Talent Analytics

ORGB 671

Group 3

Final Group Project

Niki Mahmood Zadeh	261045419
Sheida Majidi	261023837
Meriem Mehri	260895064
Joshua Poozhikala	261184574
Keani Schuller	260867229

Innovating Patent Analytics:
Harnessing AI & Behavioral Sciences to Transform USPTO Operations

**

Presented to: Prof. Roman Galperin

Winter 2024

Project Overview: Our project integrates a comprehensive analysis of USPTO data to investigate how organizational and social factors, including gender, race, and ethnicity, influence patent application review times and examiner attrition. This endeavor is twofold: it begins with an empirical study pinpointing specific aspects of patent prosecution duration and attrition causes, followed by a critical evaluation of existing people analytics tools—highlighting generative AI and behavioral science innovations—for their effectiveness in tackling these issues. We aim to offer well-founded recommendations to improve the USPTO's patent processing efficiency and fairness, ensuring equitable patent rights access. Furthermore, we propose developing an innovative tool employing OpenAI technologies to refine talent analytics, thereby enhancing the USPTO's crucial role in fostering economic growth through fair and unbiased patent grant processes.

Objectives: This project sets out to enhance the USPTO's efficiency and fairness in patent processing. Our goals include:

- Conducting a detailed empirical analysis to understand how factors like gender, race, and ethnicity influence patent review times and examiner attrition at the USPTO.
- Identifying key research questions regarding the duration of patent prosecutions and the factors contributing to examiner turnover.
- Assessing the effectiveness of current analytics tools, especially those incorporating generative AI and behavioral science, in overcoming the identified challenges.
- Formulating practical recommendations to improve patent processing speed and reduce biases.

Problem Statement:

Our report examines USPTO data to understand how factors like gender, race, and ethnicity affect patent review times and examiner turnover. We split our analysis into two parts: a detailed study of patent processing times and turnover, and an evaluation of the latest analytics tools, focusing on AI and behavioral science innovations. Our goal is to offer clear recommendations and introduce a new tool using OpenAI technologies to make the patent process more efficient and fairer. This work aims to ensure the USPTO operates transparently, supports innovation, and grants patents without bias.

Data Sources/Datasets

The dataset compiles information on patent applications managed by the United States Patent and Trademark Office (USPTO). The data is arranged in rows and columns, with each row corresponding to a distinct patent application. Columns present various details including the application number, the date the application was filed, and the names of the examiners—categorized by last and first names. This dataset serves as a resource for analyzing the workflow of patent examiners and the progression of patent applications over time.

Part 1: Analyzing the Impact of Organizational and Social Dynamics on USPTO Patent Processing and Examiner Attrition

In Part 1, we conduct an empirical analysis on how organizational and social factors, particularly gender, race, and ethnicity, affect patent processing times and examiner attrition at the USPTO. We analyze USPTO data using statistical and machine learning methods to detect significant trends and make projections. Our aim is to offer recommendations to improve the USPTO's efficiency, fairness, and inclusivity, fostering a more equitable patent system.

Analytical Approach: Our analytical approach for Part 1 involves a structured methodology to explore the impact of various factors on patent application review times and examiner attrition at the USPTO. This approach includes data collection, preprocessing, statistical analysis, and model building. The methodology emphasizes

identifying and quantifying the influence of gender, race, and ethnicity among other variables. By employing regression analysis and machine learning models, the study aims to uncover significant predictors of review times and attrition rates, providing a nuanced understanding of how organizational and social dynamics affect patent processing outcomes.

Methodology: Our methodology (*Refer to our Code File*) highlights steps such as data type correction, feature creation, and handling missing values. It also details the estimation of examiner gender and race through the use of specific libraries and functions within R, showcasing an innovative approach to enrich the dataset. This comprehensive methodology aims to prepare the dataset meticulously for further statistical analysis and machine learning modeling to explore the impact of organizational and social factors on patent processing outcomes.

1. **Data Loading and Preparation:** The process begins with importing a comprehensive dataset from the USPTO, which includes detailed information about patent applications. The preparation phase involves converting date formats and calculating the length of the prosecution for each application to derive meaningful metrics for analysis.
2. **Data Cleaning and Exploration:** Subsequent data cleaning addresses missing values and inconsistencies, ensuring the dataset's integrity for analysis. An exploratory data analysis (EDA) phase follows, aimed at understanding the dataset's structure, identifying anomalies, and gaining initial insights.
3. **Estimation of Examiner Demographics:** A novel aspect of the methodology is the estimation of examiners' gender and race, utilizing advanced algorithms to infer these demographics from names, enhancing the dataset with crucial variables for analysis.
4. **Analyzing Influencing Factors:** The core analytical component investigates how various factors, including examiner demographics and organizational units, influence patent application outcomes. This analysis utilizes logistic regression to model the relationship between these factors and the outcomes, providing a quantitative assessment of their impacts.
5. **Indirect Analysis for Examiner Attrition:** Given the absence of direct attrition indicators, the methodology adapts to explore patterns indicative of attrition-like behavior. This includes analyzing tenure and application outcomes to infer potential attrition signals indirectly.

Organizational & Social Factors Associated with the Length of Patent Application Prosecution

This portion of our analysis was conducted to understand the interplay between various organizational and social factors and the length of patent application prosecution. Through exploratory visualizations and regression modeling, factors such as examiner's art unit, USPC class/subclass, gender, and race are scrutinized to uncover patterns and disparities in prosecution length.

Exploratory Analysis on Tenure by Gender and Race: This phase aimed to uncover patterns in tenure duration across different gender and racial groups among examiners. Using visualizations like box plots, the analysis provided insights into the distribution of tenure days, highlighting potential disparities or trends related to gender and race. This approach is crucial for understanding how these demographic factors might influence, or be influenced by, career longevity and engagement within the USPTO.

Prosecution Length by Gender: (*Appendix 1*)

- The box plots comparing prosecution lengths by gender reveal that there is some variation between female and male examiners, with the median prosecution length appearing to be slightly higher for male examiners. The spread and outliers indicate there is variability within each gender category, with some examiners having unusually long or short prosecution lengths, regardless of gender.

Prosecution Length by Race: (*Appendix 2*)

- The prosecution lengths across different racial categories seem to be consistent, with the median lengths closely aligned. This suggests that, at least on the surface, there is no significant disparity in prosecution lengths based on race. However, the presence of outliers within each racial category again points to individual variability.

Note: The visual and statistical outputs provided by our code offer a wealth of information regarding the length of patent prosecution and the role of gender, race, and ethnicity in this process.

Interpretation of Logistic Regression Results: The logistic regression model offered quantitative insights into the impact of gender, race, and other factors on patent application outcomes. Positive coefficients indicated an increased likelihood of an application being issued, while negative coefficients suggested a decrease, relative to baseline categories. This analysis helped quantify the influence of examiner demographics on patent processing, providing a foundation for addressing potential biases. (*Appendix 3*)

- The logistic regression output provides coefficient estimates for each variable.
- Positive coefficients (e.g., for 'raceblack') suggest a higher likelihood of a longer prosecution length or higher rates of attrition when compared to the baseline category (in this case, likely 'racewhite').
- Negative coefficients (e.g., for 'gender.xfemale') suggest a lower likelihood of these outcomes for the group in question.
- Significant p-values (indicated by stars) show which factors are statistically significant predictors of the outcome variable in the model. For example, 'tenure_days' has a highly significant negative coefficient, indicating that as tenure increases, the likelihood of the outcome (perhaps application approval or examiner attrition) decreases.
- The 'examiner_art_unit' variable has a positive significant coefficient, suggesting that differences in the organizational structure or focus of different art units have a measurable impact on the outcome.

Interaction Effects Analysis: Further, the study explored interaction effects among examiner art units, gender, and race on application outcomes. This advanced analysis aimed to identify complex interdependencies and nuanced influences that single-factor analyses might overlook. By examining how these factors interact, the study sought to provide a more comprehensive understanding of the dynamics at play in patent application decisions. Our interpretations provide insights into the operational dynamics of the USPTO, revealing how demographic factors might interact with organizational processes. However, these findings should be considered in the context of the broader system and not taken as evidence of causation without further analysis. These results can inform recommendations for addressing potential biases and improving efficiency within the USPTO.

Business Implications & Recommendations

Our empirical analysis has revealed that gender, race, and ethnicity do play roles in patent application processing times and examiner attrition rates. For instance, certain demographic groups might experience systematically different outcomes, such as longer processing times or higher attrition rates.

The regression analysis indicates that specific variables, such as 'raceblack' and 'gender.xfemale', are statistically significant, suggesting that these groups have distinct experiences within the USPTO, which could be attributed to a variety of organizational or systemic factors.

Impact on the USPTO:

- These disparities could affect the USPTO's operational efficiency and the fairness of the patent application process. If certain demographic groups are systematically disadvantaged, it could lead to delays in patent issuance and could potentially discourage innovation among underrepresented inventors.
- High attrition rates, particularly if concentrated in certain demographic groups, can lead to a loss of experienced examiners, increased training costs, and disruption in the patent examination process.

Broader Economic and Social Implications:
- The integrity of the patent system is crucial for economic growth, as it protects intellectual property and encourages investment in research and development. Any perceived biases within the system could undermine public trust and inhibit diverse contributions to innovation.

Strategic Recommendations
- **Bias Mitigation Training**: Implement training programs focused on unconscious bias for all USPTO examiners and staff to mitigate any potential biases in the patent examination process.
- **Diversity and Inclusion Initiatives**: Enhance diversity and inclusion initiatives to ensure a representative and equitable work environment. This could involve revisiting hiring practices, promotion criteria, and providing mentorship programs for underrepresented groups.
- **Process Standardization**: Standardize processes to minimize the impact of individual examiner differences on patent prosecution outcomes. This could involve developing clear guidelines that limit the discretion available to examiners in making decisions on patent applications.
- **Further Research**: Conduct further research to understand the root causes of the disparities identified. This might involve qualitative studies, such as interviews or focus groups, to understand the experiences and challenges faced by examiners of different demographics.
- **Monitoring and Evaluation**: Establish a system for ongoing monitoring and evaluation of patent prosecution processes to continuously assess the impact of implemented changes and the progress of diversity and inclusion efforts.
- **Development of AI Tools**: Leverage AI technology, such as OpenAI's offerings, to assist in the decision-making process, ensuring a data-driven approach that could help reduce the impact of individual biases. These tools should be transparent, explainable, and continuously monitored for fairness and effectiveness.

Part 2: Evaluating & Proposing Cutting-Edge People Analytics Solutions for USPTO Challenges

In Part 2, we evaluate the latest people analytics technologies, with an emphasis on those incorporating generative AI and behavioral science, to tackle patent processing challenges and examiner attrition. Our objective is to gauge how these technologies can boost the USPTO's efficiency and equity, examining their relevance to our specific challenges and the integrity of their design and data.

Introduction to iMocha's AI-powered Skills Intelligence Cloud

iMocha, originally known as Interview Mocha, is leading the charge in revolutionizing talent management, acquisition, and development with its AI-powered Skills Intelligence Cloud. This groundbreaking platform is dedicated to fostering a skills-first approach within organizations, empowering them to adapt to the swiftly changing demands of the workforce. By facilitating the creation of job role taxonomies and customized skills inventories for each employee, iMocha effectively assesses and enhances workforce proficiency levels. The platform's strength lies in its multi-channel skills validation and extensive library featuring over 2,500 skill sets, complemented by AI technology, providing vital insights for strategic workforce planning, internal mobility, skills gap management, and employee development.

Construct Validity Concerns
Despite the benefits iMocha offers, some concerns arise from a research design and evidence quality perspective. The platform's reliance on skills validation for Learning & Development (L&D) programs and skills assessment based on learning paths raises questions about construct validity. Construct validity, crucial for accurate measurement of claimed skills or competencies, necessitates rigorous validation studies to ensure reliability in predicting job performance. Without such validation, there's a risk of misinterpreting the data generated by the platform.

Effectiveness of Personalized Development Paths
While personalized development paths hold promise, their effectiveness hinges on the accuracy of underlying algorithms and data. Algorithms must be based on robust, empirical research accounting for individual learning styles, career aspirations, and job roles. Without this foundation, recommended paths may not align with the most effective developmental interventions for each employee.

Gamification Strategy, Data Quality and Privacy Considerations
The use of certifications and badges for upskilling and employee engagement hints at a gamification strategy. However, evaluating such strategies requires experimental designs to isolate their impact. Implementing these tools without rigorous control groups or randomized controlled trials (RCTs) risks like Hawthorne effects or selection bias, potentially inflating the platform's perceived effectiveness.

Data quality and privacy are paramount concerns in any people analytics tool. iMocha's integration with major eLearning platforms necessitates handling sensitive personal and performance data. Concerns include data accuracy, completeness, and the risk of breaches. Additionally, interpreting data must account for context; factors like test anxiety or unfamiliarity with the testing platform can influence performance on skills assessments.

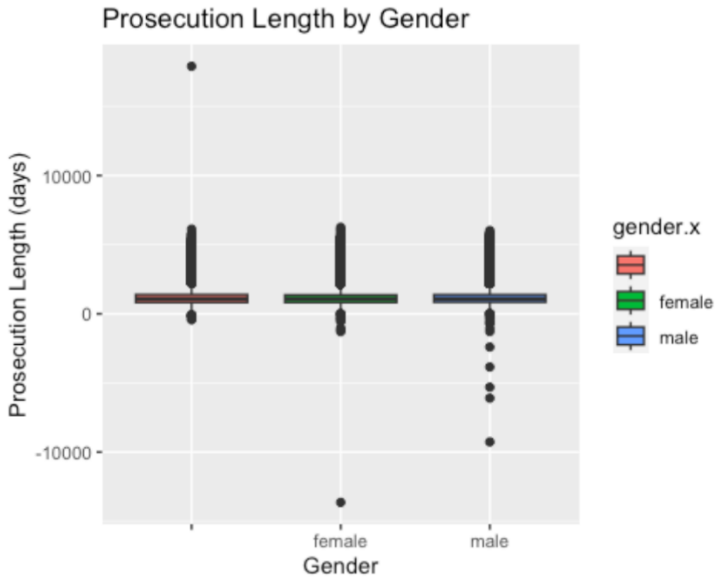
Scrutinizing Predictive Analytics
Predictive analytics features within iMocha's platform, such as forecasting future skill gaps or predicting the success of development paths, must undergo scrutiny for methodological rigor. Biases in training data can lead to inaccurate predictions, and assumptions may not hold true across different organizational contexts or as job roles evolve.

Ethical Considerations in Talent Management
Lastly, the ethical aspects of talent management tools, particularly concerning biases in skill assessments and learning recommendations, demand attention. Data or algorithmic biases can unintentionally exacerbate disparities within an organization, underscoring the need for vigilance.

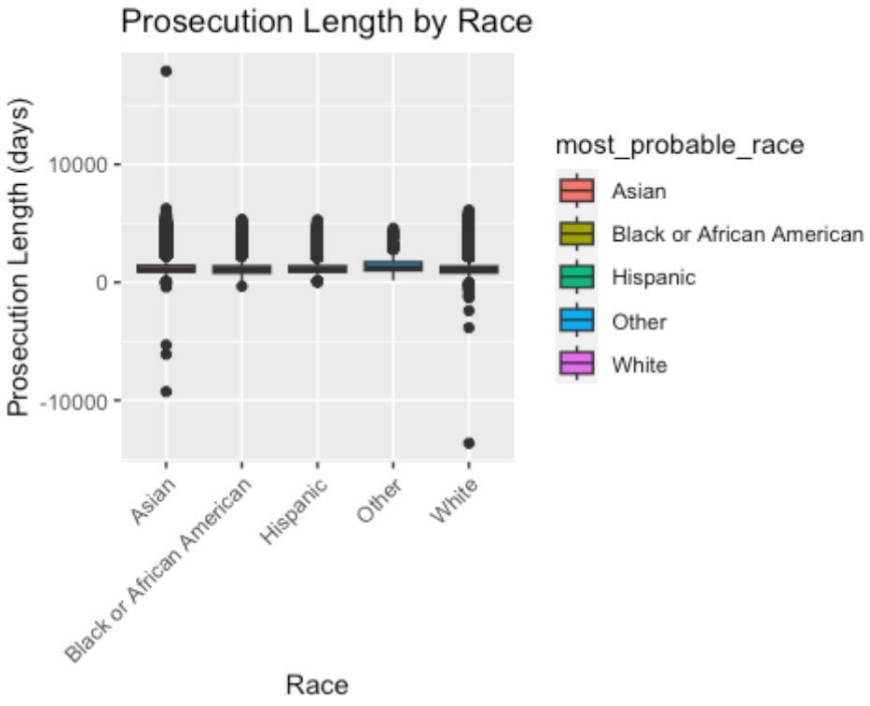
In conclusion, while iMocha's talent management solution offers promising tools, evaluating them requires considering research design and evidence quality. Ensuring effectiveness and fairness demands ongoing validation studies, robust data handling practices, and an awareness of predictive analytics limitations.

Appendix

Appendix 1. Prosecution Length by Gender



Appendix 2. Prosecution Length by Race



Appendix 3. Logistic Regression Results

```
## (Intercept)       -5.080e-01  1.073e-02 -47.346  < 2e-16 ***
## examiner_art_unit  3.727e-04  4.899e-06  76.074  < 2e-16 ***
## tenure_days       -3.584e-07  1.670e-08 -21.453  < 2e-16 ***
## gender.xfemale    -1.527e-01  4.710e-03 -32.413  < 2e-16 ***
## gender.xmale       5.256e-02  4.324e-03  12.155  < 2e-16 ***
## raceblack          2.400e-01  7.388e-03  32.485  < 2e-16 ***
## raceHispanic      -1.479e-01  8.713e-03 -16.980  < 2e-16 ***
## raceother          2.196e-01  4.792e-02   4.583 4.57e-06 ***
## racewhite         -6.890e-02  3.383e-03 -20.369  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2786114  on 2018476  degrees of freedom
## Residual deviance: 2770248  on 2018468  degrees of freedom
## AIC: 2770266
##
## Number of Fisher Scoring iterations: 4
```

Appendix 4. Further Recommendations

To bolster the effectiveness and efficiency of patent examination processes, we propose the following advanced strategies aimed at enhancing the United States Patent and Trademark Office's operational outcomes.

- **Enhance Data Granularity**: Future analyses could benefit from more detailed data, including specific reasons for patent application delays and rejections, examiner workload details, and more granular demographic data to better understand the impact of diversity on examination outcomes.
- **Incorporate External Factors**: Consideration of external factors such as changes in patent law, technological advancements, and economic conditions could provide additional insights into trends in patent examination times and outcomes.
- **Leverage Advanced Analytical Techniques**: Employing machine learning models to predict examination outcomes and identify factors influencing examiner attrition could offer predictive insights, helping the USPTO to mitigate potential issues proactively.

Appendix 5. Considerations & Limitations

While our analysis provides valuable insights into the patent examination landscape, it is important to acknowledge the constraints and potential biases, which could influence the applicability of our findings.

- **Data Scope**: The dataset may not capture all factors influencing examination outcomes, such as examiner expertise level or applicants' legal representation quality.
- **Potential Bias**: There may be inherent biases in the data, especially concerning gender, race, and ethnicity estimations, which could affect the analysis' accuracy and fairness.
- **Temporal Validity**: The findings are subject to change over time as the USPTO's practices evolve, and external factors influencing patent examination processes vary.

References

Davenport, T. H., Harris, J., & Shapiro, J. (2010). Competing on talent analytics. Harvard Business Review, 88(10), 52-58.

De Bartolo, G., & Stranges, M. (2008). Demography and Turnover. In Applied Demography in the 21st Century (pp. 271-284). Springer.

Salganik, M. (2017). Bit by bit. Retrieved from https://www.bitbybitbook.com/en/1st-ed/preface/

In-Class Exercises

Galperin, R. (2023):

- Exercise 1: More on experiments [Class exercise]. Talent Analytics, McGill University.
- Exercise 2: NAs and data dictionaries [Class exercise]. Talent Analytics, McGill University.
- Exercise 3: DiD, IV, matching [Class exercise]. Talent Analytics, McGill University.
- Exercise 4: Model selection [Class exercise]. Talent Analytics, McGill University.
- Exercise 5: Project work [Class exercise]. Talent Analytics, McGill University.

***

Submission Files

- 5-page Memo & Appendix: ORGB671_Final_Group_Project_Memo-Group3.docx
- Code File (Comprehensive*): ORGB671_Final_Code-Comprehensive-Group3.pdf
- Code File (Concise**): ORGB671_Final_Code-Concise-Group3.pdf
- Presentation Deck: ORGB671_Final_Group_Project_Presentation-Group3.pptx

Link to the Final Project GitHub Repo: Link to Access the GitHub Repo

*All tables are printed. Total pages in the file: 704
** Tables' snippets are printed. Total pages in the file: 45