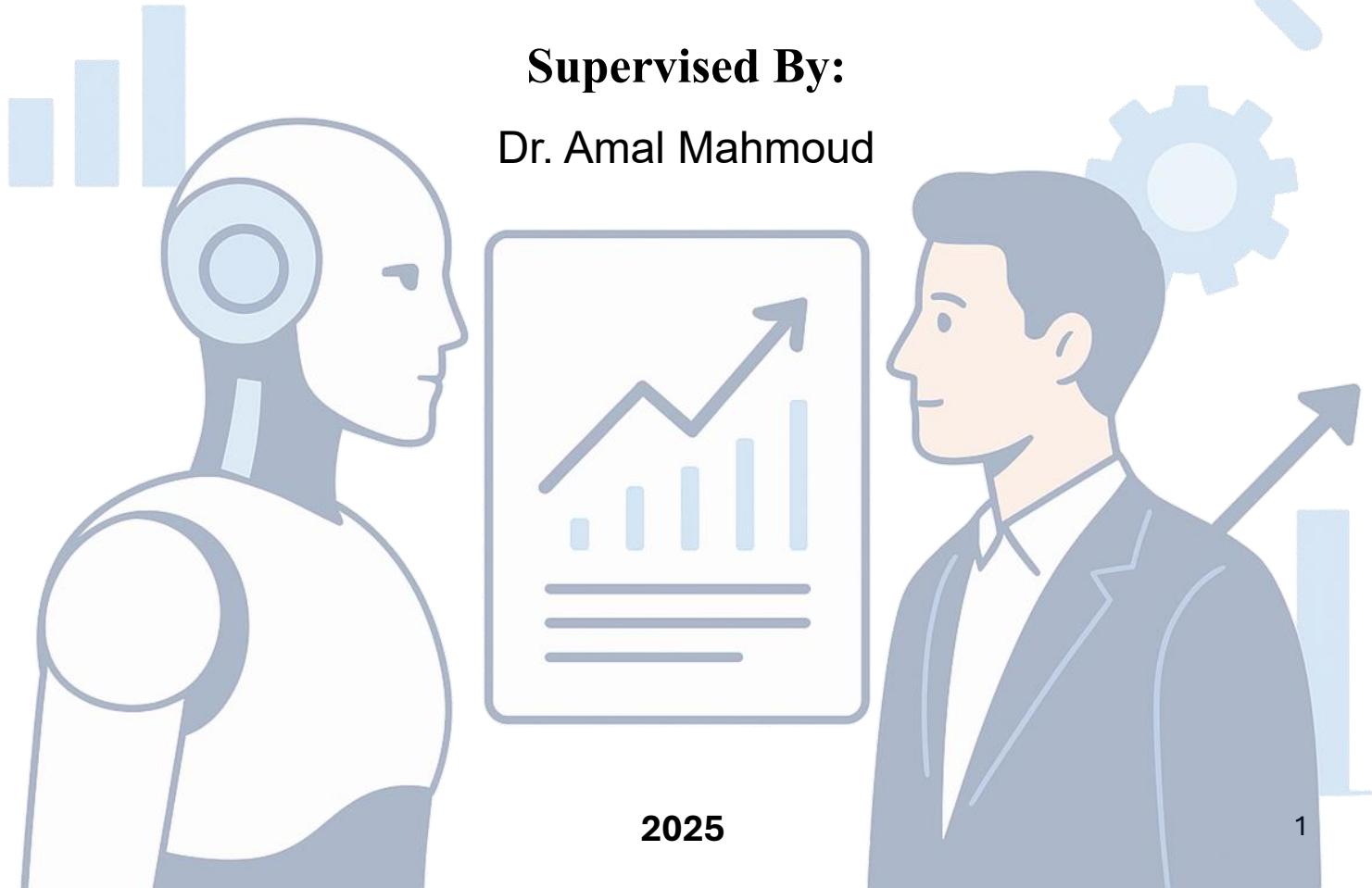# AI Jobs Analysis

Team Members:
- Abdelrhman Mamdouh Abdelhamid Ibrahim
- Ahmed Gamal Mahmoud Ismail
- Ammar Yasser Gaber
- Ethar Emad Abdelmoaty El-Maghraby
- George Mark Youssef Yacoub
- Raed Taha Erfan Mohamed

Supervised By:

Dr. Amal Mahmoud

2025

Index

1. Abstract

This project presents a comprehensive analysis of the global AI job market using a dataset of approximately 30,000 job postings collected from Kaggle. The study investigates key factors influencing AI employment trends, including salary patterns, required skills, company characteristics, work modes, and experience levels. A complete data analysis pipeline was applied, beginning with data preprocessing, relational restructuring, and exploratory analysis using SQL, Python, Excel, Power BI, and Tableau.

Through systematic exploration and interactive dashboards, the project identifies the most in-demand technical skills—such as Python, SQL, and TensorFlow—and highlights notable salary trends across job roles, industries, countries, and work environments. The analysis reveals that remote positions and certain engineering-focused roles consistently offer higher compensation. Additional insights examine the impact of education, experience level, and company size on hiring and salary decisions.

The findings contribute to a clearer understanding of current AI job-market dynamics and provide actionable recommendations for job seekers, educators, and organizations. This work serves as a practical reference for individuals planning AI career paths and for institutions aiming to align training programs with real-world industry demand.

2. Introduction

In recent years, the rapid advancement of artificial intelligence has reshaped the global job market, creating new opportunities while transforming traditional roles. As organizations increasingly integrate AI-driven systems into their operations, the demand for specialized skills continues to rise. Understanding these changes has become essential for students, researchers, and industry leaders who aim to navigate the evolving workforce.

This project focuses on analyzing job postings related to the field of artificial intelligence to identify major trends, required skills, salary patterns, and the most common roles offered in the market. By examining a large dataset of AI-related jobs, the project aims to provide a clear picture of the current landscape and highlight the factors influencing job availability and hiring practices.

The goal of this analysis is not only to extract meaningful insights but also to support decision-making for individuals planning their career paths and for institutions designing AI-focused training programs. Through data exploration, visualization, and structured analysis, this study contributes to a deeper understanding of how AI is shaping modern employment trends.

Fig.1 shows the raw dataset:

Fig.1: Raw dataset

3. Methodology

3.1 Dataset

The dataset used in this project was obtained from **Kaggle**, a reputable platform that provides publicly available datasets for analytics and machine learning. The dataset is published for educational and research purposes.

The data is provided in **CSV** format, which is widely used for structured datasets.

The dataset contains:

- Number of rows: Approximately 30,000 job postings
- Number of columns: 19 features

This makes it suitable for exploratory analysis, predictive modeling, and statistical insights.

3.2 Data Preprocessing

The preprocessing phase focused on preparing the dataset for analysis and restructuring it to enhance clarity, usability, and analytical efficiency. Since the dataset was clean and did not contain missing values, the preprocessing was centered around data refinement, renaming, transformation, and normalization.

Verification of Data Quality

An initial inspection of the dataset confirmed that:

- No missing values were present.
- All existing data types were appropriate for the variables. Therefore, no imputation or data type conversion was required.

Removing Irrelevant or Unnecessary Columns

The following columns were removed because they did not contribute significantly to the analysis or modeling goals:

- salary_currency
- posting_date
- application_deadline
- job_description_length
- benefits_score

These fields were either redundant, weakly correlated with the analysis objectives, or not required for the structural remodeling of the dataset.

Renaming and Standardizing Categorical Attributes

To improve clarity and enhance the readability of categorical variables, the following columns were transformed:

- company size: S → Small, M → Medium, L → Large
- experience level: EN → Entry, MI → Mid, SE → Senior, EX → Executive
- employment type: CT → Contract, FL → Freelance, PT → Part-time, FT → Full-time
- remote ratio: 0 → On-site, 50 → Hybrid, 100 → Remote

This normalization ensures consistent interpretation across the dataset.

Outlier Analysis

A review of numerical variables highlighted some high salary values. However, these were determined to be **valid observations**, especially since higher salaries corresponded to higher years of experience.
Therefore:

- These values were not removed
- They were treated as legitimate data points, not statistical outliers.

Splitting the Dataset into Multiple Logical Tables

To improve data organization and support a relational structure suitable for analysis and potential database modeling, the dataset was split into five separate tables, each representing a specific dimension:

1. Fact Table: (company_id, employee_id, job_id, salary_usd, remote_ratio, education_required, years_experience, industry)

2. Company Table: (company_id, company_name, company_location, company_size)

3. Employee Table: (employee_id, experience_level, employment_type, employee_residence)

4. Job Table: (job_id, job_title)

5. Skills Table: (job_id, Skill_Count)

Fig.2: Power Query, Data Preprocessing

Creating Relationships Between Tables (Data Modeling)

Keys were used to link the tables through a star-schema–like structure. This design supports cleaner analysis, reduces redundancy, and allows efficient querying across job, company, employee, and skills dimensions.
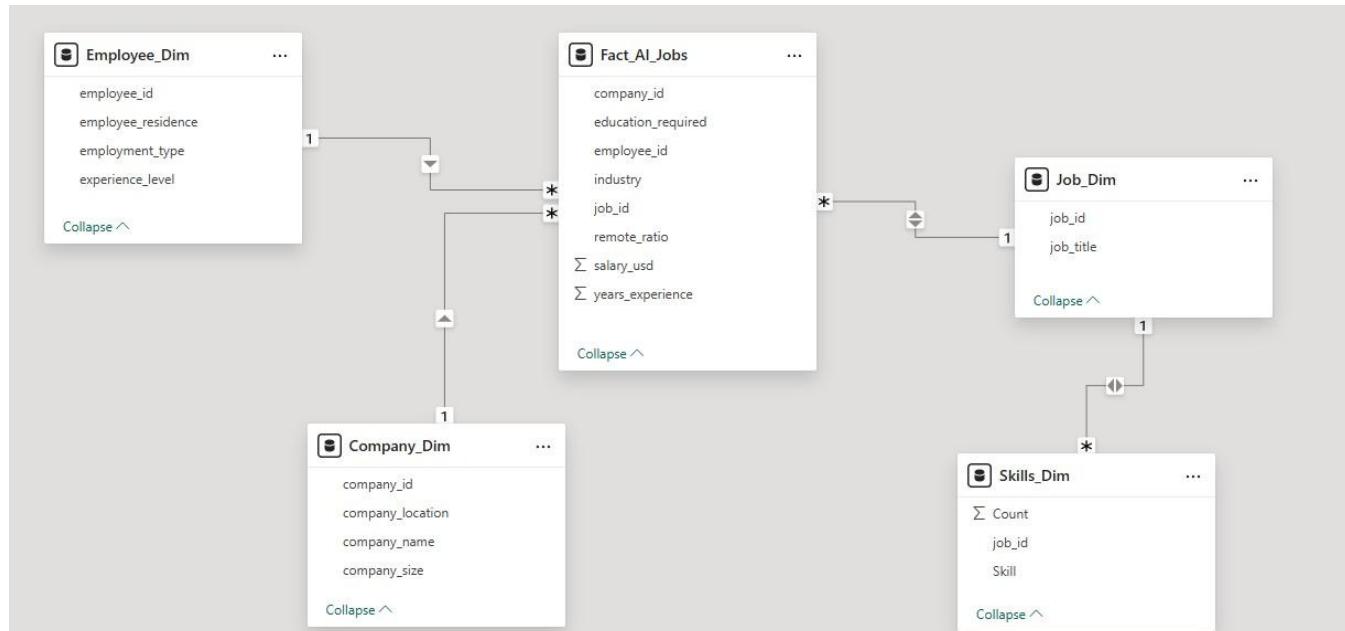
Fig.3: Data Modeling

3.3 Tools & Technologies

A combination of analytical, statistical, and visualization tools was employed throughout the project to support data preparation, transformation, exploration, and insight generation. Each tool played a specific role within the analysis pipeline, helping ensure data quality and enabling clear, reliable results.

Microsoft Excel

Excel was used during the initial data-preparation phase, where it supported data cleaning, restructuring, and organizing the dataset into meaningful segments. Through Power Query, complex transformations and

early ETL processes were performed efficiently. Power Pivot enabled advanced data modeling and relationship handling inside Excel, while Pivot Tables provided rapid summarization. An early exploration dashboard was also developed to validate data integrity and identify emerging patterns.
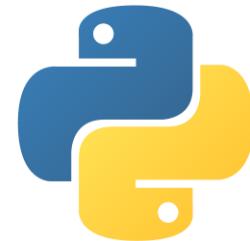
MySQL (SQL)

MySQL served as the primary relational database environment. The dataset was normalized and organized into structured tables to support efficient querying and storage. SQL was used to perform cleaning, preprocessing, and to generate aggregated views that offered analytical insights and prepared the data for downstream processing

Python

Used for advanced analytical processing, including data cleaning, preprocessing, and splitting the dataset programmatically. Python enabled the extraction of statistical insights and supported the creation of reproducible visual dashboards using Pandas, NumPy, Matplotlib, and Seaborn.

Power BI

Power BI was utilized to design interactive and user-friendly dashboards that present analytical findings clearly. The tool enabled data modeling, splitting and shaping the data, and performing custom calculations using DAX. Through Power Query, additional ETL operations were performed. The final Power BI dashboards included filters, drilldowns, KPIs, and comparative charts that enhanced the interpretability of the insights.

Tableau

Tableau was used to create high-quality visual dashboards with strong storytelling capabilities. It supported data splitting, calculated fields, and building visually compelling layouts that highlighted trends and relationships. The Tableau dashboards offered interactive elements such as filters, parameters, and multi-view compositions, providing a smooth exploration experience and clear interpretation of the dataset.

3.4 Method / Approach

This project followed a complete and structured Data Analysis Process to ensure that all analytical steps were executed systematically, leading to accurate, reliable, and actionable insights. The methodology consisted of the following phases:

Understanding the Problem

The analysis began with defining the main objectives of the project, which included examining salary patterns, identifying top AI skills, understanding industry demand, and analyzing the impact of experience level, company size, and remote work on job characteristics.

This step ensured that all subsequent work remained aligned with the project goals and overall business context.

Exploratory Data Analysis (EDA)

A detailed Exploratory Data Analysis was conducted using Python, SQL, Excel, Power BI, and Tableau to understand the structure of the dataset and reveal key relationships.
This phase involved:

- Analyzing distributions of numerical attributes (salary, years of experience, etc.)

- Studying correlations between skills, roles, industries, and compensation

- Detecting patterns in remote work and employment types

- Visualizing trends through charts and dashboards to support initial insight generation

EDA guided the later analytical decisions and highlighted the most significant variables

Business Rules Application

Domain knowledge was integrated to validate and interpret the data realistically. Examples include:

- Mapping experience levels to natural ranges (Entry → low years of experience, Senior → high years)

- Interpreting remote ratio as work mode (On-site, Hybrid, Remote)

- Verifying that high salaries align with higher experience or advanced technical skills

- Ensuring consistency between company size and job expectations

This ensured that all analyses reflected real-world job-market logic.

Insight Generation

After preparing and transforming the data, insights were generated using statistical techniques and visual analytics.

This included:

- Identifying top AI skills and their frequency

- Analyzing salary variations across roles, countries, industries, and experience levels

- Studying hiring patterns across industries

- Examining employment types and remote-work adoption

Insights were presented using interactive dashboards in Power BI and Tableau, supported by SQL queries, Python scripts, and Excel models. The results were cross-validated across tools to ensure accuracy and consistency.

4. Results and Discussion

This project analyzed a large dataset of AI-related job postings covering more than 20 job roles, 20 countries, and thousands of employees across different company sizes. The goal was to understand hiring trends, salary patterns, required skills, and how factors such as experience level, education, or work mode affect compensation.

4.1 Skills Demand

The analysis reveals that the AI job market is strongly driven by technical expertise.

The top three skills consistently required across most roles were:

- Python
- SQL
- TensorFlow

These skills appeared far more frequently than any others, confirming that companies prioritize candidates who can build machine learning models, handle data efficiently, and work with deep learning frameworks.

4.2 Salary Patterns

Several important insights emerged from the salary analysis:

- The United States has the highest overall salaries, often significantly above other countries.

- Remote jobs recorded the highest salaries, which was one of the most surprising findings. This suggests global companies are willing to pay premium rates for remote AI talent.

- Employees with a Master's degree had higher average salaries than those with a PhD. This was unexpected, but when using job-specific filters, the pattern changed across roles.

- Salary also varied strongly based on work mode, experience level, and geographic location.

4.3 Company Size and Employment Type

The distribution of company sizes and employment types appeared almost equal across the dataset, which indicates a balanced representation.

However, deeper analysis showed:

- Full-time roles are dominant in medium and large companies, reflecting more stable hiring structures.

- Contract jobs are more common in small companies, possibly due to budget limitations or flexible project-based needs.

- Freelance and part-time roles exist but represent a very small percentage of the market.

4.4 Job Roles and Average Salary

Using the dashboard, we calculated the average salary for each job role, including AI Specialist, Machine Learning Engineer, AI Architect, Data Engineer, and others.

Engineering and architecture positions usually ranked among the highest-paying roles, while analyst-type positions showed slightly lower averages.

5. Recommendations

Based on the insights extracted from the analysis, several recommendations can be made:

Prioritize Core Technical Skills

Anyone entering the AI job market should develop strong skills in:

- Python

- Machine Learning / Deep Learning (TensorFlow, PyTorch)

- SQL

These skills consistently appear across almost all high-demand roles.

Gain Expertise in Remote Collaboration

Since remote roles offer high salaries, professionals should build:

- Strong asynchronous communication skills

- Experience with cloud platforms

- A portfolio that demonstrates independent project execution

Build Role-Specific Projects

Salary varies greatly between job titles. To be competitive:

- Focus on creating end-to-end AI or ML projects

- Tailor the portfolio to the specific job (e.g., ML Engineer vs. Data Analyst)

- Include real datasets and measurable results

Continue Learning and Upskilling

The AI field evolves rapidly. Professionals should continuously learn:

- Advanced ML and DL techniques

- Cloud services (AWS, Azure, GCP)

- MLOps and deployment skills

This will greatly increase competitiveness and salary potential.

Considering Global Job Opportunities

Since countries like the U.S. and parts of Europe offer the highest salaries, candidates should consider applying globally, especially for remote roles.

For Companies

Organizations can use these findings to:

- Align salaries based on global standards

- Understand which skills are most valuable

- Prepare for the rising demand for senior expertise in AI

6. References

Kaggle: https://www.kaggle.com/datasets/nourahmed1/ai-job-dataset