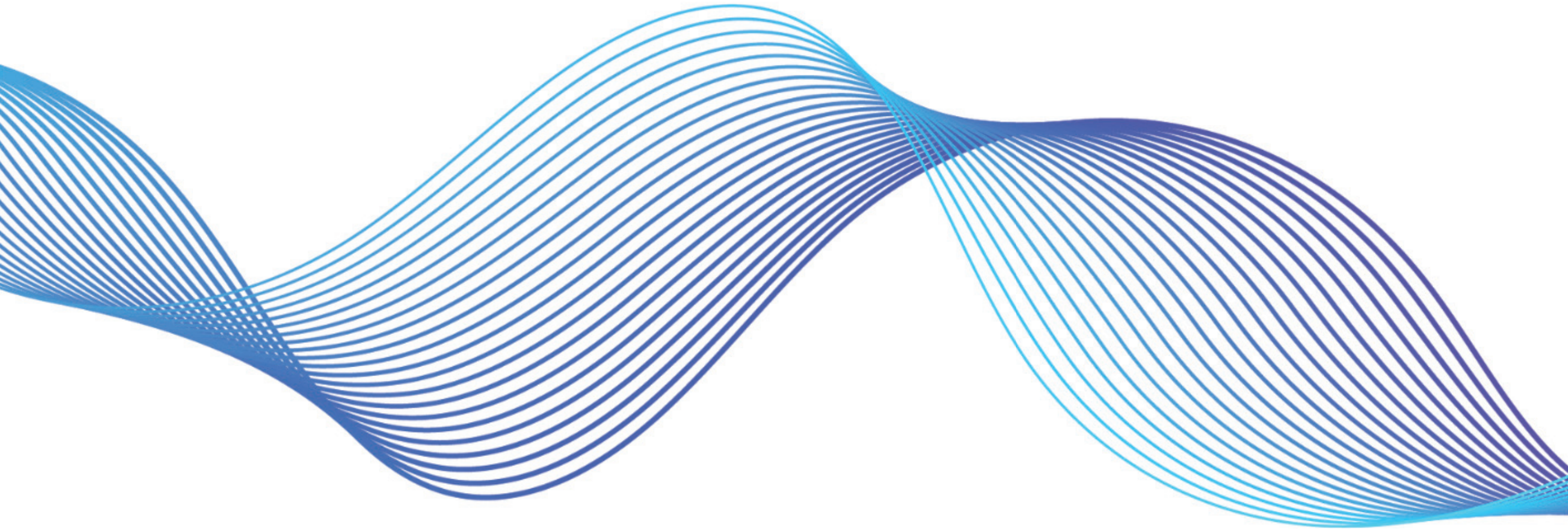
LTIMindtree

Whitepaper

Managing AI Risks

The Role of Governance in Ensuring Ethical AI Practices

Table of Contents



01 Introduction

Artificial Intelligence (AI) has seen remarkable growth in the past decade, driven by increased data availability, advancements in deep learning algorithms, enhanced GPU performance, and cloud computing. However, with AI becoming more accessible, concerns about its misuse have risen significantly. One such example is the Cambridge Analytica scandal, where data from over 87 million Facebook users was misused using AI to influence political events. AI was employed not only for unauthorized data handling but also for real-time testing and optimization of online ads based on individuals' personalities. This, along with numerous other incidents, has raised important questions about AI's social, political, and economic implications. There is an increasing need for regulatory policies and best practices to address concerns regarding AI's alignment with ethics, sustainability, and privacy.

AI governance refers to the establishment of policies and practices to address concerns surrounding AI ethics and lay the foundation for its responsible and safe usage. Such governance guidelines can be leveraged to create a governance framework that integrated governance with ethics review, bias detection, and monitoring of AI systems.

In this paper, we discuss the need for AI governance and its role within the socio-technical ecosystem. Also, by examining AI risk tolerance and evolving regulatory landscape, we learn how organizations can navigate AI governance complexities to build trust, reduce risks, and capitalize on opportunities.

What is AI Governance?

AI governance is the collection of rules and guidelines designed to ensure that AI tools and systems are safe, responsible, and ethical. It involves establishing best practices, frameworks, and processes to manage risks, including bias, privacy concerns, and potential misuse. Since AI systems are created by humans, they are prone to human errors and biases. AI governance provides a structured approach to minimizing these risks, ensuring that algorithms are constantly evaluated, monitored, and updated to prevent flawed or harmful decisions.

Effective **AI governance policies** enhance an organization's capacity to identify and manage risks while leveraging high-quality, industry-specific data from vast databases. Implementing AI governance requires leadership and executives to carefully deliberate on its applications, scalability, and accountability for ensuring smooth implementation.

02 The need for AI governance

AI was originally developed to automate repetitive laborious processes and augment system capabilities. Organizations has sought to improve their efficiency, productivity, and profitability using AI capabilities. While AI has delivered on its promise of increased productivity, cost savings, and competitive advantages, it has also raised ethical implications regarding its deployment. These concerns have led to calls for responsible AI practices and a revaluation of corporate responsibility. One of the primary concerns is transparency in AI decision-making processes. Deep learning algorithms, particularly neural networks, often operate as "black boxes," making it difficult to understand how they reach specific conclusions. This opacity increases the risk of unintended outcomes or flawed decisions if not properly supervised. Currently, there is no simple method for auditors, customers, or internal stakeholders to trace and understand the decision-making process of an AI model.

Another concern is the potential bias embedded in AI algorithms. Complex AI models may deliver biased outcomes if they are trained on unfiltered data. For example, facial recognition systems have shown considerable bias against people of color. Studies, such as those from the MIT Media Lab, have highlighted how these systems misidentify individuals with darker skin tones at much higher rates compared to those with lighter skin tones. Similarly, hiring algorithms used by companies have demonstrated gender bias, favoring male candidates over females.

While AI models may appear unbiased during deployment, they can develop biases over time. AI systems analyze historical data, and if that data contains biases, the system may perpetuate and even exacerbate those biases. A notable instance is Amazon's AI hiring tool, which was ultimately abandoned due to its bias against female candidates, resulting from training on predominantly male resumes. Current governance frameworks are insufficient in addressing algorithmic safety, particularly in monitoring AI performance for bias and "drift" over time.

Training and retraining AI models are both costly and resource-intensive. If an AI model is found to be non-compliant with the existing regulatory standards, the entire model must be retrained, leading to significant financial penalties and reputational damage. Without a robust AI governance framework, organizations risk severe consequences as AI systems may fail, provide inaccurate results, or exhibit unwanted behavior.

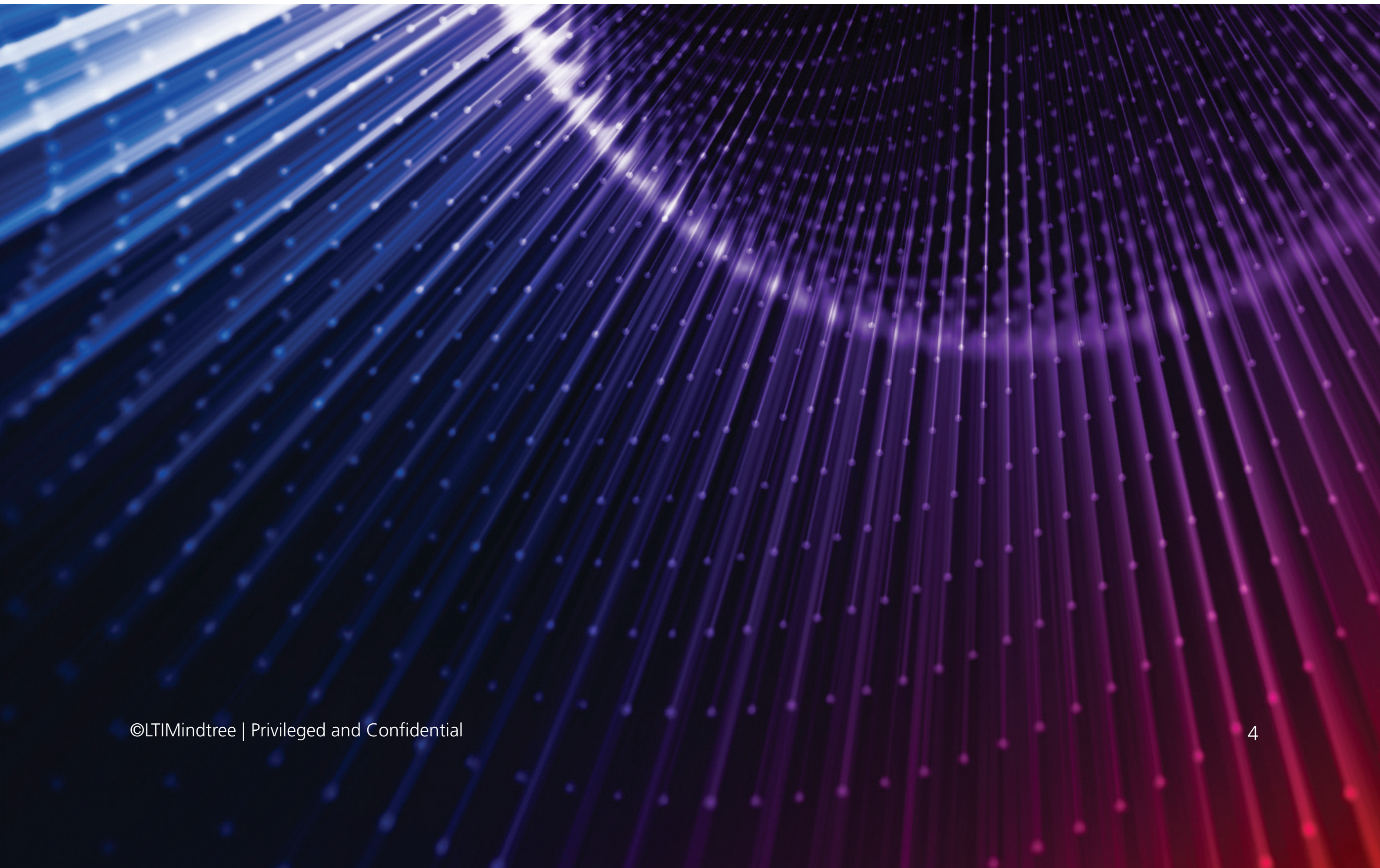
Current AI regulation, such as the EU AI Act, AI regulations in the UK, and the U.S. President's Executive Order on AI, are still in their early stages and are not yet fully equipped to address emerging risks. It remains to be seen how these regulations will be enforced, how they will align with existing norms like GDPR, and whether they will be rigorously followed.

03 AI governance: Acknowledging the socio-technical ecosystem

To fully understand AI governance, it is essential to view AI within a broader socio-technical context. Modern AI has evolved from simple automation tools to sophisticated systems capable of controlling access to various technologies. The individuals who control AI wield significant power, deciding which tasks to automate and what data to use for training. This power also extends to the relationships between decision-makers and those affected by AI-driven decisions. AI governance acts as a safeguard, protecting society from malicious uses of AI systems.

AI governance helps define the problems AI should address, the types of AI that should be developed, and how AI interacts with other technologies. It also dictates the societal values that should guide AI's development and deployment. The development of AI governance frameworks requires input from a diverse set of stakeholders, including AI systems vary in their capabilities and applications. Some systems can primarily automate (basic, minor) tasks and decision-making processes, while others can comprehend the meaning and the context of their actions and outcomes. Thus, developing AI governance will need input from various stakeholders, including AI developers, users, policymakers, and ethicists, ensuring that AI systems align with societal values.

AI operates within a socio-technical ecosystem, meaning AI governance must account for the interplay between people and technology. Strong governance structures can significantly influence societal outcomes and AI adoption.



04 Case study: AI system failure

The development of advanced AI models, including large language models, has notably heightened their risk levels. The intricate nature of these models makes them more prone to adversarial attacks, where minor, deliberate changes to input data can result in incorrect or damaging outcomes. Hence, there is a need to modify existing governance policies and simultaneously develop new ones to handle such risks and challenges. The absence of such appropriate governance policies can lead to AI failures with severe consequences.

The Dutch childcare benefit scandal illustrates the critical need for AI governance. The Dutch Tax and Customs Administration used a self-learning algorithm to classify childcare benefit claims based on perceived risk. The system included discriminatory factors like nationality, ethnicity, and postal codes, disproportionately flagging parents from ethnic minorities and low-income households. Thousands of families were wrongfully accused of fraud, leading to financial hardship, broken marriages, and even children being placed in foster care.

This case study showcases the need for an AI governance, especially in regulating AI decision-making. In this case, lack of transparency and explainability made it difficult for affected individuals to challenge the decisions. With AI governance in place, errors like these could be identified and corrected early, preventing large-scale harm. It also highlights the need for ethical guidelines to stop discrimination and ensure fairness in AI-driven decisions.

AI governance would have tracked and isolated such errors and alerted the authorities of AI system failure. This case also serves as an example highlighting the need for governance policies to monitor the use of AI in sensitive areas like public services and emphasizes the need for ethical guidelines to prevent discrimination and ensure justice for all individuals affected by AI-driven decisions. Evaluation of such risks and weighing them against organization's business objectives are of paramount importance to form a sturdy AI governance policy.

05 Regulatory landscapes shaping AI policies

Since the introduction of the GDPR in 2018, many organizations have adopted data protection frameworks to comply with stringent regulations. Now, business leaders are looking for similar regulations to ensure the ethical use of AI. Key regulations include the European Union AI Act (EU AIA), AI regulation frameworks in the UK, and the U.S. President's Executive Order on AI.

European Union AI Act (EU AIA)

The EU AIA is a comprehensive framework that categorizes AI application areas into three risk categories:

- **Unacceptable risk**: Systems or solutions that create unacceptable risks, such as the social scoring system in China or manipulative AI
- **High risk**: Use of AI for CV scanning and application profiling, unethical AI and applications where certain legal requirements must be met. Developers and service providers are held majorly accountable for implementing high-risk AI applications
- **Unregulated risk**: Solutions that do not fall under any of the above categories

This act bans using AI systems that employ unethical techniques to manipulate behavior and impair informed decision-making. It also prohibits AI systems that exhibit age, disability, or socio-economic circumstance-related biases, creating distorted outputs. Using biometric categorization systems and social scoring systems that leverage attributes like race, political opinions, or sexual orientation is prohibited.

AI systems that perform risk profiling without objective and verifiable facts are also banned under this act. Further, creating facial recognition databases using images from the internet and CCTVs is also prohibited. Meanwhile, the use of real-time remote biometric identification in public spaces for law enforcement is restricted and is only done for searching missing persons, preventing imminent threats, or identifying suspects in serious crimes.

This act gives definitive guidelines for general-purpose AI (GPAI) providers. Before this act, minimal risk applications currently available in the market were unregulated. Now, as per the EU AIA act, developers must let the user know that they are interacting with AI chatbots/deepfakes. Therefore, all GPAI model providers must:

- Provide technical documentation and usage instructions
- Follow copyright laws
- Publish a summary of the training data

All GPAI model providers that pose a systemic risk must also:

- Evaluate their modelsFollow copyright laws
- Conduct adversarial testing
- Track and report serious incidents
- Ensure cybersecurity protections

It is critical to know that most of the obligations outlined in this act are expected to become effective by the first half of 2026.

AI regulation framework in the UK

On February 6, 2024, the UK Government reaffirmed its "pro-innovation" approach to AI regulation, based on its response to the 2023 consultation on AI regulation led by the Department for Science, Innovation and Technology (DSIT). The UK’s non-statutory framework is designed to apply across sectors, aiming to balance fostering innovation while ensuring safety, while leveraging the existing technology regulatory landscape. Though the government acknowledges that legislation may be necessary in the future, particularly concerning general-purpose AI systems, it currently believes such actions would be premature. Instead, it advocates for a better understanding of AI risks, regulatory gaps, and suitable solutions before proceeding with formal legal measures. This contrasts with the more prescriptive legislative paths seen in the EU and, to a lesser degree, the US, highlighting potential divergence in global AI regulatory strategies.

The UK Government's framework is grounded in five core principles: safety, security and robustness, transparency and explainability, fairness, accountability and governance, and contestability and redress. Regulators are tasked with enforcing these principles within their specific sectors, working under current laws but with added regulatory guidance. Annual AI strategic plans are expected from key regulators, with the next release due by April 30, 2024, to offer businesses with clear expectations. Although not legally binding yet, the framework encourages voluntary adoption of safety and transparency standards for advanced AI developers. The Government has not ruled out future legislation to address regulatory shortcomings, especially in dealing with the complexities posed by general-purpose AI and its developers.

US President’s executive order on AI

The US President’s executive order on AI introduces comprehensive standards to mitigate potential AI-related risks, ensuring the development of secure and trustworthy AI systems. AI developers will be required to disclose safety test outcomes and relevant data to the government. The National Institute of Standards and Technology (NIST) has been assigned the responsibility of formulating standards and testing protocols to ensure the safety of AI systems. In critical infrastructure sectors, the Department of Homeland Security will oversee the implementation of these standards and form an AI Safety and Security Board to enforce compliance. In response to concerns about AI being used to create dangerous biological substances, new biological synthesis screening processes will be mandated, particularly for federally funded life-science research.

To combat AI-driven fraud and misinformation, the government will introduce AI-generated content detection systems and establish methods to verify the authenticity of official content. The Department of Commerce will draft content verification guidelines, including watermarking for clear identification of AI-generated content. The Biden Administration also plans to bolster its cybersecurity efforts, leveraging AI tools to detect and address vulnerabilities in critical software, complementing the AI Cyber Challenge initiative. A National Security Memorandum will further outline steps to ensure that AI use within military and intelligence operations adheres to ethical and safety standards, and counter strategies for foreign military AI use will be developed.

06 Conclusion

As generative AI continues to drive transformation in business operations, organizations are steadily moving from proof of concept to large-scale AI imFFplementation. However, this journey is not without its challenges. Issues such as inconsistent regulatory frameworks, unintended bias, AI model hallucinations, and the escalating costs of training and retraining models have surfaced, complicating the path toward AI standardization and governance.

Despite these hurdles, businesses recognize the critical need for robust AI governance frameworks that address these risks effectively. Currently, organizations are cautiously initiating low-risk AI projects while making significant investments in Proof of Concept (PoC) creation. As AI adoption expands, the focus on ensuring security, data protection, privacy, and governance will become even more vital.

Standardization and risk governance will serve as the backbone of successful AI integration. While regulatory guidelines like the EU AIA provide a foundation, they alone will not guarantee immediate success. Instead, developing tailored internal governance frameworks that promote responsible AI usage is key to long-term success. This underscores the necessity of embedding ethical, transparent, and accountable AI practices within organizational structures, ensuring sustainable growth and trust in AI-driven initiatives.

07 Authors

Bablu Lawrence

Managing Principal – Architecture, Enterprise AI

An experienced technology leader with over 22 years in application development, data, and AI/ML, Bablu Lawrence is known for driving innovation and delivering high-impact solutions. Currently, as part of the Enterprise AI service line, he helps clients unlock the potential of AI through advanced, AI-driven solutions.

Bharat Trivedi

Senior Principal – Architecture, Global Technology Office

As generative AI continues to drive transformation in business operations, organizations are steadily moving from proof of concept to large-scale AI implementation. However, this journey is not without its challenges.

Hakimuddin Bawangaonwala

Senior Executive – Consulting, Global Technology Office

A seasoned consultant at LTIMindtree, Hakimuddin has over 5 years of experience exploring beyond-the-horizon technologies. He has collaborated on a diverse range of projects, helping organizations create use cases for rapid incubation and industrialization.

08 References

1. Building trust in AI: How to overcome risk and operationalize AI governance, Preeti Shivpuri, et.al, Deloitte, 2021 https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/financial-services/ca-omnia-ai-operation-trust-pov-aoda-en.pdf
2. Perspectives on Issues in AI Governance, Google https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf
3. Vic Katyal, AI governance for a responsible, safe AI-driven future, Deloitte, 2021 https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-ai-governance-for-a-responsible-safe-ai-driven-future-final.pdf
4. Towards a Trustworthy Transformation: Empowering Governance for AI Implementation, KPMG, 2024 https://assets.kpmg.com/content/dam/kpmg/nl/pdf/2024/services/empowering-governance-for-ai-implementation-anonymous.pdf
5. Trusted AI and AI governance, EY, 2023 https://assets.ey.com/content/dam/ey-sites/ey-com/en_ca/topics/ai/ey-trusted-ai-and-ai-governance-discussion-paper.pdf
6. Generative AI Risk and Governance Way towards a responsible and trusted AI, EY, January 2024 https://assets.ey.com/content/dam/ey-sites/ey-com/en_in/topics/ai/2024/ey-generative-ai-risk-and-governance.pdf
7. The urgency of AI governance, IBM https://www.ibm.com/downloads/cas/MV9EXNV8
8. Tim Mucci, Cole Stryker, What is AI governance, IBM, October 2024 https://www.ibm.com/topics/ai-governance
9. EU Artificial Intelligence Act, EU, February 2024 https://artificialintelligenceact.eu/high-level-summary/
10. President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, press release, October 2023 https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/
11. Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST, January 2023 https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
12. The Dutch childcare benefit scandal, institutional racism and algorithms, press release, June 2022 https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html
13. Responsible Artificial Intelligence – from Principles to Practice, Virginia Dignum, Umea University, Sweden June 2022 https://arxiv.org/pdf/2205.10785
14. A global scale comparison of risk aggregation in AI assessment frameworks, Anna Schmitz, Michael Mock, Rebekka Görge, Armin B. Cremers & Maximilian Poretschkin, May 2024 https://link.springer.com/article/10.1007/s43681-024-00479-6
15. AI and ethics in business: A comprehensive review of responsible AI practices and corporate responsibility Funmilola Olatundun Olatoye, et.al. International Journal of Science and Research Archive, 2024, 11(01), 1433–1443 https://ijsra.net/sites/default/files/IJSRA-2024-0235.pdf
16. The UK's framework for AI regulation: Agility is prioritised, but future legislation is likely to be needed, Valeria Gallo, Suchitra Nair, February 2024 https://www.deloitte.com/uk/en/Industries/financial-services/blogs/the-uks-framework-for-ai-regulation.html
17. Ted Cruz using firm that harvested data on millions of unwitting Facebook users, Harry Davies, Guardian, December 2015 Ted Cruz using firm that harvested data on millions of unwitting Facebook users | Ted Cruz | The Guardian: https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data
18. Amazon's sexist hiring algorithm could still be better than a human, Maude Lavanchy, November 2018 Amazon's sexist hiring algorithm could still be better than a human - IMD business school for management and leadership courses: https://www.imd.org/research-knowledge/digital/articles/amazons-sexist-hiring-algorithm-could-still-be-better-than-a-human/

LTIMindtree