PAC

WHITEPAPER

The Responsible AI Imperative

Spencer Izard, PAC

June 2024

Infosys®

Navigate your next

Contents

This whitepaper was commissioned by Infosys. For more information, please visit www.pacanalyst.com.

Independence and data protection
This whitepaper was produced by Pierre Audoin Consultants (PAC). The ordering party had no influence over the production of the paper.

Preface

November 30th, 2022 was a profound day in the world of technology because a pivotal shift in awareness of AI occurred, unlike what had occurred before. From that point forward, due to global media coverage, all tiers of society became aware of the potential of AI to change how society behaves in our professional and personal lives. The release of ChatGPT by OpenAI drew global attention to a form of AI based on large language model (LLM) technology called generative AI (GenAI). Like many innovative technologies before it, competition amongst technology firms to create and provide GenAI solutions has occurred, but unlike prior technologies, this has happened at a highly accelerated pace. Whilst the rapid pace of innovation for many can be inspiring, PAC considers that for all the benefits of GenAI, society has entered a period of higher risk in part due to security vulnerabilities, reliability, explainability, and ethical issues both already understood and yet to be identified. As this is an Infosys-sponsored report, PAC also considers it important to mention that the company's global breadth and depth of capabilities place it at the heart of the risk exposure mentioned because it delivers internal AI deployments, co-creates solutions with partners, and, of course, delivers customer projects. By being impacted in these three dimensions, Infosys is fully aware that an answer to these risks is needed and has created a structured set of programs, policies, process interventions and technical guardrails under an overarching framework called AI3S (AI scan, shield, and steer). With a focus on providing a means to achieve strong AI governance and management across all aspects of how an organisation leverages AI and the three pillars of the framework are:

- **Scan:** A cohort of accelerators and solutions are leveraged for scanning internal compliance via reviews, assessments and audits, a centralised single source of truth for critical metrics for all AI projects as well as collecting market intelligence on recent vulnerabilities, regulatory changes and risks.
- **Shield:** Helps in building technical guardrails, and accelerators for protecting AI models from vulnerabilities. These guardrails span across all of the AI lifecycle stages from data preparation, and training to testing and inferencing.
- **Steer:** This cohort of consulting and advisory services helps in setting up a strong AI governance, via policies, process interventions and changing the ways of work by building a dedicated Responsible AI function.

Infosys has adopted AI3S internally and forms the basis for the umbrella of offerings that provides to organisations seeking help in accelerating RAI adoption.

As the direct and indirect adoption of artificial intelligence (AI) continues to spread within organisations across all industries, PAC understands from senior leaders that addressing ethical operating concerns regarding its use continues to grow in importance from both regulatory and societal perspectives. This has led to the growing awareness of and demand for the concept of responsible AI (RAI) as a critical imperative to ensure successful AI adoption at scale across the multitude of business processes and technologies within an organisation. Over the last decade, corporate social responsibility (CSR) has become a more prevalent part of an organisation's operating model as governments and industry regulatory bodies have increased their demands for transparency regarding organisations' ethical and sustainable behaviours.

Like many aspects of AI that have grown to mass awareness over the last couple of years, the concept of RAI dates back to the origins of the creation of AI within academia. However, now that AI has grown in capability and awareness to become a mass-market software tool for cost-effective use across organisations, the role of RAI has rapidly increased in importance. Broadly speaking, given that societal and governmental ethics can vary significantly between countries, PAC has seen concerns arising from organisational leaders regarding the potential impact of AI on their workforce and society relating to privacy violations, algorithmic biases, and the displacement of human labour. In recent years, this has led to the European Union and industry bodies like the Institute of Electrical and Electronics Engineers Inc (IEEE) taking proactive steps to establish ethical guidelines, sometimes called guardrails and regulatory frameworks.

Like many matters relating to AI, the concept of RAI has evolved because of the growing recognition that the development and use of AI capabilities must be guided by a set of core ethical principles that serve as a moral compass focused on upholding fundamental human rights, promoting fairness and non-discrimination by fostering trust and providing transparency. However, PAC would like to emphasise that not all cultures, societies, and governments share the same core ethical principles relating to AI. So, determining a consistent global use of RAI is unlikely to occur. PAC also considers that this challenge

relating to RAI will also create complexity as organisations operating internationally will likely have to adapt their usage of RAI to align with the ethical principles set out by each country in which they operate. This, in turn, could add complexity internally and when an organisation uses AI solutions and services adhering to core ethical principles of one country/region when engaging with an external partner organisation operating out of another country/ region that adheres to different core ethical AI principles.

However, despite the continued rapid onset of AI adoption into the operating dynamics of organisations, PAC is still witnessing a lag regarding the establishment of core ethical AI principles from both governmental and industry body levels. As the adoption and use of AI continues to evolve rapidly, PAC expects that regulations at all levels will continue to change and evolve to address the evolutionary nature of how AI as technology is used. This means that the role of responsible AI will not only be critical to an organisation but will very likely be a complex and ongoing concern where consuming it as an "ever-green" service from a trustworthy service provider partner is the most pragmatic and cost-effective means of adoption and ongoing operation. Whilst regulatory bodies and governments worldwide have recognised the need to establish guidelines and frameworks for the use of AI, PAC would argue that this is being driven in significant part by the level of public awareness and scrutiny of AI practices, with high-profile incidents fuelling demands for greater transparency, consistency, and accountability.

As senior leaders navigate the AI-driven future, PAC believes they must embrace RAI as a strategic imperative. By fostering a culture of transparency, accountability, fairness, privacy protection, and human oversight, organisations can mitigate risks and unlock AI's potential to drive innovation, enhance customer/ client/citizen experiences, and create sustainable value. Responsible AI is not a passing trend but a fundamental pillar for the future of business. Senior leaders who recognise its significance and embed it into their organisational DNA will be well-positioned to navigate the AI revolution, foster stakeholder trust, and establish a competitive/ operational advantage.

Why Governance and Ethics are Needed for Trustworthy AI

The use of responsible AI (RAI) within an organisation is achieved through a **combination of technology and,** importantly, **governance.** Whilst RAI itself can be seen as a means to govern machine and deep learning AI models ethically, there is also a need for an over-arching human-led governance framework atop an RAI solution or service focused on both the development and deployment of AI capabilities. Given the nascent use case nature of deep learning AI, for example, effective human-led governance is essential for up-holding core ethical principles, mitigating risks, and fostering ongoing stakeholder trust in an organisation's AI practices.

PAC advises that at the heart of RAI governance lies the **principle of fairness,** which demands that AI capabilities be designed and deployed in a manner that does not perpetuate or exacerbate existing biases or discrimination aligned to those determined by both the organisation and country/region in which an AI is operating. Cultural, legal, and regulatory frameworks can vary significantly from country to country, and what is considered ethical in one context may be viewed differently in another.

It is essential to adopt either a **dedicated RAI governance framework or its principles into an existing AI framework** to ensure processes exist to identify and mitigate potential sources of bias, whether they stem from the training data, algorithmic models, or human decision-making processes. Ongoing monitoring and evaluation are crucial to ensure that AI capabilities operate fairly and equitably, promoting diversity and equal opportunities for all stakeholders.

However, an RAI governance framework also must focus on accountability as a means to ensure fairness, which requires organisations to take **responsibility for the actions and consequences of how adopted AI capabilities are used, either embedded within business applications or developed in a bespoke manner.** An organisation's senior leadership team (SLT) must establish clear lines of accountability, ensuring that appropriate individuals or teams are responsible for AI technologies' ethical development, deployment, and ongoing oversight. PAC considers that the function

of accountability extends not only to legal and regulatory compliance but also to broader societal and ethical considerations, encompassing the potential impacts of AI capabilities on individuals, communities, and the environment.

From PAC's perspective, the third pillar of effective AI governance, through the use of RAI, is that of **AI output transparency,** which provides the means for an organisation to achieve the governance functions of fairness and accountability regarding overall AI usage. RAI provides transparency by enabling stakeholders to understand the decision-making processes of AI capabilities. This is achieved by ensuring AI capabilities are explainable and auditable by providing insights into data inputs, algorithmic models, and decision-making criteria for both machine and deep learning AI models. **Explainability and reproducibility** are critical functions of responsible AI to ensure that both internal and external parties can have the utmost trust in using AI across all an organisation's processes. It also allows organisations to identify and address issues quickly and at an operational scale.

Finally, **safeguarding data privacy rights** is critical to responsible AI governance, as AI capabilities often process vast amounts of personal data. Senior leaders must prioritise or further reinforce the usage of robust privacy safeguards and data governance practices to protect the privacy rights of individuals and maintain public trust. This includes adhering to relevant data protection regulations, embracing principles such as data minimisation, purpose limitation, and secure data handling practices, and implementing appropriate access controls and security measures.

From a governance perspective, the why is just as important as the how, and from PAC's engagement with Infosys, it is clear that this is core to the value of their RAI-related solutions and services regarding governance. For example, they rightly position the role of RAI as a centralised function that, to be successful, collaboratively spans a range of core departments, including legal, data, procurement, and cyber-security teams. In particular, Infosys' capabilities provide a foundation to clearly define policies that support not just the specific RAI needs of one department but a range of critical paths spanning the interactions between one or more departments, focusing on areas like crisis management and capability evaluations. For PAC, this is where the scale and capability of an organisation like Infosys become invaluable because their RAI offerings have both internal and third-party audits (i.e., ISO 42001:2023), the firm actively builds ecosystems across a range of AI-related companies and drives discourse, has market monitoring capabilities to sense externally for upcoming vulnerabilities, and ensures they are always up-to-date with regulatory compliance across all regions like the EU AI act.

> **Given the nascent use case nature of deep learning AI, for example, effective human-led governance is essential for upholding core ethical principles, mitigating risks, and fostering ongoing stakeholder trust in an organisation's AI practices.**

How to Navigate Operationalising RAI

Over the last several decades, with every new technological innovation, the term "silver bullet" has grown in usage to describe a technology that is considered to solve all the problems relating to a certain matter. To this end, PAC considers it critical that organisations planning to adopt responsible AI (RAI) do not see it as a "silver bullet" to manage AI ethically. This is because, as has been seen time and time again, a technology solution or service in isolation rarely solves all, but rather, it requires a range of human-led activities to be created or realigned to gain the most business value.

In this respect, integrating RAI into business operating models requires a holistic approach encompassing the entire AI lifecycle and business workflows, from ideation to development and ongoing usage. As discussed, to operationalise RAI, a clear governance framework and decision-making process must be embedded at every stage to assess, protect, and advise on the usage of all forms of AI across the whole organisation.

When using RAI operationally, it is crucial to conduct rigorous **risk assessments and impact analyses** to identify potential ethical concerns, biases, and unintended consequences during AI development. Such a proactive approach enables organisations to address issues before they become entrenched in an AI model behaviour, reducing the occurrence of costly remediation efforts down the line, often referred to as 'regret costs'. PAC considers it essential that organisations adopt privacy-by-design and fairness-aware AI development methodologies in conjunction with the use of RAI. Given the still relatively nascent nature of AI development globally, these types of development methodologies are still not widely used, but they ensure that privacy protection, fairness, and non-discrimination are built into the heart of AI capabilities from the ground up, rather than being, as all too often occurs, treated as afterthoughts which are then harder to implement.

As with any AI technology, RAI requires a process of tuning both prior to adoption and on an ongoing basis to validate the underpinning AI models it uses to ensure it is functioning as intended and adhering to the organisation's ethical standards at any point in time. Operationally this includes the ongoing need for **rigorous bias testing, explainability checks, and scenario-based evaluations** to identify and mitigate potential RAI issues before they impact real-world decision-making. Of course, the approach needed to maintain the operational effectiveness of RAI is akin to what is needed to occur should the results of an RAI solution or service indicate that another AI capability requires tuning.

As mentioned in the prior section, effective oversight and governance mechanisms are critical to managing the use of responsible AI (RAI) and determining how to resolve any issues it indicates in the function of other AI models. From an operational perspective, PAC advises that an organisation create a **dedicated ethics board or committee** to bring relevant, diverse perspectives and expertise to guide the responsible adoption of AI capabilities and decision-making models underpinning RAI. Such an operational capability provides a means for proactive communication with stakeholders, regulators, and the general public to help build trust in how AI is used, address concerns, and foster a collaborative approach where appropriate. Continuous monitoring and evaluation processes are of paramount importance for how RAI is applied to analysing AI capabilities used across an organisation as well as to an RAI solution or service. As mentioned earlier in this report, shortages of skilled talent mean that it is important for senior leaders to prioritise internal education and training on responsible AI practices to upskill their workforce as needed.

PAC considers Infosys to have a strong posture for operationalising RAI because they have embedded a responsible-by-design (RBD) approach to provide a wide and sophisticated range of AI guardrails. Their quoted position is that "Integrating RAI into business operating models requires a holistic approach encompassing the entire AI lifecycle and business workflows, from ideation to development and ongoing usage". This statement reflects the requirements and challenges many of the organisation's PAC engages with and demonstrates why a company like Infosys can provide assistance when considering how to address RAI. The use of RBD practices when operationalising RAI is essential because it provides key capabilities like AI-model training to ensure data provenance and traceability; when engineering an AI model, it ensures robust model evaluation and validation and provides an approach that supports adversarial testing. Operationally securing AI is a critical part of the company's RBD approach, and one of the areas PAC feels is highly valuable is its use of purple teaming. This is where, in the context of cyber-security testing, a team of experts perform both the red and blue roles, which, from the PACs perspective, provides a stronger and deeper assurance activity for the ongoing operationalisation of RAI.

What Responsible AI Technologies and Techniques are Needed

Whether an organisation is adopting machine or deep learning forms of AI, the current models, broadly speaking, have a significant weakness that is one of the key driving forces for responsible AI (RAI) adoption. To PACs' knowledge, no AI solutions or services are one hundred percent explainable and completely free of any "black box" elements. Over the coming years and decades, PAC expects this to be addressed, but for the current moment and following years, the most important technology that is critical to the successful adoption of an RAI solution or service is another form of AI called **explainable AI (XAI).** This emerging and rapidly maturing field of AI aims to support AI capabilities in explaining the decision-making processes of AI models in an interpretable manner that is understandable to humans. By providing interpretable explanations for predictions or decisions made by AI models, XAI techniques aid in demystifying the "black box" nature of the majority of current AI capabilities. Which, in turn, when used as part of RAI, plays a crucial role in mitigating risks and fostering stakeholder trust by enabling greater transparency and accountability. Given the ethical and legal ramifications an organisation is responsible for from the output of AI models, the role of XAI as part of an RAI solution or service facilitates the identification and mitigation of potential biases or unintended consequences.

While the demand for XAI is broadly required for all forms of AI, federated learning is another AI technique used for RAI and is specifically focused on addressing **privacy concerns** in AI models. The use of federated learning has grown in importance because of risks that have arisen relating to machine learning models often requiring centralised access to large data sets, leading to privacy risks when dealing with sensitive data. The value of federated learning, however, is that it enables training AI models on decentralised data, where the data remains on individual devices or servers, and only the model updates are shared rather than all the data. This decentralised approach significantly reduces privacy risks and enables organisations to leverage data from multiple sources while maintaining strict control over sensitive information.

As discussed in this report, data privacy remains paramount for the ethical use of AI. To this end, whilst not a specific technology, there is a powerful mathematical technique called **differential privacy** that is used for training AI models to protect the privacy of individuals within a dataset. This technique is an essential component of an RAI solution or service because it introduces carefully calibrated noise or perturbations to data sets to ensure that the presence or absence of any individual data has a negligible impact on the overall results provided by an AI model. This, in turn, provides organisations with a means to share and analyse data while providing robust privacy protections associated with risks relating to data breaches and unauthorised access.

In conjunction with the use of differential privacy is an adjacent technique called **secure multi-part computation (MPC)** that allows multiple parties to collaboratively compute an AI model without revealing their input data to each other. MPC enables organisations to leverage combined datasets for model training or analysis, maintaining strict confidentiality and preventing unauthorised data sharing. Regarding RAI, this technique is particularly valuable in scenarios where data is distributed across multiple business entities or country jurisdictions, each of which may have its own privacy and data protection requirements to operate an AI capability or share data with one another.

All of the prior technologies and techniques mentioned have been focused on their role as part of operating a RAI solution or service. However, as with any technology, RAI is at risk of being hacked or influenced by bad actors. As such, **securing an RAI solution or service and the AI capabilities it supports is critical.** An organisation's AI capabilities will continue to be integrated into critical decision-making processes, exposing them as points of interest for security threats and adversarial attacks. To this end, adopting an RAI solution or service that provides multiple layers of security protection focused specifically on protecting models from direct and indirect malicious manipulation, such as adversarial examples or model poisoning attacks, is essential. To enhance the resilience and integrity of AI capabilities, using RAI, a focus on providing secure training methodologies/ techniques, input sanitisation, and defensive distillation will ensure a strong posture against potential threats.

To ensure their RAI solutions and services address explainability and security, Infosys addresses this through a range of sophisticated run-time guardrails, which is especially important for generative AI (GenAI), ranging from retrieval-augmented generation (RAG) to chain-of-thought/forward (COT/F) to support their responsible AI (RAI) solutions and services. PAC considers this approach, which Infosys continues to improve upon as the technologies evolve iteratively, to

be very comprehensive in addressing AI explainability's current challenges as organisations adopt GenAI capabilities. As with any technology, GenAI is not immune to security threats ranging from adversarial to jailbreaking attack styles. To address this, Infosys has focused on incorporating a range of frameworks into their RAI solutions and services like, for example, Bergeron to address adversarial attacks and SmoothLLM for jailbreaking, which is the first algorithm to mitigate such attacks on GenAI LLMs. Infosys pulls all the telemetry from its explainability and security capabilities into a centralised control centre to ensure organisations have a single view of all the factors involved. From PACs perspective, responsibly managing AI across all parts of an organisation for the ever-growing range of use cases it will be applied to means that the need for a centralised view to manage AI both securely and ethically at scale can only be achieved currently this way and it is encouraging to see Infosys provide such an approach.

How to Accelerate the Path to Responsible AI Adoption

With every new profound technological innovation, the time to adoption compresses into shorter and shorter windows. AI is at the start of its journey, where it can be adopted as a scalable and commodity-style technology, but its broad global adoption is accelerating fast, and the underlying technologies relating to AI are also innovating and maturing at a meteoric pace. The rapid adoption of AI across all parts of an organisation, irrespective of industry, can be overwhelming for many and their senior leadership team (SLT) from a change culture perspective. PAC continues to hear from SLT members that the onset of AI makes them feel exposed because of a lack of understanding of the related risks, operational challenges, talent requirements, and unforeseen consequences of the use of AI.

When the SLT of an organisation is considering how to adopt AI at scale and operate it responsibly now and over the coming years, it should be clear that adopting responsible AI is a complex task. However, over the past decade, PAC has seen an organisation's SLT transition from a preference for running technologies in-house to consuming technologies as **cloud-based services.** It is this mindset that PAC strongly advises the senior leadership teams of organisations to consider when looking to accelerate their path to responsible AI adoption, even if their existing AI capabilities have been built in-house. This is because, from PAC's perspective, no organisation should be bespoke building and operating its own AI platforms unless under very specific competitive and/or regulatory circumstances. Given the rapid evolution of AI capabilities year on year, the pressure to keep a responsible AI (RAI) service perpetually up-to-date with the pace of AI innovation, commonly known as the practice of ever-green, will become very costly and operationally complex given the value needed to be derived from its use.

In addition, organisations PAC engages with are challenged by a range of pain points as they adopt ever more advanced forms of AI at scale. It is important for organisations to understand that there is no "one size fits all" approach to operating AI responsibly at scale. It is a multi-faceted operational challenge that requires organisations to iteratively adapt to the regulatory demands of each country they operate AI services in and across, to leverage AI guardrails that are contextual to the dynamics of how their industry operates, and to perform perpetual AI market

intelligence to be alerted to threats and vulnerabilities. All of these, in PACs' opinion, are not just highly complicated to create but also to evolve as the fast pace of AI evolution continues in the coming years.

This is where engaging with a partner like Infosys becomes an invaluable conduit to accelerate the path to RAI adoption and ensure the function of RAI within an organisation is always up-to-date to support the latest innovations in AI. Infosys is keenly aware of how the regulatory landscape globally is evolving on an ongoing basis and the societal importance of the ethical use of AI. PAC considers the Infosys approach to developing a responsible AI service to be very pragmatic as the company understands that the many moving parts of AI require a robust ecosystem of collaborators ranging from technology partners to academia and government/institutional bodies across the world, which they manage through their **Infosys Responsible AI Office.**

From PAC's perspective, this multi-tiered approach by Infosys provides a highly effective suite of capabilities to allow an organisation's senior leaders to accelerate their adoption of RAI whilst ensuring they get the value they need from a service-based model that Infosys will maintain as ever-green by evolving it as AI does. ■

About Infosys

Infosys is a global leader in next-generation digital services and consulting. We enable clients in more than 50 countries to navigate their digital transformation. With over four decades of experience in managing the systems and workings of global enterprises, we expertly steer our clients through their digital journey. We do it by enabling the enterprise with an AI-powered core that helps prioritize the execution of change. We also empower the business with agile digital at scale to deliver unprecedented levels of performance and customer delight. Our always-on learning agenda drives their continuous improvement through building and transferring digital skills, expertise, and ideas from our innovation ecosystem.

Visit www.infosys.com to see how Infosys (NSE, BSE, NYSE: INFY) can help your enterprise navigate your next.

About PAC

We are a content-based company with a consulting DNA. PAC is the leading European consulting and analyst firm supporting software & IT service vendors worldwide.

Since 1976, we have helped our clients to understand market dynamics, grow their revenue and raise their profile. Our unrivalled understanding of European markets, and deep research coverage help key market players to define their strategy, optimize their go-to-market and increase market share.

PAC is an analyst-led consultancy with a team of over 100 experts across Europe. We provide market research and analysis on more than 30 countries worldwide, delivered through our portfolio pillars, Guidance, Insights, and Visibility, and our renowned SITSI® research platform.

More on www.pacanalyst.com.

Contact:

Spencer Izard
Research Director
s.izard@pacanalyst.com
+44 (0)7398 433071

PAC UK
2 Minton Place,
Victoria Road,
Bicester, Oxfordshire,
OX26 6QB, United Kingdom

www.pacanalyst.com
www.sitsi.com