Detailed Outline for Research Paper: A Responsible AI Framework for Bias-Free Resume Screening in Human Resource Management

Authors: Sharon Melhi, Aditya Banerjee
Affiliation: Amity University Bengaluru

I. Introduction (600 words)

- **Sharon Melhi**
  - **Context and Problem Statement** (300 words):
    - Discuss the rise of AI in Human Resource Management (HRM), focusing on automation in recruitment processes.
    - Explain how AI-driven resume screening can perpetuate biases (e.g., based on names, demographics).
    - Highlight societal and organizational risks of biased AI, such as unfair hiring practices and legal implications.
    - Reference Tambe et al. (2019) for AI in HRM and Raghavan et al. (2020) for algorithmic bias.
  - **Research Gap** (200 words):
    - Identify the lack of scalable, auditable frameworks for Responsible AI in recruitment.
    - Discuss limitations of existing tools, particularly in bias mitigation during preprocessing and output stages.
    - Reference Jobin et al. (2019) to support the gap in Responsible AI frameworks.
  - **Objective** (100 words):
    - Introduce FairHire as a bias-free, scalable AI model for resume screening.
    - List three objectives:
      1. Develop a framework to extract critical resume attributes (name, qualifications, experience, skills) while removing sensitive data.
      2. Ensure compliance with anti-discrimination standards (e.g., EEOC guidelines).
      3. Evaluate fairness and performance against traditional ATS tools.
  - **Topics to Cover**:
    - Growth of AI in recruitment and its challenges.
    - Specific examples of bias in resume screening (e.g., name-based discrimination).

        - Need for auditable, fair AI systems and FairHire's role in addressing this.

II. Literature Review (800 words)

- **Aditya Banerjee**
    - **AI and Automation in HRM** (300 words):
        - Explore AI's role in recruitment, including talent acquisition and resume screening.
        - Discuss limitations of current ATS tools (e.g., lack of transparency, potential for bias).
        - Reference Meijerink et al. (2020) to support AI's growing role in HRM.
    - **Bias in AI Systems** (300 words):
        - Identify sources of bias: training data, algorithmic design, and output interpretation.
        - Highlight challenges in NLP models, particularly with handling names or identities.
        - Reference Mehrabi et al. (2021) for bias sources and Binns (2018) for NLP fairness issues.
    - **Responsible AI Frameworks** (200 words):
        - Define Responsible AI principles: transparency, explainability, non-discrimination.
        - Discuss gaps in HR-focused AI tools regarding auditable fairness and explainability.
        - Reference Floridi et al. (2018) for Responsible AI principles.
    - **Topics to Cover**:
        - AI's impact on recruitment processes and ATS limitations.
        - Specific sources of bias in AI systems and their impact on hiring.
        - Need for Responsible AI frameworks in HRM and current shortcomings.

III. Methodology (1,000 words)

- **Sharon Melhi**
    - **System Architecture: FairHire Framework** (400 words):
        - Describe the FairHire pipeline:
            - **Preprocessing**: Use python-docx to parse .docx resumes.
            - **NER**: Apply Hugging Face's dslim/bert-base-NER to extract entities (name, organization, skills, education).
            - **Sensitive Data Filtering**: Use regex to remove attributes like gender, religion, nationality.
            - **Report Generation**: Generate concise HR reports using distilgpt2.
        - Discuss current .docx parsing limitations and plans to enhance robustness for diverse resume formats.

    - **Bias Mitigation Approach** (300 words):
        - Detail techniques:
            - Detection and removal of sensitive attributes (e.g., ethnicity, age).
            - Contextual masking and rewriting for anonymization.
            - Compliance with anti-discrimination standards (e.g., EEOC, EU AI Act).
        - Explain how FairHire ensures merit-based evaluation.
    - **Topics to Cover**:
        - Detailed workflow of FairHire's components.
        - Specific regex patterns and NER processes for bias removal.
        - Future improvements for .docx parsing to handle varied formats.
- **Aditya Banerjee**
    - **Evaluation Metrics** (200 words):
        - Define metrics:
            - Entity extraction accuracy (% of correctly identified entities).
            - Bias removal success rate (% of sensitive data filtered).
            - Processing time per resume (seconds).
            - Fairness uplift (% improvement over baseline ATS).
        - Compare FairHire's performance with traditional ATS tools (e.g., Taleo, Workday).
    - **Tools and Environment** (100 words):
        - List tools: Python, Hugging Face Transformers, SpaCy, Pandas, Regex.
        - Describe the development environment (local setup, potential for cloud scalability).
    - **Topics to Cover**:
        - Quantitative metrics for evaluating FairHire's performance.
        - Comparison with commercial ATS tools.
        - Technical tools and environment setup.

IV. Results (800 words)

- **Aditya Banerjee**
    - **Quantitative Outcomes** (400 words):
        - Present key metrics:
            - Entity extraction accuracy: 93.8%.
            - Bias removal success: 98.7%.
            - Processing time: 2.1 seconds per resume.
            - Fairness improvement: 35% over baseline ATS.
        - Use tables or graphs to summarize data clearly.
        - Discuss the significance of these metrics for ethical hiring.
    - **Topics to Cover**:
        - Detailed performance metrics and their implications.
        - Visual representation of results (e.g., bar charts for fairness uplift).
- **Sharon Melhi**

    - **Qualitative Insights** (300 words):
        - Present sample anonymized HR reports generated by FairHire.
        - Summarize feedback from HR professionals on usability and trustworthiness.
        - Discuss how reports support merit-based decision-making.
    - **Error Analysis** (100 words):
        - Identify issues: partial attribute masking, challenges with unstructured resumes.
        - Suggest improvements for future iterations (e.g., better handling of non-standard formats).
    - **Topics to Cover**:
        - Sample HR reports and their practical utility.
        - HR feedback on FairHire's effectiveness.
        - Specific errors and proposed solutions.

V. Discussion (800 words)

- **Sharon Melhi**
    - **Interpretation of Findings** (300 words):
        - Explain how FairHire aligns with Responsible AI principles (fairness, transparency, privacy).
        - Highlight contributions to merit-based hiring and ethical automation.
        - Discuss how FairHire addresses societal risks of biased AI.
    - **Comparison with Existing Systems** (200 words):
        - Benchmark FairHire against commercial ATS tools (e.g., Taleo, Workday).
        - Emphasize advantages in fairness, speed, and transparency.
    - **Topics to Cover**:
        - Alignment with ethical AI principles.
        - FairHire's superiority over existing systems.
- **Aditya Banerjee**
    - **Scalability and Generalization** (300 words):
        - Discuss FairHire's applicability across industries (e.g., tech, healthcare) and geographies.
        - Address adaptability to diverse resume formats and languages.
        - Highlight potential for cloud-based scaling.
    - **Topics to Cover**:
        - Industry and geographic flexibility.
        - Handling diverse resume formats and future scalability.

VI. Implications filho(600 words)

- **Aditya Banerjee**
    - **Theoretical Implications** (200 words):
        - Position FairHire as an advancement in Responsible AI for HRM.
        - Discuss contributions to algorithmic hiring ethics frameworks.

    - **Policy Implications** (200 words):
        - Align FairHire with regulations: EU AI Act, India's Digital Personal Data Protection Act (2023).
        - Advocate for standardized fairness metrics in AI-driven recruitment.
    - **Topics to Cover**:
        - Theoretical contributions to Responsible AI.
        - Policy alignment and advocacy for fairness standards.
- **Sharon Melhi**
    - **Practical Implications** (200 words):
        - Present FairHire as an open-source, modular tool for global organizations.
        - Highlight support for diverse hiring contexts and reduced audit risks.
    - **Topics to Cover**:
        - Practical benefits for HR teams.
        - Open-source potential and audit compliance.

VII. Limitations and Future Research (400 words)

- **Sharon Melhi**
    - **Limitations** (200 words):
        - Discuss challenges in multilingual resume parsing.
        - Highlight sensitivity to informal or non-standard resume formats.
        - Acknowledge current model's limitations in handling diverse .docx inputs.
    - **Topics to Cover**:
        - Specific parsing challenges (e.g., multilingual, unstructured resumes).
        - Impact of limitations on FairHire's performance.
- **Aditya Banerjee**
    - **Future Research** (200 words):
        - Propose enhancements:
            - Improved .docx parsing for diverse formats.
            - Integration with HRIS platforms.
            - Expansion to job-matching and interview fairness evaluation.
            - Real-world A/B testing with hiring outcomes.
    - **Topics to Cover**:
        - Specific areas for model improvement.
        - Potential for broader HR applications.

VIII. Conclusion (250 words)

- **Sharon Melhi** :
    - Summarize FairHire's contribution: a robust, transparent, scalable AI pipeline for bias-free resume screening.
    - Emphasize alignment with Responsible AI principles and ethical hiring.

- Call for responsible deployment and collaboration between HR and AI developers.
- **Topics to Cover**:
  - Recap of FairHire's key features and benefits.
  - Vision for ethical AI in recruitment.
  - Encouragement for HR-AI co-creation.