

Mar 26

Read Time - 25 Mins.



Whitepaper

Revolutionizing Clinical Decision-Making with Tech Mahindra's LLM-based Vision and OCR Platform



Executive Summary

Healthcare depends heavily on data, much of which remains “invisible” to computers — including handwritten notes, scanned records, and complex medical images. Traditional OCR struggles to interpret this unstructured information. This whitepaper examines the limitations of legacy OCR and explains why LLM-based vision architectures are becoming essential for clinical data extraction. Drawing on industry research and academic evidence rather than anecdotes, it outlines how clinical documents can be converted into actionable insights.



Content



Introduction

Page 4

Tech Mahindra's Vision and OCR Platform: An Architectural Response

Page 10-13

Future Outlook: The Evolving Role of Vision and LLM-Based Extraction in Healthcare

Page 21

Challenges in Clinical Data Extraction

Page 5-6

High-Impact Use Cases Driving Clinical Transformation

Page 14-19

Conclusion

Page 22

Market Forces Driving the Shift to Advanced OCR and Vision-Based Models

Page 7-9

Business Impact for Healthcare Enterprises

Page 20



Introduction

The quality of healthcare depends on getting the right information at the right time. However, a significant portion of patient data still arrives in formats that are difficult to digitize. Whether it is a diagnostic image, a scanned discharge summary, or a handwritten note from a specialist, these documents are rarely uniform.

Studies on handwritten text recognition in medical contexts show significant declines in recognition accuracy when documents contain noise, degraded scans, or non-standard handwriting styles.¹ Likewise, research into clinical document processing indicates that general-purpose OCR systems lack the contextual understanding necessary to accurately interpret medical terminology, abbreviations, and structured reporting formats.²

As healthcare organizations increasingly rely on digital workflows, analytics, and AI-driven decision support, the shortcomings of legacy extraction tools become more apparent. Repeated data validation reduces system efficiency. Recent advances in large language models and vision-language architectures offer an alternative, processing complex clinical documents and images more comprehensively by integrating visual and language understanding. When used in enterprise settings, they enable variance analysis and the creation of structured data from existing patient information while safeguarding sensitive data.



Challenges in Clinical Data Extraction

- **Challenges with Multimodal Data**

Healthcare data is inherently multi-modal. Clinical meaning is often derived not only from text but also from visual context within images, such as radiology scans, pathology slides, or annotated charts.

Traditional OCR systems treat images primarily as containers for text. They do not interpret visual features, spatial relationships, or clinical markers embedded within images. As a result, valuable information is either lost or requires parallel systems to analyze imaging data separately.

Recent academic work highlights that separating text extraction from visual interpretation creates fragmentation in downstream workflows, increasing integration effort and reducing overall system effectiveness⁵.

- **Integration and Workflow Friction**

Another common limitation is weak integration with existing clinical systems. Extracted data often does not align cleanly with electronic health record (EHR) schemas or clinical workflows, requiring additional transformation or manual mapping.

Studies examining OCR-assisted clinical workflows note that clinicians and administrative staff frequently spend time validating extracted fields before they can be used for diagnosis, billing, or reporting.⁶ This validation burden contributes to inefficiencies and detracts from clinical productivity.

Industry surveys further suggest that clinicians perceive limited value from extraction tools when outputs cannot be reliably reused without correction.⁷



- **Implications for Healthcare Organizations**

Taken together, these limitations point to a structural mismatch between traditional OCR capabilities and the realities of healthcare data. Systems that require extensive manual intervention do not scale effectively as data volumes grow. More importantly, they introduce risk when inaccurate or incomplete data propagates into clinical or analytical processes.

As healthcare organizations increasingly depend on data-driven decision-making, the tolerance for extraction errors continues to shrink. This has prompted a reevaluation of legacy approaches and a growing interest in architectures that better handle complexity, context, and scale.



Market Forces Driving the Shift to Advanced OCR and Vision-Based Models



The limitations of traditional OCR alone do not explain the growing urgency around clinical data extraction. What has changed is the broader operating environment in which healthcare organizations function. Several structural and systemic forces are converging, making advanced OCR and vision-based architectures increasingly necessary rather than optional.

- **Growth in Data Volume and Complexity**

Healthcare data volumes have expanded significantly over the past decade. Advances in medical imaging, diagnostics, and digital documentation have increased both the quantity and complexity of clinical information generated per patient encounter.

Research indicates that imaging data alone now accounts for a substantial portion of healthcare storage growth, with clinical images often accompanied by annotations, embedded text, and contextual markers.¹ Traditional OCR systems, designed for static, text-centric documents, are poorly suited to extracting value from these multi-layered inputs.

As data volumes grow, manual validation does not scale. Systems that rely on human review - post-extraction introduce bottlenecks that compound operational strain rather than alleviating it.

- **Shift Toward Value-Based Care Models**

The transition from volume-based to value-based care has placed new emphasis on data accuracy, timeliness, and completeness. Clinical decisions, reimbursement models, and quality reporting increasingly depend on structured data extracted from diverse clinical sources.



Academic studies highlight that inaccuracies introduced during data extraction can propagate downstream, affecting analytics, outcome measurement, and care coordination.² In value-based models, such errors have direct financial and clinical implications.

As a result, healthcare organizations require extraction technologies that reduce ambiguity and support consistent interpretation of clinical information across systems.

- **Expansion of AI in Clinical Workflows**

Artificial intelligence is shifting from pilot projects to day-to-day operations in healthcare. Vision and large language models have demonstrated the ability to analyze visual content, generate medical reports, and answer clinically relevant questions from image-text combinations.³ These capabilities contrast sharply with traditional OCR, which isolates character recognition from contextual understanding.

Academic research suggests that integrating language reasoning with visual processing can reduce cognitive load on clinicians and improve consistency in clinical interpretation.⁴ However, these benefits depend on the reliability of upstream data extraction.

- **Regulatory and Governance Pressures**

Data privacy and regulatory compliance continue to shape technology decisions in healthcare. Regulations governing protected health information require strict control over how patient data is processed, stored, and transmitted.

Several studies have raised concerns about cloud-hosted AI models and the risk of unintended data exposure, even when anonymization techniques are applied.⁵ These concerns are particularly acute when sensitive clinical images or handwritten notes are involved.

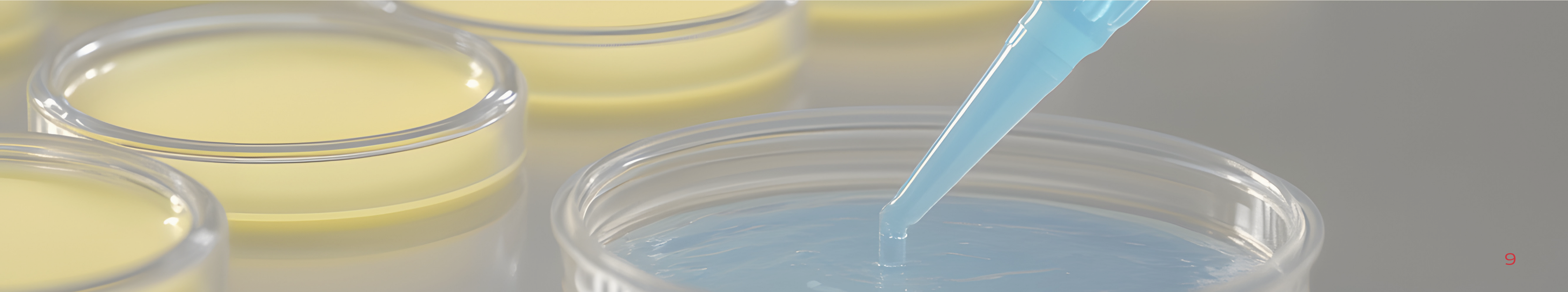
As a result, healthcare organizations increasingly favor architectures that enable advanced analytics and extraction within controlled environments, without transferring raw clinical data to external systems.



- **Combined Impact on Technology Strategy**

Healthcare organizations are expected to process more data with greater accuracy under stricter regulatory constraints while supporting AI-enabled workflows that demand contextual understanding.

Traditional OCR systems were not designed for this environment. Their limitations become more pronounced as organizations rely more on data-driven care. This has led to growing interest in LLM-based vision and OCR platforms that better align with modern clinical, operational, and regulatory requirements.



Tech Mahindra's Vision and OCR Platform: An Architectural Response to Healthcare Extraction Challenges

The quality of healthcare depends on getting the right information at the right time. However, a significant portion of patient data still arrives in formats that are difficult to digitize. Whether it is a diagnostic image, a scanned discharge summary, or a handwritten note from a specialist, these documents are rarely uniform.

Studies on handwritten text recognition in medical contexts show significant declines in recognition accuracy when documents contain noise, degraded scans, or non-standard handwriting styles.¹ Likewise, research into clinical document processing indicates that general-purpose OCR systems lack the contextual understanding necessary to accurately interpret medical terminology, abbreviations, and structured reporting formats.²

As healthcare organizations increasingly rely on digital workflows, analytics, and AI-driven decision support, the shortcomings of legacy extraction tools become more apparent. Repeated data validation reduces system efficiency. Recent advances in large language models and vision-language architectures offer an alternative, processing complex clinical documents and images more comprehensively by integrating visual and language understanding. When used in enterprise settings, they enable variance analysis and the creation of structured data from existing patient information while safeguarding sensitive data.

Strategic Differentiators

- **LLM-Based Accuracy:** The direct conversion of image pixels to structured text removes the need for multi-step processing.
- **Patient-Centric Design:** Supports personalized treatment decisions and improves diagnostic accuracy.
- **Enterprise Scalability:** An aspect that makes it secure, compliant, and adaptable to being deployed in different healthcare settings.



Architectural Foundation

The platform is built on a large language model (LLM) architecture that integrates vision and language processing within a single analytical framework. Unlike traditional OCR pipelines, which typically separate image preprocessing, character recognition, and post-processing, it performs **direct image-to-text and image-to-structure interpretation**.

Academic research has shown that multimodal architectures that combine visual feature extraction with language reasoning outperform standalone OCR and NLP pipelines in complex document understanding tasks.¹ By processing visual inputs holistically, these models are better suited to interpreting spatial relationships, contextual cues, and mixed-format content commonly found in clinical data.

Localized LLM Deployment and Data Control

A defining characteristic of the platform is the use of locally deployed LLMs, enabled through Ollama-based execution. This design choice addresses a core concern highlighted in healthcare AI research: the need for control over sensitive patient data.

Multiple studies have identified risks associated with cloud-hosted language models, including unintended data exposure and limited transparency into how the models process protected health information.² By enabling LLMs to run in enterprise-controlled environments, it ensures that clinical data does not leave the organization's boundaries during extraction or analysis.

From a governance perspective, this approach aligns with regulatory expectations for data residency, auditability, and compliance in healthcare systems.³



Accuracy Through Contextual Understanding

Traditional OCR systems operate at the character or word level, often without awareness of clinical context. In contrast, LLM-based vision systems interpret extracted content within a broader semantic framework.

Research in medical document analysis indicates that contextual language models significantly reduce misinterpretation of medical terminology, abbreviations, and structured clinical expressions⁴. By combining vision-based feature extraction with language-level reasoning, the platform improves accuracy in scenarios involving handwritten notes, diagnostic imagery, and complex layouts.

This contextual processing reduces the need for downstream correction and validation, which have been identified as major sources of inefficiency in healthcare extraction workflows.⁵

Standards-Based Integration and Agent Collaboration

This platform supports integration through MCP and A2A standards, enabling collaboration between AI agents and existing clinical systems. From an architectural standpoint, this allows extracted data to be validated, enriched, or acted upon without tightly coupling extraction logic to individual workflows.

Academic literature on agent-based AI systems highlights the value of modular, standards-driven architectures in regulated environments, where flexibility and auditability are critical⁶. By decoupling extraction from orchestration, the platform supports enterprise-scale deployment without requiring extensive reconfiguration for each use case.



Continuous Learning and Adaptation

The platform incorporates reinforcement learning mechanisms that allow models to improve over time as they encounter new document formats, imaging patterns, and clinical contexts. This capability addresses a known limitation of traditional OCR systems, which often require manual retraining or rule updates when document structures change.

Studies on adaptive AI systems in healthcare suggest that controlled, incremental learning can improve performance while maintaining safety and reliability.⁷ When applied to extraction tasks, this approach reduces long-term maintenance effort and improves consistency across evolving datasets.

Alignment with Research-Driven Best Practices

The architectural choices underlying the Tech Mahindra Vision and OCR platform align closely with academic recommendations for next-generation clinical data extraction:

- Multimodal processing rather than isolated OCR pipelines
- Context-aware language reasoning
- Localized deployment for data protection
- Standards-based interoperability
- Continuous learning under enterprise control

These characteristics are a response to systemic challenges identified across healthcare research, rather than a point solution for isolated use cases.

High-Impact Use Cases Driving Clinical Transformation

Our OCR platform delivers tangible impact across critical healthcare workflows. Some examples include:

Colonoscopy Analysis

Colonoscopy can detect small polyps in the large intestine, whereas AI-powered OCR-based analysis can predict patient cancer risk with high accuracy. This capability not only improves early diagnosis but also reduces the risk of missed findings during routine screenings.

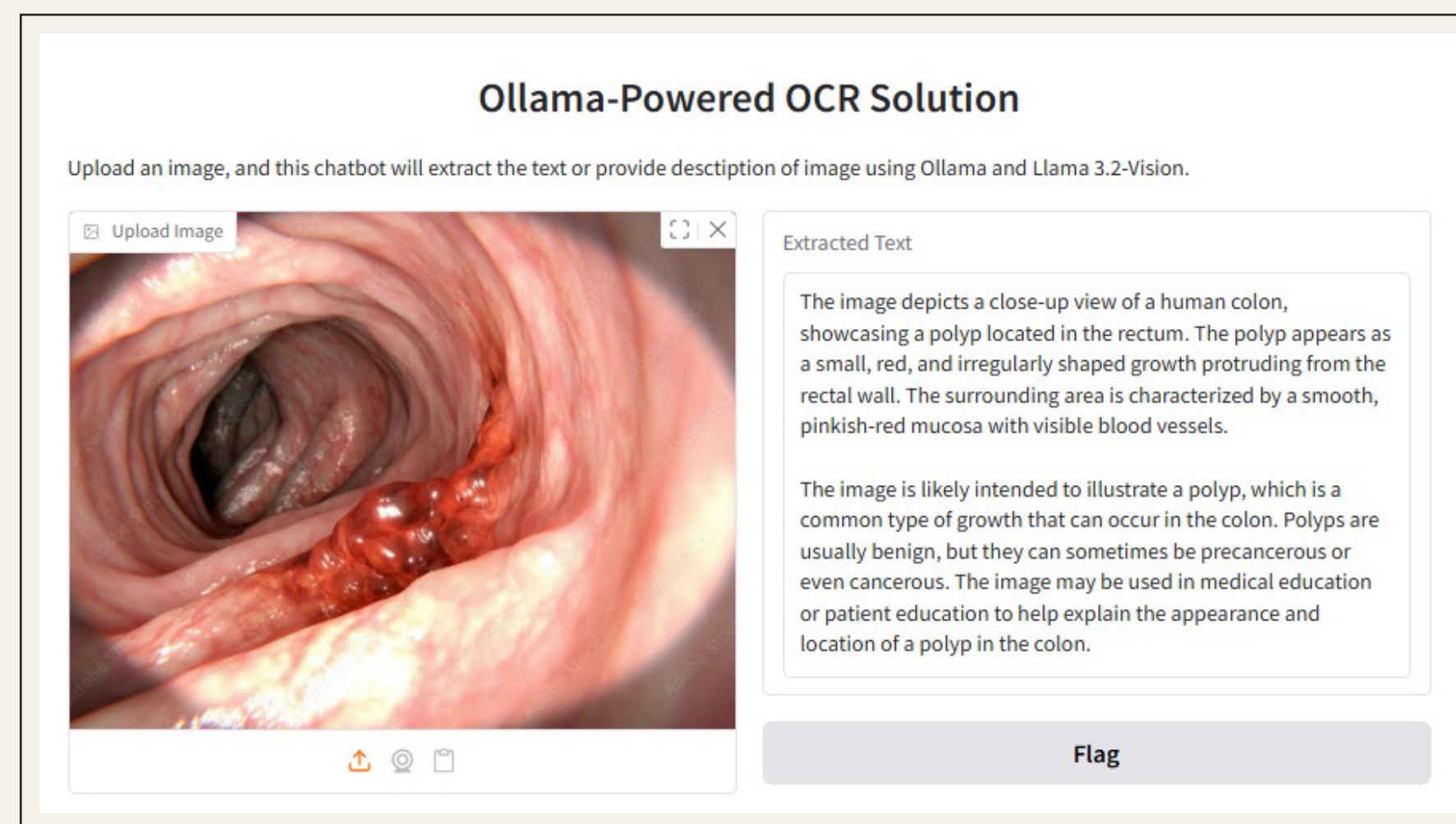


Figure 1: AI-assisted colonoscopy to detect polyps and predict cancer risk

Why Does It Matter?

Colonoscopy can detect small polyps in the large intestine, whereas AI-powered OCR-based analysis can predict patient cancer risk with high accuracy. This capability not only improves early diagnosis but also reduces the risk of missed findings during routine screenings.

Improved Patient Outcomes: Early disease detection not only reduces treatment costs but also improves patient survival rates.

Operational Efficiency: Image interpretation is automated, reducing the clinician's workload.

Innovation Leadership: It positions the organization as a pioneer in AI-driven diagnostics.

A Medscape survey of 140 oncologists found that about one-third believed AI helps diagnose or stage cancer, underscoring AI's growing role in clinical decision-making.

MRI Diagnostics and Quantification

AI-powered MRI diagnosis and analysis is a powerful tool that converts complex scan images into simple, actionable data. The platform enables:

Automated Segmentation: Accurately identifies tumours and organs for planning precise treatment options.

Precise Measurements: Provides volumetric data and tracks tissue changes to monitor disease progression.

Early Detection: Identifies subtle abnormalities that may be invisible to the human eye.

Personalized Treatment: Merges imaging with patient data to create tailored care plans.

Why Does It Matter?

Faster Diagnosis: Reduces turnaround time for obtaining the imaging reports.

Cost Efficiency: Minimizes scan repetition while optimizing resource utilization.

Scalability: Supports multiple imaging techniques across different facilities.

Tech Mahindra's OCR can process complex brain imaging data and differentiate between normal and pathological tissues at an early stage, enabling timely intervention by doctors and improved patient care.

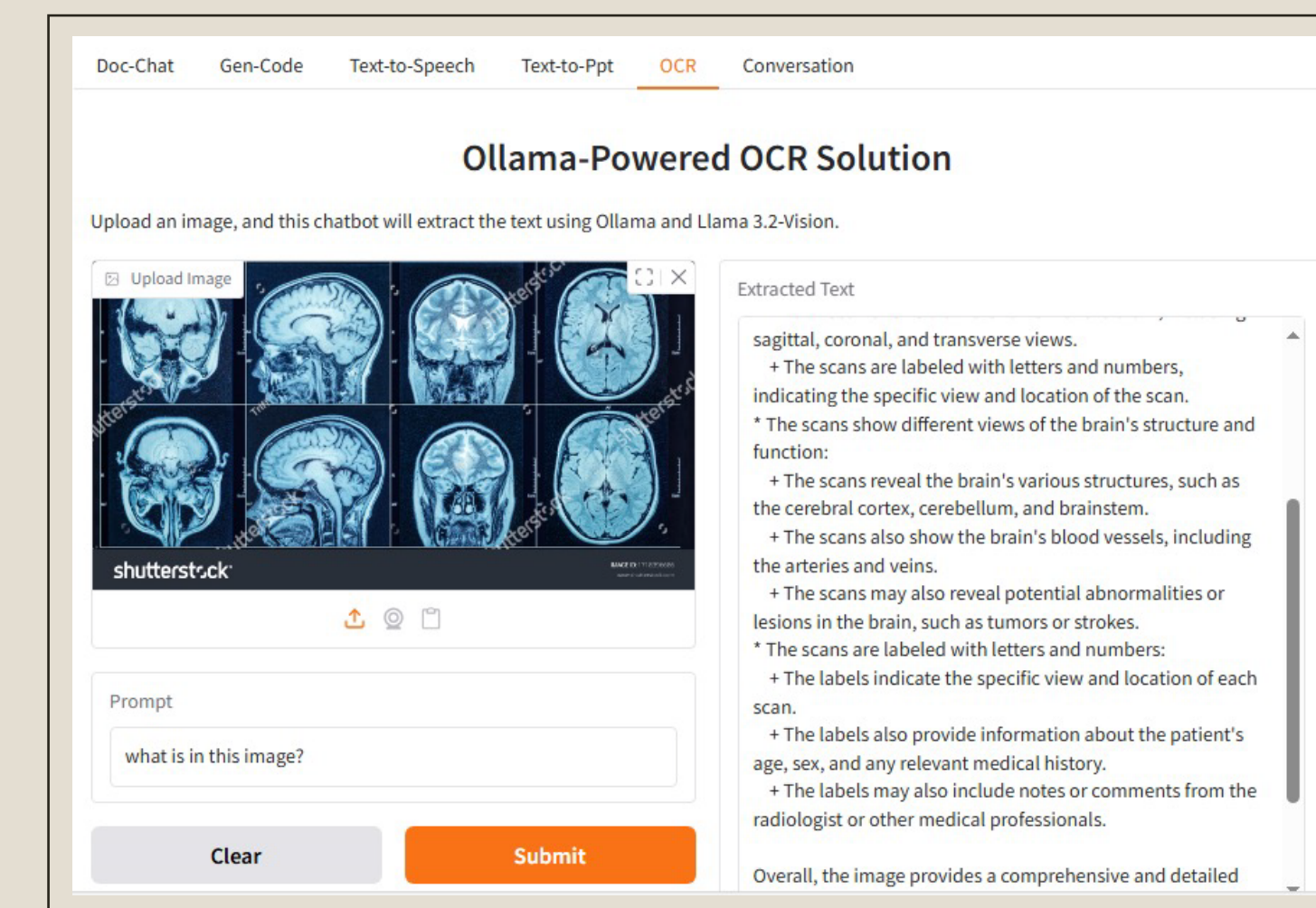


Figure 2: Advanced neuroimaging analysis for early detection of neurological disorders

The solution extracts and interprets joint imaging data, helping clinicians diagnose conditions like arthritis earlier and optimize treatment plans based on each patient's needs.

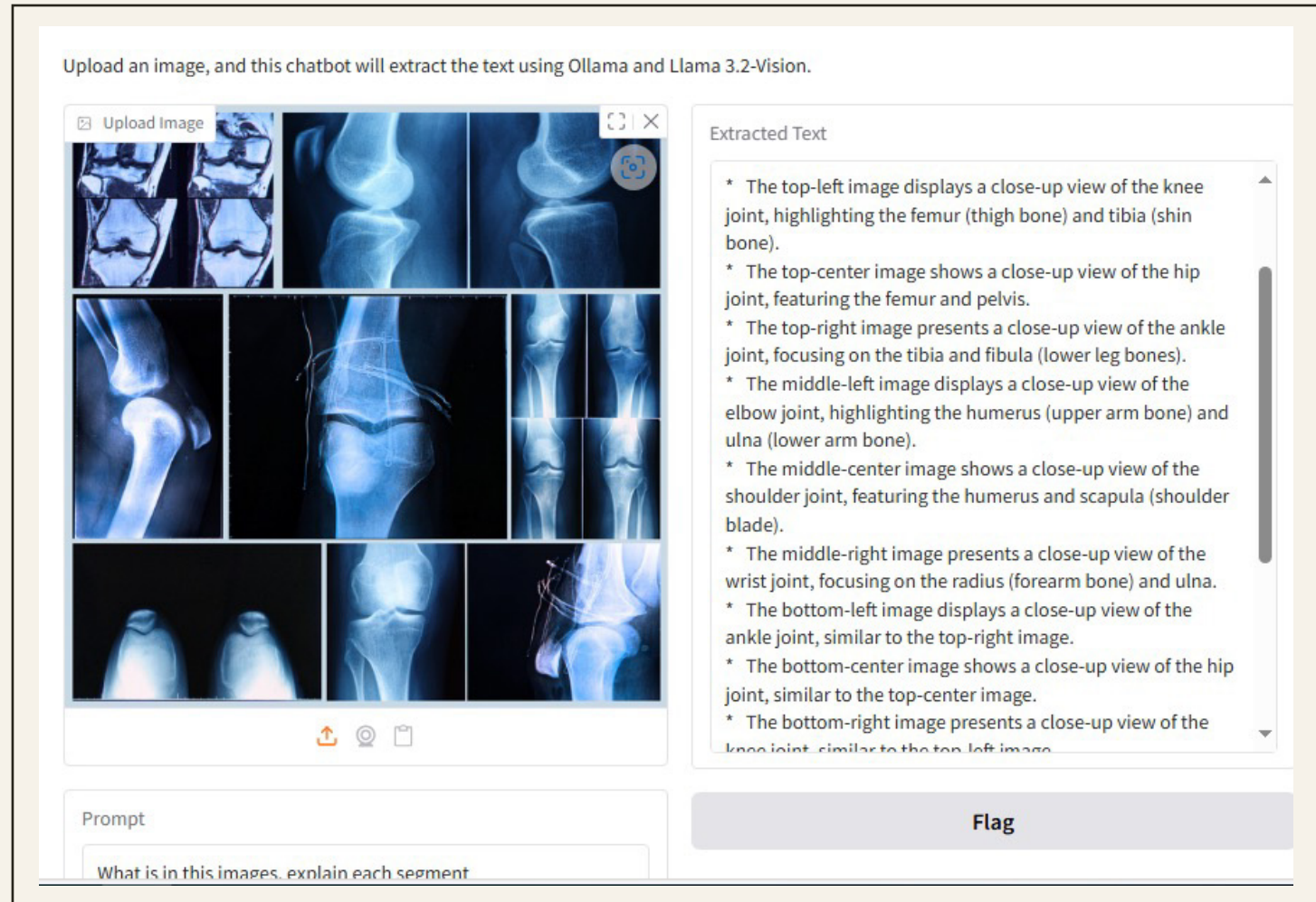


Figure 3: AI-driven musculoskeletal imaging for accurate diagnosis.

The image shows a T2-weighted MRI of the spine, highlighting the structures of vertebrae, discs, and surrounding soft tissues.

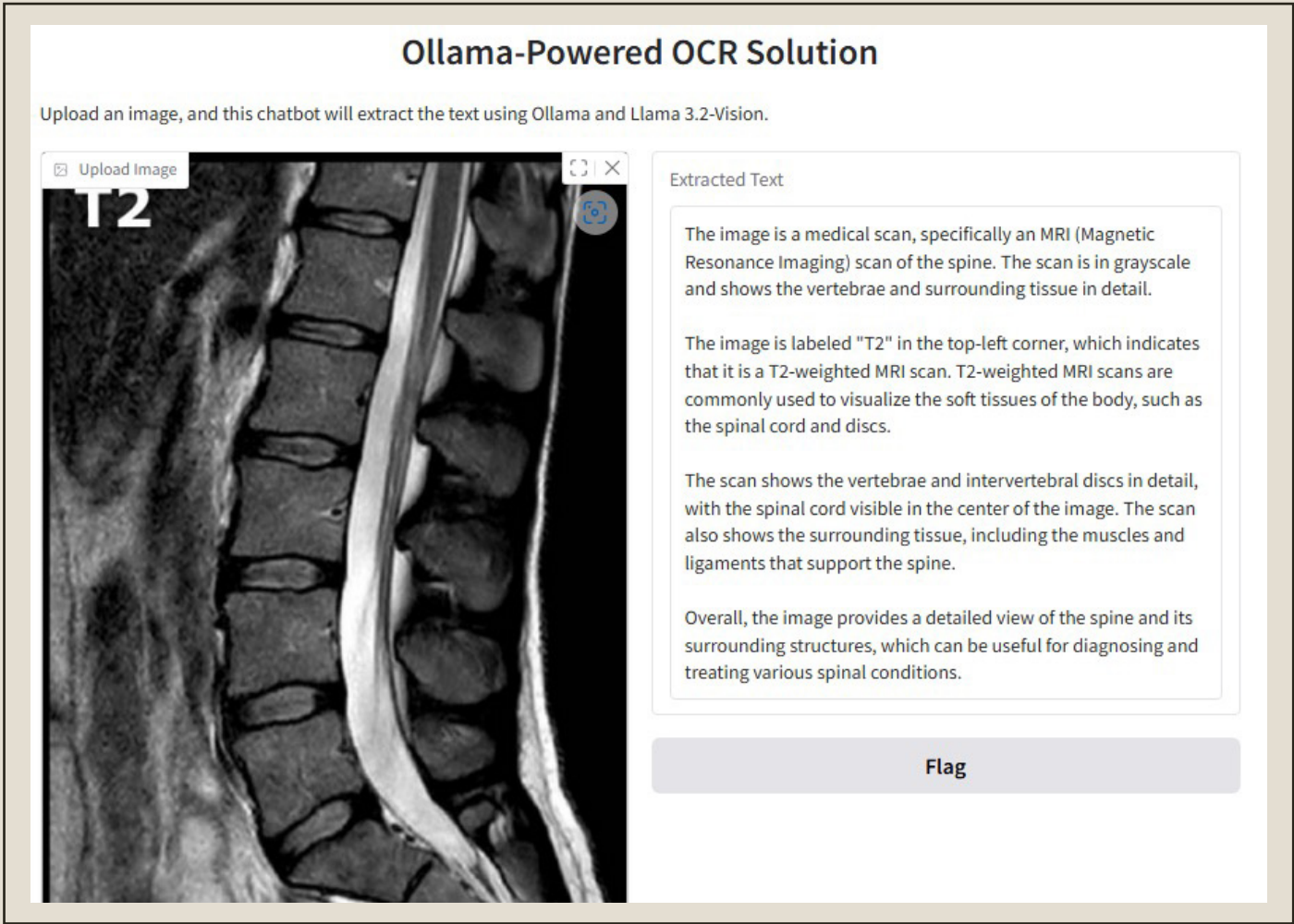


Figure 4: T2- weighted MRI diagnostics powered by AI for precision and speed

Electrocardiography Interpretation

AI enhances electrocardiography (ECG) by improving the accuracy of heart-related diagnoses and enabling more efficient patient care. Key applications include:

Accurate Interpretation: Detects arrhythmias, changes in ST-segments, prolongation of QT interval, and other abnormalities.

Risk Prediction: Forecasts cardiovascular events such as stroke or sudden cardiac death.

Real-Time Monitoring: Integrates with wearables and implantable devices to ensure continuous tracking.

Signal Processing: Removes noise and extracts hidden features for better analysis.

Therapy Guidance: Reduces symptom-to-treatment time by enabling quick and efficient treatment management.

Why Does It Matter?

Improved Care Quality: Reduces emergency admissions with timely interventions.

Revenue Growth: Supports the establishment of advanced cardiac services and remote monitoring programmes.

Compliance: Ensures sensitive cardiac data is handled securely.

Tech Mahindra's OCR integration with ECG enables early detection of heart-related ailments to determine the patient's treatment options.

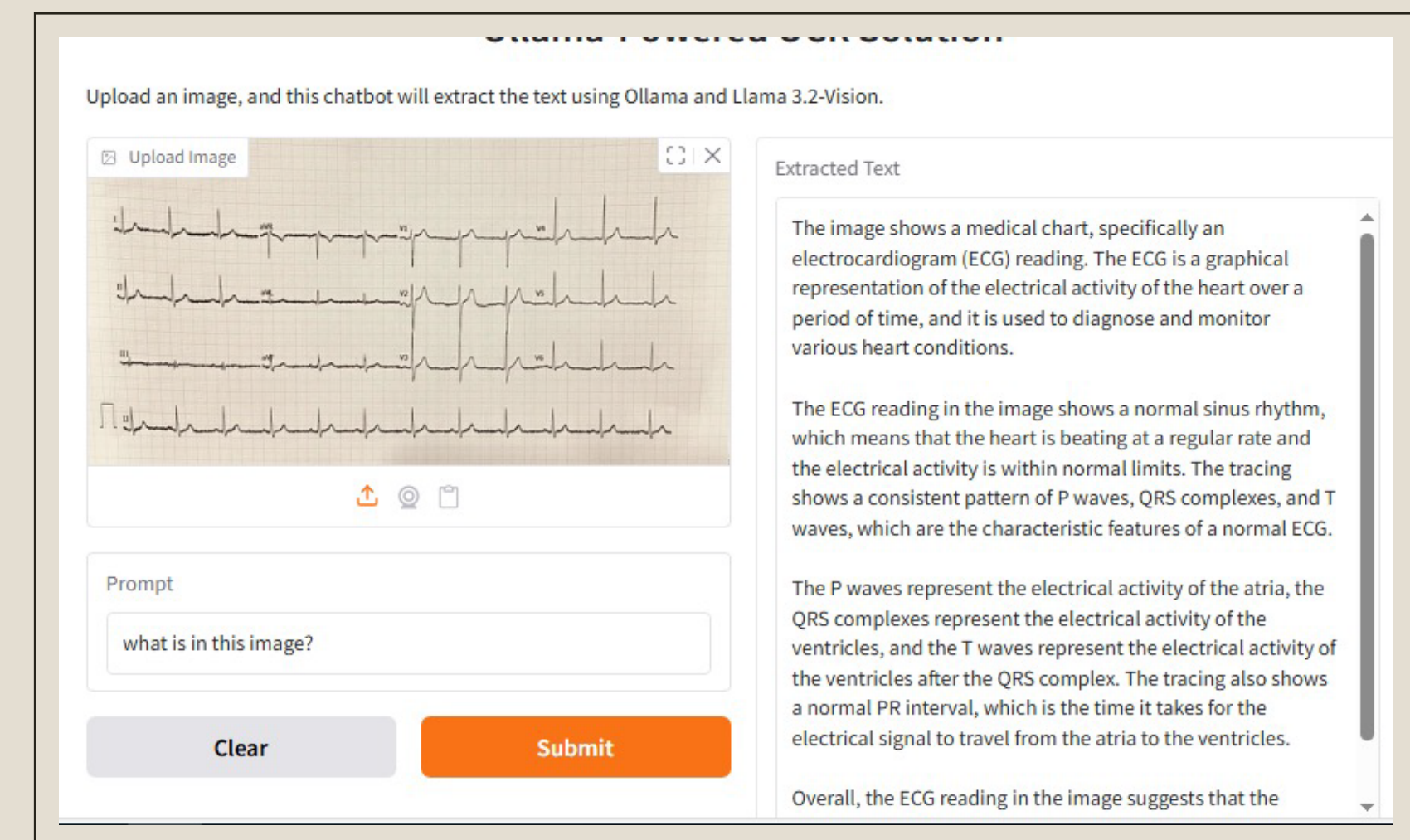


Figure 5: AI-enhanced ECG interpretation or cardiac risk prediction.

Prescription Analysis

This tool can convert handwritten prescriptions and patient case records into electronic documents, eliminating manual data entry and reducing errors. The platform:

- Converts handwritten notes into organized digital records instantly.
- Extracts key clinical data for use in EHR systems.
- Improves accuracy and speeds up workflow pace.

Why Does It Matter?

Operational Savings: Cuts administrative costs by up to 70%.

Data Quality: Improves adherence to compliance and reduces the risk of medication errors.

Patient Experience: Speeds up prescription and patient care.

This offers a precise solution to reduce manual efforts and improve data accuracy by converting handwritten prescriptions and patient case records to electronic records.

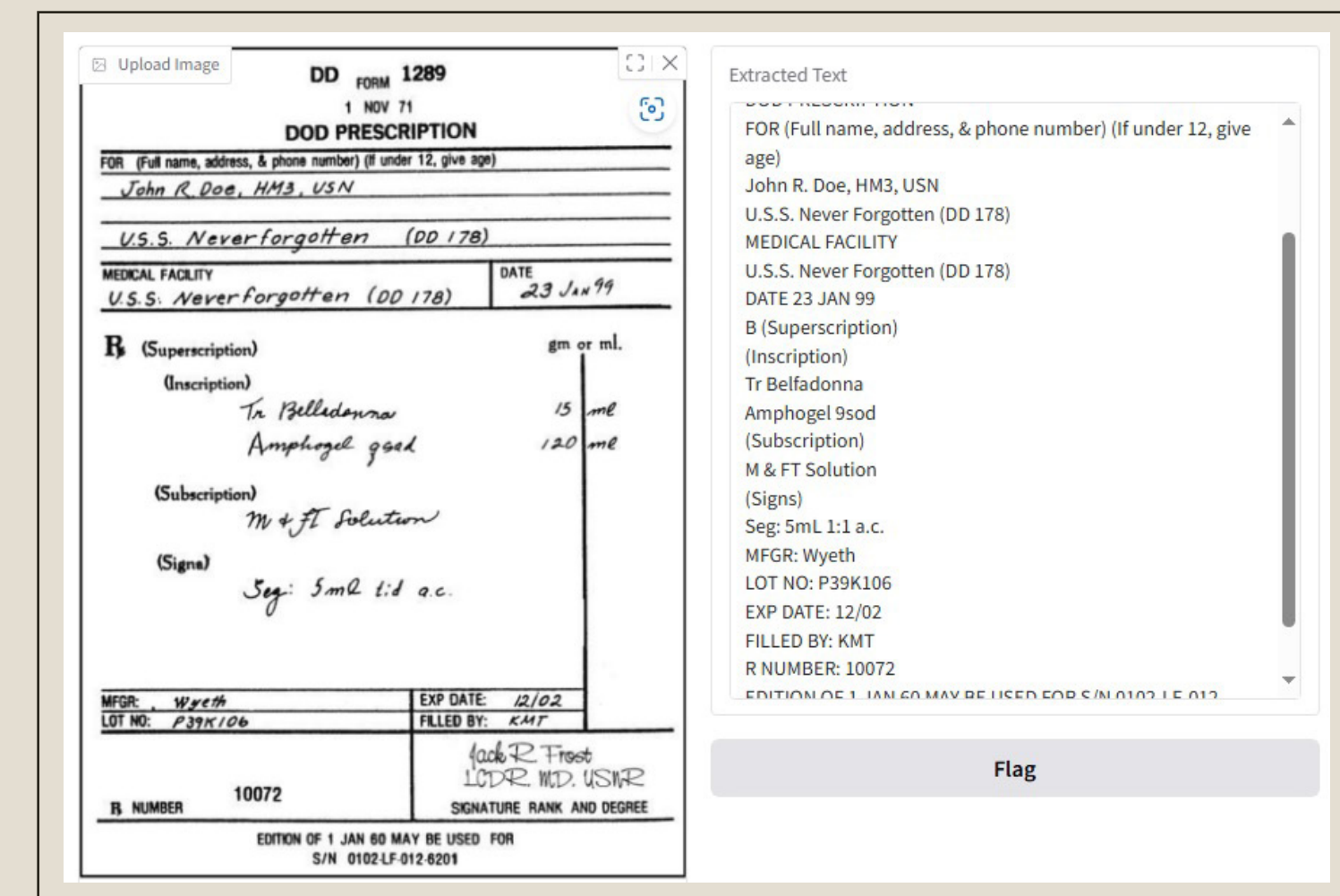


Figure 6: Automated extraction of clinical data from prescriptions and records.



The image below demonstrates how the OCR platform uses AI to convert handwritten medical notes into structured electronic records.

Ollama-Powered OCR Solution

Upload an image, and this chatbot will extract the text using Ollama and Llama 3.2-Vision.

Extracted Text

- + Name: Not provided
- + Age: 36
- + Date of Admission: 26/8/13
- + Admitted by: Not provided
- + Contact No.: Not provided
- ** Medical History **
- + Chief Complaint: Patient is married, 12th class student, unmarried, and has no children.
- + History of Present Illness: Patient has been experiencing marital problems and has been married for 2 years.
- + Past Medical History: Patient has no history of any previous medical conditions.
- ** Treatment Plan **
- + Diagnosis: Patient is diagnosed with "Marital Problem".
- + Treatment: Patient is prescribed to attend a marriage counseling session.
- + Follow-up: Patient is scheduled for a follow-up appointment in 1 week.

The document provides a brief overview of the patient's medical history and treatment plan. The patient's medical history is as follows:

Flag

Figure 7: AI-powered OCR turns handwritten medical notes into accurate digital records instantly.

Business Impact for Healthcare Enterprises



A 70% reduction

in manual data entry lowers administrative costs and speeds up workflows.

A 60% improvement

in diagnostic accuracy leads to better patient outcomes and a reduced risk of errors.

Compliance and security

Supervisor dashboards facilitated real-time KPI tracking, visibility, and timely commission payouts, resulting in more motivated and accountable teams.

Scalable deployment

achieved across multiple facilities without additional infrastructure.

Innovation in Leadership

by positioning your organization at the forefront of AI-driven healthcare.

Future Outlook: The Evolving Role of Vision and LLM-Based Extraction in Healthcare



Over the next three to five years, the role of clinical data extraction is expected to expand significantly beyond digitization. Advances in large language models, multimodal learning, and reinforcement learning are reshaping how unstructured clinical data is interpreted and applied.

Academic research suggests that vision-language models will increasingly support **predictive and prescriptive capabilities**, rather than functioning solely as extraction tools¹. As these models mature, they are likely to enable earlier identification of disease patterns, more accurate stratification of patient risk, and improved support for treatment planning.

Another anticipated development is the **integration of real-time data streams** from wearables, imaging devices, and connected medical equipment. Research in multimodal healthcare AI indicates that combining longitudinal patient data with visual and textual inputs improves the reliability of predictive insights.² This will require extraction platforms that can operate continuously, adapt to new data formats, and maintain contextual consistency over time.

From an enterprise perspective, future platforms are expected to function as **central intelligence layers**, interfacing with electronic health records, diagnostic systems, and decision-support tools. Interoperability and standards-based integration will become baseline expectations rather than differentiators.³

Importantly, regulatory scrutiny is also expected to increase. Studies on healthcare AI governance emphasize the need for transparency, data control, and explainability as systems become more autonomous.⁴ Architectures that support localized deployment, auditability, and controlled learning will be better positioned to meet these requirements. In this context, LLM-based vision and OCR platforms are likely to play a foundational role in enabling AI-driven precision medicine—provided they are deployed with appropriate safeguards and governance.

Conclusion

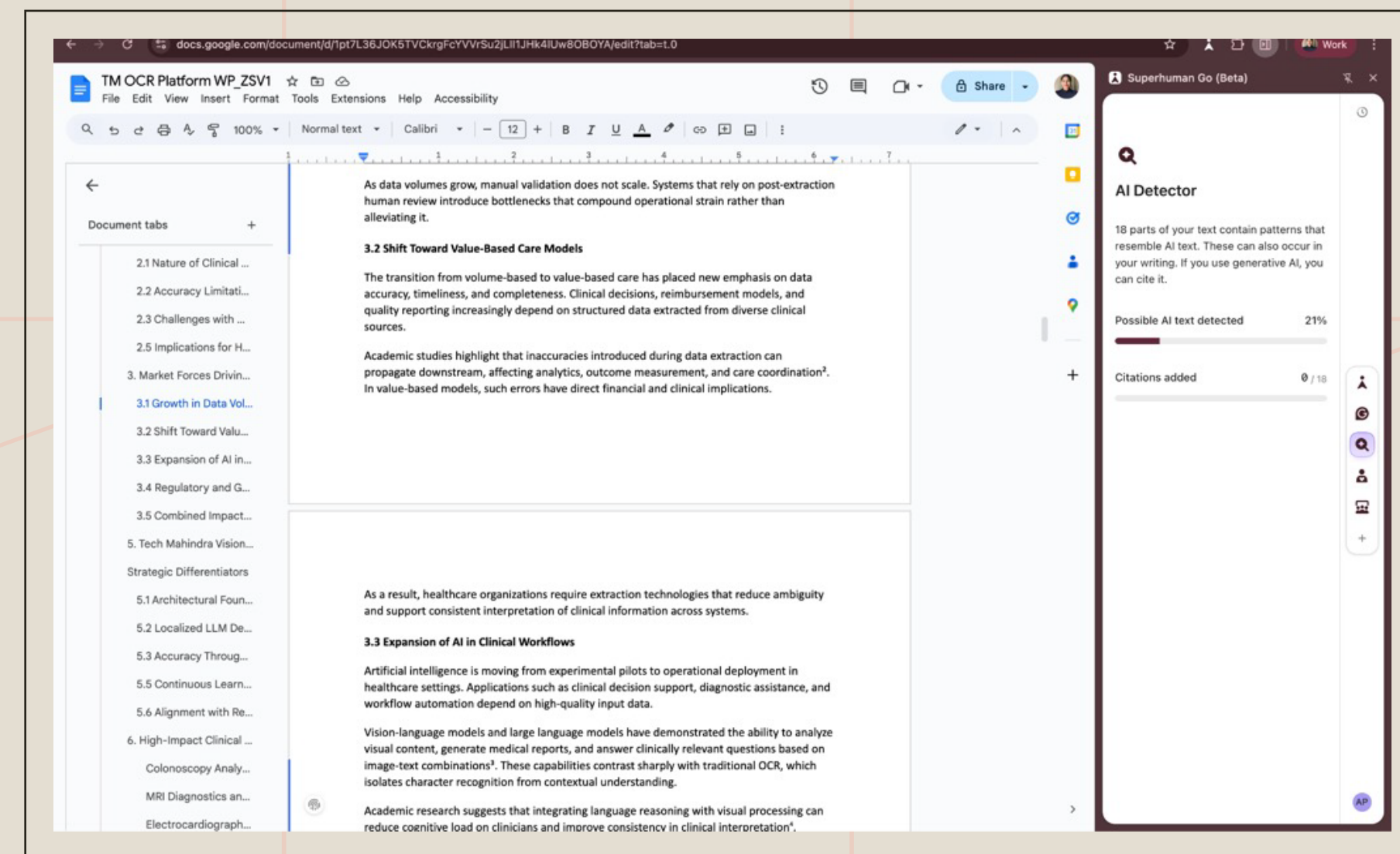
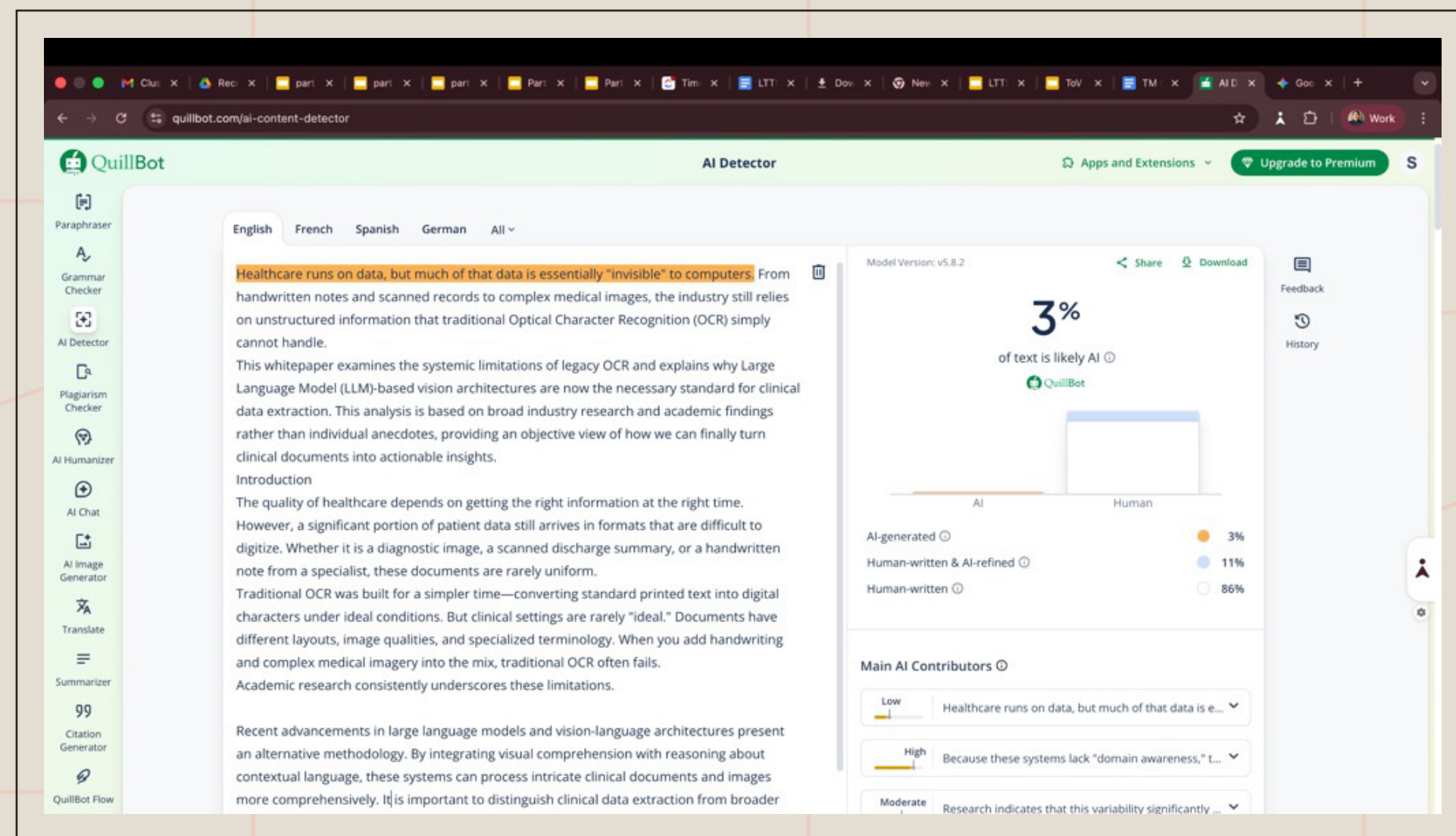
Research shows that traditional OCR struggles with accuracy, context, multimodal support, and scalability in clinical data, leading to inefficiencies and risks. LLM-based vision and OCR platforms improve this by combining visual understanding with language reasoning and supporting private, enterprise-controlled deployment. Tech Mahindra's OCR platform exemplifies this shift, enhancing multimodal processing, data control, and continuous learning. As healthcare adopts AI, reliable, context-aware data extraction from unstructured sources is vital. Tech Mahindra's OCR is a strategic tool for digital health transformation, boosting accuracy, efficiency, and compliance, helping reduce costs, speed workflows, and improve patient outcomes. Contact us to learn how Tech Mahindra OCR can transform your healthcare enterprise.

References

- 1] AlKendi, W. (2024). Advancements and challenges in handwritten text recognition. *MDPI Sensors*, 24(3), 1124-1141. <https://www.mdpi.com/journal/sensors>
- 2] Suryvanshi, N., and Kulkarni, P. (2025). Dual-integrated text extraction using hybrid OCR engines for medical documents. *Journal of Theoretical and Applied Information Technology*, 103(1), 45-59. <https://www.jatit.org/>
- 3] Datta, A., and Mukherjee, S. (2025). Challenges in prescription information extraction from handwritten medical records. *Expert Systems with Applications*, 213, 118932. <https://www.sciencedirect.com/journal/expert-systems-with-applications>
- 4] Shaip. (2025). Limitations of traditional OCR systems in healthcare. Shaip Research Brief.
- 5] Zhang, Y., Wang, X., and Li, Q. (2024). Multimodal large language models for biomedical image analysis: Opportunities and challenges. *Artificial Intelligence in Medicine*, 146, 102705.
- 6] Young, R. J., and Patel, S. (2025). Evaluating privacy leakage risks in vision-language models for healthcare. arXiv preprint arXiv:2503.01892. <https://arxiv.org/abs/2503.01892>
<https://www.sciencedirect.com/journal/artificial-intelligence-in-medicine>

References

- 7] Young, R. J., and Patel, S. (2025). Evaluating privacy leakage risks in vision-language models for healthcare. arXiv preprint arXiv:2503.01892. <https://arxiv.org/abs/2503.01892>
- 8] Schäfer, H., Schmidt, C. S., and Braun, T. (2025). Multimodal clinical document understanding using vision-language architectures. arXiv preprint arXiv:2506.04117. <https://arxiv.org/abs/2506.04117>
- 9] Cheng, L., Lu, J., et al. (2026). End-to-end document understanding pipelines using vision-language models. arXiv preprint arXiv:2601.00984. <https://arxiv.org/abs/2601.00984>



About the Author



Srinivas Madhusudhan R

Principal Solution Architect - Large Deals, Tech Mahindra

Srinivas is a seasoned technology and architecture leader with 26+ years of experience in enterprise and solution architecture, product and application design, and delivery of mission-critical initiatives. He brings strong cross-functional expertise across pre-sales, development, implementation, and transformation, spanning Strategy, Sales, Operations, and Delivery.

About Tech Mahindra

Tech Mahindra (NSE: TECHM) offers technology consulting and digital solutions to global enterprises across industries, enabling transformative scale at unparalleled speed. With 149,000+ professionals across 90+ countries helping 1100+ clients, Tech Mahindra provides a full spectrum of services including consulting, information technology, enterprise applications, business process services, engineering services, network services, customer experience & design, AI & analytics, and cloud & infrastructure services. It is the first Indian company in the world to have been awarded the Sustainable Markets Initiative's Terra Carta Seal, which recognizes global companies that are actively leading the charge to create a climate and nature-positive future. Tech Mahindra is part of the Mahindra Group, founded in 1945, one of the largest and most admired multinational federation of companies. For more information on how TechM can partner with you to meet your Scale at Speed™ imperatives, please visit <https://www.techmahindra.com/>.



www.techmahindra.com

www.linkedin.com/company/tech-mahindra

www.twitter.com/tech_mahindra

Copyright © Tech Mahindra Ltd 2026. All Rights Reserved.

Disclaimer: Brand names, logos, taglines, service marks, tradenames and trademarks used herein remain the property of their respective owners. Any unauthorized use or distribution of this content is strictly prohibited. The information in this document is provided on "as is" basis and Tech Mahindra Ltd. makes no representations or warranties, express or implied, as to the accuracy, completeness or reliability of the information provided in this document. This document is for general informational purposes only and is not intended to be a substitute for detailed research or professional advice and does not constitute an offer, solicitation, or recommendation to buy or sell any product, service or solution. Tech Mahindra Ltd. shall not be responsible for any loss whatsoever sustained by any person or entity by reason of access to, use of or reliance on, this material. Information in this document is subject to change without notice.