SMART INDIA HACKATHON 2024

- Problem Statement ID: [Your PS ID]
- Problem Statement Title: Automated PII Detection and Protection in Document Uploads
- Theme: [Your Theme]
- PS Category: Software
- Team ID: [Your Team ID]
- Team Name: [Your Registered Team Name]

IDEA TITLE: Cross-Platform PII Shield: Intelligent Document Protection

Slide 1: Proposed Solution

Detailed explanation of the proposed solution:

- Browser extension for desktop and SDK/API for mobile platforms
- Automatic PII detection during document uploads
- Server-side advanced processing with OCR and ML
- Privacy-preserving verification through secure hashing
- User-friendly redaction tools with intelligent replacement

How it addresses the problem:

- Prevents inadvertent sharing of sensitive information
- Enhances user awareness of PII in their documents
- Provides tools for users to control their data
- Assists organizations in compliance and risk mitigation

Innovation and uniqueness of the solution:

- Cross-platform approach ensuring consistent protection
- Privacy-preserving verification using secure hashing
- Intelligent redaction with context-aware replacement
- Real-time detection and alerting during upload process

Slide 2: Technical Approach

Technologies to be used:

- Frontend: React.js for web interface, React Native for mobile SDK
- Backend: Node.js with Express.js
- Database: MongoDB for document metadata storage
- OCR: Tesseract.js for text extraction
- Machine Learning: TensorFlow.js for advanced PII pattern recognition
- Cryptography: bcrypt for secure hashing

Methodology and process for implementation:

[Insert a flowchart here showing the following process]

1. User initiates document upload
2. Client-side initial PII check (Extension/SDK)
3. Secure transmission to server
4. Server-side processing (OCR + ML detection)
5. PII verification through secure hashing
6. User alerted and offered redaction options
7. Document uploaded with user-approved modifications

Slide 3: Feasibility and Viability

Analysis of the feasibility:

- Leverages widely-used, open-source technologies
- Modular architecture allows for scalability and maintenance
- Cross-platform approach ensures wide applicability

Potential challenges and risks:

- Ensuring accuracy of PII detection across various document types
- Maintaining performance with large-scale adoption
- Keeping up with evolving PII regulations across regions

Strategies for overcoming challenges:

- Continuous model training with diverse document datasets
- Implementing efficient caching and load balancing strategies
- Regular compliance audits and modular design for quick updates

Slide 4: Impact and Benefits

Potential impact on the target audience:

- Individuals: Enhanced control over personal data sharing
- Organizations: Reduced risk of data breaches and regulatory fines
- Government: Secure verification without exposing entire databases

Benefits of the solution:

- Social: Increased trust in digital document sharing
- Economic: Reduced costs associated with data breaches and compliance violations
- Technological: Advancement in privacy-preserving technologies

Slide 5: Research and References

- NIST Guide to Protecting PII: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf (https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf)
- Privacy-Preserving OCR Systems: https://arxiv.org/abs/1908.09887 (https://arxiv.org/abs/1908.09887)
- Secure Multi-Party Computation for Privacy-Preserving Data Mining: https://dl.acm.org/doi/10.1145/1125663.1125722 (https://dl.acm.org/doi/10.1145/1125663.1125722)
- [Add more relevant research papers and industry reports]