TruEVAL:

Streamlined, secure, and insightful evaluation of AI models and agentic solutions for smarter model selection

TruEVAL makes AI evaluation systematic, cost-effective, and repeatable—so you can focus on choosing the best AI model and mode for your needs, at an optimal price-point. Evaluate AI models and agentic solutions—quickly, securely.

- **No Expertise Needed:** Designed for non-technical and technical users, with a guided UI and intuitive workflows.

- **Customisable Criteria:** Use built-in evaluation parameters or define your own with ease.

- **Flexible Scheduling:** Run evaluations on demand or set up recurring assessments with minimal setup.

- **insightful Reporting:** Get clear, customisable dashboards and reports covering performance, accuracy, and cost.

- **Natural language interface:** Powerful, natural language conversation-based analysis of results, through MCP client access (such as Claude Desktop)

- **Repeatable & Scalable:** Evaluate across multiple datasets with minimal effort and maximum consistency

- **Plug-and-Play Evaluation:** Load test and ground-truth data effortlessly. Store and reuse predefined prompts with a single click.

- **Typical scenarios**:
  - Non-technical users need to evaluate AI models and modes (eg chat, instruct, specific size), to select the most appropriate for the use case in hand, from a performance, accuracy and cost perspective.
  - Need to evaluate the performance of alternative prompts on a specific AI Model and mode, against a specific set of data.
  - Evaluation of AI models that drive Agentic AI solutions, to optimise process accuracy, speed and execution costs.

- **Scalability, Availability and Security built-in by design:**
  - An enterprise-ready SaaS solution that uses cloud-native AWS services for unparalleled scalability, availability and reliability
  - Cloud based levels of security and observability, underpinned by best-practice design

- **Integrations:**
  - Full out–of-the-box integration with our 'True' range of solutions for enhanced and continuous AI model and TrueAGENT performance evaluation.
  - Integration with MCP clients (such as Claude Desktop), for natural language-based analysis of the evaluation findings and reports.