



STATATHON 2025



- **Problem Statement ID: 5**
- **Problem Statement Title: AI-enabled Semantic Search for National Classification of Occupation (NCO)**
- **PS Category: Software**
- **Team ID: 5524**
- **Team Name (Registered on portal): Penguins**

STATATHON 2025

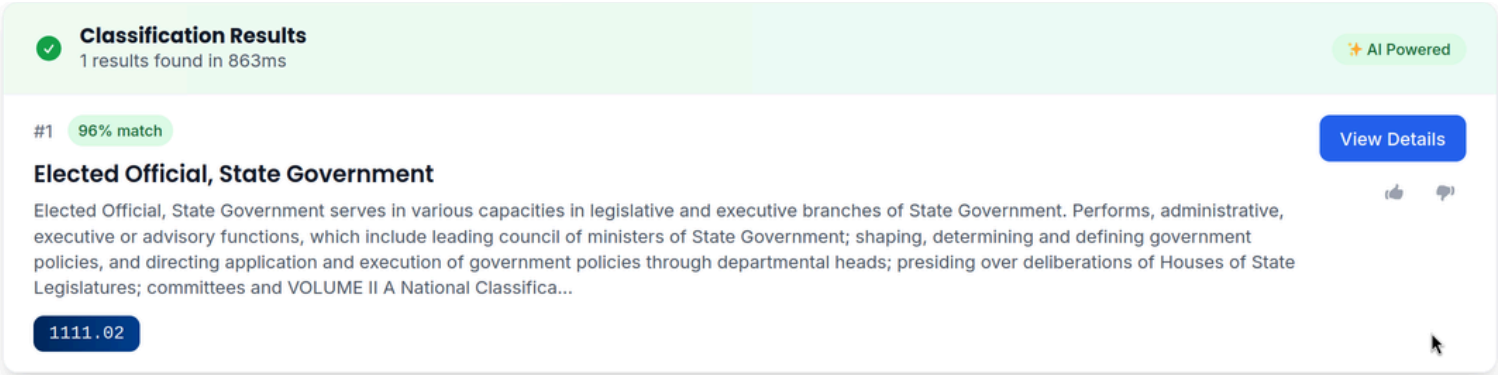
Solution: An AI-powered, multilingual semantic search engine for NCO-2015 that enables fast, accurate occupation code retrieval using natural language, voice input, and synonym/context understanding.

Detailed Explanation

- Convert NCO-2015 dataset into structured JSON/CSV preserving hierarchy.
- Use NLP embeddings (BERT/Sentence Transformers) for semantic understanding.
- Rank search results by relevance & confidence scores.
- Support synonyms, spelling variations, and related term mapping.
- Admin panel for database updates & search audit logs.

Problem & Innovation

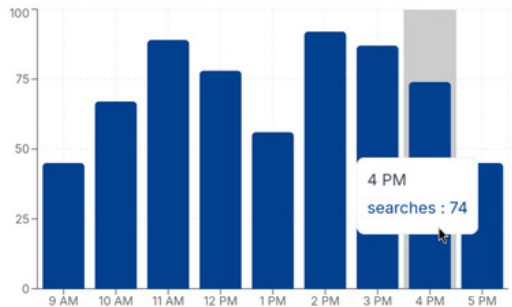
- Eliminates dependency on exact keyword search in PDFs.
- Reduces classification errors & speeds up survey workflows.
- Improves accuracy for 50,000+ enumerators nationwide.
- Multilingual + voice search for real-time field usability.
- Domain-specific synonym/context mapping for Indian occupations.
- API-ready for integration with MoSPI systems.

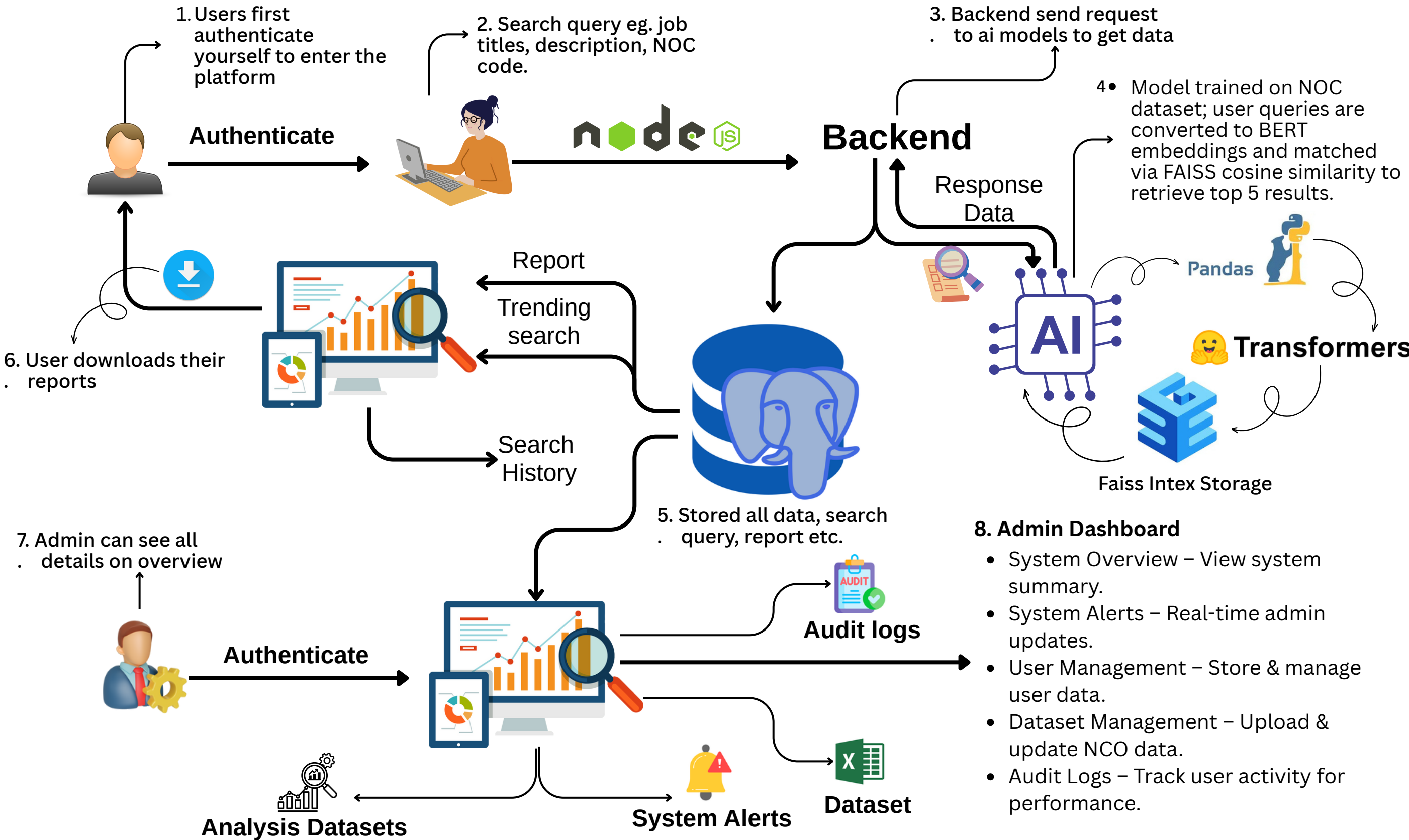


System Alerts

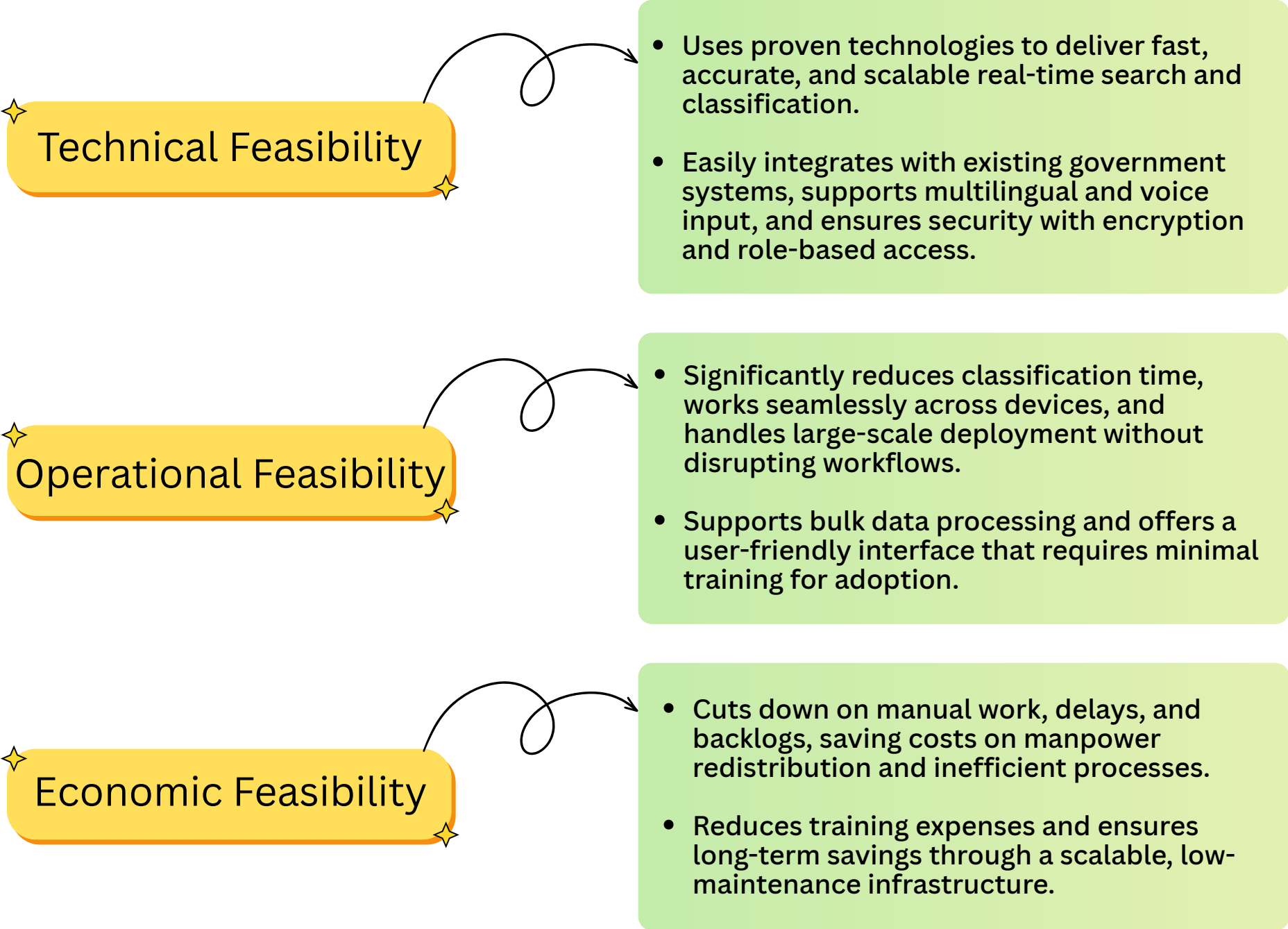
- ⚠️ AI service response time above 300ms threshold
15 minutes ago
- ⚠️ Database maintenance scheduled for tonight
2 hours ago

Daily Activity Pattern





Feasibility Analysis:



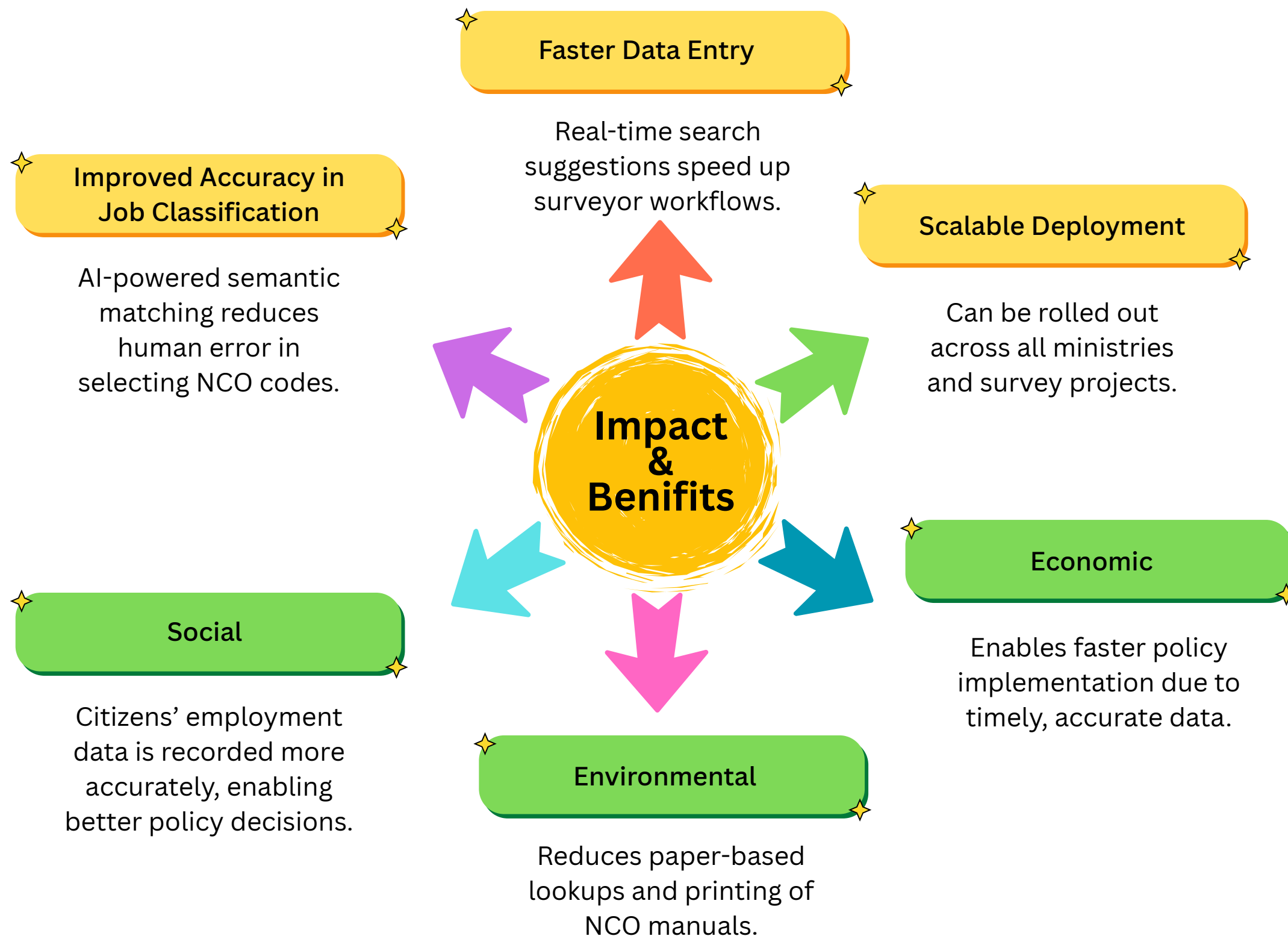
Challenges & Risks

- **Performance & Accuracy :**
Fast searches with high classification accuracy.
- **Multilingual & Context Handling :**
Accurate, error-free translations across all languages.
- **Security :**
Safeguarding sensitive data and ensuring compliance.

Strategies to Overcome Challenges

- **Enhance Performance :**
Optimize search with FAISS, retrain AI regularly.
- **Strengthen Multilingual Capabilities**
Get expert-checked translations with context-aware AI.
- **Ensure Security :**
Use secure access, encryption, and audits..

IMPACT AND BENEFITS



Revenue Model

Dual Monetization Approach

- Government Contracts** (high-ticket, B2G sales)
- SaaS Model for Private Sector** (recurring, scalable)
- Advertising Revenue** (from high-traffic public-facing tools).
 - Implementation Contracts: Full system deployment in each state.
 - Annual Maintenance & Support: Ongoing bug fixes, updates, scaling.
 - Customization Services: Adapting system to state-specific needs.

RESEARCH AND REFERENCES

Innovation and Uniqueness

- First AI-based occupation classification tool in India aligned with NCO-2015.
- Built for scale, capable of handling nationwide census level data without manual bottlenecks.
- Field-centric design ,optimizes for on-ground conditions (mobile use, offline capabilities, voice input for low-literacy users).
- High Accuracy + High Speed combination, rare for government AI tools (99%+ accuracy).
- Continuous improvement through user feedback, manual override tracking, and adaptive AI models that incorporate suggested terms over time.

Research Sources

- NCO-2015 – Ministry of Labour & Employment, India
<https://labour.gov.in/nco>
- Devlin, J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Google AI).
<https://arxiv.org/abs/1810.04805>
- Singapore GovTech – NLP for citizen services.
<https://www.tech.gov.sg>
- The Economic Impact of Artificial Intelligence on Public Services – OECD
<https://www.oecd.org>
- McKinsey & Company – AI in Public Sector reports.
<https://www.mckinsey.com>



Video Explanation

[Website Link](#)

[Github Link](#)