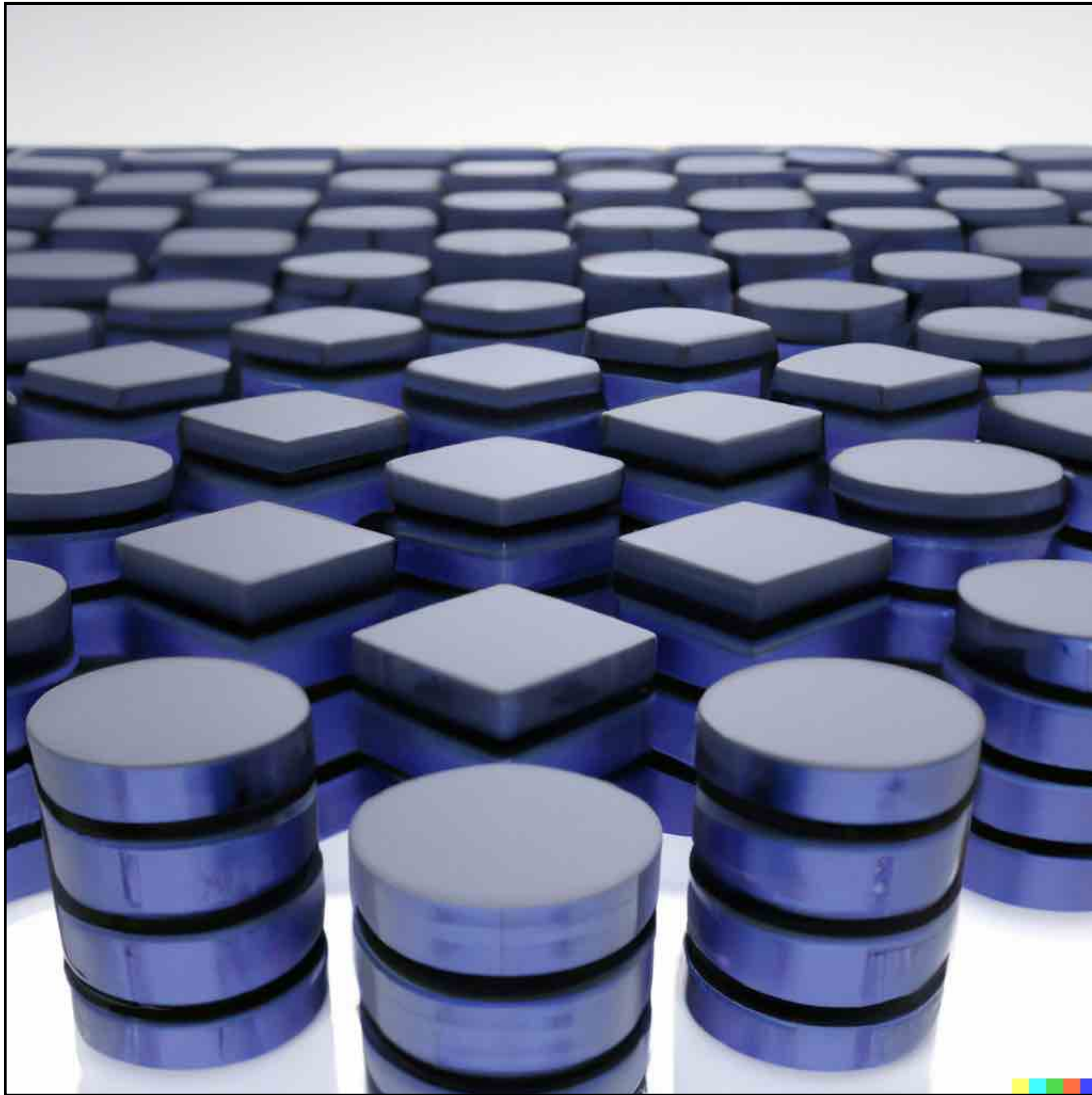
**“students at bigdata course at tel aviv university”**



“students at bigdata course at tel aviv university”



**“a 3d render of distributed database”**



“a 3d render of distributed database”

**“a 3d render of bigtable”**



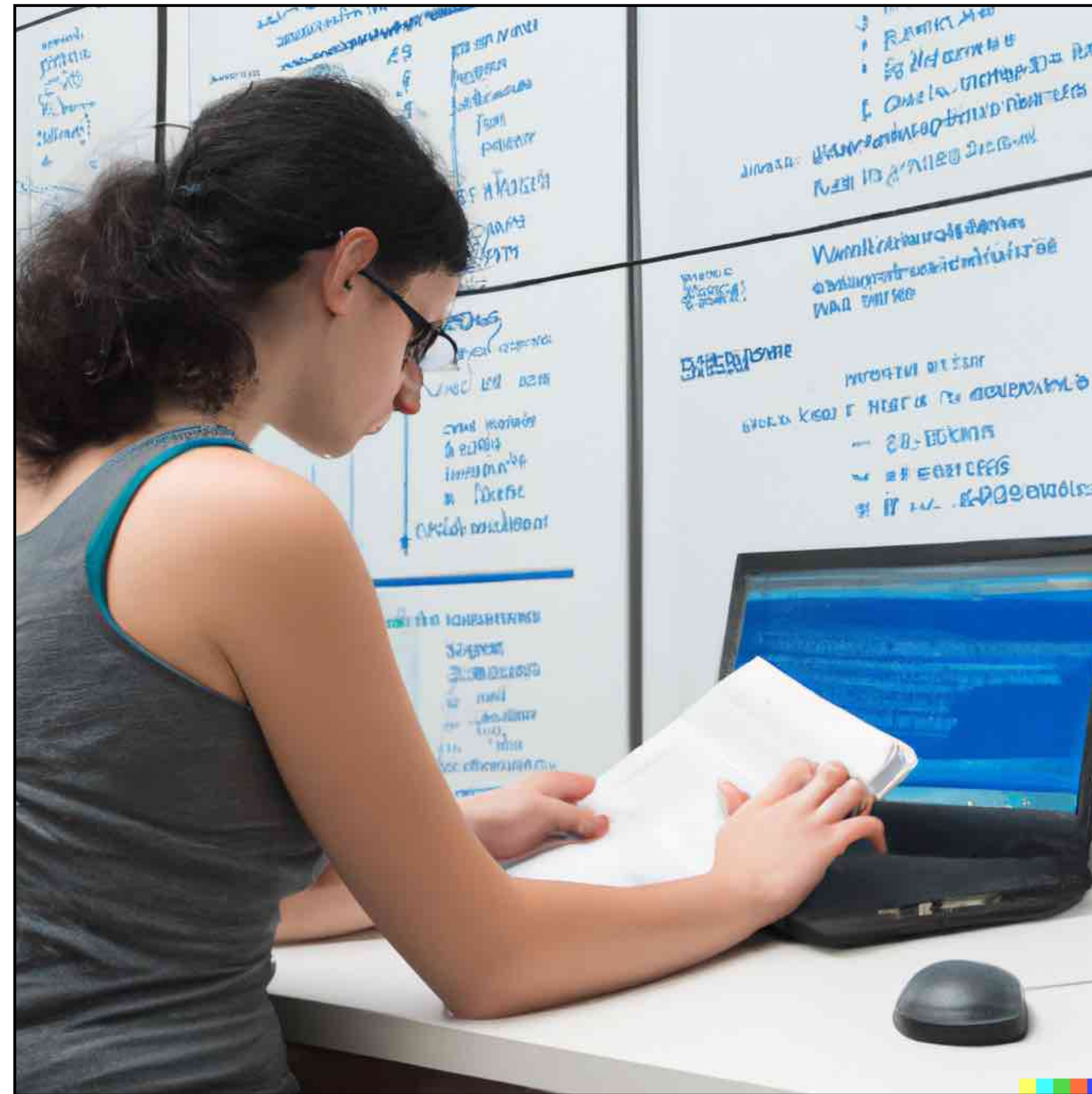
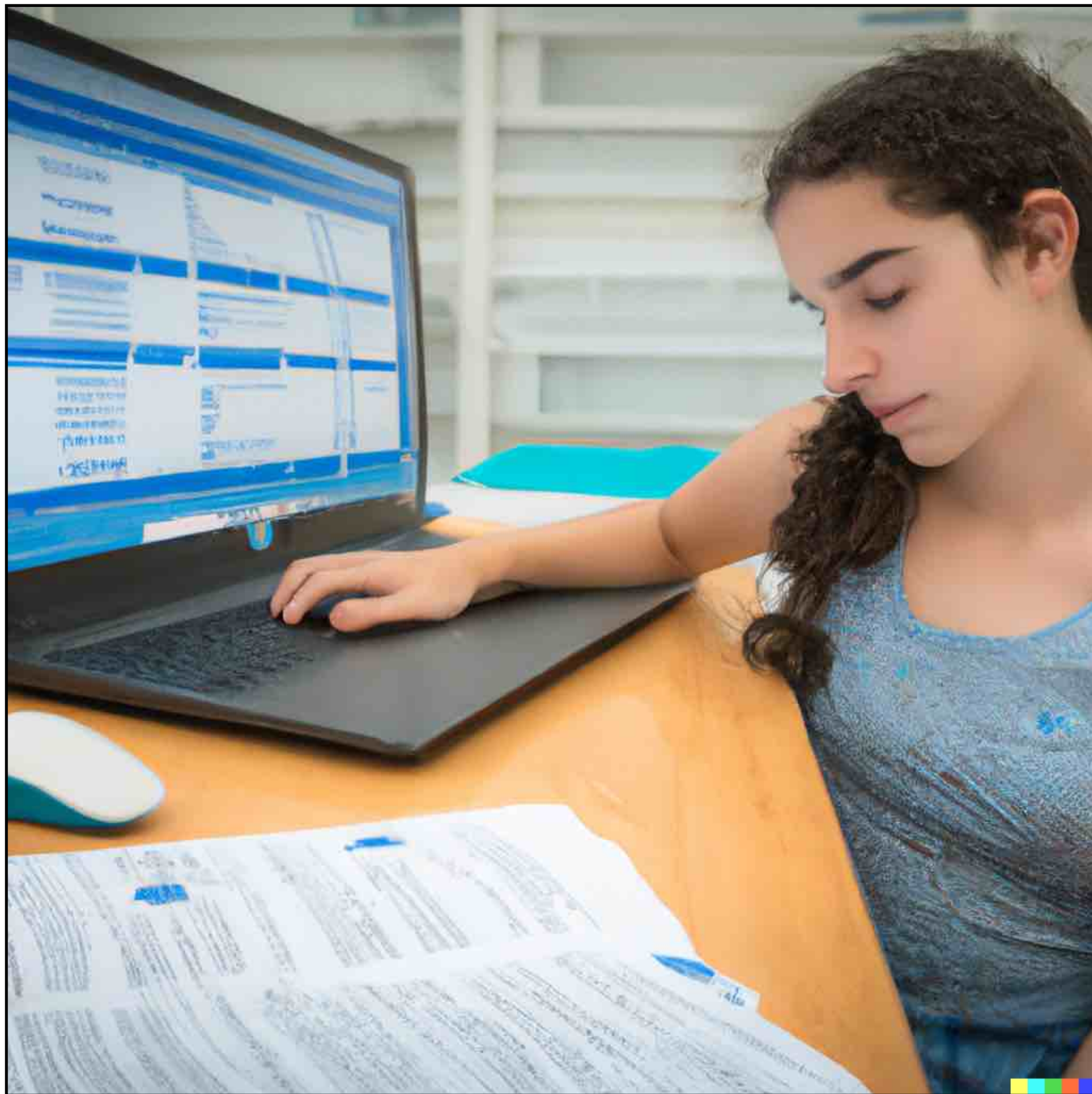
“a 3d render of bigtable”

**“a 3d render of cassandra”**



“a 3d render of cassandra”

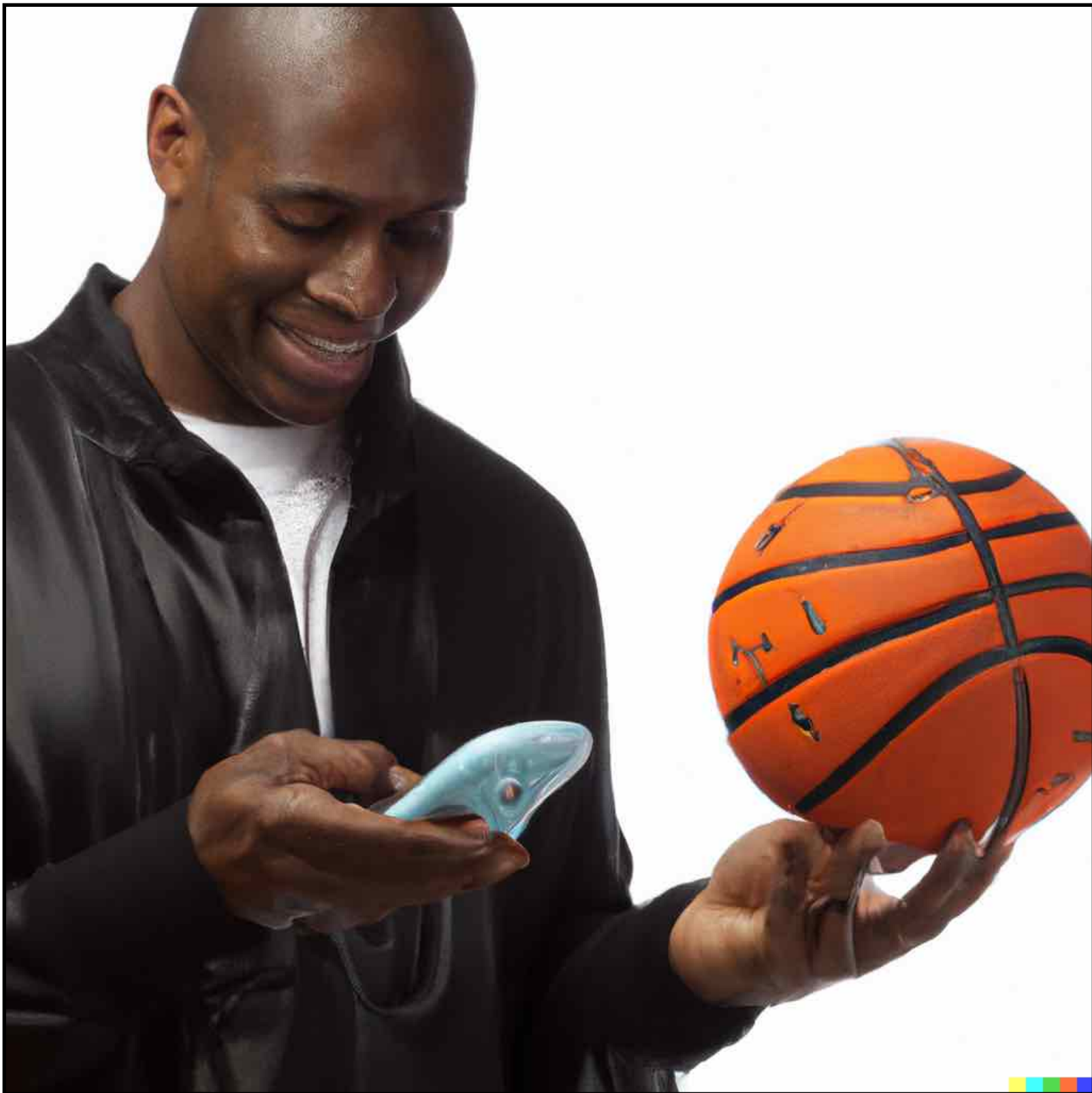
**“a student using cassandra and bigtable databases while doing her homework for the course big data system in tel aviv university”**



“a student using cassandra and bigtable databases while doing her homework for the course big data system in tel aviv university”



**“michael jordan using twitter to tweet about  
a new basketball he bought”**



“michael jordan using twitter to tweet about a new basketball he bought”

# Big data is like teenage sex

- everyone talks about it
- nobody really knows how to do it
- everyone thinks everyone else is doing it
- so everyone claims they are doing it...

*(Dan Arieli, 2013)*

# Course goal

provide the theoretical as well as the practical hands on knowledge required for designing and developing internet scale based data applications

**(from a data management and data modeling perspective)**

# privilege to teach and learn

failing in this course is not an option.

# understanding the motivation

examples before theories

# Buzzwords to be covered (tentative)

- Relational databases
  - SQL and normalized DB
- Distributed databases
  - NoSQL (wide column)
  - CAP theorem
  - Dynamo
  - BigTable
  - **Cassandra**
  - Advance data modeling
- Other
  - Kafka
  - Data warehouse

# Staff

## Dr. Rubi Boim

- boim AT **cs**.tau.ac.il
- <https://www.cs.tau.ac.il/~boim/>
- Office hours: by appointment

## Nadav Magar (TA)

- nadavmagar AT mail.tau.ac.il



# Communications

- **Course website:**  
<https://courses.cs.tau.ac.il/bigdata/>
- **Moodle**  
<https://moodle.tau.ac.il/enrol/index.php?id=368327601>

# Requirements

- Written test
- 2-4 HW assignments
  - in pairs
  - Java as programming language
- Database system course  
highly recommended prerequisite or in parallel

# Reference texts

- Principles of Distributed Database Systems (*M. Tamer Özsu, Patrick Valduriez*)
- Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services (*Seth Gilbert et al.*)
- CAP Twelve Years Later: How the “Rules” Have Changed (*Eric Brewer*)
- Bigtable: A Distributed Storage System for Structured Data (*Fay Chang et al.*)
- The Google File System (*Sanjay Ghemawat et al.*)
- Cassandra - A Decentralized Structured Storage System (*Avinash Lakshman et al.*)
- Dynamo: Amazon’s Highly Available Key-value Store (*Giuseppe DeCandia et al.*)
- Kafka: a Distributed Messaging System for Log Processing (*Jay Kreps et al.*)

# Course plan (tentative)

- Intro to big data
- Relational databases
- **Distributed databases and techniques**
- **Cassandra (wide column databases)**
- **Advanced modeling**
- Data warehouse
- Streaming

# Course schedule



<https://courses.cs.tau.ac.il/bigdata/>