Advai

Assurance of Third-Party AI Systems.

Report for National Security, but with learnings for the private sector.

The Alan Turing Institute

The report arrives at a recommended System Card template.

It’s designed for the high standards of the national security sector.

However, there is excellent insight for private sector organisations seeking the advantages of external AI vendors.

Benefits and risks from working with third party vendors.

Benefits:

1. Time Savings
2. Cost Reduction
3. Skill Gap Filling
4. Interoperability
5. Talent Attraction
6. Safety Identification
7. Increased Compute Power
8. Access to Innovation

Risks:

1. Dependency Risks
2. Supply Chain Complexity
3. Data Representation Issues
4. Adversarial Vulnerability
5. Legal Knowledge Gap
6. System Compatibility
7. Ethical Alignment
8. Reputation Risk
9. Skill Erosion

Let’s define ‘AI Assurance’

In the report, AI assurance is defined as the portfolio of processes required to evaluate and communicate, iteratively throughout the AI lifecycle, the extent to which a given AI system:

- Does everything it says it is going to do, and nothing it shouldn’t do.
- Complies with the values of the deploying organisation.
- Is appropriate to the specific use case and envisioned deployment context.

THE CHALLENGE

Understanding the vendor’s technology

Transparency from suppliers on the features (and weaknesses) of their AI systems

Clear responsibilities in the assurance process.

How to implement AI Assurance:

Build infrastructure for a sustainable assurance ecosystem, including a portfolio of assurance techniques.
#advai

Invest in skills for reviewing assurance cases (technical, ethical, and legal).

Connect assurance techniques from academic work to people solving practical challenges.
(#advai)

Showcase ample examples of how assurance recommendations apply in context.

Draft bespoke legal contracts to ensure vendor transparency.

The report compared strengths and weaknesses of 3 methods of documenting AI system properties.

Model Cards.

Straightforward and standardised, but perhaps too simple and subjective.

System Cards.

More encompassing, but less structured and hard for non-technical audiences.

Argument-based assurance.

Emphasises evidence and clarity, but it’s complex, difficult and time consuming.

In result, they propose an ideal method to document AI system properties.

Balance the level of detail, be interpretable to non-experts, ensure consistent structure, accommodate context-specific flexibility, build on industry practice but push for more transparency, clarify integration with other processes (i.e. legal and procurement)

System Card Template

The full template is several pages long. Here are the categories.

1. Summary information
2. Mission properties and legal compliance
3. The supply-chain
4. Performance and security
5. Ethical principles
6. Iterative requirements

Bring AI Assurance to your business with Advai.

We can align your AI adoption with the best practices, frameworks and industry standards.

PROPIETARY APPROACHES

State of the art AI assurance techniques wrapped into development environments. Novel methods to test, stress and break AI Models.

AUTOMATE STRESS TESTS

Our assurance and adversarial capabilities allow users to automate the discovery of model vulnerabilities.

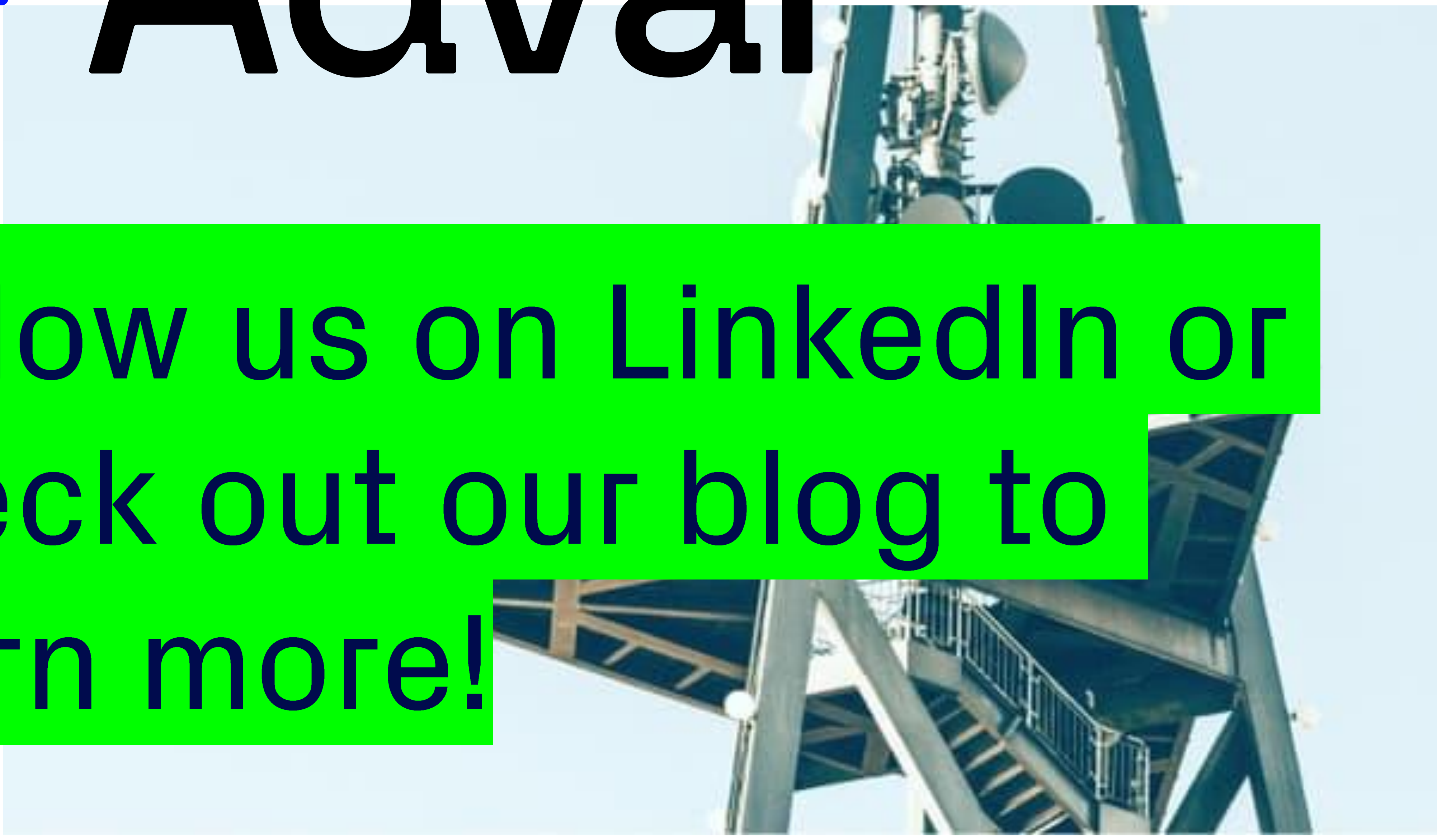
BRIDGE STAKEHOLDERS

Connect assurance test results to high-level insights. Enable management to interpret technical intricacies and make informed decisions.

GUARDRAIL CREATION

Use ‘Advai score’ breakdowns to isolate weak spots to inform when a model shouldn’t be used. Customise model selection for a mission or purpose.

Advai



Follow us on LinkedIn or check out our blog to learn more!

Advai.co.uk/journal

Or get in touch if you'd like to discuss AI safety at your organisation.
contact@advai.co.uk