Australian Government | GovAI

AI Peer Review

AI-assisted evaluation of documents like research papers or proposals against review rubrics, quality standards, or acceptance criteria.

CSIRO

Australian Government | GovAI

Executive Summary

Problem Faced

CSIRO and other government agencies regularly need to review large volumes of documents—such as research papers, tenders and grant proposals—against specific evaluation rubrics or criteria. Currently, evaluations are often conducted manually by human reviewers, which presents key challenges in terms of low efficiency, poor scalability and inconsistency. These challenges impact CSIRO's and APS's ability to deliver high quality, consistent, timely and scalable evaluations.

Solution Overview

AI Peer Reviewer provides an AI-assisted, customisable evaluation system designed to support reviewing complex research papers, tenders or grant proposals—against predefined criteria. Features include:

- Grounded Review Generation: Produces structured review clearly linking to content in document.
- Criteria-Based Assessment: Evaluates documents against predefined criteria.
- Marginal Risk Assessment: Assess increased or decreased risk from AI adoption *without* fully assessing existing process risks or performance.
- Human oversight to ensure quality, risk management and trust

Benefits and Impact

- Efficiency Gains: Automate the reviewing work and enable timely, consistent and scalable assessment.
- Cost Savings: Lower labour costs by reduces the need for multiple reviewers per documents while maintain human oversight.
- Improved Decision-Making: Reduces variability due to human bias or fatigue.
- Broder Impacts: Enables scalable, fair, and transparent reviews through explainable AI, while laying the foundation for responsible AI adoption across public sector evaluation tasks.

Target Audience and Stakeholders

- Research units inside CSIRO
- Research funding bodies
- Policy evaluation departments
- Grant administration offices
- Regulatory and ethics committees
- Government agencies that have needs for document review, researchers.

Risks and Mitigation Overview

Key risks of this AI use case include unreliable performance (e.g. hallucinations, inconsistency), privacy leakage, external security threats, and fairness or bias in assessments. To manage these risks, the system includes context-specific robust marginal risk assessment, human oversight, outcome monitoring mechanisms across multiple dimensions (e.g. accuracy, fairness), and a risk register established prior to deployment. Customised process monitoring and security practices, such as decision process and reasoning trace logging for future improvement and audit, as well as access control, can also be enabled post-deployment.

Use Case Status

Pilot

Use case timeline

This tool has been internally piloted by Data61 research teams for research paper review and grant/proposal pre-submission analysis. In addition, we have provided recommendations to government in-design/pilot "AI evaluator" use cases, including refinements to the Marginal Risk Assessment methodology.

Additional Information

Data Considerations

- Types of Data: Includes confidential data, IP content, organisational policies, documents, and user inputs.
- Data Governance: Ensures compliance with legal and regulatory standards through privacy/sensitivity-preserving data treatment and security practices.

Lessons Learned

Users may reject an AI solution due to incorrect configuration or use. Many solutions lack a systematic way for users to evaluate and incrementally improve quality through their own involvement. A highly user-customisable and evaluation-friendly design was developed. AI risk and performance evaluations often, inevitably, assess the performance of existing processes or individuals, leading to resistance. We apply Marginal Risk Assessment, focusing on changes in risk without evaluating the absolute performance of existing processes.

Contact information

Responsible Entity Name

CSIRO's Data61

Area of Entity

Software & Computational Systems

Use Case Website/s

N/A

Open for Collaboration?

Interested in collaboration with other Commonwealth entities

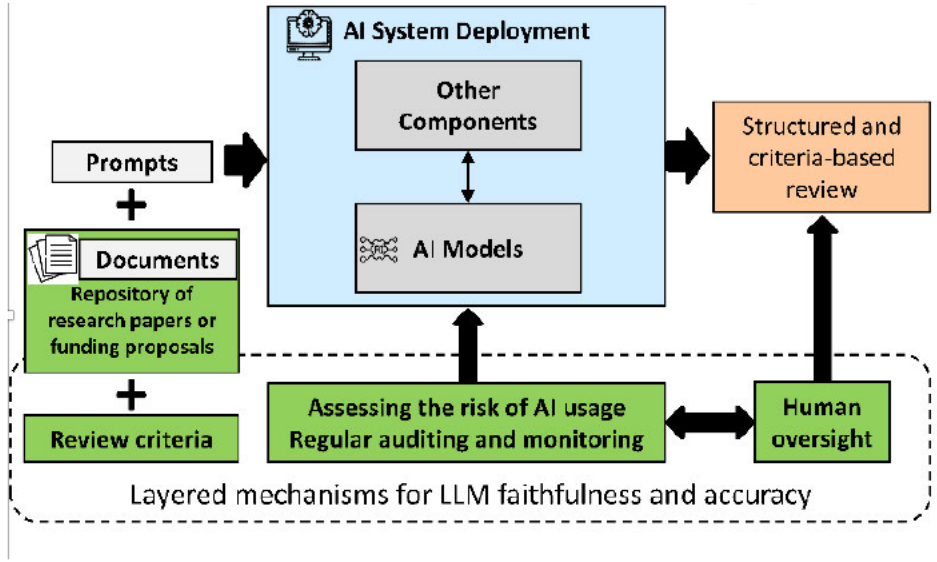
Use Case Contact

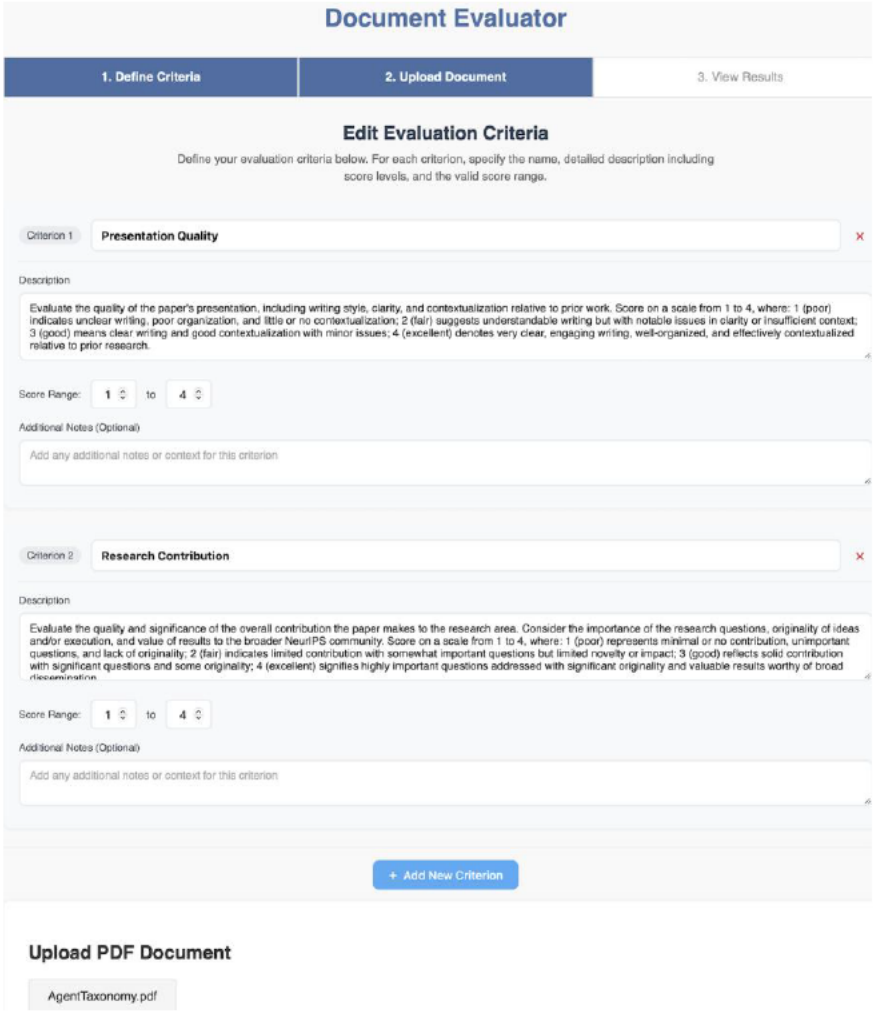
Liming Zhu: Liming.Zhu@data61.csiro.au
Qinghua Lu: Qinghua.Lu@data61.csiro.au

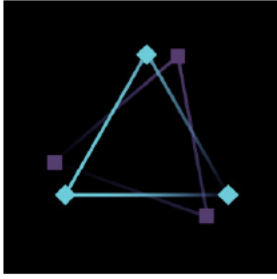
Use Case Owner

Liming Zhu: Liming.Zhu@data61.csiro.au

Screenshot/s







Detailed Overview

Version Control

Version	Date	Author	Description of Changes
1.0	3 Feb 2025	GovAI	Version 1 created
1.1	17 Mar 2025	GovAI	Modified based on feedback

Index

Note: For details about category items in the detailed overview, see *APS AI Use Case Repository Guidance- Guidance for Use Case Owners and Editors*.

Responsible Organisation Category

Select the Classification of the Functions of Government – Australia (COFOG-A) 3-digit category that best identifies the functional area associated with your AI use case.

☒ **01 - General Public Services**	019 - General public services (other)
☐ **02 - Defence**	Choose an item.
☒ **03 - Public Order and Safety**	035 - Research and development in public order and safety
☐ **04 - Economic Affairs**	Choose an item.
☒ **05 - Environmental Protection**	055 - Research and development in environmental protection
☒ **06 - Housing and Community Amenities**	065 - Research and development in housing and community amenities
☒ **07 - Health**	077 - Research and development in health
☒ **08 - Recreation, Culture, and Religion**	085 - Research and development in recreation, culture, and religion
☒ **09 - Education**	096 - Research and development in education

☒ **10 - Social Protection**	107 - Research and development in social protection
☒ **11 - Transport**	118 - Research and development in transport

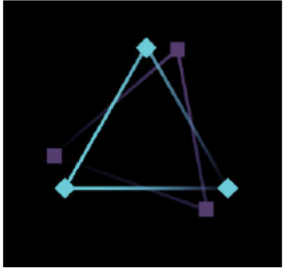
Scope of the Use Case

Use the dropdown menus below to identify the scope of your use case.

Geographical focus Choose the region for implementation from the dropdown list	Across countries
Primary type of government interaction Choose the type of government interaction from the dropdown list	Government-to-government (G2G)
Cross-features - Sector Indicate if the use case describes a solution that can be used across sectors or in cross-sector scenarios (Yes/No).	Yes
Cross-features – Jurisdiction Indicate if the use case describes a solution that can be used across State/Federal borders or in cross-border scenarios (Yes/No)	Yes

Ethical Considerations

Accuracy, Fairness, Accessibility, Bias and Discrimination	We ensure accuracy, fairness and prevent bias through comprehensive evaluation and real-world validation and regular audit and monitoring. The system is accessible.
Privacy	No personal data will be stored or shared, and all processing follows relevant legal and regulatory requirement.
Rights of Users	Users will be informed their rights through clear documentation. Feedback mechanisms are built into the system, allowing users to report issues or provide input. Users can challenge AI decisions through designated review channels involving human oversight.



Value of the Use Case

Identify the public value that the solution provides or is expected to provide. Select from the multi-select options.

Improved public service This category refers to solutions that enhance the services provided to end users, whether they are citizens or businesses.	☐ Personalised services ☐ Public (citizen)-centred services ☒ Increased quality of public information and services ☒ More responsive, efficient and cost-effective public services ☐ New services or channels
Improved administrative efficiency This category refers to solutions that increase efficiency, effectiveness, and quality while reducing costs within administrative processes, systems, and services.	☒ Cost reduction ☒ Responsiveness of government operation ☐ Improved management of public resources ☒ Increased quality of processes and systems ☐ Better collaboration and better communication ☐ Reduced risk of corruption and abuse of the law by public servants ☒ Greater fairness, honesty and equality enabled
Open government capabilities This category refers to solutions that enhance the level of openness, transparency, engagement, and communication within public organisations.	☐ Increased transparency of public sector operations ☐ Increased public participation in government actions and policymaking ☒ Improved public control of and influence on government actions and policies

AI Process Type

Select the types of tasks within government operations that the AI solution is performing or expected to perform

Supporting Decision Making- Tasks that support formal or informal agency decision-making on benefits or rights.	☒ Taking decisions on benefits ☐ Managing copyright and intellectual property rights

Analysis, monitoring and regulatory research – Tasks that collect or analyse information that shapes agency policymaking.	☒ Information analysis processes ☐ Monitoring policy implementation ☐ Innovating public policy ☐ Prediction and planning
Enforcement – Tasks that identify or prioritise targets of agency enforcement action.	☒ Smart recognition processes ☒ Management of auditing and logging ☐ Predictive enforcement processes ☒ Supporting inspection processes ☐ Improving cybersecurity ☐ Registration and data notarisation processes ☐ Certification and validation processes
Internal management – Tasks that support agency management of resources, including employee management, procurement, and maintenance of technology systems.	☐ Internal primary processes ☒ Internal support processes ☐ Internal management processes ☒ Procurement management ☐ Financial management and support
Public services and engagement – Tasks that support the direct provision of services to the public or facilitate communication with the public for regulatory or other purposes.	☐ Engagement management ☐ Data-sharing management ☐ Governance and voting ☐ Payments and international transactions ☐ Supporting disintermediation ☐ Authentication of self-sovereign digital ID services ☐ Service integration ☒ Service personalisation ☐ Tracking of goods and assets along the supply chain

AI Technologies Utilised

Select the types of AI technologies proposed / utilised to deliver the use case.

Reasoning or Knowledge Representation AI systems that store, structure, and process knowledge to make inferences, derive conclusions, or support decision-making.	☒ Knowledge Representation ☒ Automated Reasoning ☒ Commonsense Reasoning
Planning and Optimisation AI techniques that generate, refine, and optimise action sequences or resource allocation to achieve specific goals efficiently.	☒ Planning and Scheduling ☐ Searching ☒ Optimisation
Learning and Adaptation AI systems that identify patterns, extract insights, and improve performance over time based on data.	☐ Machine Learning ☐ Deep Learning ☒ Generative AI
Communication and Natural Language Processing AI systems that process, interpret, and generate human language for interaction, comprehension, and automation.	☒ Natural Language Processing (NLP) ☒ Text Generation ☐ Text Mining ☐ Machine Translation
Perception through the Senses AI systems that process and interpret sensory data, such as visual, auditory, or tactile inputs, to understand and respond to their environment.	☐ Computer Vision ☐ Audio Processing
Integration and Interaction with the Environment AI systems that interact with physical or digital environments, including autonomous agents, robotics, and interconnected systems.	☒ Multi-agent Systems ☐ Robotics and Automation ☐ Connected and Automated Vehicles (CAVs)
AI as a Service AI capabilities delivered through cloud-based platforms, offering tools, models, and infrastructure for AI-powered applications.	☒ AI Services (e.g., cognitive computing, machine learning frameworks, bots) ☐ Infrastructure as a Service (IaaS) ☐ Platform as a Service (PaaS) ☐ Software as a Service (SaaS)

Additional Comments or Explanation:	If you have selected any of the subcategories above, feel free to provide more detailed comments or a description of how these elements apply to your specific use case.

Technical Elements

Platform implementation	Designed to be deployed on secure cloud platforms like AWS, Azure, or GCP, enabling scalability and high availability. *On-premises deployment*: For organizations requiring strict data governance, an on-premises deployment option is supported.	
Model / Algorithm used	We use large language models like GPT, Claude with Retrieval Augmented Generation (RAG) framework. The system will implement different level of mechanism to ensure the LLM faithfulness and perform input filtering to defend adversarial inputs.	
Data Sources Select the types of data sources used and provide relevant details.	☐ Internal ☒ Public	☐ Third-party ☒ Synthetic
	Details: Training and testing will be done on the hybrid of public and synthetic dataset.	
Risk Assessment and Mitigation Details	Challenges: • Risk of data leakage when using cloud computing services. • Model hallucinations or misinterpretation of protected information. • Reliability and AI bias Mitigations: • Use local/managed LLMs • Implement prompt sanitization by applying pattern-based filters and entity redaction (e.g., replacing confidential data with placeholders).	

	• Apply human oversight to selectively validate and monitor • Monitor the performance and outcomes across diverse dimensions. • Decision process and reasoning trace logging for auditing	
Security and Compliance Frameworks Select the security and compliance frameworks and measures implemented. Provide details or additional artifacts if relevant.	☐ Authority to Operate (ATO) ☐ System Security Plan (SSP) ☐ Security Risk Management Plan (SRMP)	☐ Information Security Registered Assessors Program (IRAP) ☐ Penetration Testing
	Details:	
Assurance and Government Frameworks	Identify the assurance or government frameworks you have evaluated your use case against so far (e.g., DTA AI Assurance Framework, AI in Government Policy, National AI Assurance Framework).	
Record maintenance	Provide an overview of documentation practices for AI decisions, testing, and data assets.	
Disengagement	Describe any contingency plans for disengagement in the event of critical failures (N/A if in early planning stage)	
Performance Metrics and Results	Quantitative Metrics • Correlation, Agreement Metrics, Accuracy, precision, recall, Argument Match rate, to evaluate the quality of the generated reviews • Pass rates to understand the impacts on the final results • Time saving Qualitative Metrics • User Satisfaction: Based on user study and feedback	