# Functional Programming with Scala

Project Title: Scala Project

Group Name: Group11

Léon BOUDIER

IG5_Polytech

November 19, 2024

# About the Project

- **Your Specific Contribution:**
  - Get data from Neo4j, Print the output of the query in a text file (query: top 5 impactScores for each year) (years: 2024, 2023)

- **Methodology or Approach:**
  - Set Up Spark  Neo4j: Install the Spark-Neo4j connector. Configure SparkSession with Neo4j details (URL, username, password).
  - Sample Data in JSONL: Create a JSONL file with CVE IDs, descriptions, and impact scores.
  - Push Data:
    - Read JSONL with spark.read.json().
    - Create CVE, Description, and ImpactScore nodes.
    - Insert into Neo4j using the Neo4j-Spark connector.
  - Define Relationships: Create relationships (HAS) between CVE and Description, and CVE and ImpactScore.
  - Query: Use Cypher to query the nodes and relationships in Neo4j.

# Challenges and Learning

- **Challenges Faced and Their Resolutions:**
  - Handling large datasets efficiently.
  - Ensuring correct relationship creation between nodes.
  - Optimized Spark configurations for better performance.
  - Used proper Cypher queries for defining relationships.
- **Learning Gained:**
  - Gained experience in integrating Spark with Neo4j.
  - Learned how to manage data flow and relationships in graph databases.
- **Future Improvements:**
  - Could improve data processing speed by tweaking Spark settings.
  - Implement more complex relationships and data models.