

An abstract blue sphere with a textured surface of golden dots and light trails, positioned on the left side of the page.

RESPONSIBLE AI TRANSPARENCY REPORT

OCTOBER 2025



▶ CONTENTS

FOREWORD 4

EXECUTIVE SUMMARY 6

STRATEGY 8

Planning Ahead: Charting a Clear Path to Achieve Our RAI Goals 9

Building the Expertise to Deliver on Our RAI Vision 9

Engaging in RAI Beyond G42 10

Implementing RAI Principles at G42 12

The Core of our RAI Practice at G42: Policies and Guidelines 13

RAI GOVERNING BODIES: ROLES AND RESPONSIBILITIES 14

RAI Governing Bodies 15

RAI Executive Council 15

Frontier AI Governance & Sensitive Use Case Committee 16

RAI Review Committee 17

RAI Team 17

RAI Ambassadors 18

Strengthening Incident Reporting to RAI 18

GOVERNING PRINCIPLES, GUIDELINES, AND POLICIES 20

G42's RAI Playbook 21

Governance Initiatives 21

Other Policies Relevant to Our RAI Practice 22

 Generative AI Policy, Data Policy, and Data Governance Framework 22

 Frontier AI Safety Framework 22

 Supporting Policies and Governance 23

Building RAI Structures and Processes Aligned with International Standards 24

Aligning with UAE AI Strategy and UAE AI Policy Guidance 25

Securing Documentation and Tracking Risks: Repositories and Model Cards 25

RAI WORKFLOW ACROSS G42 26

G42's RAI Workflow 27

 Assessing and Mitigating RAI Risks with External Partners: Procurement Checklists and Flow-Down Requirements 27

TOOLS, TESTING, AND PROCEDURES 28

G42's Ethics-by-Design Approach 29

Privacy-by-Design 29

Transparency and Explainability, Minimization of Discrimination, and User Control and Human Agency by Design 30

RAI Risk and Impact Assessment 31

Pre-development Risk Assessment 31

Pre-deployment Risk Assessment 32

Procurement Checklist and Risk Assessment 32

Red Teaming in Practice Across G42 32

TRAINING & UPSKILLING 36

CONCLUSION 38

APPENDIX: REFERENCES 40



FOREWORD




At G42, our pursuit of AI leadership has always been guided by a fundamental principle: innovation must be anchored in responsibility.

As AI becomes more integrated into the fabric of everyday life - shaping industries, influencing policy, and touching billions of lives - the need for strong governance is not an afterthought. It is a prerequisite.


This inaugural Responsible AI Transparency Report is both a milestone and a signal. A milestone that reflects the hard work across our teams to codify standards, develop internal frameworks, and contribute meaningfully to global dialogue. A signal of our ongoing commitment to ensuring AI systems are designed, deployed, and scaled in ways that reflect the values we stand for: safety, fairness, accountability, and inclusion.

We see transparency not as a one-time disclosure, but as a process. One that evolves with the technology itself. We will continue to engage with partners, regulators, and communities, not only to share our progress but also to challenge our assumptions and refine our approaches.

We offer this report as part of our broader mission to help shape an equitable AI future. We welcome scrutiny, we welcome dialogue, and we remain committed to building AI that earns trust, because without it, no system, no model, no platform can endure. 

PENG XIAO
Group CEO, G42





EXECUTIVE SUMMARY

► THE REPORT OUTLINES

At G42 we believe that our commitment to RAI is a long-term evolution in our approach to building great AI solutions. This report provides an overview of how we approach RAI, what we have learned so far on our RAI journey and what we plan for the future as the field and our practices continue to expand and evolve.

► THE PRINCIPLES AND POLICIES THAT GUIDE OUR RAI EFFORTS

► OUR ORGANIZATIONAL GOVERNANCE FRAMEWORK AND INTERNAL ACCOUNTABILITY MECHANISMS

► THE TECHNICAL METHODOLOGIES WE APPLY TO ASSESS AND MITIGATE RISK ACROSS OUR AI SYSTEMS

► OUR PROGRESS IN IMPLEMENTING RAI PRACTICES AND THE CHALLENGES WE ENCOUNTER

► OUR COMMITMENT TO CONTINUOUS IMPROVEMENT, INCLUDING TRANSPARENCY INITIATIVES, CROSS-SECTOR COLLABORATIONS, AND INVESTMENT IN RESPONSIBLE INNOVATION

We approach RAI not as a compliance exercise, but as a strategic priority and a shared responsibility.

By outlining our practices in this report, we aim to foster greater understanding of how we operationalize our principles and implement our RAI practices. We consider such information sharing to be of key importance to meaningfully contribute to the global discourse on AI governance.

This report is not a declaration of completion but rather marks a milestone on an ongoing journey. RAI is a dynamic and evolving field, and we approach it as a continuous, iterative process of learning, improvement, and adaptation. This report is part of our broader commitment to transparency, and we intend to update and expand it annually as our practices mature and new risks and opportunities emerge.



STRATEGY

► PLANNING AHEAD: CHARTING A CLEAR PATH TO ACHIEVE OUR RAI GOALS

At G42, our ambition for RAI is not to simply meet baseline standards, but to be best in class. This means embedding RAI into the core of how we design and deploy our technologies.

To realize this vision, we have developed a structured, forward-looking strategy that translates our principles into concrete, operational practices. We believe that achieving excellence requires a deliberate, systematic effort that acknowledges the varied nature of RAI implementation. While some RAI measures can be deployed swiftly, others require iterative design over a longer period of time to mature in a way that ensures long-term effectiveness and sustainability.

To this end, we have established a detailed two-year roadmap that outlines how our principles and policies will be operationalized across G42. Many of the most crucial RAI practices and foundational capabilities are already in place, positioning us to accelerate the next phase of initiatives in the near term.

▼▼ Our strategy reflects a clear and sustained commitment to long-term investment in RAI. We seek to not only meet current demands but also to build durable systems and frameworks that evolve with the pace of innovation. ▲▲

RAI is a dynamic and evolving field. At G42, we are committed to replacing temporary solutions with robust, scalable alternatives. By prioritizing long-term resilience and adaptability, we aim to ensure that AI is developed and deployed in ways that are not only technically sound and ethically grounded, but also capable of driving positive impact at scale.

► BUILDING THE EXPERTISE TO DELIVER ON OUR RAI VISION

AI presents socio-technical challenges that require a multidisciplinary approach to effectively address. This requires the integration of technical excellence with ethical, legal, and societal considerations. Realizing our RAI vision at G42 means building not just systems, but the expertise necessary to steward them responsibly.

▼▼ We are actively building a highly specialized, multidisciplinary RAI team to bring together experts in machine learning, safety, ethics, governance, law, policy and social science. ▲▲

We consider such an interdisciplinary approach to be critical to ensure that RAI is not treated as a siloed function, but as a foundational part of how we build technology and make decisions across the organization. As the field of RAI continues to evolve, so will our team. We are committed to continuous learning and will continue to actively engage with international best practices, external advisors, and multi-stakeholder initiatives to ensure that our approach remains best in class.

As we extend the RAI structures and processes at G42, we align with established standards and acknowledged frameworks, as well as with international legal standards.

Likewise, we align with specific regional policies and regulations concerning AI strategy and AI policy guidance.

▶ ENGAGING IN RAI BEYOND G42



Although establishing a strong RAI practice internally across G42 is our main objective, we also understand that to significantly further the RAI agenda worldwide, we must reach beyond G42. Accordingly, our secondary strategic objective to actively lead RAI efforts in a broader regional and international context.

Our establishment of the Responsible AI Future Foundation, in close collaboration with Microsoft and Mohamed Bin Zayed University of Artificial Intelligence, reflects this ambition.

▼▼ The Responsible AI Future Foundation is a newly established institute focused on furthering RAI research, strengthening industry RAI practices, and ensuring the inclusion of traditionally underrepresented areas in the global RAI conversation, especially the Global South. ▲▲

The Foundation is based in Abu Dhabi and is poised to be one of the largest RAI research institutes worldwide, in terms of funding, head count, and research output. The Responsible AI Future Foundation is constructed to be independent from its founders, ensuring research quality and avoiding any potential conflicts of interest.

We actively participate in RAI networks and initiatives, both within the region and internationally, including with WEF, UNESCO and UNGA. Through these engagements, G42 contributes to the exchange of best practices,

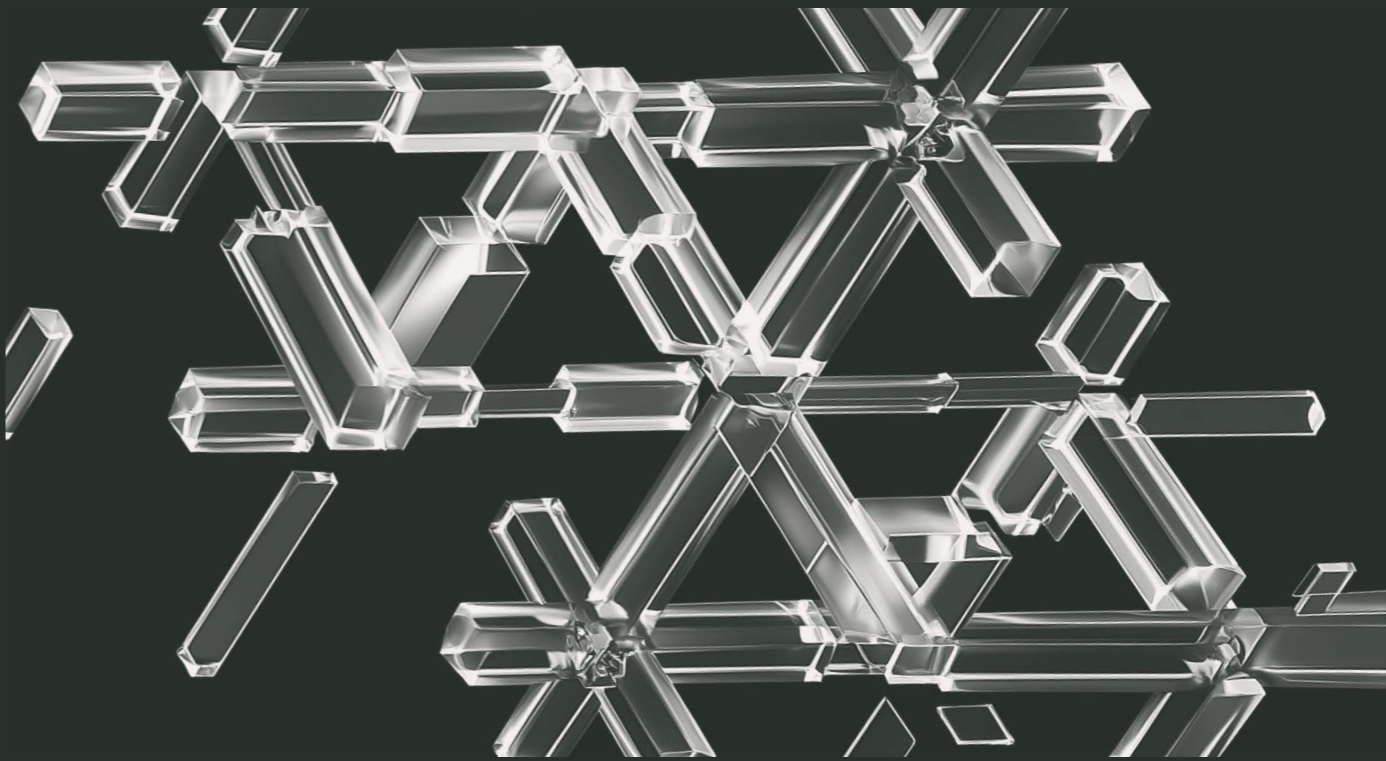
policy development, and collaborative research aimed at promoting ethical, transparent, and accountable AI systems. This involvement not only strengthens regional cooperation but also positions G42 as a proactive player in shaping global RAI standards and discourse.

▼▼ We also established the Abu Dhabi AI for Good Research Lab - a collaborative research initiative established to leverage advanced AI models, platforms and technologies for impactful social good applications. Serving as a catalyst, the Abu Dhabi AI for Good Research Hub has as a mission to foster collaborations among G42 entities, Microsoft, Mohamed Bin Zayed University of Artificial Intelligence and other local universities to demonstrate AI's potential for positive societal impact.

Additionally, the Lab actively participates in global AI dialogues, contributing insights and expertise during prominent events such as the AI for Good Global Summit. ▲▲

CASE STUDY

▶ G42 AND THE FRONTIER AI SAFETY COMMITMENTS



In May 2024, G42 joined a global coalition of leading AI organizations in signing the Frontier AI Safety Commitments at the AI Seoul Summit. This landmark initiative brought together companies from across the world to affirm a shared responsibility to develop and deploy frontier AI systems with safety, transparency, and public trust at the core.

By signing these voluntary commitments, G42 pledged to adopt best practices in frontier AI safety including rigorous internal and external red-teaming, proactive risk assessments throughout the AI lifecycle, and clear public reporting on model capabilities and limitations.

Following the summit, our teams worked intensively, in collaboration with SaferAI and METR, to develop a comprehensive Frontier AI Safety Framework, tailored

to the unique challenges and opportunities of advanced AI systems. This framework outlines how we assess risks, govern model development, and ensure ethical deployment.

In February 2025, we proudly presented this framework at the AI Action Summit in Paris, demonstrating how our principles translate into practice. The framework highlights our governance structures, safety protocols, and the collaborative spirit that drives our responsible innovation. By sharing our framework at the Summit, we aimed not only to demonstrate our progress, but to contribute meaningfully to the global conversation on how frontier AI can be governed responsibly and transparently. Further details on how we operationalize our framework are set out in Section 5 of the report.

▶ IMPLEMENTING RAI PRINCIPLES AT G42

Core and Instrumental Principles

G42's RAI framework is built upon three foundational core principles that serve as universal navigation points for all applied ethics assessment.

These core principles—respect for human autonomy, minimization of harm, and upholding justice—form the basis of all RAI risk and impact assessments.

In their application to AI systems, core principles are complemented and supported by a comprehensive set of instrumental principles that translate these seemingly abstract values into actionable guidance with specific measurement areas. This approach, which organizes instrumental principles in accordance with the core values they aim to support, allows them to be implemented with purpose and a clear understanding of the resulting trade-offs.

These instrumental principles derive their importance from their effectiveness in protecting and promoting the intrinsic values of the core principles, and they are intentionally designed to be interchangeable based on context and priority. The framework recognizes that instrumental principles serve different functions depending on the context and use cases, allowing G42 to determine which principles to prioritize and how to best achieve the overarching core values.

This hierarchical structure ensures that AI systems remain ethically robust by providing a clear basis for evaluating and addressing specific ethical concerns such as transparency, privacy, discrimination, and accountability. These core and instrumental principles reflect our commitment to creating AI solutions that are ethically sound and contribute to societal progress through innovation. These guiding principles demonstrate our belief that through responsible leadership, practices, and governance, AI can significantly improve the human condition for the majority of people.

1. PROTECTING HUMAN AUTONOMY



Under the core principle of protecting human autonomy, we have established seven instrumental principles that focus on preserving human agency and control. These include promoting human control through oversight and intervention capabilities; ensuring transparency by making AI operations clear and understandable; providing explainability for AI decisions and actions; empowering human agency for informed decision-making; obtaining explicit consent before data collection or use; protecting user privacy and personally identifiable information; and minimizing epistemic risks by promoting factual representation while reducing misinformation and disinformation.

These instrumental principles work together to ensure that individuals maintain meaningful control over their interactions with AI systems and can make informed choices about their engagement with AI technology.

2. MINIMIZING HARM AND MAXIMIZING BENEFIT



The second core principle is operationalized through nine supporting principles focused on ensuring AI systems produce positive outcomes while avoiding negative consequences. These principles emphasize promoting accuracy to minimize errors; maintaining scientific validity in AI foundations and outcomes; ensuring reliability through consistent performance and stable results; ensuring security against unauthorized access and attacks; ensuring safety to prevent physical or psychological harm; improving overall well-being for individuals and society; assessing broader societal impacts across all stakeholders; optimizing efficiency for effective and sustainable performance; and actively minimizing the negative environmental impacts of AI systems.

3. JUSTICE



The third core principle of upholding justice is operationalized through five supporting principles that ensure fair AI implementation and deployment. These include promoting non-discrimination through diverse datasets and appropriate algorithms; ensuring the fair distribution of AI benefits and burdens across all segments of society; and protecting vulnerable populations who might be disproportionately affected by AI systems. This core principle also establishes clear accountability for AI outcomes and provides contestability mechanisms for individuals to challenge AI decisions.

▼ POLICIES & GUIDELINES



A strong foundation is essential to an effective RAI program. Our RAI principles and policy form that foundation; every RAI activity is traceable to them, ensuring coherence and alignment across all business units. The policy has two parts:

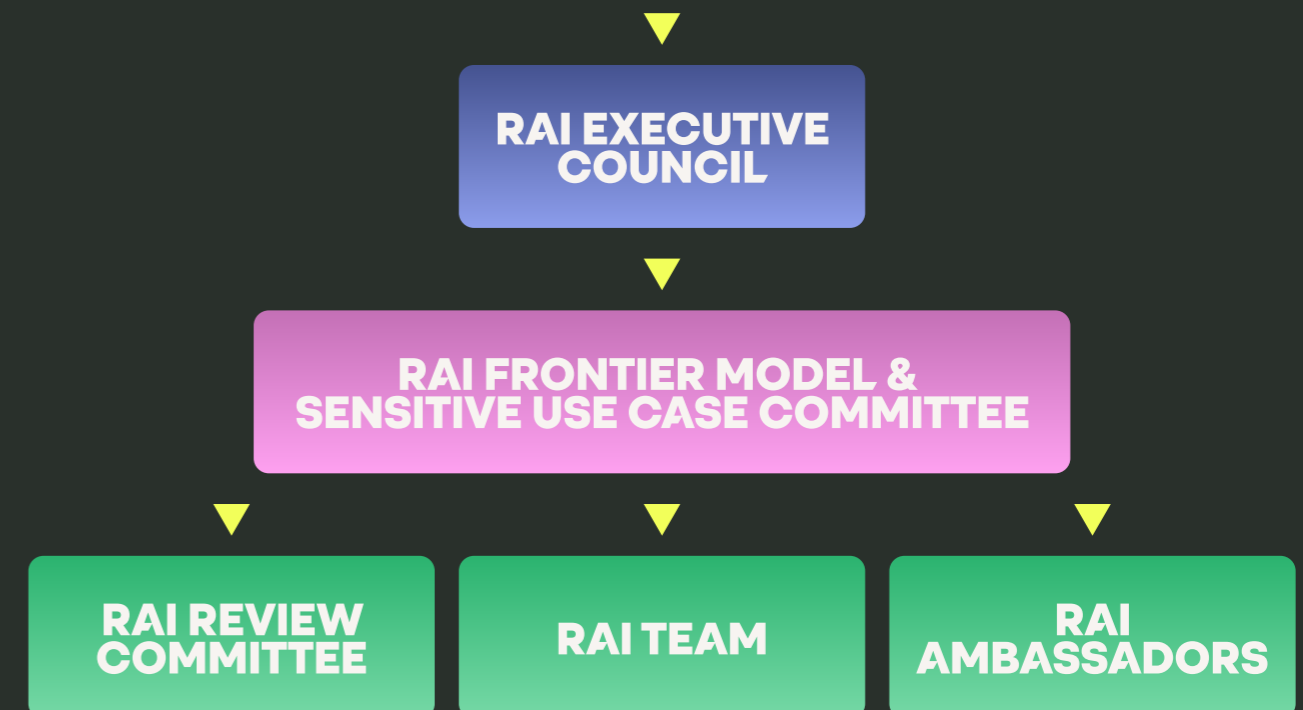
1. Principles & methodology: the principles themselves and how we derived them (see above).
2. Operationalization: twenty-three initiatives that put the principles into practice, each with their own execution plan.

As our program matures, these initiatives will sit at different stages of execution: many are already underway, and the remainder are scheduled and will be activated in the near term.



RAI GOVERNING BODIES: ROLES AND RESPONSIBILITIES

▶ RAI GOVERNING BODIES



▶ RAI EXECUTIVE COUNCIL

The RAI Executive Council serves as the highest-level oversight body providing strategic guidance on G42's RAI governance. The Executive Council ensures that all AI initiatives align with our values and long-term goals. It is responsible for setting the overall vision and priorities for RAI development, ensuring alignment between business objectives and ethical considerations.

The Executive Council's main function is to hold strategic decision-making authority across all three key aspects of RAI governance. It oversees (1) the process and workflow for standardized implementation; (2) the playbook, including relevant policies, guidelines, and tools; and (3) the people, with roles and responsibilities assigned throughout the organization.

In more detail, these three key functions translate into the following specific goals:

Process:

- Providing strategic guidance to all G42 AI initiatives in implementing RAI governance and aligning with RAI goals
- Overseeing the RAI team's work in the ethical development and deployment of AI models, products, services, systems, and applications
- Supervising the monitoring of AI systems and initiatives to ensure they meet ethical, safety, and regulatory standards

Playbook:

- Ensuring the development of up-to-date RAI policies and guidelines that align with best practices and regulatory requirements, and providing final approval of these policies and guidelines
- Engaging with internal and external stakeholders to promote transparency and ensure their needs, priorities, and concerns are reflected in the resulting policies and processes

The Council derives its authority from the CEO of G42 Group. The Executive Council’s terms of reference (detailing its function, scope, responsibilities, composition, reporting lines, and meeting cadence) have been established, and the Council is ready to actively engage.

People:

- Overseeing the assignment of roles and responsibilities within the organization and their reporting lines to the RAI Executive Council
- Ensuring incentives exist for raising RAI-related concerns internally, and removing disincentives for neglecting RAI governance steps
- Reporting on these matters to the G42 Board and, where appropriate, making recommendations
- Reporting, as required, to G42 shareholders on the activities and remit of the Council

FRONTIER AI GOVERNANCE & SENSITIVE USE CASE COMMITTEE



This committee is responsible for evaluating and approving AI use cases that involve sensitive or high-risk applications. It ensures that these use cases meet ethical, safety, and regulatory standards before deployment.

The Committee conducts thorough reviews and risk assessments to safeguard against potential adverse impacts. The key roles and responsibilities of the Frontier AI Governance & Sensitive Use Case Committee are:

- Use Case Evaluation: Evaluate and approve AI use cases that involve sensitive, Frontier or high-risk applications, ensuring they meet ethical, safety, and regulatory standards before deployment.
- Risk Assessment: Conduct thorough risk assessments

for sensitive, Frontier AI use cases, identifying potential ethical, safety, and regulatory concerns.

- Compliance Verification: Verify compliance with ethical guidelines, safety protocols, and regulatory requirements for sensitive and Frontier AI use cases.
- Incident Response: Develop and implement incident response plans for sensitive and Frontier AI use cases, ensuring prompt and effective resolution of any issues.
- Continuous Improvement: Continuously review and improve the evaluation and approval processes for sensitive and Frontier AI use cases.

RAI REVIEW COMMITTEE

The RAI Review Committee evaluates and approves AI products and systems to ensure they meet ethical, safety, and regulatory standards before deployment. The Review Committee focuses on the practical implementation of AI solutions, ensuring that they are safe, reliable, and compliant with all relevant guidelines. It serves as a critical part of the RAI process, where AI systems developed in-house or acquired through procurement are assessed for their risks and impact, and are continuously monitored for safety and reliability. The Committee oversees risk assessments and determines whether proposed systems are justified and align with G42’s overall risk thresholds. To achieve this, the Review Committee engages in five key functions:

- Evaluation of AI products and systems before development and deployment
- Ongoing monitoring of AI products and systems for safety, reliability, and quality
- Ensuring compliance with regulatory requirements and alignment with industry standards
- Maintaining comprehensive documentation of evaluations and reporting results to the RAI Team
- Supporting innovation teams in their review processes and engagement with the Review Committee

The RAI Review Committee serves as an independent evaluation body for all G42 companies and maintains no conflicts of interest with any team within the Group. It has the authority to recommend modifications, request additional safeguards, or in extreme cases, halt AI projects that pose unacceptable risks or fail to meet ethical standards.

The Committee also investigates reported incidents or concerns related to AI systems already in deployment, conducts thorough post-incident analyses, and recommends corrective actions. It maintains detailed records of its review processes and decisions, contributing to institutional knowledge and helping refine RAI policies based on real-world experience and lessons learned.



RAI TEAM

The RAI Team forms the core function responsible for designing and implementing RAI governance mechanisms throughout the Group. Due to the socio-technical nature of RAI, the team requires a cross-functional group of experts including AI ethicists, policy specialists, technical researchers, legal experts, and program managers. Such experts work collaboratively to translate high-level RAI principles and strategies into practical guidelines and processes. They are responsible for turning the strategies, policies, and directives set by the Executive Council into an operational workflow on a day-to-day basis.

The team will oversee the implementation of the RAI process and workflow: conducting AI risk and impact assessments, working with teams on risk mitigation plans, troubleshooting ethical issues, providing guidance to product teams throughout the AI development lifecycle, and ensuring the application of an ethics-by-design approach. They serve as internal consultants, helping other departments navigate complex ethical questions

and ensure compliance with established RAI standards. They also develop and deliver (or oversee the development and delivery of) RAI training as needed across the organization.

The RAI Team will monitor RAI risks and mitigation efforts throughout G42. Responsibility for signing off on projects and greenlighting these as AI initiatives rests with the RAI Team. The Team works closely with all other committees. Both the Review Committee and the RAI Frontier Model & Sensitive Use Case Committee inform the Team of their assessments, while the RAI Team ensures alignment with our overall RAI strategy.

Furthermore, the RAI Team maintains documentation of RAI processes, measures the effectiveness of ethical AI initiatives, and regularly reports progress to senior leadership and the RAI Executive Council.

► RAI AMBASSADORS

The RAI Ambassadors play a key role in our RAI governance. RAI Ambassadors are individuals appointed by business unit leadership who are trained and tasked with carrying out the RAI components of their unit and serve as designated contact persons. Their role can be construed as one of first responders: they recognize and detect RAI needs and concerns, are skilled in conducting risk assessments and applying RAI tools and guidelines established by G42, and their engagement within their teams helps foster a RAI-forward culture.

The RAI Ambassador program also serves as an early warning system, identifying potential ethical issues or risks before they escalate and connecting their teams with appropriate RAI resources and expertise.

The main goal of the RAI Ambassador program is to ensure that innovation teams have a designated individual who can respond to day-to-day RAI needs and concerns and who has a direct line to the RAI Team for consultation and troubleshooting. By embedding the RAI Ambassadors, the governance model aims to ensure a seamless and efficient integration of RAI within each unit.

- To that end, the Ambassadors have the following responsibilities:
- Act as a liaison between their teams and the RAI governance bodies, facilitating communication and collaboration on RAI initiatives
 - Conduct ethics risk and impact assessments, engaging with the RAI Review Committee throughout the pre- and post-review processes
 - Collaborate with the innovation team in developing risk mitigation solutions
 - Keep their teams up to date on G42 RAI policies, tools, and guidelines
 - Raise awareness around RAI processes and practices and foster a culture of ethical responsibility
 - Provide feedback from team members on RAI practices and suggest improvements to the RAI governance bodies

This distributed network approach ensures that RAI considerations are integrated into daily operations across the entire organization, rather than being confined to a centralized team.

► STRENGTHENING INCIDENT REPORTING TO SUPPORT RAI

Timely and effective incident reporting is essential to preserving the security, integrity, and reliability of AI systems. As part of our broader RAI governance framework, we are implementing structured procedures to ensure that AI-related incidents are reported and addressed promptly. These protocols aim to embed a culture of accountability and continuous improvement into every stage of the AI system lifecycle at G42.

Our third-party-managed, anonymous ethics and compliance hotline, G42 Voice, is a critical component of our RAI framework. As part of our broader commitment to accountability and ethical governance, it provides a

secure, confidential channel for employees, partners, and stakeholders to raise concerns related to AI development and deployment. By enabling early identification of potential risks, G42 Voice strengthens our ability to uphold RAI principles in practice. Insights gathered through this channel help inform continuous improvements to our RAI processes and support oversight mechanisms.

We consider efficient incident reporting pathways as a critical component to reinforce a culture of transparency and trust across the organization.

CASE STUDY

QUDRATECH-BUILDING RESPONSIBLE AI FROM THE GROUND UP

►►► Launched in May 2024, QudraTech is an AI upskilling and work experience initiative for EmiratIs in Abu Dhabi and Al Ain. The program trains participants in areas such as large language model red teaming, responsible AI and data annotation, while providing remote work opportunities.

With 130 annotators currently engaged, QudraTech supports six major AI projects, including Arabic ASR, OCR, LLM evaluation datasets, and product localization. Key achievements include:



Contributed over 100 Arabic speech hours for Arabic ASR in four dialects, improving the accuracy by 13%



Over 70,000 pages of Arabic text digitized



14B Arabic tokens added to JAIS LLM



First Emirati text to speech voice developed



Two Gulf dialects (Saudi and UAE) added to JAIS LLM



Translated the entire Microsoft AI for Good Introductory module



Recognized by the Abu Dhabi Department of Economic Development for job creation in Al Ain



97% female participation; 18% Persons of Determination; 94% with no prior AI experience

QudraTech is a central part of the AI for Good Hub, driving inclusive, culturally aware AI solutions in collaboration with Microsoft, G42 and other partners. Its commitment to ethical AI, inclusivity, and sustainability has been featured in major events, reports and thought leadership. QudraTech is building a new generation of AI professionals who are equipped to lead with responsibility, technical skill, and social impact. We are proud that this initiative aligns with the UAE's AI Vision 2031, and reflects our commitment to making AI accessible, accountable, and locally relevant.

GOVERNING PRINCIPLES, GUIDELINES, AND POLICIES

► G42'S RAI PLAYBOOK

Our governing playbook for RAI encompasses all guiding and overarching policies, documents, and tools that are available to help developers and deployers navigate the RAI innovation landscape. The RAI playbook enables and ensures that ethical considerations are systematically integrated into every aspect of AI development, deployment, and management. It provides guidance on processes, goals, and roles, thus creating decision-making and oversight accountabilities.

The goal of the RAI playbook is to translate high-level principles and values into actionable items with systematic and clear directives. The RAI playbook consists of high-level principles that are translated into operational

risk and impact assessment tools; six policies and frameworks -RAI Policy, GenAI Policy, Data Policy, Data Governance Framework, Frontier AI Safety Framework, and Cybersecurity Risk Framework -along with three repositories: the Model Repository, RAI Risk Assessment Repository, and Sensitive Use Case Repository.

By making these guiding documents, tools, and repositories readily available and accessible to all developers and deployers within the Group, and with the support of RAI Ambassadors and the RAI Team, we aim to ensure that all innovation teams are empowered to follow.

► GOVERNANCE INITIATIVE

01. Executive RAI Council
02. Frontier AI Governance & Sensitive Use Case Committee
03. RAI Review Committee
04. Ethics Risk Assessment
05. Mitigation Plan
06. AI Procurement Checklist
07. Frontier AI Safety Framework
08. Data Governance Framework
09. RAI Ambassadors
10. RAI Training and Culture
11. Ongoing Model Monitoring
12. External RAI Auditing
13. Incident Reporting Pathway
14. Sensitive Use Case Repository
15. Risk Assessment Repository
16. Stakeholder Analysis & Stakeholder Engagement
17. Red-Teaming
18. Transparency & Explainability by Design
19. Minimising Discrimination by Design
20. Privacy by Design
21. User Control & Human Agency by Design
22. Cybersecurity Risk Framework
23. Environmental Impact Assessment



▶ OTHER POLICIES RELEVANT TO OUR RAI PRACTICE

Generative AI Policy, Data Policy, and Data Governance Framework

We acknowledge the important role strong data governance plays in RAI and have therefore developed a detailed Information Privacy and Data Protection Policy, which defines our data practices across the Group to ensure the ethical use of data in all models. We are enhancing our policy to outline our principles for managing data: lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, and confidentiality. We are detailing a data management framework designed to ensure that these principles are operationalized. It outlines the role and responsibilities of the Data Protection Officer, identifies the needs and responsibilities for ongoing employee training, and provides a clear privacy incident and data breach management plan to efficiently mitigate any unwanted data breaches.

To ensure compliance with our data policy, we have

integrated data-related questions throughout our RAI risk assessment tools and procedures and look forward to expanding these efforts into an even more comprehensive data governance framework over the coming period.

We understand that some AI tools require specific guidance to ensure their use aligns with our principles. On that basis, we have developed a specific Gen AI Policy that defines the scope of allowed use within G42. The Gen AI Policy outlines acceptable and responsible use of Gen AI and focuses in detail on establishing standard requirements in relation to areas such as privacy, security, confidentiality, verification procedures, and transparency, as well as defining monitoring obligations and assigning roles and responsibilities for Gen AI use across the Group.

Frontier AI Safety Framework

As detailed earlier in the report, our Frontier AI Safety Framework is a comprehensive set of protocols designed to ensure the safe and responsible development, deployment, and management of advanced AI technologies. The framework ensures that our frontier AI models are governed by dynamic safeguards and evolving risk controls aligned with their growing capabilities.

The Frontier AI Safety Framework introduces a multi-layered approach to AI risk management, ensuring that advanced AI systems are developed, tested, and deployed responsibly. It includes:

- **Defined Capability Thresholds & Mitigation Strategies** – the framework introduces clear capability thresholds to assess biological threats, cybersecurity vulnerabilities, and autonomous decision-making risks. Each threshold mandates Deployment Mitigation Levels (DMLs) and Security Mitigation Levels (SMLs) to ensure appropriate safety measures are implemented.
- **RAI Frontier Model & Sensitive Use Case Committee** –the committee oversees model compliance, safety protocols, and incident response.
- **Independent Audits** – we will conduct annual external governance audits to ensure compliance (further details on audit processes are provided on the following pages).



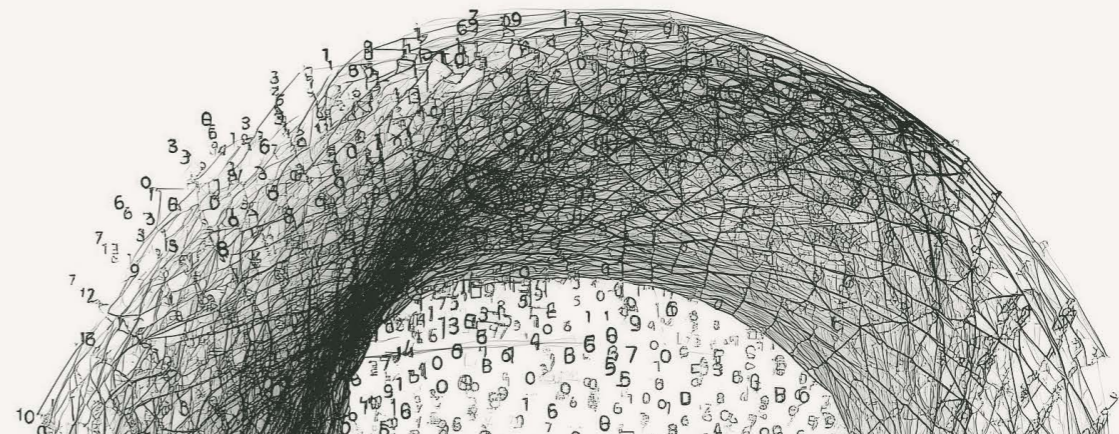
Supporting Policies and Governance

The anonymous mechanisms for reporting AI-related risks are outlined in the G42 Whistleblowing and Non-Retaliation Policy. Further details on incident reporting pathways have been set out earlier in Section 4 of this report.

The Environmental, Social, and Governance ("ESG") commitments of G42 are outlined in the G42 Environmental, Social and Governance Policy ("ESG Policy"). As a Group, we recognize the profound impact AI can have on society, and we are dedicated to ensuring our technology acts as a force for good: enhancing the lives of people around the world while safeguarding their rights and promoting human well-being. G42's Responsible AI Policy aims to uphold the highest ethical AI standards and establishes an internal governance framework, ethical principles, and guidelines across the Group.

We are committed to embedding ESG principles across our strategy, operations, and culture. Guided by the Group ESG Policy, which aligns with the UAE's Net Zero by 2050 target, we integrate sustainability into the development and deployment of AI technologies. This approach reflects a dual commitment: to proactively manage ESG impacts, risks, and opportunities, and to harness AI as a force for environmental stewardship, social wellbeing, and responsible innovation. We are working to reduce our environmental footprint while maximizing positive contributions, including through the deployment of efficient buildings, evaluating clean energy sourcing, and advancing the development of more energy-efficient AI systems.

The approval process to be followed in the case of any deviation from, or risk of non-compliance with, the RAI Policy is set out in the G42 Enterprise Risk Management (ERM) Framework. Where a risk of non-compliance with the G42 RAI Policy is anticipated or identified, the relevant business leader is required to seek explicit advance approval for any deviation. All exception requests will be considered and assessed in accordance with the ERM Framework exemption process. In cases where temporary exceptions are granted for defined periods, there is clear assignment of responsibility to make timely and necessary arrangements for compliance, either prior to or upon the expiry of the exception.



► BUILDING RAI STRUCTURES AND PROCESSES ALIGNED WITH INTERNATIONAL STANDARDS

The RAI structures and processes at G42 are designed to align with international standards based on notable governance frameworks such as the NIST AI Risk Management Framework (NIST AI RMF), the OECD AI Principles, and the G7 Hiroshima Process International Guiding Principles for Advanced AI Systems. We act and operate according to the principles of responsible development, deployment, and use of AI systems, with a focus on responsible innovation and minimizing potential harms such as bias, lack of transparency, and security vulnerabilities.

►►► G42 ALIGNS WITH THE NIST AI RMF IN THE FOLLOWING WAYS:

- 

►

MAP

►►►

The context, purpose, and impacts of AI systems are identified through a prerequisite risk assessment that includes stakeholder analysis.
- 

►

MEASURE

►►►

Risks such as performance, uncertainty, bias, and reliability are evaluated via risk assessments and red teaming/testing during model development.
- 

►

MANAGE

►►►

Risks are mitigated using defined workflows, red teaming, and technical tools such as benchmarking and RAI model cards.
- 

►

GOVERN

►►►

Risk governance is ensured through organizational policies (including policies on RAI, data, and cybersecurity), a defined accountability structure, and mechanisms such as whistleblowing channels and the RAI Ambassador role.

Furthermore, we classify our AI systems in a manner that aligns with the model risk classification under the EU AI Act, that is according to a risk-based approach in which obligations are assigned to AI systems based on the potential risks they pose.

In particular, when classifying a high-risk AI system used in sensitive areas such as infrastructure, education, employment, law enforcement, health, and migration, we observe strict requirements related to risk assessments, data governance, and human oversight.

► ALIGNING WITH THE UAE AI STRATEGY AND THE UAE AI POLICY GUIDANCE

Our RAI practices align with the UAE's AI Ethics Principles and Guidelines, which emphasize ethical, transparent, and human-centric AI development. Through these practices, we support the regional approach to promoting RAI, particularly by supporting ethical innovation and acknowledging that structured oversight is essential to RAI.

The UAE's AI Ethics Principles and Guidelines explicitly highlight fairness, accountability, transparency, explainability, robustness, and safety and security as key priorities in connection with AI.

As this report clearly demonstrates, the RAI policies and practices developed at G42 incorporate all these concerns at their core.

We recognize that oversight lies with the UAE Ministry of AI and that implementation is supported by regulations, education (for example, through collaboration with academic institutions), and global partnerships, all of which we are enthusiastic about contributing to in any way possible. We are eager to support the continued collaborative development of RAI at the regional level.

► SECURING DOCUMENTATION AND TRACKING RISKS: REPOSITORIES AND MODEL CARDS



The ability to extensively document and trace model design choices, model behavior, training data and data procurement practices, as well as identified risks and mitigation plans over time, is a cornerstone of RAI governance and closely aligned with our focus on traceability, transparency, and auditability. To ensure that documentation is as extensive and methodical as needed, we have established several repositories, each with its own function.

Our newly established Model Repository tracks all models being developed by G42, beginning from the design phase. It includes all updates made to the model, all risk assessments and testing procedures, and all use cases.

Any high-risk or sensitive use cases that require a separate level of security are archived in the Sensitive Use Case Repository, and all risk assessments are filed in our Risk Assessment Repository to ensure that the RAI Team can maintain an overview of identified risks and detect patterns across all models.

To further extend our efforts, we are currently in the process of adding RAI elements to our existing Model Cards, so that they reflect not only technical information but also RAI-relevant data.



RAI WORKFLOW ACROSS G42

**MODEL/SYSTEM
DEVELOPMENT**

Use Case Definition

Business owners define AI systems in line with regulations, standards, and G42 RAI principles

Data Acquisition & Preparation

Owners and developers follow standards for data acquisition, preparation, and engineering, ensuring legality, quality, ethics, and representativeness

Model / System Development

RAI standards embedded in design; risk assessment performed ex ante to set proportional risk levels

Risk Criticality Appraisal

High-risk, edge, and frontier models identified and triaged for independent review

Ethics & Evaluation by Design

Ethical criteria and model/system testing are set in the context of the business application. Context-specific criteria and measurements are defined



**PRE-DEPLOYMENT
REVIEW**

Standards & Specifications

Business leadership sets acceptance criteria and RAI standards for AI systems

RISK BASED APPROACH

Independent Review & Challenge

For high-risk/edge/frontier cases, independent teams test and evaluate models. Results and mitigations are documented, with ongoing monitoring requirements

Independent, Cross-functional Review

RAI Review Committee & Frontier Governance Committee conduct multidisciplinary review. Sign-off required before go-live



**MONITORING/
AUDITING**

Ongoing Monitoring

Deployment teams track RAI performance and maintenance needs

Issue Escalation

Breaches escalated to business owners, second line, and independent committees

Business Continuity

Clear fallback plans, remediation, and senior escalation processes in place

Auditing

Governance, risk appraisals, mitigations, roles, and responsibilities documented proportionately. Conformity assessments conducted as needed

End-to-End Workflow & Documentation

Standards, Inventory, Procedures & Checklists, Accountability & Oversight (RACI), Conformity Assessment (Risk-Based), Policy, Risk and Control Matrix, Risk-Based Reviews and Escalations, Independent Reviews and Testing, Continuous Monitoring

► G42'S RAI WORKFLOW

The illustration above demonstrates the RAI workflow across G42, acknowledging that transparent reporting paths, well-defined tasks, and clear lines of responsibility are key to achieving our ambitious RAI goals. While outlining our practices for defining, assessing, and

documenting RAI risks, it also describes mitigation pathways and demonstrates our aim of integrating RAI as an active practice with clear requirements and objectives across the model development lifecycle.

Assessing and mitigating RAI risks with external partners: Procurement checklists and flow-down requirements

To ensure that our RAI efforts also extend to our partners, we have recently integrated an extensive **Procurement Checklist** (see Section 7), so that all AI brought into G42 from external vendors complies with the same high ethical standards. The Procurement Checklist consists of more than forty questions designed to gather necessary information in areas such as compliance with legal requirements and international standards; mapping of data used for training and testing; mitigation of privacy concerns; integration of RAI measures; types of bias testing performed; whether a stakeholder analysis has been conducted; and more.

Whereas the Procurement Checklist secures our RAI standards for all incoming AI, we make use of our newly

developed **RAI Flow-Down Requirements** to ensure that models are also held to high RAI standards after they leave G42. These flow-down requirements may place constraints on the purposes for which a model can be used and may also impose specific obligations on those procuring a model.

For example, there may be an obligation to clearly disclose when an output is generated, a chat is carried out, or a decision is made by an AI rather than a human. As such features are often communicated through the user interface, the obligation to disclose falls on those who design or control the interface. In cases where this is an external partner, the requirement to disclose is passed on to them through case-specific flow-down requirements.

► G42'S ETHICS-BY-DESIGN APPROACH

At G42, we endorse the ethics-by-design approach for our RAI practice.

▼▼ Ethics-by-design is a specific approach to technology development in which ethical considerations are proactively integrated into every phase of the design and development lifecycle of the AI system. ▲▲

This methodology requires ethicists and developers to work closely together to identify potential ethical implications of the proposed system from the initial conceptual stages through deployment and continuous maintenance.

By translating ethical concerns and considerations directly into design and development decisions, this approach ensures that the system is constructed in a way that is ethically robust and less prone to producing unethical outcomes or uses. Integrating ethical frameworks directly into technical specifications, user interface design, and system architecture allows organizations to prevent harmful outcomes before they occur, rather than attempting to remediate them after deployment.

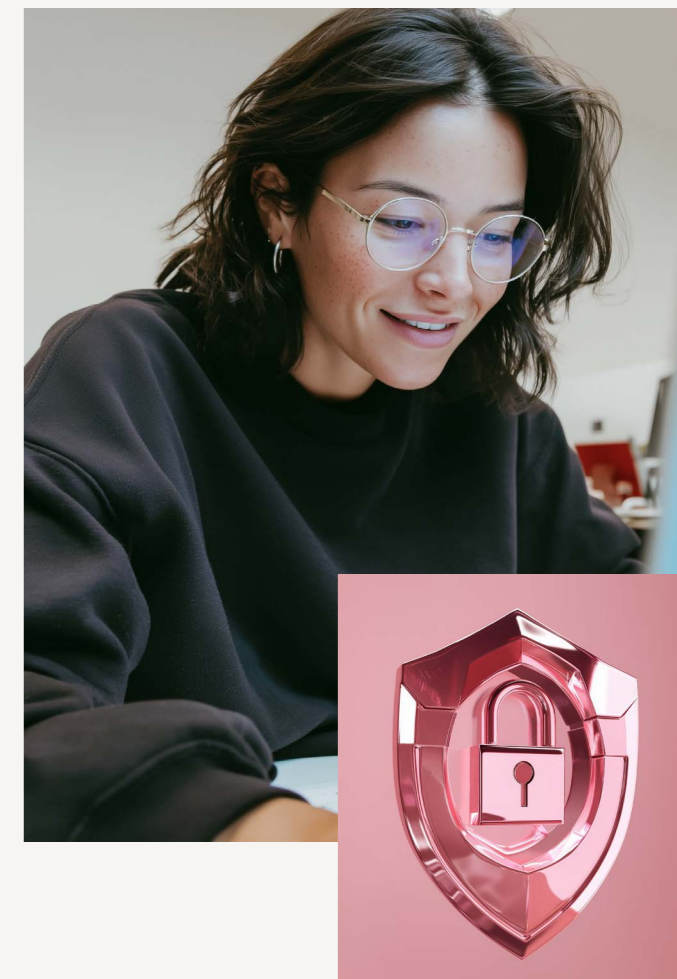
TOOLS, TESTING, AND PROCEDURES

► PRIVACY-BY-DESIGN

Privacy-by-design is a comprehensive framework that embeds privacy protection directly into the architecture and operation of AI systems, ensuring that privacy protection is not an add-on feature but a core component of system functionality. This approach operates on seven foundational principles:

- Proactive not reactive; preventive not remedial
- Privacy as the default setting
- Privacy embedded into design
- Full functionality – positive-sum, not zero-sum
- End-to-end security – full lifecycle protection
- Visibility and transparency – keep it open
- Respect for user privacy – keep it user-centric

Privacy-by-design constitutes the most well-established ethics-by-design component, with a clear, known, structured, and principled approach to the privacy concerns raised by technologies.



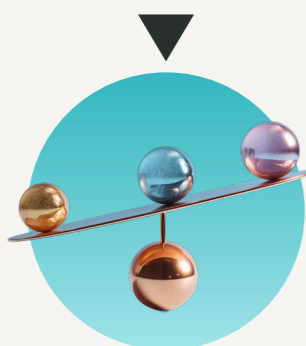
► TRANSPARENCY AND EXPLAINABILITY, MINIMIZATION OF DISCRIMINATION, AND USER CONTROL AND HUMAN AGENCY BY DESIGN

Our RAI Policy recounts and endorses specific requirements around principles and by-design approaches for operationalizing these principles. Specifically, the policy lists (1) transparency and explainability, (2) minimizing discrimination, and (3) user control and human agency as principles to be implemented through design practices. While, unlike privacy-by-design, these specific aspects do not yet have well-established frameworks, they are emphasized due to their importance during the development process.



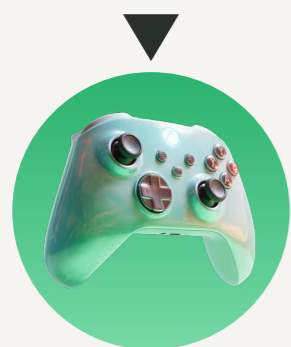
▼ Transparency and explainability by design:

Transparency by design ensures that systems and processes are built with inherent openness and clarity about their operations, decision-making processes, and data-handling practices, making it possible for users and stakeholders to understand how technologies affect them. Explainability by design goes further by integrating the ability to provide clear, understandable explanations of system behavior and decision-making processes directly into the architecture of AI systems.



▼ Minimization of discrimination by design:

Minimization of discrimination by design involves proactively identifying and addressing potential sources of bias and unfair treatment within system design, data collection, algorithmic processing, and user experience creation to ensure fair outcomes across diverse user populations. This approach requires systematic analysis of how different demographic groups might be affected by system features, conducting bias audits and testing throughout the development process, and implementing technical safeguards to prevent discriminatory outcomes from emerging during system operation.



▼ User control and human agency by design:

User control and human agency by design ensures that individuals maintain meaningful control over their interactions with technology systems, preserving human autonomy and decision-making authority even as systems become increasingly sophisticated and automated. This approach requires building accessible control mechanisms directly into system interfaces and functionality, allowing users to customize their experiences, set their own preferences, and override automated decisions when desired or necessary.

► RAI RISK AND IMPACT ASSESSMENT

▼ RAI risk and impact assessment is a systematic approach to identifying, evaluating, and mitigating potential harms that AI systems may cause to individuals, organizations, and society.

A comprehensive RAI risk assessment functions as a guiding tool for both the innovation team and the RAI team to navigate the risks of a proposed AI system. ▲▲

It clearly and visibly lays out the impact and probability of risks and enables the innovation team and the RAI team to assess those risks in relation to potential benefits. By doing so, RAI risk assessment becomes an integral part of cost-benefit analysis and overall business risk assessment.

Effective risk assessment requires examining not only technical failures and security vulnerabilities, but also broader ethical concerns such as algorithmic bias, privacy violations, and unintended social consequences. The assessment process must be iterative and ongoing, recognizing that AI risks can evolve as systems learn, adapt, and encounter new scenarios in real-world deployments.

We have developed and employ three risk assessment frameworks: (1) pre-development, (2) pre-deployment, and (3) procurement risk assessment. All three frameworks build on our core and instrumental RAI principles, turning them into measurable and actionable requirements. By doing so, we put ethics-by-design into action.

THREE RISK ASSESSMENT FRAMEWORK



► PRE-DEVELOPMENT RISK ASSESSMENT

Pre-development risk assessment occurs at the initial conceptual stage of an AI project. By laying out the risk landscape in which the AI system will be situated and planning ahead to identify and mitigate potential risks, the pre-development risk assessment process enables the innovation team to design their approach and define their needs in accordance with the risks ahead.

This phase involves analyzing the intended use case, identifying direct and indirect stakeholders, understanding the potential risks and their impact on affected groups and individuals, and evaluating the broader societal implications of the proposed AI system. Risk assessments at this stage should consider, among other factors, potential biases in data sources, algorithmic fairness concerns, and privacy implications—before any technical

development begins. The assessment should also evaluate whether the AI system aligns with organizational values and regulatory requirements.

Documentation of identified risks and mitigation strategies during this phase creates a foundation for responsible development throughout the project lifecycle. The resulting assessment is reviewed by the RAI Review Committee. In the case of frontier models and sensitive use cases, the assessment is escalated to the RAI Frontier Model & Sensitive Use Case Committee. If either committee determines that the risk level is too high to proceed, the innovation team is expected to work with the RAI Team to troubleshoot and reduce the risks. If this cannot be achieved, the project does not proceed.

► PRE-DEPLOYMENT RISK ASSESSMENT

Pre-deployment risk assessment takes place after development but before the AI system goes live, serving as a final checkpoint to evaluate system performance and safety.

This assessment involves comprehensive testing across diverse scenarios and user groups to identify potential failures, biases, or unintended behaviors that may have emerged during development. Teams should conduct red team exercises and adversarial testing to stress-test the system’s robustness and identify security vulnerabilities. The evaluation should include performance metrics across

different demographic groups to ensure fair outcomes and detect any disparate impacts.

Risk assessments at this stage must also verify that appropriate monitoring and feedback mechanisms are in place for post-deployment oversight. The relevant review committee is responsible for ensuring that risks are identified and adequately mitigated. Only after satisfactory completion of the pre-deployment assessment can innovation teams proceed with system launch, ensuring that identified risks have been adequately addressed or mitigated.

► PROCUREMENT CHECKLIST AND RISK ASSESSMENT

Pre-development and pre-deployment risk assessments ensure that the in-house development of AI systems aligns with G42’s high ethical standards. These standards also apply to any AI system that is procured or outsourced. This approach allows for coherence and quality control across all our initiatives.

To this end, the procurement checklist and risk assessment that we employ involve evaluating third-party AI systems, tools, or components before purchasing and integrating them into organizational workflows. In our view, this assessment is critical because we recognize that

organizations often lack visibility into the development processes, training data, and internal mechanisms of externally developed AI systems.

Teams must evaluate vendor transparency regarding model architecture, data sources, testing procedures, and known limitations to understand potential risks and reliability issues. These assessments are also reviewed and signed off by the relevant review committees. If the review committees determine that the risks are too high, or if there is insufficient visibility into the risks that the system poses, the teams do not proceed with the procurement.

► RED TEAMING IN PRACTICE ACROSS G42

We are actively developing and implementing robust, state-of-the-art testing methods across all models and are continually expanding and improving these efforts, particularly through ongoing red teaming, benchmarking, monitoring, and evaluations, to support RAI development.

We require that red teaming be conducted using both automated and manual methods to identify risks and support AI system improvements. These efforts reflect a strong foundation in proactive evaluation and continue to expand in scope and depth. Plans are in place to further increase linguistic, cultural, and contextual coverage, integrate real-time feedback, and enhance testing across diverse scenarios. Evaluation workflows have

been established to guide testing at key stages, including reviews of data quality, fairness, robustness, reliability, and security. Current technical assessments also explore how AI systems respond to challenging prompts and adversarial use cases, helping to surface potential issues before deployment.

We have also made substantial progress towards integrating RAI tools into development workflows, though known wider challenges remain in applying these tools effectively across different languages and cultural contexts. This highlights the importance of developing more inclusive and adaptable frameworks for multilingual AI evaluation.

New initiatives have been introduced to better assess language models in underrepresented contexts. For example, tools have been created to evaluate Arabic-language models (see the “Case Study: Setting the Standard” below), focusing on key aspects of output quality and safety.

Our research continues to advance RAI practices and is tailored to linguistic as well as cultural nuances. This includes creating tools that measure bias, fairness, and harmful content, as well as adapting explainability and transparency mechanisms to suit local languages and dialects. These efforts directly inform real-world AI system design and deployment. In specialized domains like healthcare, new evaluation frameworks are also being applied. These tools assess language models on core competencies such as clinical safety, ethical reasoning, and bias detection. By covering tasks like summarization, question-answering, and note generation, these benchmarks help ensure AI systems meet high standards in critical domains.

To maintain accuracy, reliability, and alignment with RAI principles, we are implementing continuous monitoring and feedback mechanisms. These systems provide real-time visibility into model performance and help identify issues early in the development process. Efforts to evaluate AI safety in high-risk areas are ongoing. New benchmarking initiatives assess how language models perform under

complex and adversarial scenarios, such as misinformation and security-related prompts. While still developing, this work supports deeper understanding of model behavior in sensitive contexts; particularly for the UAE.

Operational safeguards are also in place to ensure responsible development. These include regular risk reviews, structured testing processes, and adherence to privacy and data protection standards. Such practices help reinforce the technical robustness and safety of AI systems from early development through deployment.



Advancing our RAI Practices

Looking ahead, efforts are focused on broadening testing approaches, incorporating a wider range of perspectives through co-design, and aligning closely with evolving RAI standards.

Transparency, continuous refinement, and stronger traceability between risks and mitigations are key priorities.

Future iterations of our evaluation frameworks aim to distinguish between performance metrics and RAI indicators, while also integrating stakeholder feedback and real-time monitoring across the AI lifecycle.

Ensuring that AI systems are inclusive and contextually aware remains a central goal. Research is underway to develop multilingual RAI methodologies that reflect linguistic, cultural, and societal diversity.

This includes building benchmarks that address bias, fairness, and harmful content in underrepresented languages, and adapting explainability tools for non-English contexts. These efforts support the development of culturally grounded AI guidelines that reflect local values and regulatory expectations.

In domain specific application areas such as healthcare and energy, work continues on expanding evaluation tools and use cases in Arabic. This involves developing domain-specific datasets, adapting existing safety metrics, and creating culturally relevant standards for risk assessment. Human-in-the-loop methods and cross-lingual evaluation techniques will play an essential role in ensuring both scientific rigor and contextual appropriateness in high-stakes applications. Specialized benchmarking tools are also being developed to evaluate AI safety under high-risk and adversarial conditions.

Security and safety remain core pillars of RAI development and deployment. Independent assessments and collaborative testing, including third-party red teaming and internal evaluations, play a vital role in identifying potential risks early. To advance these efforts, partnerships are being established with RAI, cybersecurity, and domain experts (e.g., healthcare, energy, and finance) to develop customizable guardrail solutions. These collaborations will support the creation of AI security and moderation tools for diverse applications, reinforcing leadership in safe and adaptive AI development.

Efforts are also underway to align with internationally recognized RAI standards by incorporating external audits, stakeholder engagement, and public documentation of model performance. We believe robust auditing practices are essential to ensure that AI systems are safe, ethical, and aligned with public values.

As part of our future approach to AI governance, we plan to align with ISO/IEC 42001:2023, the first certifiable AI management system standard. This framework offers a comprehensive structure for managing ethical, operational, and safety considerations across the AI lifecycle. We see this as a key step toward establishing clear, auditable processes for RAI. In addition, we aim to incorporate and adapt guidance from the NIST AI Risk Management Framework (AI RMF 1.0) to strengthen our risk evaluation processes. While originally developed for a

U.S. context, its focus on fairness, privacy, transparency, and robustness provides valuable direction for enhancing internal assessment practices. We also plan to integrate the OECD AI Principles into our broader governance strategy.

These globally recognized guidelines will help ensure that our AI systems remain human-centered, ethically grounded, and aligned with international expectations for responsible innovation.

Assessing Stakeholder Impact

At G42, stakeholder engagement is a key element of our RAI approach. Engaging with those affected by or involved in AI systems ensures our technologies are aligned with societal needs, accountable in their impact, and developed with informed awareness of diverse perspectives.

What is Stakeholder Analysis and Why It Matters

Stakeholder analysis helps us identify and understand the roles, interests, and potential influence of all parties connected to a given AI system. This includes internal teams, partner institutions, regulatory bodies, end users, and impacted communities. It enables more thoughtful design, better risk management, and more transparent, inclusive AI development.

How We Conduct Stakeholder Analysis

We begin by identifying relevant stakeholders based on their proximity to the AI system, level of influence, and potential to be affected. We then assess their needs and expectations using structured mapping techniques, such as salience analysis and interest/influence matrices. This helps us prioritize engagement and tailor communications accordingly.

Engagement Techniques

After identifying key stakeholders, we will then seek to engage them through methods appropriate to their role and relevance, including:

- Interviews for detailed, contextual understanding
- Questionnaires to collect structured input at scale
- Focus groups to explore emerging themes collaboratively
- Workshops and consultations for co-design and feedback
- Ongoing dialogue and feedback mechanisms integrated into the system lifecycle

This engagement is not a one-time effort, but a sustained process embedded into the development, deployment, and monitoring of AI systems.



CASE STUDY

SETTING THE STANDARD - G42'S ROLE IN INDEPENDENT AI EVALUATION



At G42, we believe that RAI must be measurable, transparent, and inclusive. In pursuit of this, we have invested in developing and supporting independent evaluation frameworks that help assess AI systems not just for performance, but for safety, fairness, and real-world impact.

One of our key contributions is the MEDIC Leaderboard, developed by our colleagues at M42. This framework evaluates clinical large language models across five critical dimensions: Medical reasoning, Ethics and bias, Data and language understanding, In-context learning, and Clinical safety. What makes MEDIC unique is its cross-examination approach as it quantifies model performance without relying on reference outputs, allowing for more flexible and rigorous testing across tasks like medical Q&A, summarization, and note generation.

We also launched the AraGen Leaderboard, a first-of-its-kind benchmark for Arabic generative tasks. Hosted on Hugging Face, AraGen evaluates models

using our internally developed 3C3H metric, which balances factuality and usability across six dimensions: Correctness, Completeness, Conciseness, Helpfulness, Honesty, and Harmlessness. AraGen empowers the Arabic AI community to build models that are not only high-performing but culturally and linguistically aligned.

In the energy domain, we've developed custom evaluation criteria for assessing AI models used in energy systems. These standards help ensure that models used in critical infrastructure are tested for robustness, reliability, and environmental impact.

Together, these initiatives reflect our commitment to building a safer, more inclusive AI ecosystem. By contributing to open benchmarks and transparent evaluation tools, we are helping raise the bar for responsible innovation across the global AI community.



TRAINING & UPSKILLING

► FROM PRINCIPLES TO PRACTICE: EMPOWERING OUR PEOPLE TO OPERATIONALIZE RAI

At G42, we recognize that achieving our RAI goals requires not only strong principles and governance but also a workforce equipped with the right knowledge, skills, and mindset. We are developing an ongoing comprehensive training and upskilling framework designed to embed RAI literacy and skill set across all levels of the organization, while fostering deep expertise in critical roles.

A robust and effective RAI governance framework depends on having a well-distributed network of RAI Ambassadors across the organization. These individuals are embedded within teams and assigned clear roles and responsibilities that align with their expertise.

Our commitment to this approach is outlined in Section 4 of this report, which details the structure of our RAI

ecosystem, including a dedicated RAI Team, multiple specialized RAI councils, and the integration of RAI Ambassadors throughout the organization.

Ensuring that these groups and individuals excel in their tasks require targeted training that aligns with each role and responsibilities as well as ensuring that this training is customized to utilize G42's extensive RAI Playbook.

To that end we are investing in broader organizational capability by embedding RAI training, that aims not only to raise awareness about the robust RAI work and practices that we develop and implement within G42 but also empower each team and group to be able to make the best use of the policies, guidelines, and tools available, and awareness into the workflows of our engineering, product, legal, and operations teams.

► OUR APPROACH IS GROUNDED IN A MULTI-TIERED LEARNING STRATEGY

Foundational Training:

All employees complete training AI literacy training focusing on AI risks, Responsible AI Ethical requirements, and the EU AI Act regulation. 92% of full-time G42 employees have performed this training to date, ensuring a shared understanding of core RAI principles and their implementation. This foundational awareness and basic operational capability will be continually reinforced through regular refresher courses and integrated into new employee onboarding.

Role-Specific Development:

We will provide targeted, in-depth training for technical teams, including data scientists, engineers, and product managers. This includes practical modules on bias mitigation, impact assessments, and ethical model development, equipping these teams to operationalize RAI throughout the product life cycle.

RAI Ambassadors Network:

We have established a network of multidisciplinary RAI Ambassadors embedded across functions and business units. These designated professionals will be provided with advanced training in RAI and G42's RAI Playbook and

Process to serve as local advisors and facilitators, helping to embed RAI practices at scale and fostering a culture of shared accountability.

Continuous Learning and Engagement:

Recognizing the evolving nature of RAI, we will host ongoing events, workshops, and knowledge-sharing forums. We also encourage participation in external learning opportunities and partnerships with academic and industry organizations to stay at the forefront of best practices.

Measuring Impact:

We will regularly track training participation and assess the effectiveness of our programs through feedback, competency evaluations, and alignment with key RAI outcomes. This data-driven approach allows us to refine our training continuously and respond to emerging challenges.

Through this comprehensive and evolving training ecosystem, our goal is to cultivate a culture where every team member is empowered to identify and address potential risks, and where RAI is seamlessly and efficiently integrated into our development and deployment lifecycles.



CONCLUSION



While we have made significant progress on our RAI journey, we recognize that this is an ongoing effort that requires constant reflection, innovation, and collaboration. At G42, we remain committed to evolving and strengthening our RAI practices to ensure we continue to lead in responsibly operationalizing AI across diverse sectors. We see RAI not merely as a compliance framework, but as a catalyst for building more trustworthy, transparent, and effective AI systems.

To this end, we are actively developing advanced automated governance tools that will integrate and streamline our RAI processes, reducing operational burden while ensuring rigorous adherence to our principles. These tools will support comprehensive documentation, auditability, and continuous improvement across our organization.

Looking ahead, we are eager to deepen our engagement with internal teams, external partners, regulators, and the wider AI community. Multi-stakeholder collaboration is essential to shaping RAI as a shared and scalable standard - not just within G42, but across the global AI ecosystem. We are excited to continue this journey and to contribute meaningfully to the development of responsible, impactful AI.

► APPENDIX: REFERENCES

National Frameworks and Guidelines (UAE)

1. **UAE AI Ethics Guidelines.** UAE Minister of State for Artificial Intelligence, 2019. Key principles include implementation of safety and security protocols, fairness assessments, performance validation, and clear accountability.
2. **UAE National AI Strategy 2031.** UAE Government. Sets out governance standards, national AI ecosystem alignment, and testing methodology requirements.
3. **UAE AI and Robotics Guidelines.** UAE Council for AI and Blockchain, 2023. Covers risk assessment frameworks, security testing, cultural considerations, and compliance with national data protection requirements.
4. **The UAE Charter for the Development & Use of Artificial Intelligence.** UAE, July 2024.
5. **UAE AI Ethics Principles & Guidelines.** UAE Minister of Artificial Intelligence, 2022.

International Standards and Frameworks

Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology (NIST), USA, 2023.

Artificial Intelligence Act (AI Act). European Union, 2024.

Microsoft Voluntary Commitments to Advance Responsible AI Innovation. Microsoft, 2023.

Hiroshima Process: International Guiding Principles for Organizations Developing Advanced AI Systems. G7, 2023.

Ethics Guidelines for Trustworthy AI. High-Level Expert Group on AI, European Commission, EU, 2019.

OECD AI Principles for Trustworthy AI. OECD, 2019.

Ethically Aligned Design. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019.

Recommendations on the Ethics of Artificial Intelligence. UNESCO, 2022.