BIG DATA & BUSINESS INTELLIGENCE
WEEK 1

Winter Semester 2025-2026

Lecturer: Narges Chinichian
SRH University of Applied Science

ABOUT ME

Dr. Narges Chinichian

PhD in Physics (Computational Neuroscience), with 8+ years experience in data science & machine learning.

Hobby teacher, I also teach rope climbing at Der Kegel in Berlin!

ABOUT YOU?

Please share briefly:

1. Your name & if you'd like: where you're from.
2. What is one area of data or technology you find exciting?
3. What is your educational background.
4. How would you rate your Python and Git skills?
5. What are your expectations for this course? What do you hope to learn?

Bonus: If data were a person, how would they look and what would they be like?

ORGANIZATIONAL ASPECTS

- Six weeks (6 days of 8 units)
- There is no class on the 6th of November.
- Classes start at 12:30 and are until 8pm.

- From 5:30-8 presence in the class environment is optional (if you wish and don't need assistance from me, you could leave the room and work in the common area or your dorm).

ORGANIZATIONAL ASPECTS

- Course Hub:
- All up-to-date links and info can be found here:

ORGANIZATIONAL ASPECTS

- We form teams during the course based on the activities.

- Of course there is always help from lecturer but try to test this hirarchy for better efficiency:

  - Think yourself. (there is one of you for EACH of you)
  - Consult GenAI. (there is one GenAI for EACH of you)
  - Ask your neighbors. (there is at least one neighbor for MOST of you)
  - Ask the lecturer. (there is one lecturer for ALL of you)

COURSE OBJECTIVES

- Understand data lifecycles and architectures
- Learn principles of data warehousing and ETL
- Apply visualization and BI tools for insights
- Explore some aspects of data ethics and governance in practice

COURSE STRUCTURE & ASSESSMENT

- Weekly sessions combining theory and hands-on work
- GitHub-based materials and assignments
- Quizzes and mini-projects for each topic
- Final group project with real data

SETTING UP YOUR GITHUB

Please all add your GitHub handle here:

It's what you get in your url when you are on your profile page:

So if you see:

https://github.com/NoCh-Git

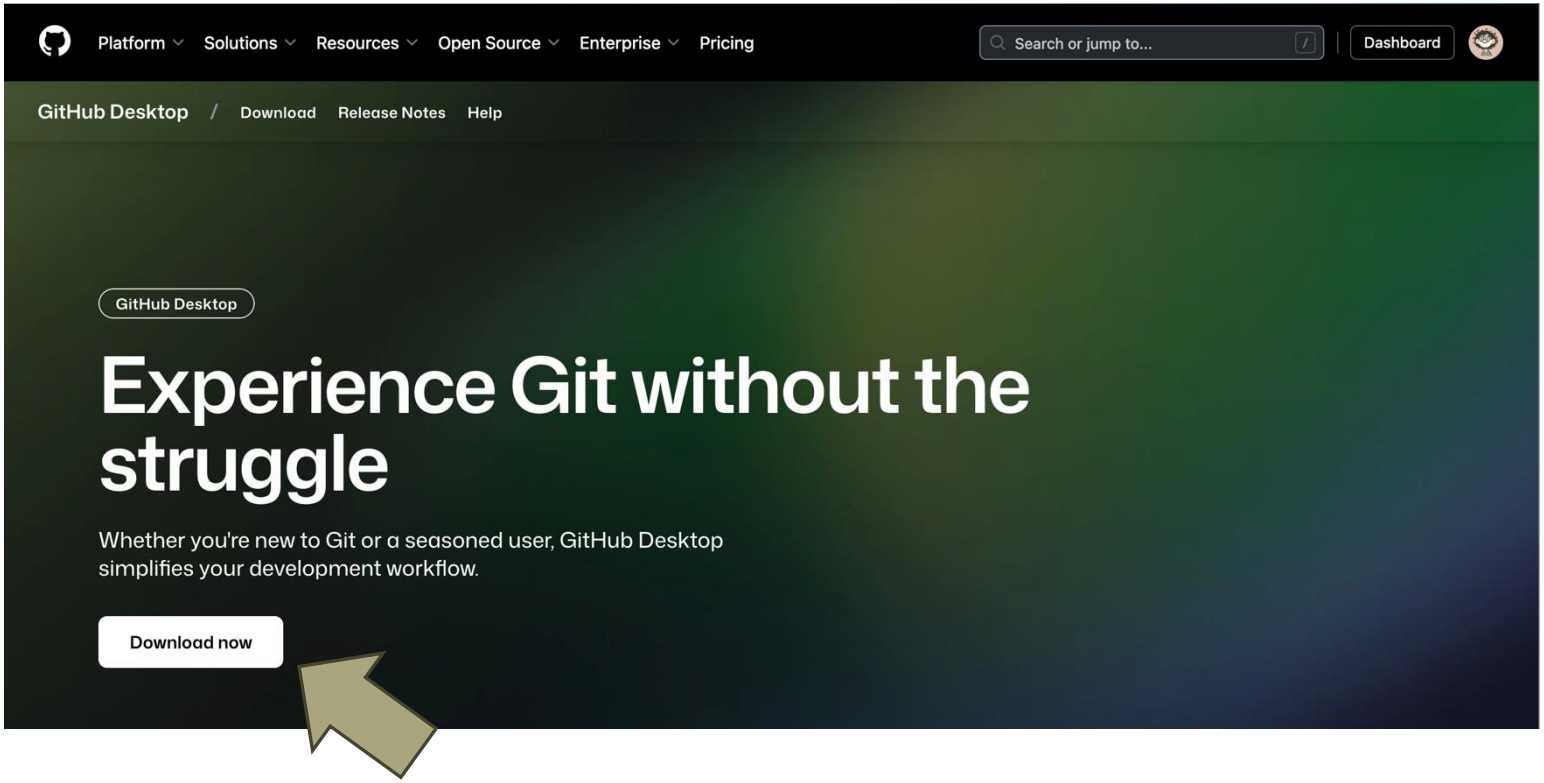
Your handle is NoCh-Git.

Please enter your name and handle here:

https://shorturl.at/OT035

INSTALL GITHUB DESKTOP (OPTIONAL)

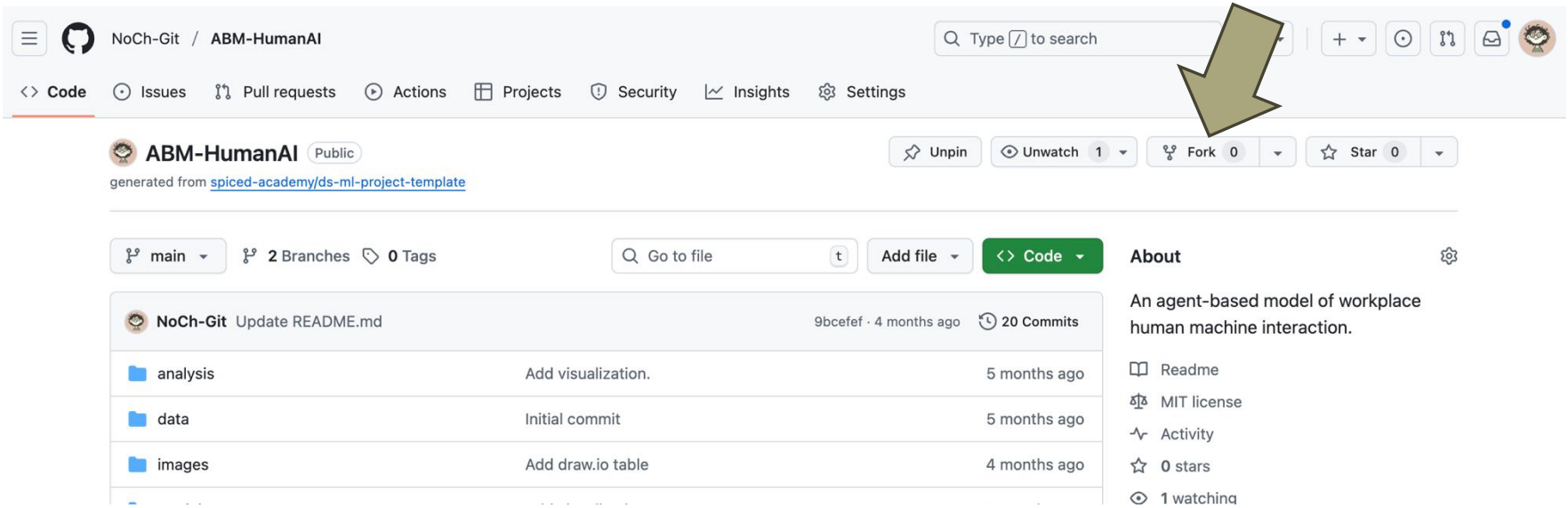
https://github.com/apps/desktop



FORK REPO OF TODAY

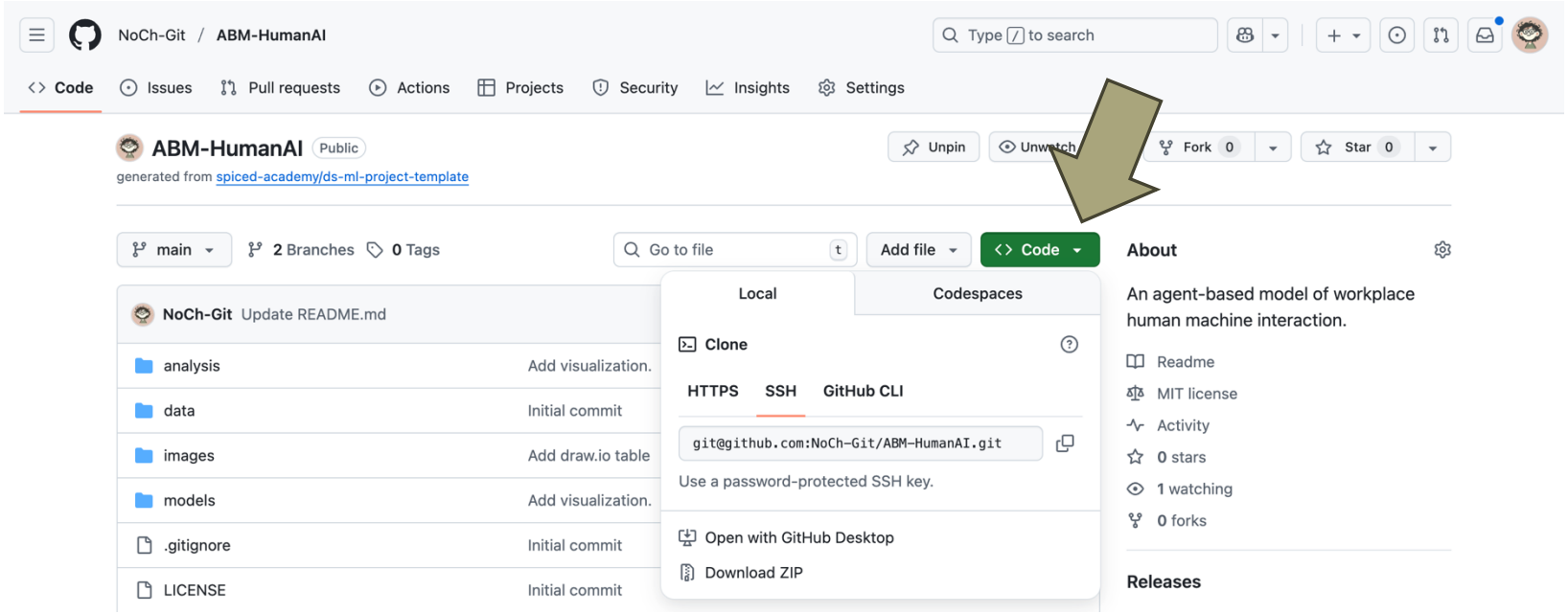
Forking a repo would create a copy of that repo for you that you can play with.

Choose yourself as the owner and untick the “Copy the main branch only” box.



CLONE THE COPY OF REPO TO YOUR MACHINE USING GITHUB DESKTOP OR CLT

You need to have a local copy of the Python notebooks.



JOIN COURSE ORGANIZATION

After you add your handle, I will invite you to join the course organizion.

https://github.com/AppliedNLP-SRH

This is where you will keep your project repo to be evaluated.

WHAT IS DATA?

Data = recorded observations or measurements.

Foundation for information, knowledge, and decision-making.

Examples: temperature readings, purchase records, images, social media posts.

"Big Data" → often described by the **3 Vs**:

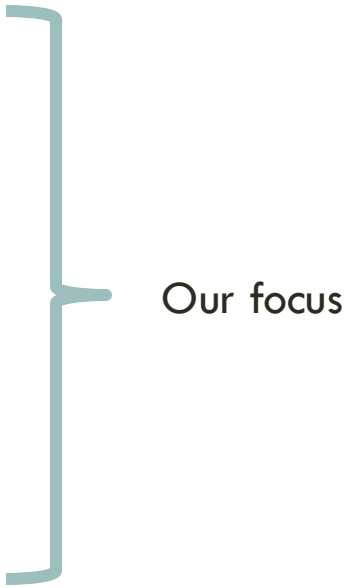
- **Volume** – large amounts
- **Velocity** – fast generation
- **Variety** – many formats

TYPES OF DATA

Type	Description	Examples
Structured	Organized in fixed rows/columns	Excel tables, SQL databases
Semi-Structured	Flexible but still has some structure	JSON, XML, web logs
Unstructured	No fixed format	Text, audio, video, images

DATA LIFECYCLE

1. Collection – sensors, forms, apps
2. Storage – files, databases, cloud
3. Processing – cleaning, transforming
4. Analysis – statistics, visualization
5. Decision – reports, dashboards
6. Archival – backup, deletion



DATABASES

Organized systems to store and retrieve data.

Two main families:

- **Relational Databases** – structured tables with schema (SQL).
- **NoSQL Databases** – flexible storage (key-value, document, graph).

Used for day-to-day business transactions.

Examples:

Relational → MySQL, PostgreSQL

NoSQL → MongoDB, Cassandra, Neo4j

DATA WAREHOUSE

A **data warehouse** is usually built from many databases.

If databases are notebooks for daily operations, the data warehouse is the bookshelf where we organize summaries and reports.

DATA LAKE

A **data lake** is even broader; it stores *all* raw data before it's cleaned or structured.

May contain structured data (from databases), semi-structured (logs, JSON), or unstructured (images, video).

Data from the lake can later be transformed into a warehouse.

If the warehouse is the organized library, the lake is the big storage room where all kinds of information arrive before sorting.

LETS CHECK OUR SQL KNOWLEDGE

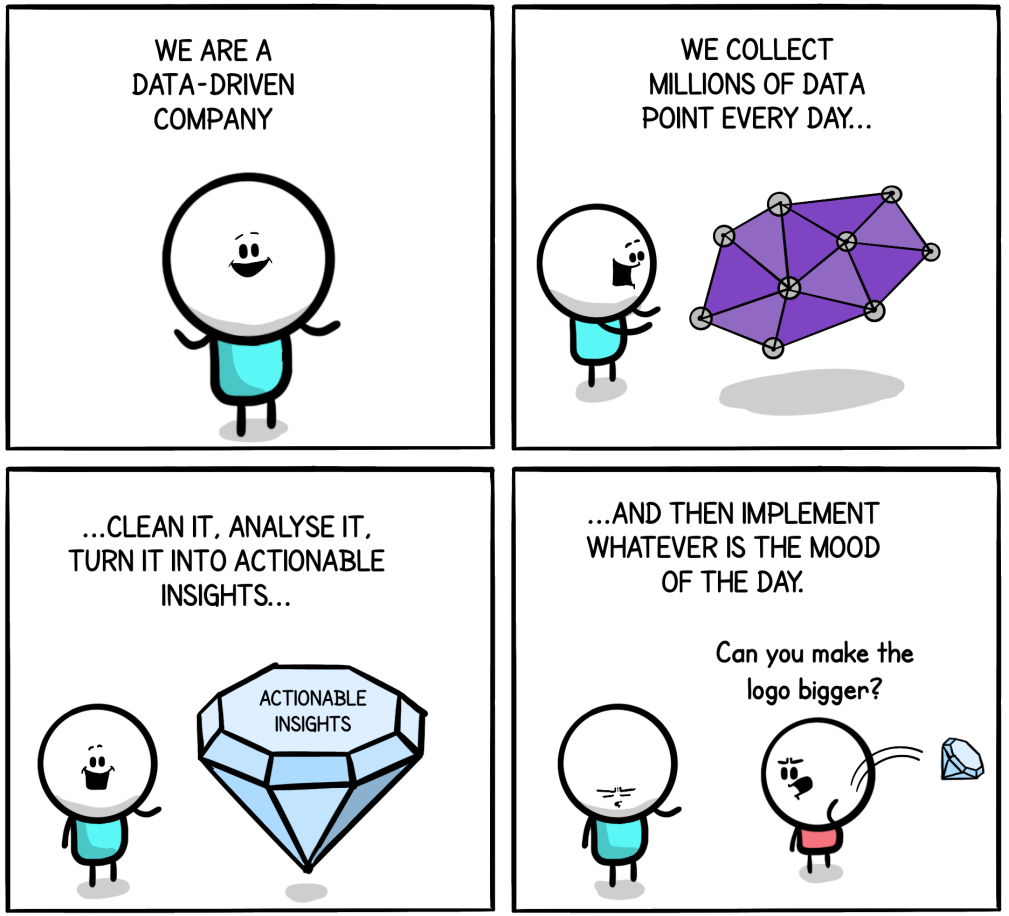
WHAT IS BUSINESS INTELLIGENCE (BI)?

- Processes and tools for turning data into actionable insights
- Includes data visualization, reporting, and decision support
- Bridges technical data work and business strategy

KPI: A **KPI (Key Performance Indicator)** is a *specific, measurable value* that indicates how well a company, department, or project is meeting its strategic or operational goals.

WHY IT MATTERS

- Data-driven decisions improve efficiency and innovation
- Big Data enables personalization, automation, and forecasting
- BI empowers all roles to access and interpret insights



workchronicles.com
follow on Instagram / Twitter / Facebook

REAL-WORLD EXAMPLES

- Netflix: recommendation systems using Big Data pipelines
- Amazon: real-time inventory and pricing analytics
- HelloFresh: optimization of logistics and menu planning
- Government: open data and smart city dashboards

NOW WE ARE GOING TO DO SOME HANDS-ON EXPERIMENTS WITH PYTHON.

UNTIL NEXT SESSION:

Play some SQL games:

WRAP-UP & EXPECTATIONS

- Be curious, collaborative, and hands-on
- Use provided GitHub and notebooks for exploration
- Ask questions and connect topics to your interests