LTIMindtree

Whitepaper

Introduction to Human-Centered AI

LTIMindtree

Table of Contents

Executive Summary

Researchers started discussing Human-Centered AI (HCAI) around 2019. What Is Human-Centered AI? The philosophy of HCAI is to synchronize and find a balance between humans, AI technology, and ethics. By 2024, HCAI has evolved into a novel approach that fosters collaboration between humans and AI8. Unlike traditional AI, which is seen as a tool for assistance, HCAI envisions AI as a cognitive teammate capable of effective collaboration, always under human oversight.

HCAI focuses on designing AI systems that enable computer automation while ensuring appropriate human oversight. It can recognize scenarios and tasks where human control is necessary or whether AI should autonomously execute the instructions. Over time, HCAI can be trained to mimic the information processing capabilities of human agents, thus laying an architectural foundation for the development of AI agents.

This whitepaper aims to define and conceptualize HCAI, help the readers understand what Is Human-Centered AI, showcasing its benefits through various use cases. It provides guidance for organizations aiming to adopt HCAI and offers strategies for successful integration.

Introduction

Technological advancements have always influenced the interaction between humans and machines. Initially, machines were designed to adapt to human needs, but over time, this relationship evolved into a collaborative one. In this age of Artificial Intelligence (AI), the human-machine relationship is poised to evolve into a dynamic human-AI collaboration where AI systems function as collaborative teammates and as supportive tools.

As AI advances, we are also bombarded with scary studies and urgent warnings to watch AI technologies closely and take the right steps. From AI systems showing biases while hiring to concerns around AI replacing human jobs, there are intensified calls for AI systems that are more human-centric. Such cases have accelerated the efforts toward making AI safer and aligned with human values and needs.

Researchers and organizations have taken initial steps to address these challenges and have come up with an alternative form of AI called “Human-Centered Artificial Intelligence (HCAI).” HCAI has multiple different definitions, but its widely recognized definition was articulated by Ben Shneiderman in 2020. He states that HCAI systems will be designed to amplify, augment, and enhance human performance, thus ensuring that systems are reliable, safe, and trustworthy. These systems are designed to support human self-efficacy, foster creativity, clarify responsibility, and facilitate social participation.

Traditionally, the design and development of AI systems prioritizes creating algorithms proficient in specific tasks, such as image recognition, natural language processing, or automation. The focus was on enhancing performance with limited emphasis on implications for human values or societal impact. However, HCAI recognizes the influence of AI on individuals, society, and the broader human condition, placing humans at the forefront. Its research strategies underscore that the next frontier in AI is not solely technological but also encompasses humanistic and ethical dimensions.

Technology Insights

Contemporary literature has given us a conceptual understanding of HCAI. However, there is a lack of research that offers a practical definition, application, and implementation methods. For instance, various HCAI frameworks have been proposed and examined, but these frameworks focus largely on high-level strategic design philosophies. While AI governance and ethical AI design principles are being established, the framework and methodologies to facilitate their integration into AI system development remain a challenge.

In 2024, researchers developed a conceptual framework for HCAI that leverages situation awareness cognitive engineering theory, joint cognitive systems theory, and AI agent theory to address the challenge.

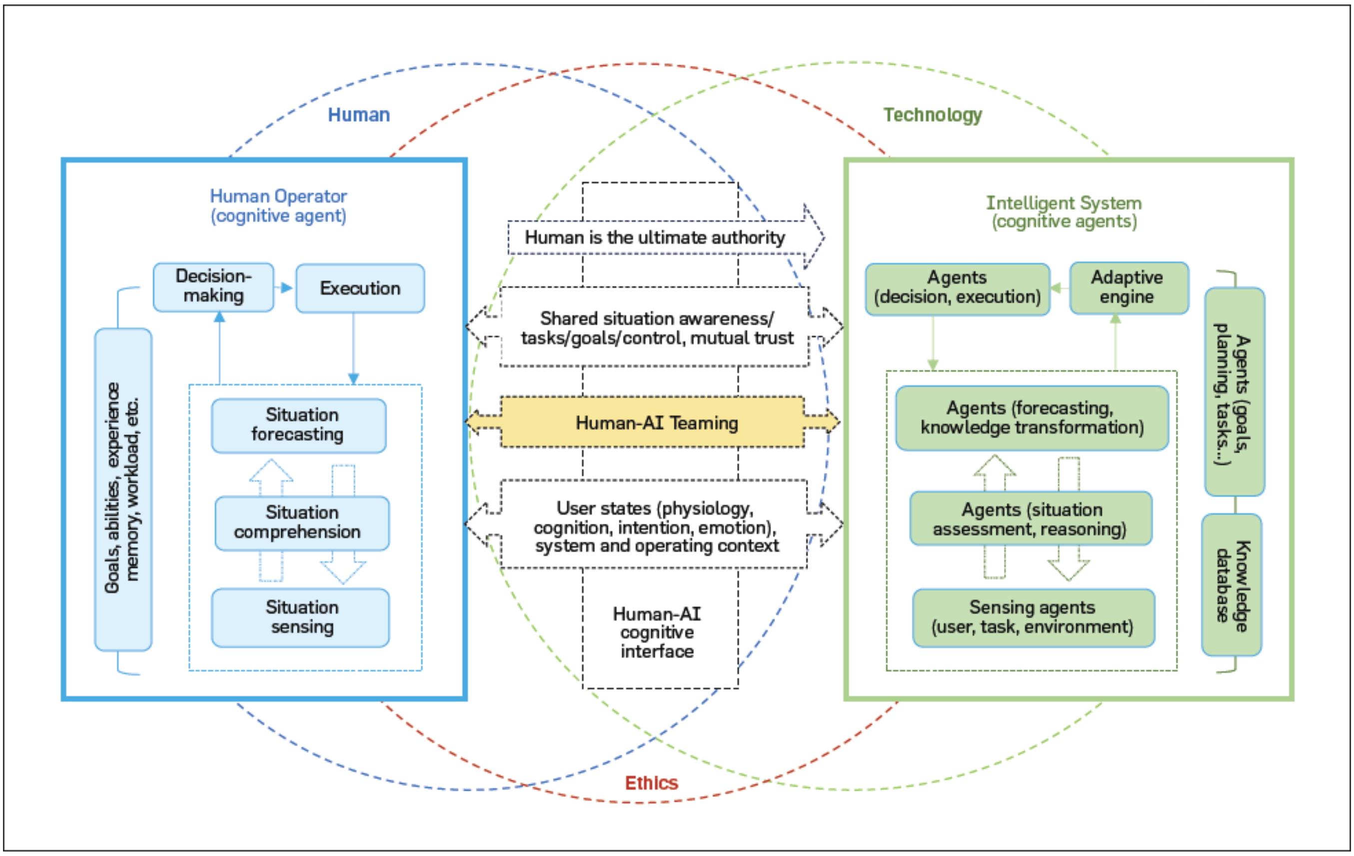

Figure.1: Applying HCAI in developing effective human-AI teaming: A perspective from human-AI joint cognitive systems, Wei Xu, Zaifeng Gao, November 2023: https://arxiv.org/pdf/2307.03913

As illustrated in Figure 1, the HCAI framework views AI systems (with one or more AI sub-systems) as cognitive agents capable of performing specific tasks autonomously. This framework utilizes Endsley's situation awareness theory to depict the information-processing mechanism of both cognitive agents. The model outlines how human users perceive and comprehend the current state of the environment and predict future situational states using Endsley's three levels of information processing. This includes cognitive interactions involving situation awareness, memory, experience, and knowledge. HCAI replicates the architecture employed by human agents to represent the information processing mechanism for AI cognitive agents.

Both cognitive agents work together through a human-AI cognitive interface powered by multimodal technology, maintaining shared situation awareness, tasks, goals, trust, decision-making, and control, all under Human-Centered AI (HCAI) principles, where the human remains the ultimate authority. This HCAI framework elevates the capabilities of AI agents (sensing, reasoning, decision-making, and execution) by integrating its human-driven information processing mechanism. This will elevate AI agents to a programmable entity with autonomous features that don't discriminate between human and AI capabilities.

Human-centered AI is not a standalone system but a collaboration of multiple models and human-in-loop. There are user-centric models like Dynamic Personalization Model (DPM) and User Experience Modeling (UEM), which are used to create human-AI cognitive interfaces with a focus on improving user experience and offering personalized recommendations. Collaborative models like Human-AI Teaming (HAT) and Collaborative Filtering Algorithms (CFA) enable seamless interaction and cooperation between humans and AI systems through joint decision-making algorithms and shared control mechanisms.

The ethical engine forms a key aspect of the HCAI framework. These engines comprise agents like ethical models, explainable AI models, and reinforcement learning algorithms. The ethical models employ techniques like privacy-preserving data analysis techniques and fairness-aware ML algorithms to ensure responsible AI decision-making. Explainable AI models make the model outputs interpretable and track and trace the reasoning process of the AI system, thus ensuring transparency throughout the AI decision-making process.

Industry Use Cases

The following figure encompasses some of the use cases where human-centered AI systems can yield positive and beneficial results for both their immediate users and society.

Banking and Financial Services	• Reduce the risks of high frequency trading, which can cause sudden market crashes and huge financial losses. • Keep track of trading logs to monitor and analyze trades in real-time, allowing them to take quick action to prevent or fix problems, thus protecting financial markets and investors. • In combination of domain and context-specific prebuilt multilingual natural language processing (NLP) chatbots (both voice and text), HCAI can develop dynamic customer profiles and suggest the next best actions. • Can be deployed with machine vision for biometric authentication and document ingestion, and task bots with microservices to facilitate seamless transactions.
Healthcare	• Analyze vast quantities of healthcare information, including images, clinical trials, and medical claims, uncovering patterns and insights that might be overlooked by humans. • Adapt existing drugs to fight newer illnesses by analyzing scientific literature. • Create customized treatment plans by analyzing data from clinical trials, medical claims, and scanned images. • Enable the development of intelligent devices that employs image and facial recognition, text summarizations, etc. to aid individuals in overcoming both physical and cognitive disabilities.
Manufacturing	• Enhances requirements gathering in design for manufacturing process through predictive analytics and data mining. • Identify inefficiencies in operator as well as process and suggests optimizations in production. • AI-based simulations evaluate adjustments prior to implementation, forecasting potential impacts on performance and safety. • Optimize raw material combinations and Additive Manufacturing product geometry by analyzing design variables and their interactions.
Retail	• Interpret non-selection as a form of input to develop an extensive model to understand why users do not choose a recommended item. • HCAI can use eye tracking to discern the probable order in which users scan the screen. It can then integrate survey data, familiarity estimates, and the recommendation context to forecast whether a recommendation should be repeated soon or withheld for a period.
Media and Entertainment	• Scrutinize users' listening patterns and preferences to offer highly customized suggestions. This user-centric strategy improves the user experience by personalizing content to match individual tastes, showcasing how HCAI can effectively understand and cater to user needs.

Figure.2: HCAI Use Cases

Other use cases

Hyper-personalization: HCAI-enabled conversational ads enable meaningful, personalized interactions between brands and customers. These ads also include actionable options, such as call-to-action buttons, to engage the conversation further. Consequently, these dynamic ads allow digital marketers to go beyond traditional static and video display assets, enhancing direct-to-consumer interactions and creating new purchase pathways without needing third-party navigation. Leveraging these AI-powered conversational ads is essential, as failing to do so would be a missed opportunity.

EdTech: HCAI can automate repetitive tasks like evaluating student performance to help educators save valuable time. It can also derive insights into each student's specific needs, helping teachers customize their courses to address critical knowledge gaps or problem areas before students fall too far behind.

HCAI can be used to tailor training programs to meet each student's unique needs. It can also predict learning outcomes beforehand, enabling tutors to develop personalized content aligning with each learner's objectives and achievements.

LTIMindtree

How to design a human-centric AI-based system?

The integration of AI components in software solutions has experienced substantial growth in recent years. However, these solutions often prioritize technical dimensions, overlooking critical human considerations. Incorporating these human-centered elements during Requirements Engineering (RE) can lead to more responsible, unbiased, and inclusive AI-based software.

To address this, we have developed a new framework to integrate human-centered AI guidelines into the requirements-gathering process for AI-based software. This framework will enable the project team to comprehensively understand and cater to human-centered needs. Additionally, it will distinguish which requirements should be captured during the preliminary stages versus the later stages of the AI-based software engineering lifecycle.

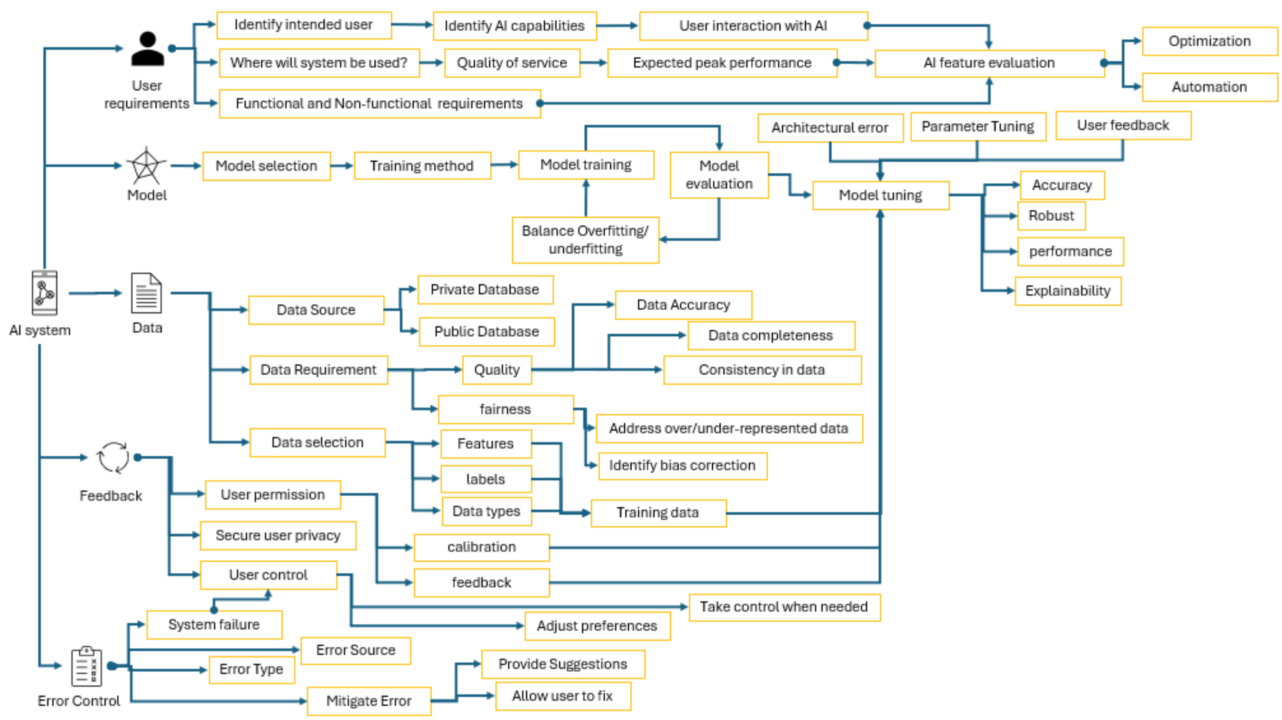

Figure.3: HCAI framework for software development

User requirements: Capturing user requirements is the first step in developing human-centered AI-based software. This involves documenting the system's capabilities, understanding its functions, evaluating its performance, and identifying the users. Consideration must also be given to how users interact with the system and select suitable approaches to meet their needs. It is essential to decide whether the system will automate or optimize the user's tasks. AI-based software can automate tasks that do not require human oversight, at the same time AI can augment tasks that users find enjoyable or necessitate human involvement to enhance the user experience and boost efficiency.

Finally, evaluating AI features and outlining the rationale behind the selected function is important. It is necessary to know how AI will discern correct from incorrect predictions within a learned model. AI functionality and output assessment will vary according to the task and domain of the AI application in use. The choice between precision and recall is determined by weighing their respective trade-offs. Consequently, when assessing the reward function, it is important to document these trade-offs to justify the chosen function.

Models: The second area involves detailing the requirements for models. It deals with questions like, Should the system be more explainable or more accurate? Various algorithms yield accurate results but are poor in terms of explainability. Conversely, other algorithms might offer better explanations but with less confident predictions. Additionally, it is essential to define the potential settings for each variable of a given AI algorithm when applied to a specific task in a certain context.

While selecting models, it is important to define how the Learning Model (LM) will improve with new data: should it be a dynamic LM that updates and trains online or a static system that advances only with updates? Model selection hinges on the algorithm choice, encompassing supervised, unsupervised, and reinforcement learning methods. During system training, it is essential to set a threshold to prevent overfitting or underfitting and to choose appropriate tools for evaluating the LM. When tuning the LM, consider the types of feedback, user behavior, and training data used in model tuning and adjust the parameters accordingly.

Data: The third area pertains to data requirements and collection strategies. This includes determining the type and volume of data necessary. After specifying the data collection methods, it is crucial to consider data requirements such as quality, security, privacy, and fairness. Data quality comprises three elements: accuracy, completeness, and consistency. However, it is also important that data is highly credible. Credibility is achieved by ensuring the data is truthful.

Data quantity emphasizes the importance of data diversity, while data quality focuses on ensuring the data is complete, consistent, and correct. When selecting data, it is important to identify the features, labels, and sampling rates. Labels are used to identify the necessary features for training the machine learning model, such as labeling a scanned image of a tumor as malignant. Explicit labeling is manually performed, whereas implicit labeling occurs when the model independently learns the pattern. Examples, representing rows of data, contain features, and labels provide descriptions for the data. Having more samples in the dataset enhances diversity but also increases costs. Therefore, it is necessary to set a threshold to determine the required amount of data within the available budget.

Recognizing and addressing biases in data is crucial. These biases can encompass automation, selection, and group attribution. Automation bias occurs when choices are influenced by the system's automated recommendations. Selection bias typically arises when data is gathered based on the preferences of stakeholders rather than through random sampling from the target population. Group attribution bias involves the assumption that an outcome appropriate for one individual will be equally effective for all members of the group. To prevent discrimination and biases, it is essential to establish key data characteristics early in the process.

Feedback: This area focuses on incorporating user feedback into requirements engineering for AI-based systems. Feedback can be implicit, explicit, and calibration. It is crucial to identify the timing and usage of feedback in model tuning and its impact on the AI system. Privacy measures must be implemented to protect user feedback, and users should be allowed to dismiss feedback requests. Providing users with control over the system, particularly in case of failure or by offering multiple choices, is vital for fostering a sense of control, which is central to human-centered AI.

Error correction: Errors can be categorized into background, contextual, and system limitations. Background and context errors are particularly challenging to detect and are often invisible to the end user. These errors typically occur when the outcome is technically correct but does not meet the user's requirements. Error sources can be system errors, incorrect predictions, data errors, and input/output errors. Data errors can result from mislabeled data, poor training, or inaccurate labeling. Prediction errors may occur when an inappropriate model is used, the data lacks comprehensiveness, or critical elements are missing. Input errors happen when users input unexpected data, often due to outdated user habits or abusive behavior. Output errors occur when the system produces low-confidence predictions or irrelevant high-confidence outputs. Additionally, system errors can arise when multiple AI-integrated systems interact or depend on one another. Identifying error types and sources in RE and formulating an action plan to address and rectify them is crucial.

Conclusion

Human-centered AI necessitates a steadfast commitment to ethical principles that guide AI development. Ensuring that AI systems are equitable, accountable, and transparent is crucial in building trust and reliability. HCAI does not aim to replace human roles but to enhance human capabilities, fostering a collaborative environment where individuals can achieve more together than they could alone. This approach acknowledges the unique strengths of both humans and AI, promoting a synergistic relationship that maximizes their combined potential.

The success of HCAI hinges on integrating insights from diverse disciplines, including computer science, psychology, ethics, and design. Such interdisciplinary collaboration ensures that AI technologies are developed with a comprehensive understanding of human needs and societal impacts.

Establishing clear accountability and governance structures is vital in HCAI. This involves creating systems for human oversight of AI and ensuring that AI decisions can be explained and justified appropriately.

As we look to the future, the potential for dystopian scenarios is evident. However, HCAI stands as a beacon of hope and progress. It offers a path to harnessing the power of AI in a manner that respects and amplifies human potential. The journey of HCAI is just beginning, and its ultimate success will depend on a steadfast commitment to its principles and a vision of a technology-enhanced world that prioritizes human well-being and potential above all else. For any questions on what is human-centered AI.

Authors

Bablu Lawrence

Managing Principal – Architecture
Enterprise AI

Bharat Trivedi

Principal – Architecture
LTIMindtree Research

Hakimuddin Bawangaonwala

Senior Executive - Consulting
LTIMindtree Research

References

- Human-Centered Artificial Intelligence: a review, Emmanuel Adjei Domfeh et.al, University of Energy and Natural Resources, March 2022: https://www.researchgate.net/publication/358936588
- Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, Ben Shneiderman, March 2020: https://www.tandfonline.com/doi/full/10.1080/10447318.2020.1741118
- Human-centered AI: The role of Human-centered Design Research in the development of AI, Jan Auernhammer Stanford University, United States of America, August 2020: https://dl.designresearchsociety.org/cgi/viewcontent.cgi?article=1178&context=drs-conference-papers
- What is human-centered AI?, IBM, March 2022: https://intgovforum.org/sites/default/files/webform/igf_2024_workshop_proposal_form/252910/HCAI%20%26%20innovation%20-%20IGF%202024%20%20proposal%20documents-compactado.pdf
- Overview on the current techniques to display contextual information with adaptive AR, Daniel Onumbu, Ludwig-Maximilians-Universität München, München, Deutschland, 2020: https://www.um.informatik.uni-muenchen.de/aktuelles/ubiaction2021_proceedings/ubiaction2020_2.pdf#page=68
- Human-Centered AI,' An In-Depth Study Of The Current State Of The Artificial Intelligence Concept, David A. Teich, Forbes, Feb 2022: https://www.forbes.com/sites/davidteich/2022/02/15/human-centered-ai-an-in-depth-study-of-the-current-state-of-the-artificial-intelligence-concept/
- Human-centered AI: The ethics of designing new technology, Mindful Studio, Oct 2023: https://medium.com/@mindful_studio/human-centered-ai-the-ethics-of-designing-new-technology-984ea5333012
- Ensuring a human-centered approach to AI, Deloitte. June 2023: https://www.deloittedigital.com/us/en/insights/perspective/human-centered-ai.html
- Defining Human-Centered AI: A Comprehensive Review of HCAI Literature, Stefan Schmager et.al, September 2023: https://www.researchgate.net/publication/373019807_Defining_Human-Centered_AI_A_Comprehensive_Review_of_HCAI_Literature

- The Rise of Human-centered AI: A General Overview of HCAI, Morgan Hopper, Medium, Apr 11, 2024: https://medium.com/@morgan.hopper/the-rise-of-human-centered-ai-a-general-overview-of-hcai-c919eeaca4d5
- Requirements engineering framework for human-centered artificial intelligence software systems, Khlood Ahmad et.al, Applied Soft Computing, Vol.143, August 2023: https://www.sciencedirect.com/science/article/pii/S1568494623004738?ref=cra_js_challenge&fr=RR-1
- Applying HCAI in Developing Effective Human-AI Teaming: A Perspective from Human-AI Joint Cognitive Systems, Wei Xu et.al, Interactions, Volume 31, Issue 1,Pages 32 – 37, 10 January 2024: https://dl.acm.org/doi/abs/10.1145/3635116
- An HCAI Methodological Framework: Putting It Into Action to Enable Human-Centered AI, Wei Xu et.al, IEEE, November 2023: https://arxiv.org/pdf/2311.16027:
- What is Human-Centered AI (HCAI)? Interaction Design Foundation, October 2024: https://www.interaction-design.org/literature/topics/human-centered-ai.
- Human-Centered AI (HCAI) Carnegie Mellon University, 2021: https://insights.sei.cmu.edu/documents/610/2021_019_001_735364.pdf
- Human-Centered and Sustainable Artificial Intelligence in Industry 5.0: Challenges and Perspectives, Martini, B.; Bellisario, D.; Coletti, P., Sustainability 2024, 16, 5448: https://doi.org/10.3390/su16135448
- Human-centered recommender systems: Origins, advances, challenges, and opportunities, Joseph A. Konstan, Loren G. Terveen, Association for the Advancement of Artificial Intelligence, 42:31–42, 2021
- Applying HCAI in developing effective human-AI teaming: A perspective from human-AI joint cognitive systems, Wei Xu, Zaifeng Gao, November 2023: https://arxiv.org/pdf/2307.03913
- Human-centered AI development in practice—insights from a multidisciplinary approach, Friedrich, J., Brückner, A., Mayan, J. et al., Z. Arb. Wiss. 78, 359–376 (2024): https://doi.org/10.1007/s41449-024-00434-5

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by 82,000+ talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit **https://www.ltimindtree.com**.