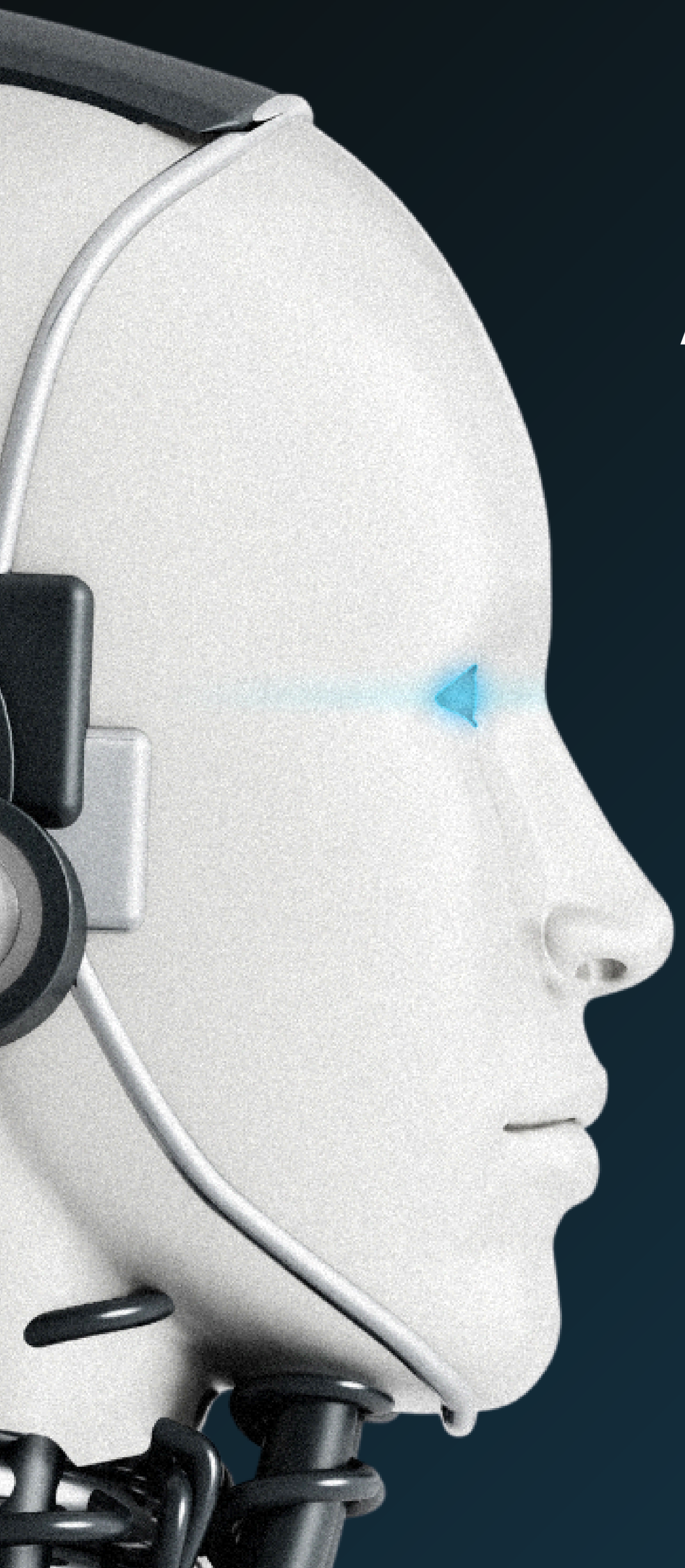


THE IMPACT ON JOBS BY

Artificial Intelligence

Quantitative Data Analytics

Team:

- Sanjil Kokare
- Aum Tailor
- Vaishnavi Chaudhari
- Gaurav Chavan
- Devansh Vashist
- Adarsh Thandayuthapani

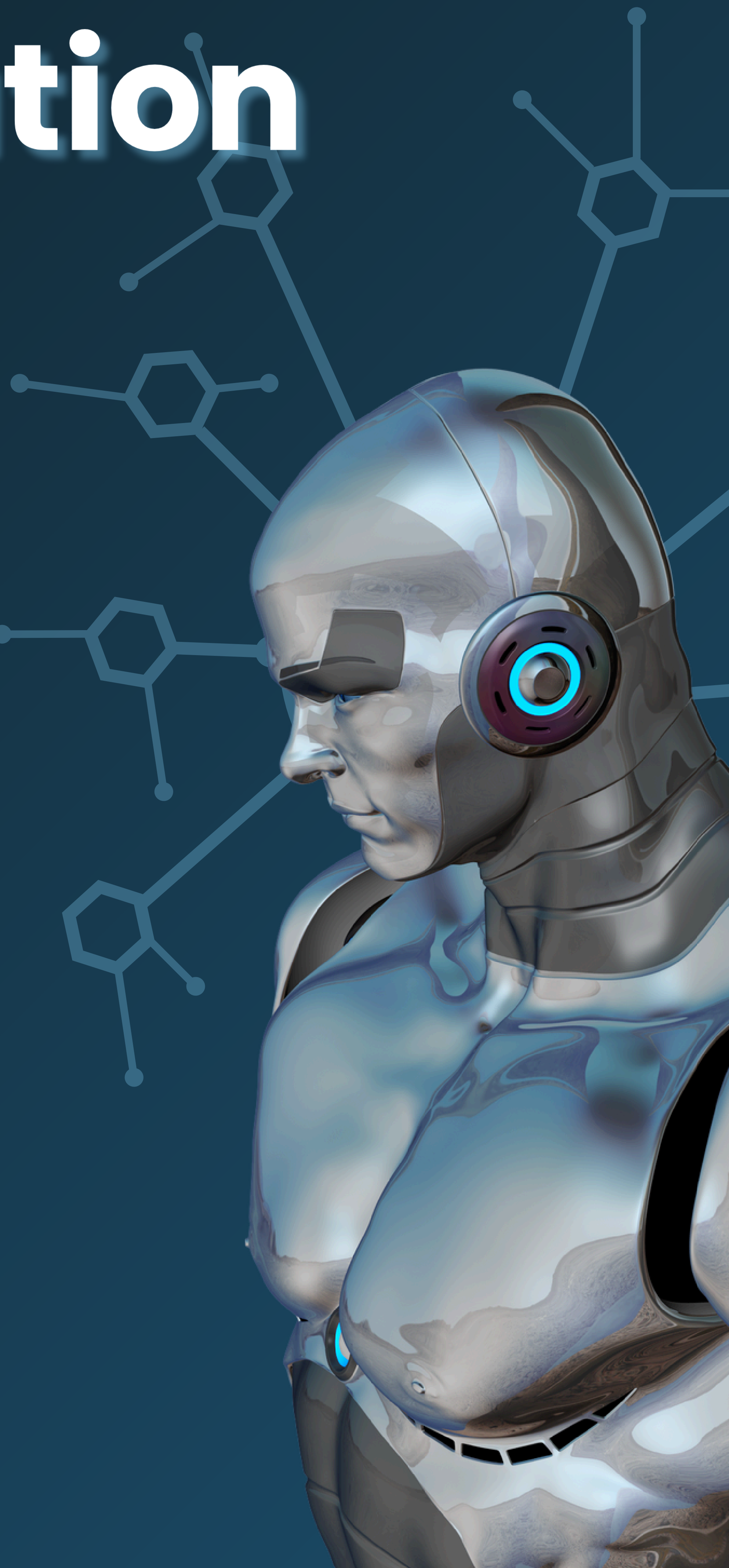
Prof. Dr. Alexander Maximilian Röser

Content

Introduction & Motivation

Are machines replacing jobs or simply changing how we work ?

Artificial Intelligence is transforming the workplace automating tasks, redefining roles, and reshaping industries.
This project explores which jobs are most at risk and why, using real-world data from Kaggle.
We analyze features like task complexity, workload ratios, and AI model usage to uncover patterns.
Our goal is to understand whether jobs are being eliminated or merely transformed.
This analysis helps workers and policymakers prepare for the future of employment.



Research Objectives

- To identify which job roles are most vulnerable to AI automation.
- To analyze how factors like task complexity, model usage, and workload ratio influence AI impact.
- To classify jobs into low, medium, and high AI-risk categories.
- To determine whether jobs are being eliminated or simply transformed.
- To apply regression modeling to predict AI risk based on job traits.

Dataset Overview

📍 Source:
Kaggle – "AI Impact on Jobs" (by UnclePablo)

🧾 Dataset Description:
A comprehensive dataset capturing the estimated AI automation risk across nearly 4,700 job titles.

Key Variables:

- Job Title – Role or profession name
- Tasks – Average number of distinct tasks per job
- AI Models – Types of AI models relevant to each role
- AI Workload Ratio – Proportion of tasks considered automatable
- AI Impact – Estimated probability of AI impact

💡 Goal of Dataset Use:

- To identify which job characteristics (e.g., task complexity, AI model exposure) influence vulnerability to AI-driven disruption.

	Job titiles	AI Impact	Tasks	AI models	AI_Workload_Ratio	Domain
0	Communications Manager	98%	365	2546	0.143362	Communication & PR
1	Data Collector	95%	299	2148	0.139199	Data & IT

Data Preprocessing

Handled Infinite Values

- Replaced infinite values in the AI_Workload_Ratio column with NaN to avoid analysis distortion.

Data Type Conversion

- Converted all relevant columns like Tasks, AI Models, and AI_Workload_Ratio to numeric formats to support statistical analysis.

Missing Value Treatment

- Checked for nulls across all key columns and removed or handled rows with incomplete data.

Consistency Checks

- Ensured uniform naming, removed special characters, and validated data ranges.

```
df['AI_Workload_Ratio'].replace([np.inf, -np.inf], np.nan, inplace=True)

/tmp/ipython-input-9-1948544571.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series thro
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[c


  df['AI_Workload_Ratio'].replace([np.inf, -np.inf], np.nan, inplace=True)
```

```
#check for missing data
for col in df.columns:
    pct_missing = np.mean(df[col].isnull())
    print('{} - {}%'.format(col, round(pct_missing*100)))

Job titiles - 0%
AI Impact - 0%
Tasks - 0%
AI models - 0%
AI_Workload_Ratio - 0%
Domain - 0%
```

Exploratory Data Analysis (EDA)

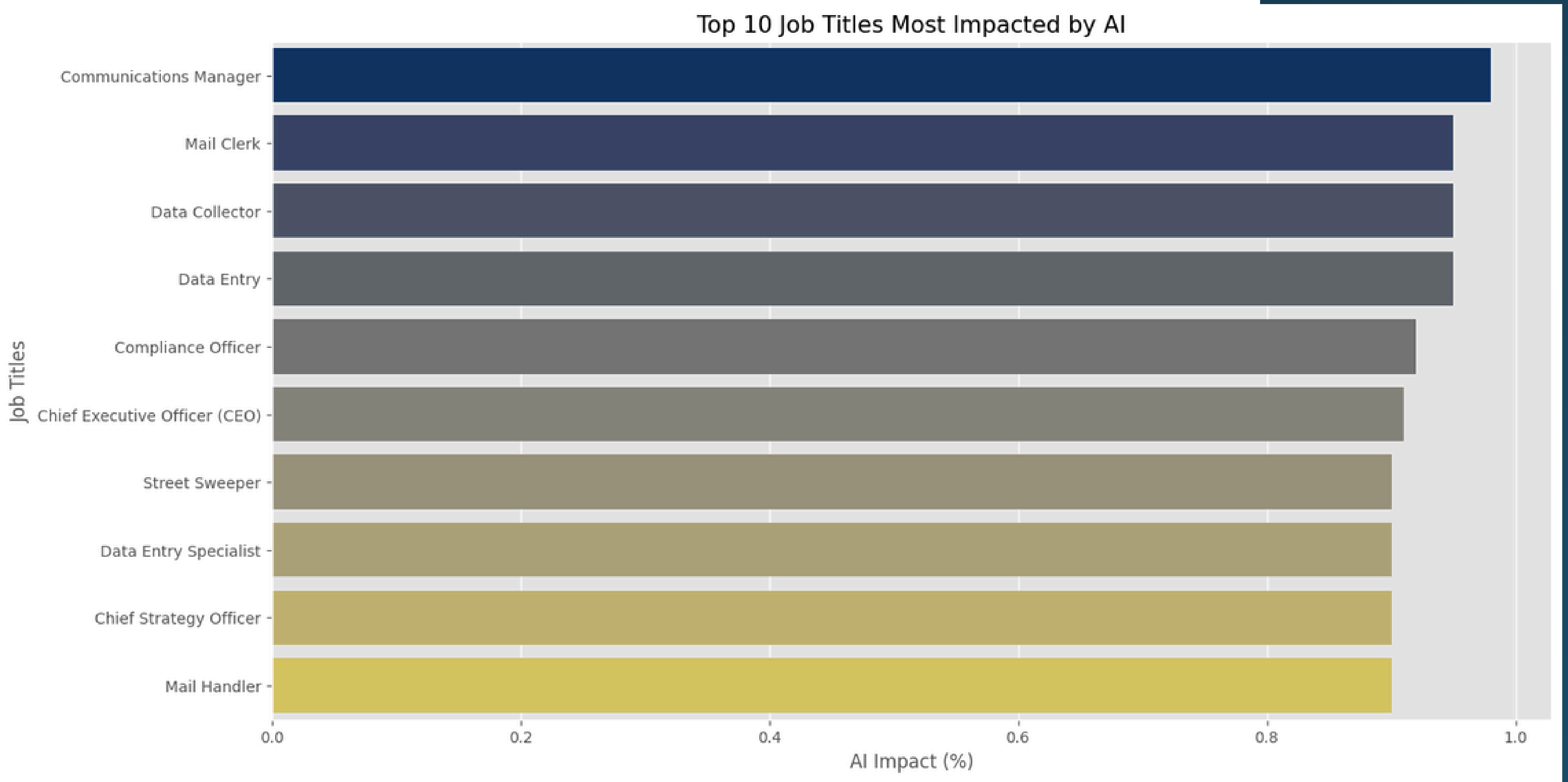
TOP Top 10 Most AI-Impacted Jobs

We sorted the dataset by AI Impact Score to identify the most vulnerable roles:

- Communications Manager
- Data Entry Clerk
- Mail Clerk
- Compliance Officer
- Administrative Services
- (Others as shown in your bar chart)

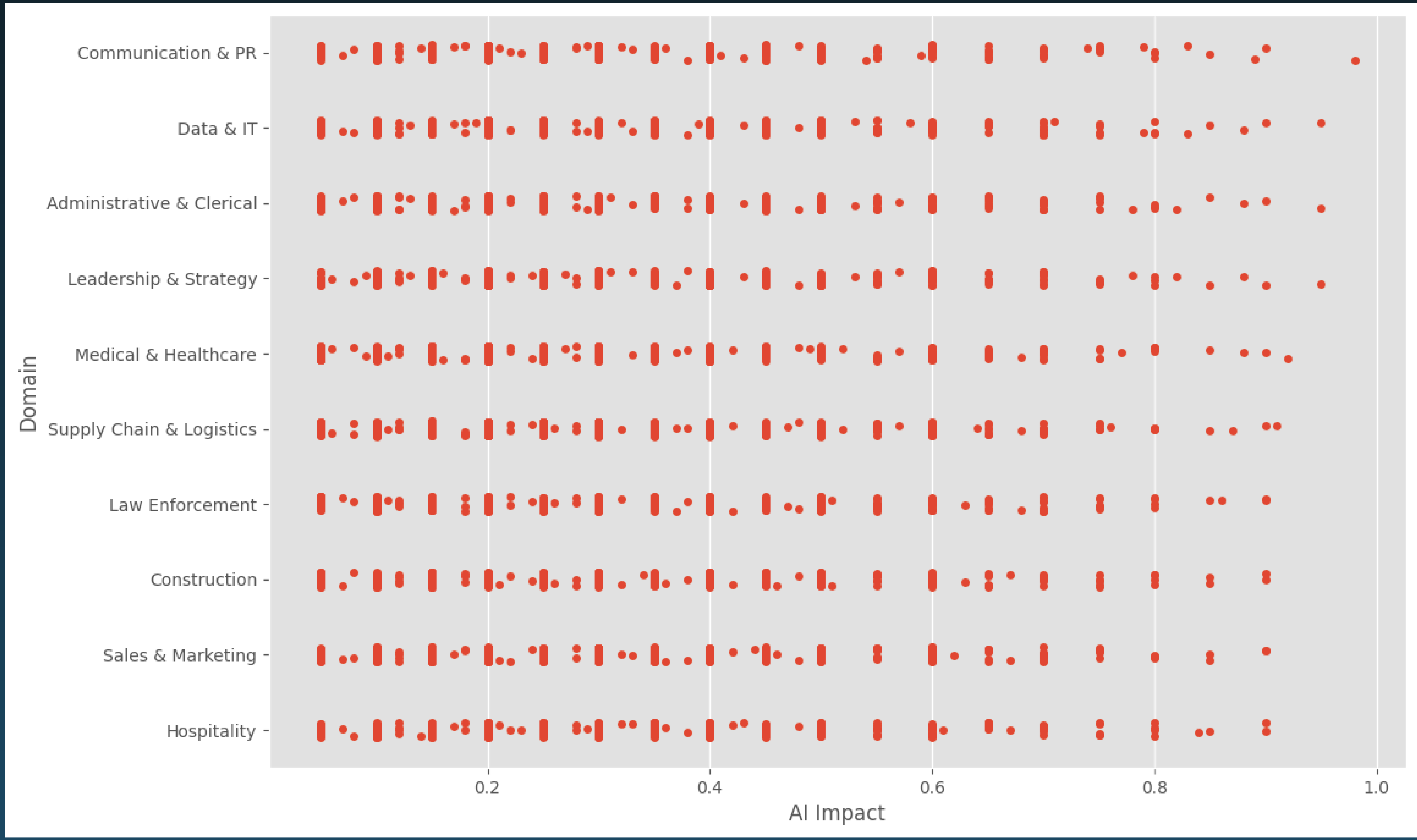
```python
df['AI Impact'] = (df['AI Impact'].str.rstrip('%').astype('int')) / 100
top_10_impacted_jobs = df.sort_values(by='AI Impact', ascending=False).head(10)

# Plotting the top 10 impacted jobs
plt.figure(figsize=(15, 8))
sns.barplot(x='AI Impact', y='Job Titles', data=top_10_impacted_jobs, palette='cividis')
plt.title('Top 10 Job Titles Most Impacted by AI', fontsize=15)
plt.xlabel('AI Impact (%)', fontsize=12)
plt.ylabel('Job Titles', fontsize=12)
plt.show()
```



```python
#This stripplot shows us which domains have the most jobs closest to AI Impact of 100%

sns.stripplot(x="AI Impact", y="Domain", data=df)
```



🔥Correlation & Heatmap

🧪 Objective:

- To understand how different features such as Tasks, AI Models used, and AI Workload Ratio relate to AI Impact.

Steps Taken:

- We used the Pearson correlation method to compute pairwise correlations between numeric features.
- Visualized the correlation matrix using a heatmap.

📊 Key Findings:

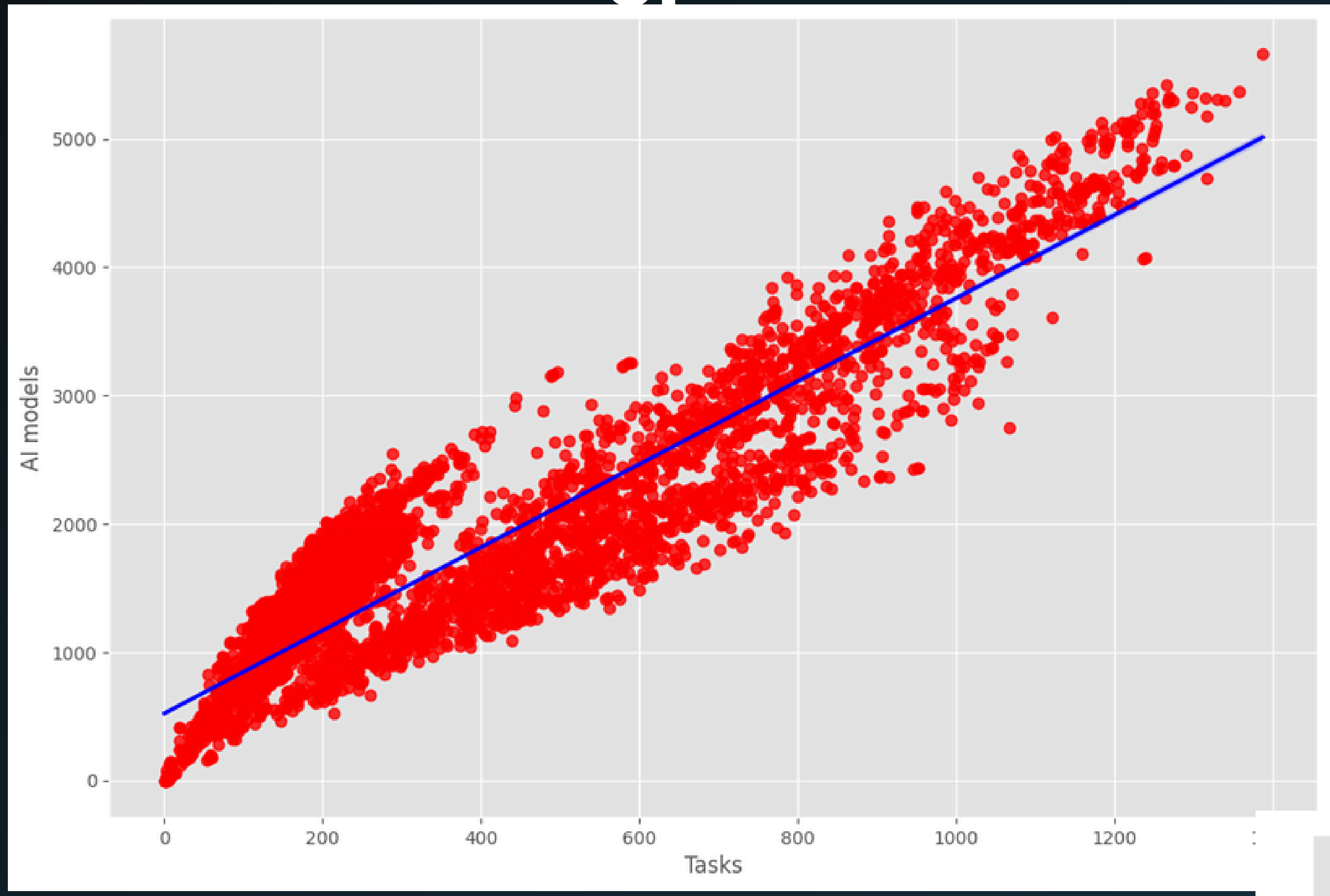
🟢 AI Workload Ratio had a moderate positive correlation with AI Impact.
🔴 Tasks and AI Models showed weak or even negative correlation with AI Impact.

🧠 Insight:

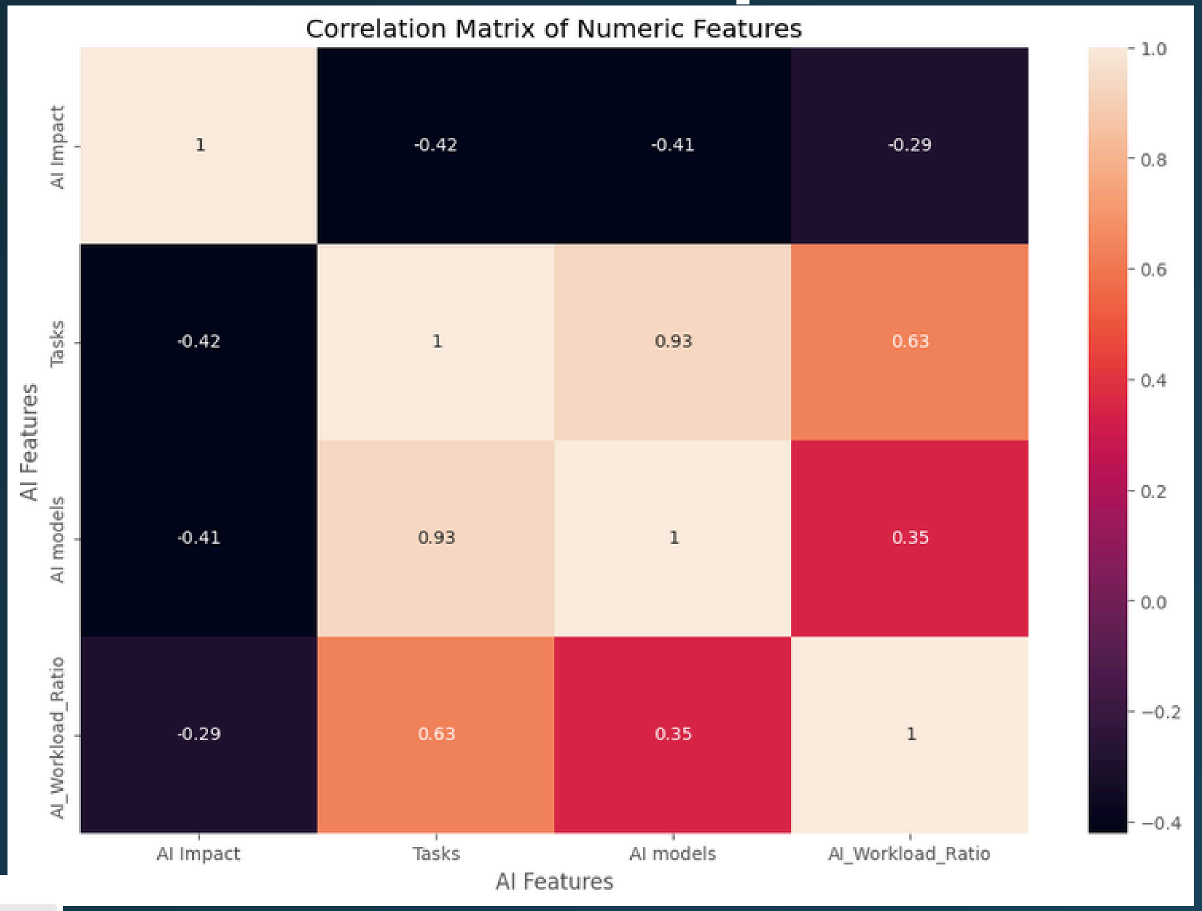
- Automatability is more dependent on the proportion of work that AI can take over, not just the variety of tasks or the AI tools being used.

🔥 Correlation & Heatmap

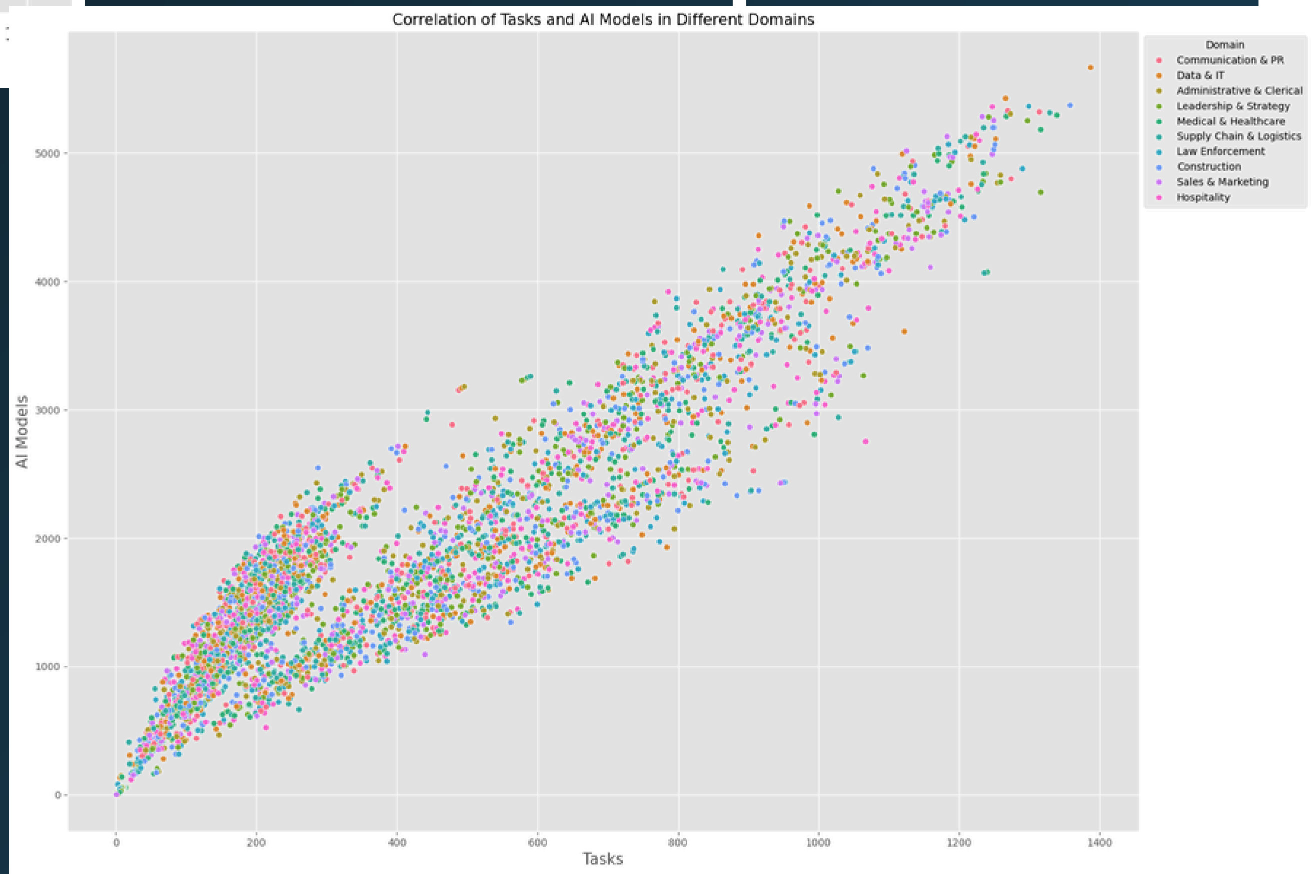
regplot



heatmap



scatterplot



AI Risk Classification

Slide Title: Categorizing Jobs by AI Risk
🧩 What You Did:
Manually segmented jobs into 3 risk levels based on their AI Impact scores:

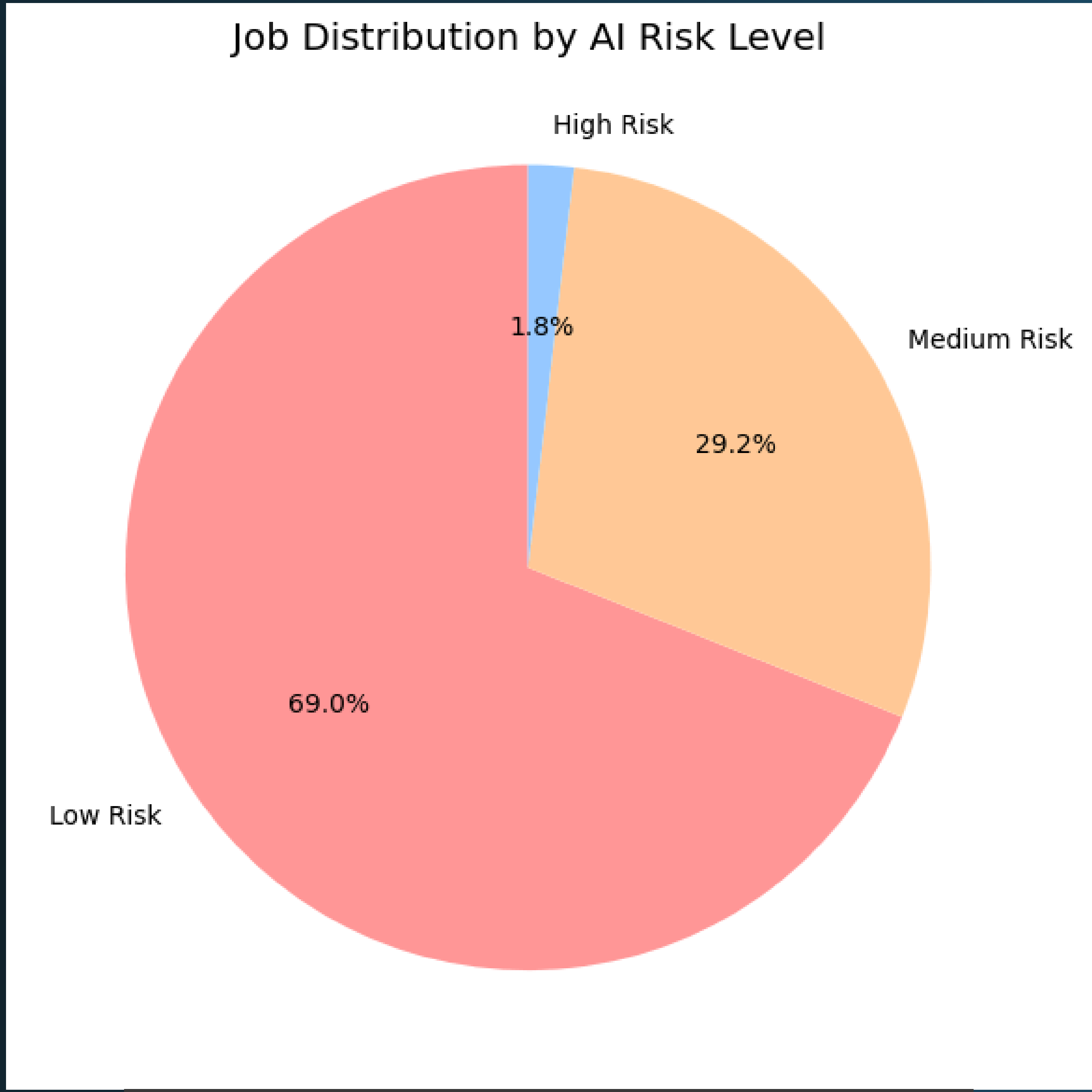
- High Risk: AI Impact > 0.75
- Medium Risk: AI Impact between 0.4 and 0.75
- Low Risk: AI Impact < 0.4

📈 Visual:
Pie chart showing the distribution:

- Low Risk → ~69%
- Medium Risk → ~29%
- High Risk → ~2%

💡 Insight:

- Majority of jobs fall under low to medium risk, indicating transformation over elimination.
- Very few jobs are in the high-risk category — mostly repetitive or automatable tasks.




```
AI_Risk_Level
Low Risk       3245
Medium Risk    1375
High Risk        86
Name: count, dtype: int64
```

Job Status Classification

📐 Classification Logic:
You defined job status based on AI Impact score:

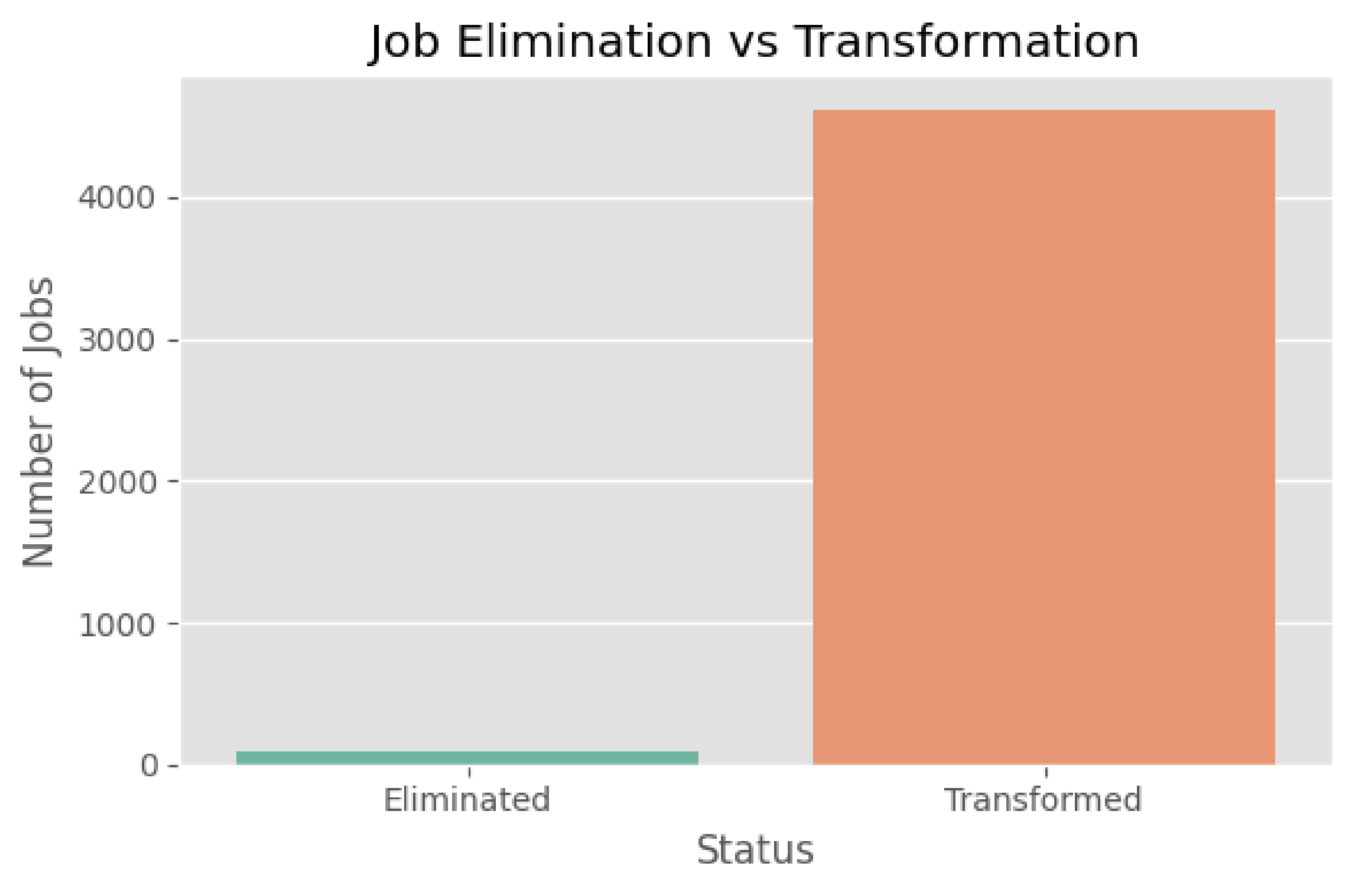
- 🟥 Eliminated → AI Impact > 0.75
- 🟩 Transformed → AI Impact ≤ 0.75

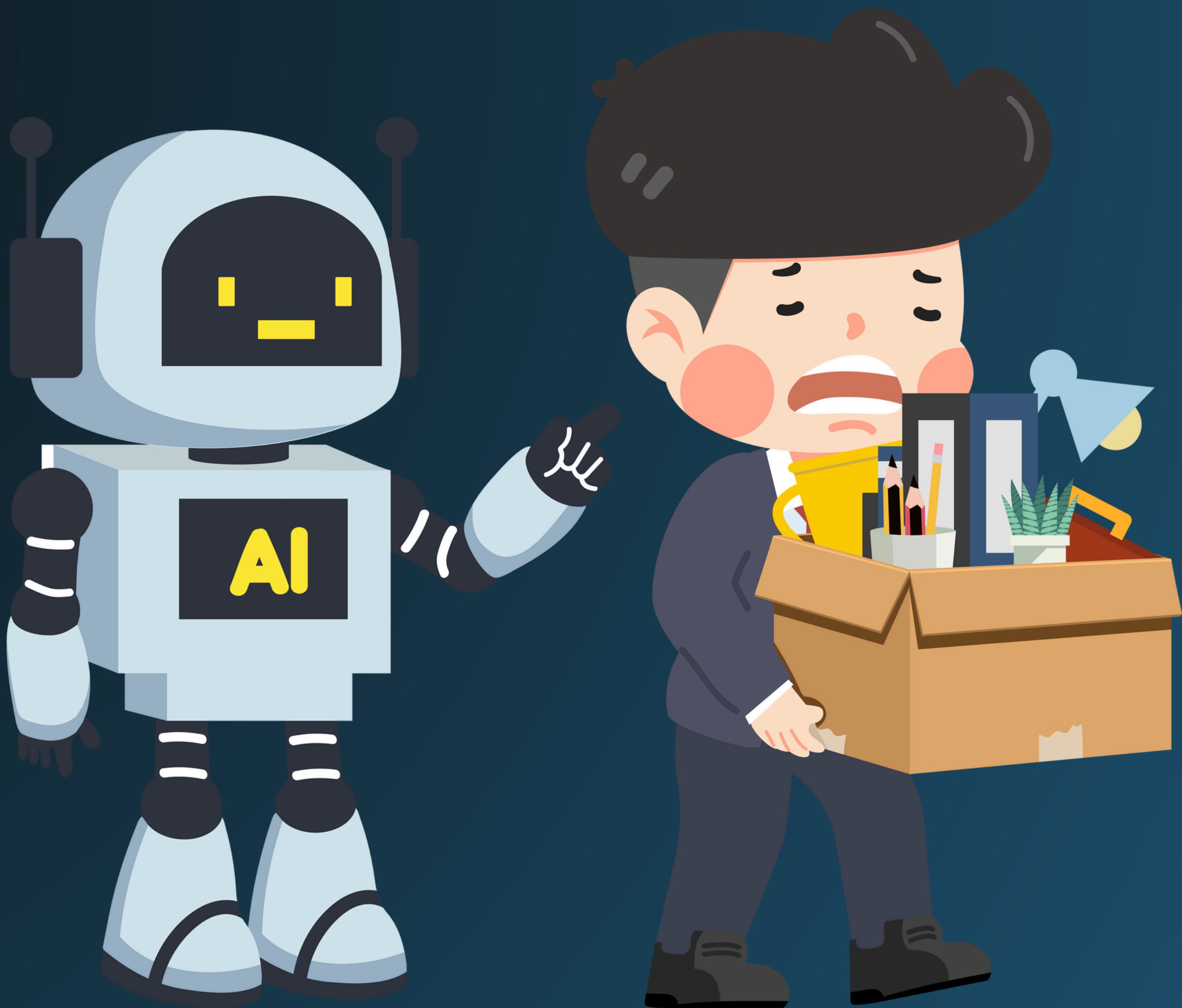
📊 Visuals Used:

- ✅ Countplot showing number of jobs in each category
- ✅ Scatterplot (Tasks vs. AI Impact), color-coded by job status

```
df['Job_Status'].head()
```

	Job_Status
0	Eliminated
1	Eliminated
2	Eliminated
3	Eliminated
4	Eliminated





📈 Predictive Modeling

🧪 Model Used:
✅ Linear Regression (from Scikit-learn)

🧮 Features:

- Tasks
- AI Models
- AI Workload Ratio

🎯 Target Variable: AI Impact

📊 Key Finding:

- The most influential predictor was the AI Workload Ratio.
- Tasks and AI Models had less explanatory power individually.

Insight:

- The more of a job's workload that AI can handle, the higher the predicted risk — regardless of how many tasks or AI systems are involved.

```python
from sklearn.linear_model import LinearRegression


# Select relevant columns
features = ['Tasks', 'AI models', 'AI_Workload_Ratio', 'AI Impact']
df_subset = df[features].dropna()

# Define X and y
X = df_subset[['Tasks', 'AI models', 'AI_Workload_Ratio']]
y = df_subset['AI Impact']

# Train the model
model = LinearRegression()
model.fit(X, y)

# View coefficients
print("Intercept:", model.intercept_)
print("Coefficients:", model.coef_)
print("Feature names:", X.columns.tolist())
```

```
Intercept: 0.5837935249621249
Coefficients: [ 2.92303988e-04 -1.27837929e-04 -8.06872082e-01]
Feature names: ['Tasks', 'AI models', 'AI_Workload_Ratio']
```

✅ Results Summary

📊 Key Takeaways from the Analysis:

🔹 Only about 2% of jobs are at high risk of full elimination by AI.

🔹 A majority (~69%) of roles fall into the low-risk category, indicating AI is more likely to support than replace them.

🔹 Repetitive, low-task, administrative roles are most vulnerable to automation.

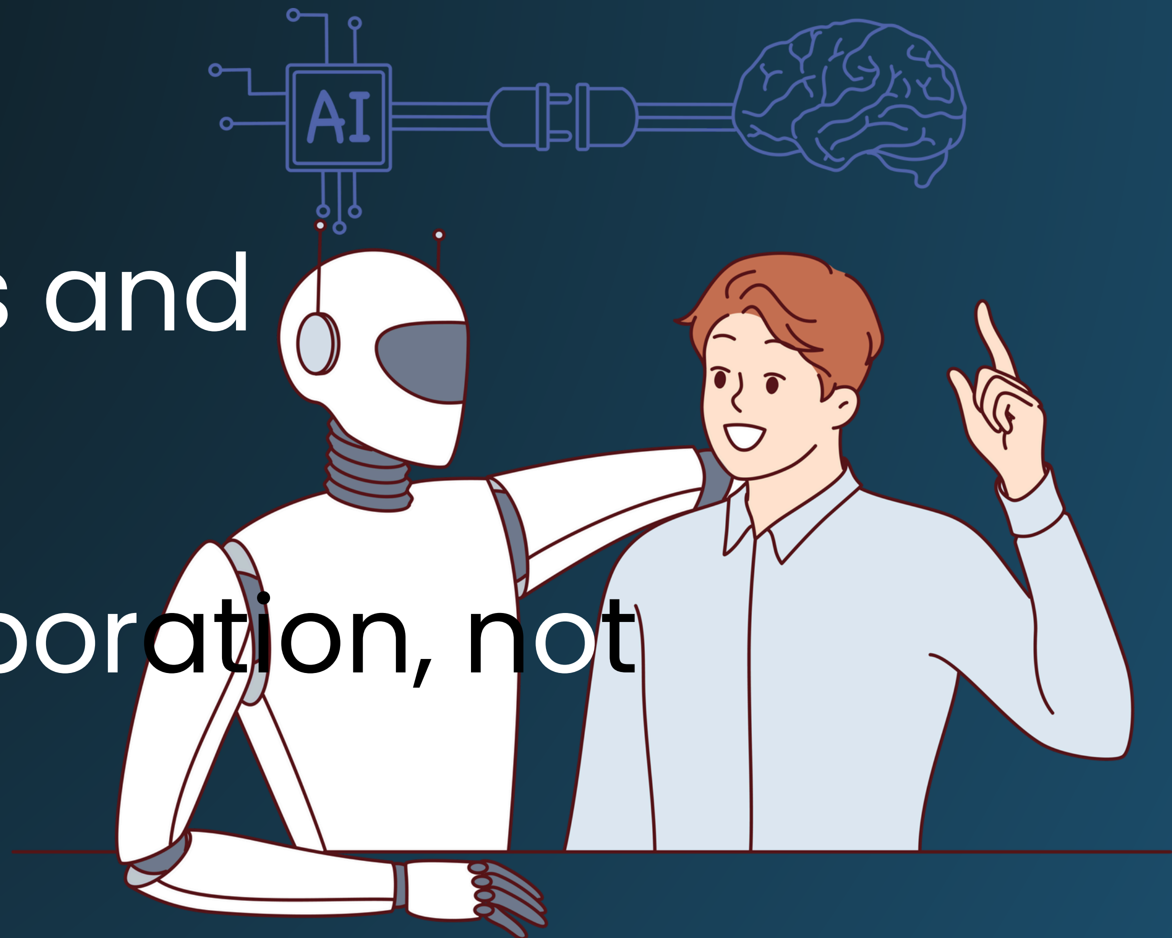
🔹 AI Workload Ratio is the most reliable predictor of AI impact, not task count or number of AI models.

🔹 Most jobs are being reshaped, not removed — AI is transforming work, not taking it away completely.

Conclusion

Rather than fearing AI, we should adapt and evolve with it.

- Artificial Intelligence is not a job destroyer — it's a powerful tool reshaping how we work.
- The majority of jobs studied are more likely to be transformed than eliminated.
- Roles involving repetitive and rule-based tasks are most vulnerable to automation.
- Data-driven classification helps policymakers and workers prepare proactively.
- The future of work will rely on human-AI collaboration, not competition.



📚 References

Dataset Source

- UnclePablo. (2023). AI Impact on Jobs – Kaggle Notebook.
  🔗 https://www.kaggle.com/code/unclepablo/ai-impact-on-jobs

Python Libraries Used

- pandas: Data manipulation and analysis
- numpy: Numerical operations
- seaborn & matplotlib: Data visualization
- scikit-learn: Machine learning (Linear Regression, train-test split)

Statistical & ML Concepts

- Pearson Correlation Coefficient
- Linear Regression
- Classification by Thresholding (Risk Levels)

Tools & Environment

- Jupyter Notebook / Google Colab
- Python 3.13
- Data sourced and cleaned using pandas; models built using scikit-learn

For full code and analysis, refer to this link
https://colab.research.google.com/drive/1WMRnZ5qFfVRWzjKc7RL2xKxSBRMTGQZ-?usp=sharing


Thank You