# APS5 Data Scientist - Expression of Interest

**Job ID:** 4152, **APS Code:** 270308

## Relevant Experience - STAR Examples

### Analytical Solution Development & Stakeholder Liaison (Duties 2, 4)

**Situation:** ATO needed to assess feasibility of using vision-language AI models to extract structured data from work-related expense documents (receipts, invoices, bank statements) for compliance verification. As sole developer, I needed to coordinate across data engineering and senior data scientists to deliver a proof-of-concept within strict timeframes.

**Task:** My responsibility was to evaluate two vision-language models against 195 business documents, extracting 17 work-related expense fields, while liaising with data engineering to deploy models on ATO infrastructure and regularly reporting progress to senior data scientists.

**Action:** I completed technical documentation and security assessments required to load large language models onto ATO systems, collaborating extensively with data engineering on infrastructure requirements. I designed a comparative evaluation framework with batch processing capabilities, implemented multiple matching strategies for different field types (exact matching for monetary values, fuzzy matching for text, position-aware matching for transaction sequences), and created ground truth datasets for validation. I maintained regular progress demonstrations with senior data scientists, meeting all predefined milestone deadlines.

**Result:** Successfully evaluated 195 documents achieving 52% overall extraction accuracy across 17 fields, identifying which field types (monetary amounts, dates) performed well versus which required further development (complex bank statement transaction tables). This feasibility analysis provided data-driven recommendations on model selection and highlighted specific technical constraints (processing speed: 2 documents/hour) that informed realistic deployment planning. The comparative analysis enabled senior stakeholders to make evidence-based decisions about next-phase investment and resource allocation.

### Technical Reporting & Data Visualization (Duty 5)

**Situation:** Complex model evaluation results from a proof-of-concept needed to be communicated clearly to senior data scientists and potential business stakeholders to inform future AI investment decisions.

**Task:** I needed to translate technical findings (field-level accuracy breakdowns, model performance differences, processing constraints) into accessible insights for non-technical decision-makers while maintaining scientific rigor.

**Action:** I developed comprehensive dashboards and automated reporting modules that visualized model performance by document type and field category, created executive summaries explaining technical tradeoffs in plain language, and generated comparative visualizations showing which expense document types were suitable for automated extraction versus manual review.

**Result:** Senior data scientists could quickly assess POC outcomes through clear dashboards showing that certain field types (GST amounts, total amounts) achieved viable accuracy while others (complex bank statement transaction tables) required additional development. The visualizations successfully communicated that while current models showed promise for pre-screening expense documents, processing speed and accuracy constraints meant automated extraction should augment rather than replace human verification. This honest assessment enabled realistic planning for incremental improvements in subsequent project phases.