Grenze International Journal of Engineering and Technology, January Issue

Versatile AI Scanner: A Conceptual Framework for Adaptive Information Retrieval and Projection using Edge AI

Jangam Poojith[1], Goura Kavita[2], Garalapati Narayana[3] and Maheshwaram Pranay[4]
[1-4]Department of Artificial Intelligence and Machine Learning, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Email: poojith1903@gmail.com, gkavita.cse@gmail.com, narayanag.1973@gmail.com, maheshwarampranay@gmail.com

***Abstract*— Versatile AI-SMART (Adaptive Information Retrieval and Projection) is a system that changes the information access based on the combined computer vision, NLP, and AR methods in a handheld device. The system uses transformer-based OCR to extract text with higher precision, GPT models to extract text according to context to create a summary, and generative AI to create visualization on the spot (flowcharts, mind maps, and diagrams) to facilitate better understanding of intricate information. Edge AI processing is capable of running in real-time and does not require a cloud, which makes it efficient and avoids data leakage. A micro projector is used to provide adaptive content projection to any surface that offers interactive learning experiences. Such an integrated system enhances the educational practices, simplifies the research process, and increases access to information by making the process of knowledge acquisition quicker, smarter, and more interactive. A system architecture and processing workflow have been depicted in Figures 2-4.**

***Keywords*— Augmented Reality, Transformer Models, optical Character Recognition, Edge Computing, Information Retrieval, General AI, Micro-Projection Systems, Adaptive Learning, Portable Computing Devices.**

I. INTRODUCTION

Advances in AI, computer vision, and NLP have transformed information retrieval, summarization, and visualization, minimizing human effort in processing printed and electronic content. The Versatile AI Scanner caters to the requirement for real-time, efficient information processing through the combination of Transformer-based OCR, GPT-driven summarization, and AR projection into a handheld device. The framework utilizes Transformer-based OCR for text extraction with high accuracy, superior to CNN and RNN models, especially in low-resolution and degraded text environments [1]. Extracted material is fed into a GPT-based NLP model, producing concise, context-sensitive summaries to facilitate accessibility and less cognitive overload [2]. For enhancing knowledge retention and interest, Generative AI transcribes text to flowcharts, mind maps, and video content, and learning and research become more intuitive [3].

In contrast to cloud-based solutions, the hardware utilizes Edge AI processing for real-time offline operation such that low latency, privacy, and high inference accuracy are achieved [4]. A embedded micro-projector supports projection based AR interactions and enables users to interact with augmented digital content in everyday environments [5].

Grenze ID: 01.GIJET.12.1.199

By removing the requirement for multiple tools, the Versatile AI Scanner simplifies information scanning, summarizing, and visualizing, for the benefit of education, research, and assistive technology. Future AI hardware and software optimizations will make it even more effective, ensuring it becomes an essential tool for adaptive information retrieval and projection.

II. Literature Survey and Related works

Since Transformer-based architectures have evolved, the processes of text extraction, summarization, and visualization have also developed. Conventional OCR systems using CRNN and LSTM architectures will run into trouble with handwritten, skewed, or degraded texts. In contrast, Transformer-based OCR models, including TrOCR, Donut, PARSeq, and Swin Transformer V2, perform much better than before due to self-attention mechanisms [1][6][7] [8] [9]. However, these are resource-heavy and need domain-specific fine-tuning, thus presenting a challenge when deploying these on constrained devices.

For summarization, GPT-based models like GPT-3, GPT-4, DistilGPT2, and Transformer-based models such as Pegasus, T5, and BERT achieved state-ofthe-art results in the cases of both abstractive and extractive summarization [2] [10] [11] [12]. These models manage to preserve the context and produce coherent summaries, yet heavy in computational requirements, thereby curbing their integration into any real-time application at the edge.

With the new Edge AI domain, the latest lightweight models have been developed for inference at low delays, say, MobileNetV4 [13] and MobiHisNet [14]. Frameworks such as XoT [15], Space4AI-D [16], and TANGO [17] also consider multi-DNN inference on heterogeneous platforms. They are all solutions for speeding up processing, but none supports single-modality tasks or does be able to place NLP, OCR, and AR all within one system.

Generative AI has made room for automated visualization with flowchart generation [18] and mind maps [19] and self-supervised vision learning-[20] (DINOv2). PhyCV [21] and augmented data methods [8] keep pushing the domain specific applications afar from their limits. On the other hand, the models, constrained by their domain, rarely address real-time multimodal information gathering. Augmented Reality has been researched as a tool for interactive education and visualization [5] [22] [23]. Projection-based augmented reality systems improve engagement but depend on the cloud for operation and lack integration with OCR and summarization engines, which prevents them from being used for fully adaptive learning and assistive technologies. In short, prior work has contributed more accurate OCR, better summarization, efficient edge inference, and AR-based interaction. Yet these works typically deal with individual components in isolation. Our Versatile AI Scanner differs from these works in that it attempts to bridge these domains into a single multi modal and edge-optimized pipeline. By combining Transformer-based OCR, GPT-based summarization, generative AI visualization, and AR projection, our framework overcomes the deficiencies of the prior work and constructs a full-fledged privacy-preserving and real-time adaptation system for education, research, and assistive applications.

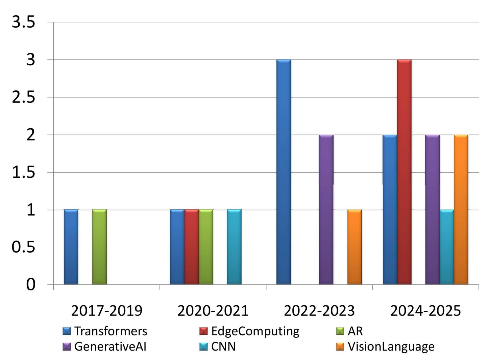

Fig.1. Technology distribution in research papers (2017–2025)

III. Survey Methodology

A. System Architecture and Hardware Integration

The AI scanner app requires a microservices architecture in which individual AI components communicate through precisely defined APIs for easy upgrade of components without system disruption. Real-time processing dictates the requirement for efficient data flow management using parallel processing, memory mapping, and rapid inter-process communication under the event-driven model. Fallback mechanisms achieve graceful degradation for the suboptimal conditions. Hardware integration is built around edge compute boards (Google Coral and NVIDIA Jetson Xavier NX) balancing power efficiency and AI acceleration by virtue of onboard dedicated neural processing units. High-resolution sensors that have excellent low-light capabilities and fast auto-focus for varied environments are needed for the camera system. A micro-laser projector (minimum 100 lumens and 720p) and depth sensors provide automatic keystone correction and surface mapping for AR projecting. Power management utilizes dynamic frequency scaling, judicious component activation, and smart battery utilization while thermal design provides sustained performance by virtue of efficient heat dissipation.

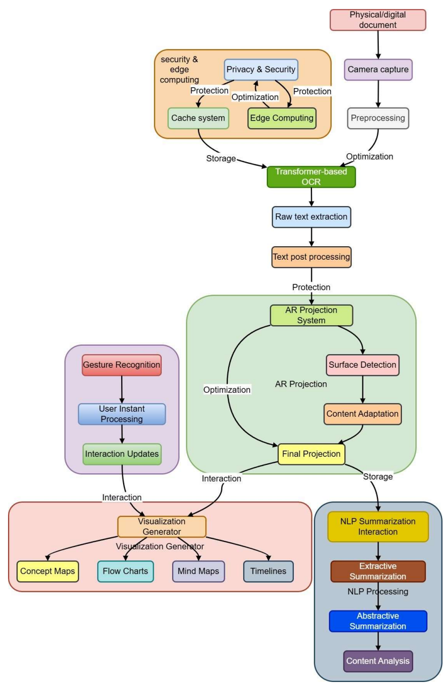

Fig.2. AI-driven document processing and AR projection framework. The pipeline integrates OCR, NLP summarization, generative visualization, and AR projection optimized for edge devices.

B. AI Processing Pipeline: OCR and Summarization

The AI pipeline integrates transformer-based OCR and NLP summarization for intelligent document processing. Training of OCR begins from base models (TrOCR, Donut, PARSeq) that are fine-tuned over domain datasets for research, legal, or technical documents. Pre-processing pipeline involves adaptive image enhancement—correction of perspective distortions, brightness/contrast adjustment, and binarization for text separation from background. Multi-step iteration processing corrects hard texts, and post-processing includes contextsensitive spell check, grammar correction, and layout analysis. Multilingualsupport is also ensured by language identifier algorithms and confidence scores guide downstream processing decisions.

The summarization engine employs edge-optimized models (DistilGPT, GPTJ) with quantization and pruning to minimize computational cost. Inference optimization uses ONNX Runtime or TensorRT. A hybrid approach combines extractive methods (identifying key sentences) with abstractive techniques (paraphrasing into coherent summaries). For long documents, hierarchical summarization processes logical sections independently before integration. RetrievalAugmented Generation (RAG) enhances factual accuracy through a local knowledge base, while user-controllable parameters enable customization of summary length, technical depth, and focus areas. Figure 2 illustrates the complete architecture integrating OCR, summarization, and AR projection for real-time processing.

C. Generative Visualization and Edge AI Optimization

The visualization system is in the form of summarized content that is translated into domain specific visual representations such as flowcharts of procedures, mind map of concepts and hierarchical data with structured diagrams. Fine-tuned lightweight generative models with edited prompting templates give robust, quality-assured, and consistent outputs due to reusable style libraries that can be configured to output a variety of content depending on the content or user preferences. The vector rendering allows one to scale to variable projections and resolution. The smart layout algorithms are also able to optimise the utilisation of space and still ensure readability, changing dynamically with projection surfaces, viewing angles, text size, and contrast. Gesture-based exploration, expansion and drilling of details are supported through interactive features. The optimization of edge AI allows real-time mobile computing using models compression methods: quantization-aware training (INT8/INT4 precision), pruning, and knowledge distillation. Inference engines (TensorRT, OpenVINO, ONNX Runtime) are hardware-accelerated to provide the highest computational efficiency. Parallelism in pipes allows parallel multi-module processing OCR on one frame as summarization operates on previous data. The memory management utilizes weight sharing, intermediate feature reuse and dynamic loading of models. Adaptive computation scales down model sizes according to the difficulty of inputs, battery constraints and latency needs, using lightweight models where the task is not challenging and heavyweight models when the task is challenging. Figure 1 above (Ai-Pipeline) shows the streamlined processing flow where the modules run in parallel with the lowest possible latency.

D. Augmented Reality Projection and User Interface

The AR projection system is a blend of digital contents with real life physical surroundings by using the real-time surface detection algorithms that detect the projection regions, scan the surface properties (texture, color, reflectivity), and change the content in response. The system is dynamic and updates on the movement of devices and surfaces. Projection optimization uses automatic keystone correction, focus control, and brightness, depending on the ambient light and the properties of the surfaces. Distortion correction normalizes the unshaped or curved surfaces to even visualization and adaptive typography algorithms change font, font weight, and font contrast based on the projection distance and the characteristics of the surface. User interaction makes use of natural motion control by using the intuitive control interfaces on the basis of gesture recognition, as well as optional eye-tracking on the basis of focus-based interaction. The interface merges advanced features with a portability limit with context-sensitive filtering to automatically adjust modes on sensor data, the position of the device, and a previous analysis of user behavior. Progressive disclosure is a strategy that unveils simple functionality then an advanced functionality thereafter. Hybrid control is a complex of touchless control (voice commands, gestures, haptic feedback) and minimum manual configuration, in response to personal patterns of use. The visual feedback visualizes system status, OCR confidence, summary visualization and custom visualization. The accessibility features are based on high-contrast mode, and customizable interaction pattern to suit various user requirements and comfort standards.

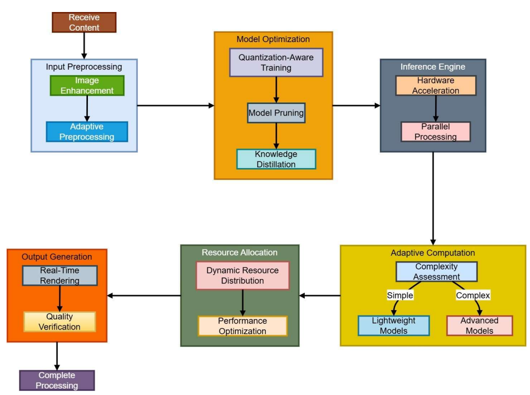

Fig.3. Optimized AI processing pipeline for efficient computation

IV. DISCUSSION

The Versatile AI Scanner combines Transformer-based OCR, lightweight GPT summarization, generative visualization, and AR projection into one device that is optimized for edge use. Unlike current state-of-the-art solutions that separate these components, our framework provides a unified, multimodal pipeline. This pipeline can perform real-time functions while also protecting user privacy. In comparison with current AR-based systems [5], [23] that solely focus on the visualization aspect, our method stands apart by additionally providing knowledge extraction and summarization. Furthermore, although edge models like MobileNetV4 [13] and MobiHisNet [14] lead to better inference efficiency, they are still single-task models. The novelty of our research is the integration of vision, language, and generative AI into one portable device. There are still some limitations such as hardware constraints for carrying out large Transformer models and the requirement for better gesture recognition and AR interaction. The next work will revolve around the optimization techniques and the domainspecific adaptations that aim to make the system more efficient and more applicable in various areas.

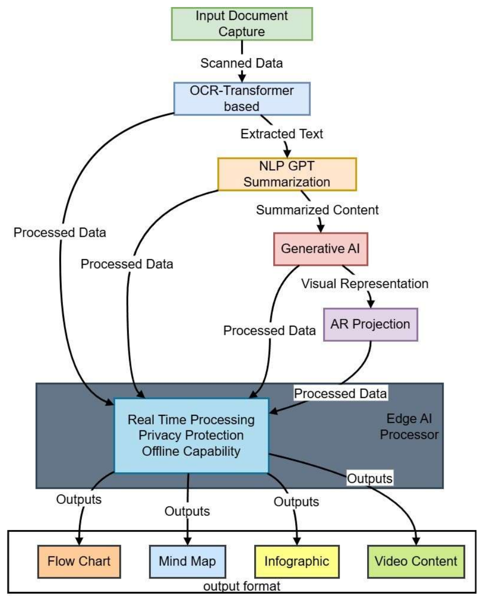

Fig.4. AI-powered document processing and visualization flow

V. Conclusion

The Versatile AI Scanner combines transformer-based OCR, GPT-based summarization and edge AI processing in a privacy-preserving miniature size to solve important information processing tasks. The accuracy of the text extraction, context-sensitive summarization and local computation are high-accuracy and allow data security to be maintained without reducing performance. Generative AI visualization and AR projection turn passive reading into a dynamic, interactive learning experience that improves the accessibility of information and knowledge comprehension in all areas of education, research, and professional practices. The areas that should be focused on the next generation development would be hardware miniaturization, optimization of domain-specific models, advanced gesture recognition, and collaboration. The Versatile AI Scanner is a major breakthrough in the natural and efficient process of information that is not only designed to address the privacy and performance needs of the targeted users but also combines the best AI with the principles of design engineering.

The system is a fundamental change in the way the users interact with text and visual information on various applications. The entire architecture, processing workflow and end-to-end adaptive document processing pipeline are shown in Figure 1, Figure 2 and Figure 3.

References

[1] D. Ganguly, A. Trivedi, B. Kumar, T. Patnaik, and S. Chaudhury, “End-to-endtransformer-based architecture for text recognition from document images,” in Proceedings of the Satellite Workshops of ICVGIP 2021. Springer, 2022, pp. 135–146.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser,and I. Polosukhin, “Attention is all you need,” Advances in neural information processing systems, vol. 30, 2017.

[3] D. G. Takale, P. N. Mahalle, and B. Sule, “Generative ai models: A comparativeanalysis,” Journal of Computer Science Engineering and Software Testing, vol. 10, no. 1, pp. 32–38, 2024.

[4] A. Sharshar, L. U. Khan, W. Ullah, and M. Guizani, “Vision-language models foredge networks: A comprehensive survey,” arXiv preprint arXiv:2502.07855, 2025.

[5] S. Prasad, S. K. Peddoju, and D. Ghosh, “Mobile augmented reality based interactive teaching & learning system with low computation approach,” in 2013 IEEE Symposium on Computational Intelligence in Control and Automation (CICA). IEEE, 2013, pp. 97–103.

[6] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Donget al., “Swin transformer v2: Scaling up capacity and resolution,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 12009–12019.

[7] Y. Wang, R. Huang, S. Song, Z. Huang, and G. Huang, “Not all images are worth16x16 words: Dynamic transformers for efficient image recognition,” Advances in neural information processing systems, vol. 34, pp. 11960–11973, 2021.

[8] W. Khallouli, M. S. Uddin, A. Sousa-Poza, J. Li, and S. Kovacic, “Leveragingtransformer-based ocr model with generative data augmentation for engineering document recognition,” Electronics, vol. 14, no. 1, p. 5, 2024.

[9] H. Zhou, L. Shao, and H. Zhang, “Srrnet: A transformer structure with adaptive 2dspatial attention mechanism for cell phone-captured shopping receipt recognition,” IEEE Transactions on Consumer Electronics, 2022.

[10] S. Pokale, K. Taware, G. Fernandes, S. Kangane, P. Bhosale, and L. Bewoor, “Textsummarization: Gpt perspective,” in 2023 3rd Asian Conference on Innovation in Technology (ASIANCON). IEEE, 2023, pp. 1–7.

[11] D. Siva Krishna, P. Srikanth, R. Srinivas Rao, and R. HOLLA M, “Data2summ: Review on enhancing summarization with transformers and attention models for speech and text compression,” Physica Scripta, 2025.

[12] H. Raza and W. Shahzad, “End to end urdu abstractive text summarization withdataset and improvement in evaluation metric,” IEEE Access, 2024.

[13] D. Qin, C. Leichner, M. Delakis, M. Fornoni, S. Luo, F. Yang, W. Wang, C. Banbury, C. Ye, B. Akin et al., “Mobilenetv4: universal models for the mobile ecosystem,” in European Conference on Computer Vision. Springer, 2024, pp. 78–96.

[14] A. Kumar, A. Sharma, V. Bharti, A. K. Singh, S. K. Singh, and S. Saxena, “Mobihisnet: A lightweight cnn in mobile edge computing for histopathological image classification,” IEEE Internet of Things Journal, vol. 8, no. 24, pp. 17778–17789, 2021.

[15] R. Nurtantyana, W.-Y. Hwang, and U. Hariyanti, “Xot (xducation of things): Harnessing ai and edge computing to educate all things,” IEEE Access, 2024.

[16] H. Sedghani, F. Filippini, and D. Ardagna, “Space4ai-d: A design-time tool forai applications resource selection in computing continua,” IEEE Transactions on Services Computing, 2024.

[17] Z. Taufique, A. Vyas, A. Miele, P. Liljeberg, and A. Kanduri, “Tango: Low latency multi-dnn inference on heterogeneous edge platforms,” in 2024 IEEE 42nd International Conference on Computer Design (ICCD). IEEE, 2024, pp. 300–307.

[18] M. Mhatre, A. Pandey, H. Rane, and S. Sahu, “A novel approach for creatingflowcharts using generative ai,” in 2024 Asia Pacific Conference on Innovation in Technology (APCIT). IEEE, 2024, pp. 1–7.

[19] L. Gao, J. Lu, Z. Shao, Z. Lin, S. Yue, C. Ieong, Y. Sun, R. J. Zauner, Z. Wei,and S. Chen, “Fine-tuned large language model for visualization system: A study on self-regulated learning in education,” IEEE Transactions on Visualization and Computer Graphics, 2024.

[20] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby et al., “Dinov2: Learning robust visual features without supervision,” arXiv preprint arXiv:2304.07193, 2023.

[21] Y. Zhou, C. MacPhee, M. Suthar, and B. Jalali, “Phycv: the first physics-inspiredcomputer vision library,” arXiv preprint arXiv:2301.12531, 2023.

[22] X. Han, J. Gao, C. Yang, Y. Yuan, and Q. Wang, “Focus entirety and perceive environment for arbitrary-shaped text detection,” IEEE Transactions on Multimedia, 2024.

[23] P. Reipschlager, T. Flemisch, and R. Dachselt, “Personal augmented reality forinformation visualization on large interactive displays,” IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 1182–1192, 2020.