



# Software Systems Research Group

Qinghua Lu

Group Leader

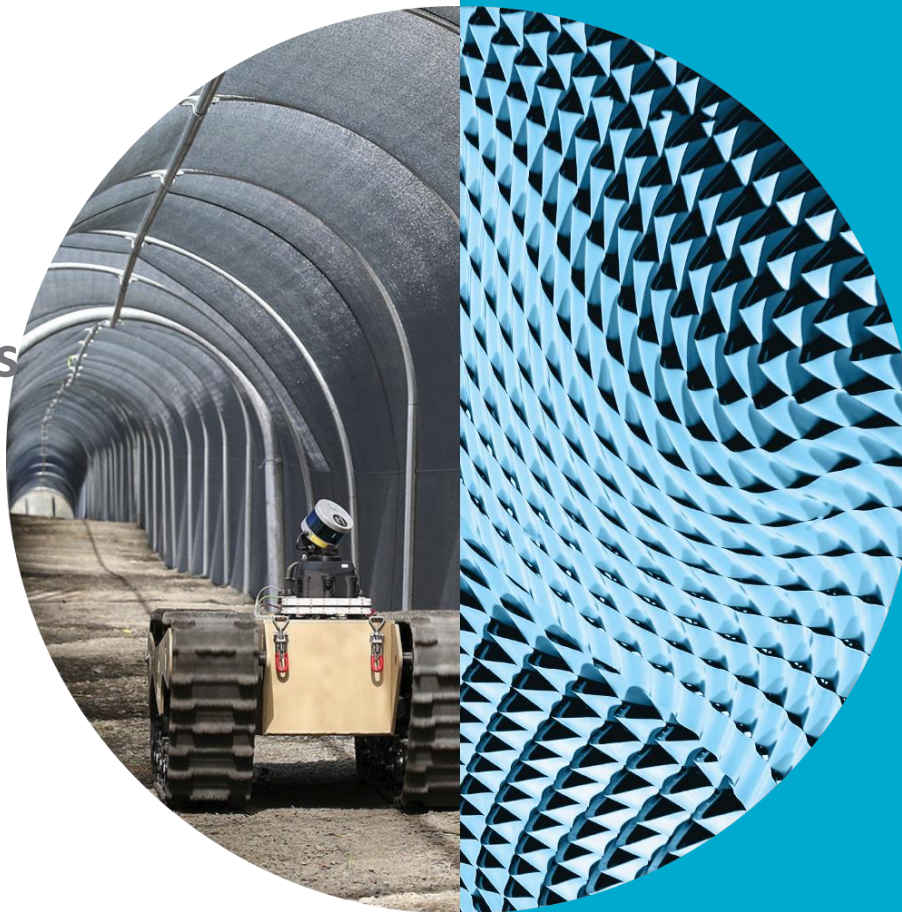
Software Systems Research Group

Data61, CSIRO

[qinghua.lu@data61.csiro.au](mailto:qinghua.lu@data61.csiro.au)

<https://research.csiro.au/ss/>

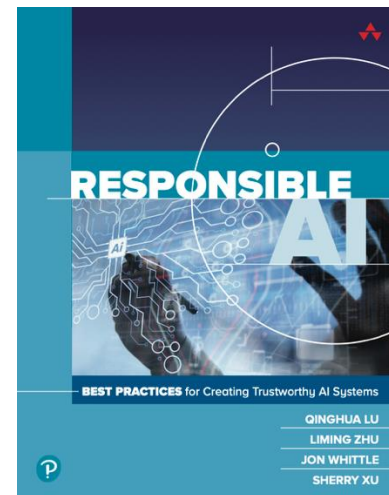
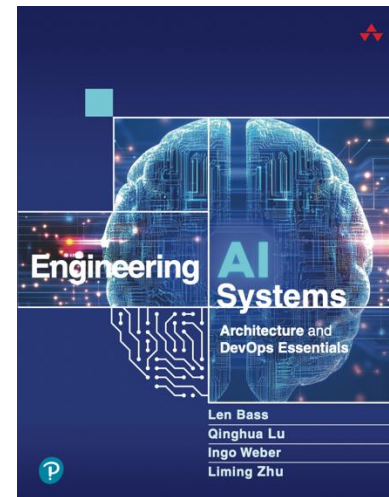
Australia's National Science Agency





# Group Overview

- ~40 full time research scientists/engineers
- 5 Research Teams
  - Software Engineering for AI Research Team
  - Applied AI Systems Research Team
  - Trustworthy Processes Research Team
  - Architecture and Analytics Platform Research Team
  - AI Diversity and Inclusion Research Team
- 6 scientists in the global top 20 for Responsible AI





# Global Leadership in Responsible AI & AI Engineering

- Contributing to International Working Groups
  - Frontier Model Forum
    - 20+ leading orgs including OpenAI, Google, Microsoft, Anthropic, Meta, AI Safety Institutes
  - International Network of AI Safety Institutes
    - Ongoing collaboration with US, UK, Canada, Japan, and Singapore AISI
  - AI Metrology
    - With Alan Turing Institute, MIT, Mila, etc
  - OECD.AI
  - EU General-Purpose AI Code of Practice
- Organising/Editorial Roles in Academic Conferences/Journals:
  - International Conferences on AI Engineering, AI Powered Software (AIware)
  - International Workshops on Responsible AI Engineering, Agentic Engineering
  - IEEE Transactions on AI, Engineering Applications of AI

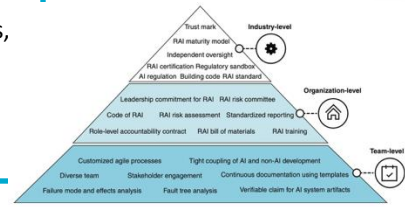


# Key Responsible AI and AI Engineering Projects

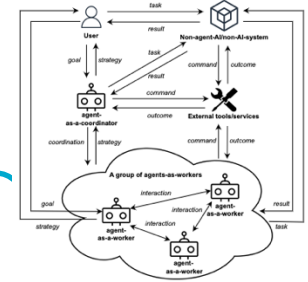
- Government Projects
  - DISR
    - Australia's AI Safety Standards (v1 and v2)
  - South Australian Gov and AIML
    - Responsible AI Research Centre
  - Responsible AI with Australian Sports Commission
- Industry Projects
  - ESG-AI project with Alphinity Investment
  - Tax copilot with Empathetic AI
  - Westpac/Cognitivo on RAG engineering



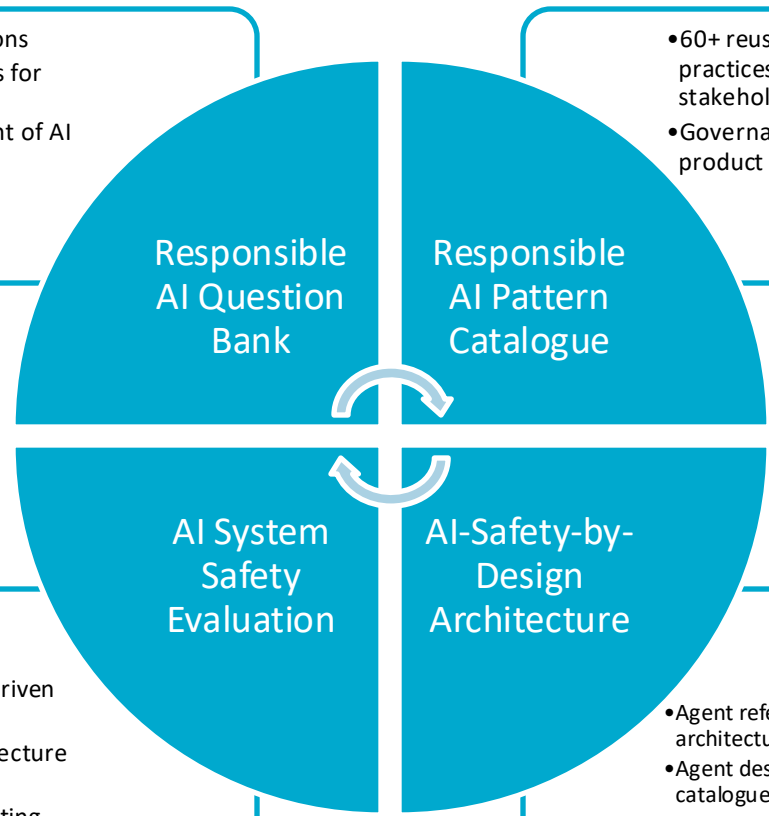
# Responsible AI Methods/Frameworks



<https://research.csiro.au/ss/science/projects/responsible-ai-pattern-catalogue/>



<https://arxiv.org/abs/2311.13148>  
<https://arxiv.org/abs/2405.10467>  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5266496](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5266496)

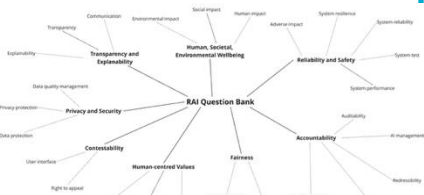


- 300+ questions
- 100+ metrics for concrete measurement of AI risks

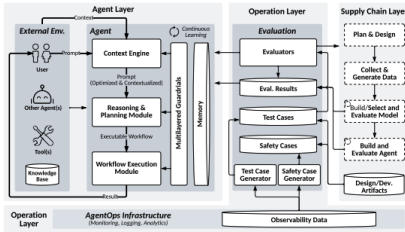
- 60+ reusable best practices for different stakeholders
- Governance, process, product

- Evaluation-driven learning
- Agent architecture evaluation
- AIS joint testing

- Agent reference architecture
- Agent design pattern catalogue
- Swiss Cheese Model for multi-layered guardrails

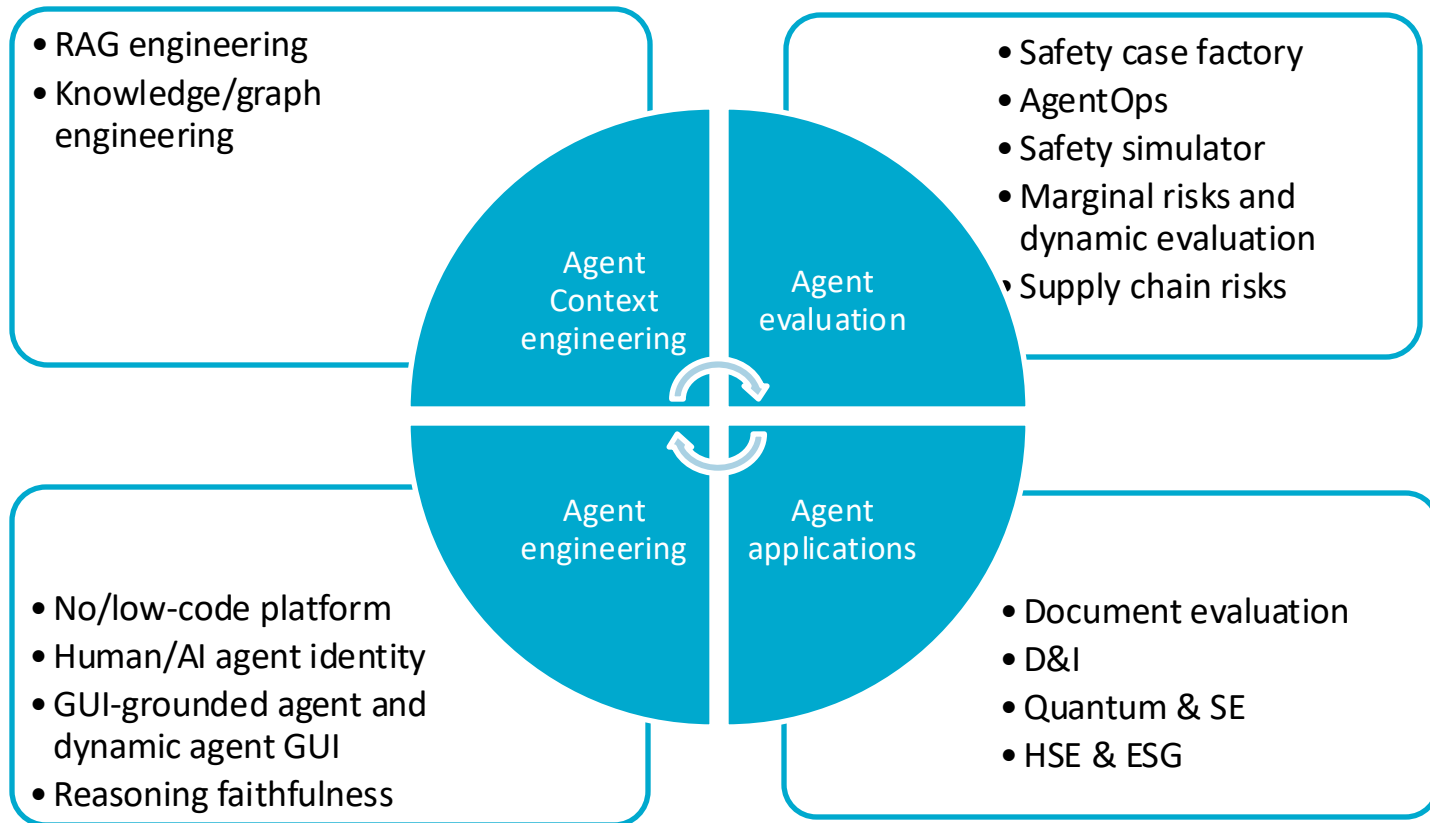


<https://arxiv.org/pdf/2305.09300.pdf>  
<https://arxiv.org/abs/2311.13158>



<https://arxiv.org/abs/2411.13768>  
<https://arxiv.org/abs/2404.05388>

# Agent Engineering





# Thank you.

Qinghua Lu

Group Leader

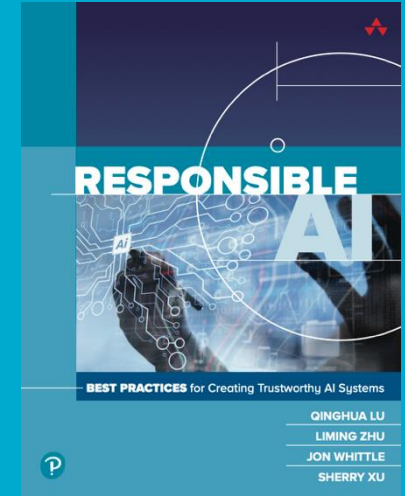
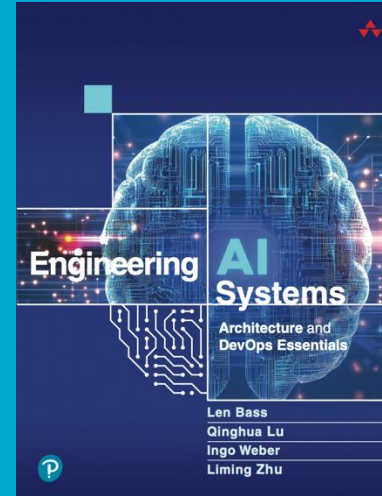
Software Systems Research Group

Data61, CSIRO

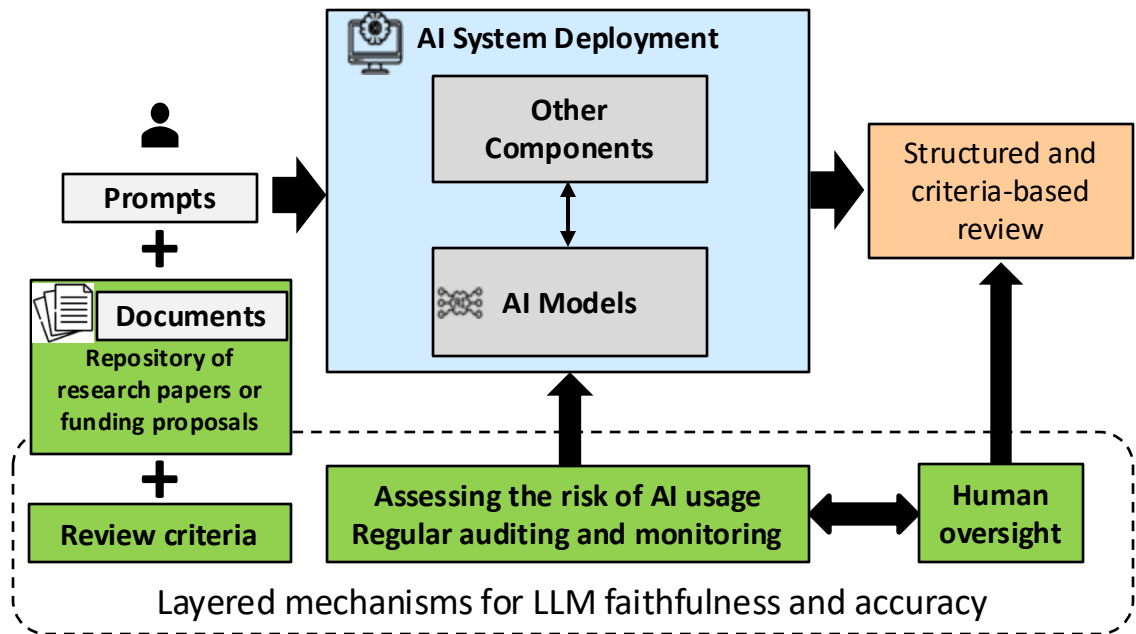
[qinghua.lu@data61.csiro.au](mailto:qinghua.lu@data61.csiro.au)

<https://research.csiro.au/ss/>

Australia's National Science Agency



# AI for Document Evaluation



## Value Proposition

1. An AI tool to **review large volumes of documents**—such as research papers or funding proposals—against specific rubrics or policy criteria.
2. **Human oversight** to ensure relevance, quality, and trust in generated reviews

## Technical Elements

1. Use large language models like GPT and Claude
2. Implement different levels of mechanisms to ensure the LLM faithfulness and accuracy



[<- Back to Evaluators](#)

## Evaluator Configuration

**Paper evaluation**
Ed

Evaluate papers based on some criteria

### Criteria

Extract from Text
Add Crit

**Reader Interest**

Evaluate what fraction of the Technical Committee membership will be interested in the subject of the paper. Score range is 0 to 10, where 0-2 means very low or no interest, 3-5 means limited or fair interest, 6-8 means good interest from a significant portion of the committee, and 9-10 means high or broad interest across the majority of the committee. Indicators include relevance of topic to the committee, appeal to diverse members, and alignment with current technical focus.

Score Range: 0 - 10

Edit
Del

**Subject Importance**

Assess the importance and timeliness of the subject matter and contribution to the field. Score 0-2 means trivial, outdated, or marginally valuable; 3-5 means somewhat important but limited in scope or value; 6-8 means important, timely, and valuable; 9-10 means critically important, timely, and offering significant value to a broad audience. Look for evidence of contribution to understanding, relevance to current developments, and significance beyond niche topics.

Score Range: 0 - 10

Edit
Del

Customizable Criteria Based Evaluation

## documents

### Add New Document

Document Name

Review guideline

Document Type

Criteria Guidance

**Purpose:** Provides detailed guidance on how to evaluate each criterion. Include specific instructions, examples of what constitutes different score levels, and any special considerations for scoring.

File

Choose File no file selected

Supported formats: PDF, Word documents, text files. Text will be automatically extracted and converted to markdown.

Cancel
Add Document

Multi-Source Document Support

[< Back](#)

## Evaluation Results

AgentTaxonomy.pdf

### Evaluation Summary

Status

Completed

Overall Score

7.8

Document

AgentTaxonomy.pdf  
1.65 MB

### Evaluation Configuration

Evaluation Persona Expert	Evaluation Logic Reasoning Before Score	Aggregation Method Weighted Average
------------------------------	--	--

Processing time: 103.02 seconds

### Detailed Criteria Evaluation

**Reader Interest**

7.5/10  
Weight: 2

**Evaluation Reasoning:**

The paper focuses on a taxonomy and decision model for foundation model-based agents, addressing both functional capabilities and non-functional qualities within AI agent architectures. Given the surge in interest and deployment of foundation models (especially large language models like GPT-4, PaLM 2, LLaMA 2), the topic is timely and relevant, especially for committees invested in artificial intelligence, software architecture, and agent system design. The document aims to unify and standardize architectural design options in this emerging and fast-evolving area, which fills a recognized gap in literature and practice. The paper also provides a comprehensive framework that could appeal not only to specialists working directly on AI agents but also software architects and system designers concerned with incorporating foundation models reliably into products and workflows. Since foundation models and their agent-based applications impact diverse fields such as healthcare, finance, autonomous systems, and software

Transparent Evaluation with Justification