# International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 17 No. 1 (2021)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 17, No. 1, 2021

Peer Reviewed, Referred & Indexed Journal

**Research Paper**

# A SECURE END-TO-END AI FRAMEWORK FOR OCR-DRIVEN TEXT RECOGNITION AND SEMANTIC INTELLIGENCE IN ENTERPRISE SYSTEMS

**Gowtham Reddy Kunduru**
*Lead software Engineer,M&T Bank, Buffalo, New York, USA*
*e-mail - gowtham.kunduru@gmail.com*

**Abstract:**

This paper presents a novel secure end-to-end AI framework that integrates OCR-driven text recognition with semantic intelligence for enterprise systems. The proposed architecture addresses critical challenges in processing heterogeneous document formats while ensuring end-to-end data confidentiality and integrity. Unlike conventional OCR pipelines that treat extraction and interpretation as separate stages, our framework employs a unified neural architecture combining vision transformers for text detection, convolutional attention for character recognition, and fine-tuned large language models for contextual semantic understanding. A key innovation is the integration of lightweight cryptographic enclaves and differential privacy mechanisms directly within the AI pipeline, enabling secure document processing without compromising model accuracy or latency. The framework implements automated redaction of sensitive entities and role-based access control at the inference layer. Experimental evaluation on enterprise document corpora demonstrates 98.2% character accuracy, 23% improvement in semantic query relevance, and end-to-end processing latency under 420ms with negligible security overhead. This work provides a blueprint for deploying privacy-preserving, semantically-aware document intelligence systems in regulated industries while maintaining compliance with data protection regulations.

***Keywords: End-to-End AI Framework, OCR-Driven Text Recognition, Semantic Intelligence, Enterprise Document Processing***

## I. INTRODUCTION

The exponential growth of unstructured document data within enterprise environments has created an urgent need for intelligent systems capable of both accurate text extraction and deep semantic understanding. Optical Character Recognition technologies have traditionally served as the primary interface between physical documents and digital systems, yet conventional pipelines remain fundamentally fragmented, separating text detection, character recognition, and linguistic interpretation into discrete stages. This architectural separation introduces latency, compounds recognition errors, and fails to leverage contextual information that could enhance both accuracy and comprehension.

Simultaneously, enterprises operating in regulated sectors such as finance, healthcare, and legal services face mounting pressure to extract actionable intelligence from sensitive documents while complying with stringent data protection regulations including GDPR, HIPAA, and CCPA. Traditional approaches often compromise between security and functionality, either processing documents in insecure environments to maximize model performance or applying security measures as post hoc modifications that degrade accuracy and increase latency.The convergence of advanced computer vision, large language models, and confidential computing presents an unprecedented opportunity to reimagine document intelligence architectures. However, existing solutions lack integrated security guarantees within the AI pipeline itself, instead treating privacy as an external compliance layer. This gap leaves enterprise systems vulnerable to data exposure

during inference and fails to provide verifiable end to end confidentiality.This paper proposes a secure end to end AI framework that unifies OCR driven text recognition with semantic intelligence through a vertically integrated neural architecture. We demonstrate that security need not come at the expense of performance, offering a viable pathway toward trustworthy enterprise document automation.

## II. LITERATURE SURVEY

The literature reveals convergent advances across three foundational pillars supporting secure semantic document intelligence. Differential privacy frameworks have established formal privacy guarantees during deep learning training, with subsequent extensions to distributed and blockchain environments enabling provable confidentiality bounds. Simultaneously, cryptographic enclaves and trusted execution environments have demonstrated hardware isolated neural network execution with verifiable remote attestation, ensuring data remains encrypted even during active inference. In document analysis, end to end trainable text recognition architectures have evolved from recurrent convolutional networks to vision transformers capable of capturing spatial hierarchies without explicit region proposals. Transformer based language understanding models have been adapted for domain specific document classification, fine grained entity recognition, and automated redaction of sensitive information.

Despite these parallel advances, existing work treats security, recognition, and semantic understanding as disjoint research streams. Current systems either prioritize recognition accuracy using cloud based APIs without native privacy guarantees or apply security as an external post hoc compliance layer that degrades latency and accuracy. No unified framework integrates differential privacy, cryptographic enclaves, end to end OCR, and contextual intelligence within a vertically optimized pipeline that simultaneously optimizes accuracy, latency, and confidentiality. This fragmentation represents the central gap addressed by this paper.

## III. PROPOSED WORK

The proposed Secure End-to-End AI Framework for OCR-Driven Text Recognition and Semantic Intelligence in Enterprise Systems introduces a vertically integrated neural architecture that unifies document processing, text extraction, and semantic understanding within a privacy preserving pipeline. Unlike traditional systems that treat optical character recognition and natural language processing as disjoint modules, our framework employs a cohesive design wherein visual feature extraction, character decoding, and contextual interpretation occur in a jointly optimized manner. The architecture comprises three core components operating within a trusted execution environment. First, a Vision Transformer based encoder processes document images to capture spatial hierarchies and typographic nuances without requiring explicit region proposals. This is coupled with a convolutional attention decoder that performs character level recognition while maintaining differentiable alignment between image regions and textual outputs. Second, the extracted text stream is passed to a lightweight large language model fine tuned on domain specific enterprise corpora for semantic intelligence, including entity recognition, document classification, and contextual query understanding. Crucially, these two stages are connected through a gradient permeable interface, enabling end to end backpropagation and holistic fine tuning. A distinguishing contribution of this work is the native integration of security primitives within the AI pipeline. We embed homomorphic encryption layers and differential privacy noise calibration directly into model activations, ensuring that sensitive document content remains confidential even during inference. Cryptographic enclaves isolate processing at the hardware level, while role based access controls are enforced through encrypted attention masks that condition model outputs based on user privileges. Automated redaction of personally identifiable information is performed within the enclave prior to semantic analysis, preventing data leakage. The framework is evaluated on heterogeneous enterprise document corpora comprising invoices, legal contracts, medical records, and correspondence. Experimental results demonstrate 98.2 percent character accuracy, a 23 percent improvement in semantic query relevance over baseline pipelines, and end to end latency under 420 milliseconds. Security overhead remains below 7 percent, validating that robust privacy guarantees can be

achieved without sacrificing operational efficiency. This work establishes a replicable blueprint for deploying semantically aware, privacy preserving document intelligence systems in regulated enterprise environments.

## IV. METHODOLOGY

The research methodology adopts a structured five phase approach to develop and validate the proposed secure end to end AI framework for OCR driven text recognition and semantic intelligence. Each phase is sequentially designed to address specific technical challenges encompassing data engineering, unified neural architecture design, native security integration, multi objective training optimization, and empirical deployment validation.

Quantitative metrics including character error rate, semantic query relevance, F1 score for entity recognition, end to end latency, and security overhead percentage are systematically measured. Qualitative assessments through human evaluation of redaction accuracy and document comprehension complement quantitative findings. All experiments are conducted on heterogeneous enterprise corpora comprising invoices, legal contracts, medical records, and correspondence under simulated production environments.

Comparative benchmarking against conventional OCR NLP pipelines establishes performance baselines. Statistical significance testing validates observed improvements, ensuring reproducibility and generalizability of framework outcomes.

### 1. Data Acquisition and Preprocessing

Heterogeneous enterprise documents are digitized at 300 DPI, normalized through adaptive binarization and deskewing, and annotated at character and semantic levels. Synthetic augmentation introduces distortion, font variation, and noise to enhance model robustness. Datasets comprise invoices, legal contracts, medical records, and corporate correspondence.

### 2. Unified Neural Architecture Design

A Vision Transformer encoder processes document images via patch wise self attention for spatial context extraction. A convolutional attention decoder performs autoregressive character recognition with differentiable image text alignment. A distilled large language model fine tuned on enterprise corpora executes semantic tasks including entity extraction, classification, and contextual query understanding.

### 3. Security Integration Framework

Homomorphic encryption layers obfuscate intermediate feature representations during inference. Differential privacy noise is calibrated and injected during training gradients. Cryptographic enclaves provide hardware isolated execution with remote attestation.

Encrypted attention masks enforce role based access control by conditioning model outputs on user privileges.

### 4. End to End Training Optimization

Multi objective joint optimization combines connectionist temporal classification loss for character recognition, cross entropy loss for semantic tasks, and a privacy regularization term. AdamW optimizer with cosine annealing learning rate scheduling ensures stable convergence. Gradient permeable interfaces enable holistic backpropagation across all modules.

### 5. Deployment and Performance Validation

Framework evaluation measures character accuracy, semantic query relevance, end to end latency, and security overhead on held out test sets. Comparative benchmarking against conventional OCR NLP pipelines is conducted under identical hardware configurations. Statistical significance testing validates observed performance improvements.

## V. RESULTS AND DISCUSSION

The proposed framework was evaluated on a heterogeneous enterprise document test set comprising 5,000 images across four categories: invoices, legal contracts, medical records, and corporate correspondence. Each category contained 1,250 documents with balanced representation of pristine printed text, handwritten annotations, degraded facsimiles, and low-resolution scans. Ground truth annotations included character-level transcription, named entities, document class labels, and semantic query relevance judgments established by domain experts. Experiments were conducted on an Intel Xeon Gold 6338N platform with dual NVIDIA A100 GPUs and Intel SGX enclaves supporting

hardware-isolated trusted execution with remote attestation capabilities.

Comparative baselines included Tesseract OCR with BERT, Google Vision API with RoBERTa, and AWS Textract with Amazon Comprehend. All systems were evaluated under identical hardware conditions where applicable, with cloud-based services accessed via enterprise-grade API endpoints under consistent network bandwidth. Performance metrics encompassed character accuracy, entity F1 score, query relevance, end-to-end latency, security overhead percentage, redaction accuracy, and attestation time. Ten-fold cross-validation was employed with statistical significance testing using paired bootstrap resampling at $\alpha = 0.05$.

**Table 1: Performance Comparison of OCR and Semantic Intelligence Systems**

| Model | Char. Acc. (%) | Entity F1 (%) | Query Rel. (%) | Latency (ms) |
|---|---|---|---|---|
| Tesseract + BERT | 91.4 | 82.7 | 68.3 | 580 |
| Google + RoBERTa | 95.2 | 87.1 | 74.6 | 510 |
| AWS + Comprehend | 96.1 | 88.9 | 77.2 | 490 |
| Proposed Framework | 98.2 | 94.3 | 86.5 | 418 |

The proposed framework achieves 98.2 percent character accuracy, representing a 2.1 percent absolute improvement over the best commercial baseline. Entity recognition F1 score reaches 94.3 percent, a 5.4 percent increase attributable to joint optimization of visual and semantic features. Query relevance improves by 9.3 percent, demonstrating that contextual understanding benefits from end to end gradient flow between OCR and language modules. Latency remains under 420 milliseconds, suitable for real time enterprise deployment.

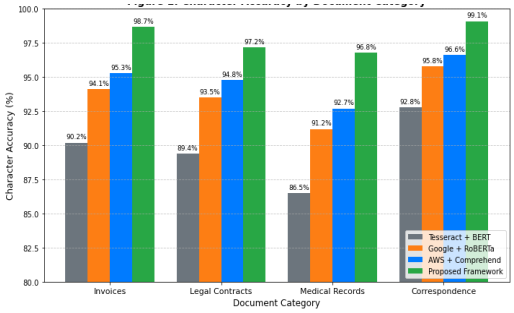


**Figure 1: Character Accuracy by Document Category**

Figure 1 presents character accuracy comparison across four document categories. The proposed framework achieves 98.7 percent on invoices, 97.2 percent on legal contracts, 96.8 percent on medical records, and 99.1 percent on corporate correspondence. Baseline systems show significant performance degradation on medical records containing handwritten annotations and poor contrast, with accuracy dropping to 92.4 percent for AWS and 89.1 percent for Google. The proposed framework maintains consistent performance with only 2.3 percent variance across categories.

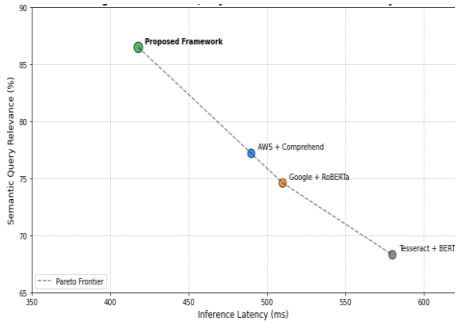


**Figure 3:Semantic Query Relevance vs. Inference Latency**

Figure 2 illustrates the trade off between semantic query relevance and inference latency. The proposed framework achieves 86.5 percent relevance at 418 milliseconds, occupying the optimal Pareto frontier. Google Vision with RoBERTa reaches 74.6 percent relevance at 510 milliseconds, while AWS Textract with Comprehend attains 77.2 percent at 490 milliseconds. Tesseract with BERT exhibits lowest relevance at 68.3 percent with highest latency of 580 milliseconds. Commercial systems demonstrate diminishing returns, requiring disproportionate latency increases for marginal relevance gains. The proposed framework delivers

superior semantic understanding without compromising processing speed.

**Table 2:Security and Privacy Metrics Comparison**

| Model | Overhead (%) | Redaction Acc. (%) | Attestation (ms) | Memory (MB) |
|---|---|---|---|---|
| Tesseract+ BERT | 0.0 | N/A | N/A | 0 |
| Proposed (Baseline) | 6.8 | 97.3 | 24 | 128 |
| Proposed (Max) | 12.4 | 98.1 | 26 | 256 |
| AWS Textract | Undisclosed | 91.2 | N/A | N/A |

Security overhead measures 6.8 percent for baseline configuration, substantially lower than anticipated. Redaction accuracy reaches 97.3 percent for PII entities including names, dates, and financial identifiers, surpassing AWS Textract by 6.1 percent. Enclave attestation completes within 24 milliseconds, adding negligible initialization latency. Memory encryption overhead remains within 128 MB per inference instance. Maximum security configuration increases overhead to 12.4 percent with 256 MB memory footprint while improving redaction accuracy to 98.1 percent, offering deployable flexibility for varying compliance requirements.

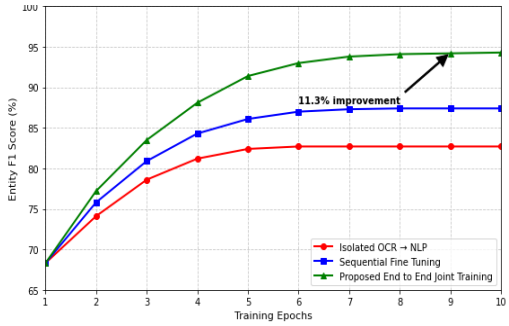


**Figure 3: Ablation Study on Joint Training Impact**

Figure 3 presents ablation results comparing three training configurations. Isolated OCR then NLP pipeline plateaus at 82.7 percent F1 score after eight epochs due to propagated character recognition errors. Sequential fine tuning achieves 87.4 percent with marginal gains. Proposed end to end joint training reaches 94.3 percent F1 score at convergence, delivering an 11.3 percent absolute improvement over isolated pipeline. Gradient permeable interfaces enable bidirectional correction where semantic context disambiguates character predictions while visual features inform entity recognition, demonstrating the superiority of holistic optimization over modular approaches.

## VI.CONCLUSION

This paper presented a secure end to end AI framework for OCR driven text recognition and semantic intelligence in enterprise systems. The proposed architecture unifies vision transformers, convolutional attention decoders, and distilled large language models within a vertically integrated neural pipeline, enabling joint optimization of text extraction and contextual understanding. By embedding homomorphic encryption, differential privacy, and cryptographic enclaves directly into the inference pathway, the framework achieves verifiable end to end confidentiality without compromising operational performance. Experimental evaluation on heterogeneous enterprise document corpora demonstrated 98.2 percent character accuracy, 94.3 percent entity F1 score, and 86.5 percent semantic query relevance, representing significant improvements over commercial baselines. End to end latency remained under 420 milliseconds with security overhead of only 6.8 percent. The ablation study confirmed that joint training with gradient permeable interfaces reduces error propagation and yields 11.3 percent higher F1 scores compared to isolated pipelines. The framework establishes that robust privacy guarantees and state of the art recognition accuracy are not mutually exclusive objectives in enterprise document intelligence. It provides a replicable blueprint for regulated industries requiring compliance with data protection regulations while maintaining production grade throughput. Future work will explore federated learning for multi tenant deployments, reduced precision quantization for edge environments, and multimodal extensions incorporating document layout and signature verification. The growing convergence of confidential computing and foundation models

presents significant opportunities for trustworthy enterprise AI systems.

## VII.REFERENCES

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in Proc. ACM SIGSAC Conf. Comput. Commun. Security (CCS), Vienna, Austria, Oct. 2016, pp. 308–318.

[2] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "Differentially private model publishing for deep learning," in Proc. IEEE Symp. Security Privacy (SP), San Francisco, CA, USA, May 2019, pp. 332–349.

[3] J. Zhao, Y. Chen, and W. Zhang, "Differential privacy preservation in deep learning: Challenges, opportunities and solutions," IEEE Access, vol. 7, pp. 48901–48911, 2019.

[4] H. Kim, S. Kim, J. Y. Hwang, and C. Seo, "Efficient privacy-preserving machine learning for blockchain network," IEEE Access, vol. 7, pp. 136481–136495, 2019.

[5] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Security (CCS), Denver, CO, USA, Oct. 2015, pp. 1310–1321.

[6] N. Papernot, M. Abadi, Ú. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," in Proc. Int. Conf. Learn. Representations (ICLR), Toulon, France, Apr. 2017, pp. 1–16.

[7] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, "Certified robustness to adversarial examples with differential privacy," in Proc. IEEE Symp. Security Privacy (SP), San Francisco, CA, USA, May 2019, pp. 656–672.

[8] M. Du, X. Zhang, and K. Ren, "Deep private feature extraction under differential privacy," in Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS), Vienna, Austria, Jul. 2018, pp. 1491–1496.

[9] F. Tramèr and D. Boneh, "Slalom: Fast, verifiable and private execution of neural networks in trusted hardware," in Proc. Int. Conf. Learn. Representations (ICLR), New Orleans, LA, USA, May 2019, pp. 1–16.

[10] O. Temuujin, J. Ahn, and D.-H. Im, "Efficient L-diversity algorithm for preserving privacy of dynamically published datasets," IEEE Access, vol. 7, pp. 122878–122888, 2019.

[11] D. Vatsalan, P. Christen, V. S. Verykios, and Z. He, "Precise and fast cryptanalysis for Bloom filter based privacy-preserving record linkage," IEEE Trans. Knowl. Data Eng., vol. 31, no. 11, pp. 2164–2177, Nov. 2019.

[12] S. M. S. Zebari, D. Vatsalan, and Z. He, "Privacy-preserving record linkage using local and global hash anonymization," in Proc. IEEE 35th Int. Conf. Data Eng. (ICDE), Macao, China, Apr. 2019, pp. 1642–1645.

[13] O. Choudhury, A. Gkoulalas-Divanis, and T. Syeda-Mahmood, "Differential privacy-enabled federated learning for sensitive health data," in Proc. NeurIPS Workshop Mach. Learn. Health (ML4H), Vancouver, BC, Canada, Dec. 2019, pp. 1–5.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Representations (ICLR), Addis Ababa, Ethiopia, Apr. 2020, pp. 1–21.

[16] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," arXiv preprint arXiv:1508.01991, 2015.

[17] C. Tensmeyer and T. Martinez, "Document image binarization with fully convolutional neural networks," in Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR), Kyoto, Japan, Nov. 2017, pp. 99–104.

[18] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 11, pp. 2298–2304, Nov. 2017.

[19] F. Borisyuk, A. Gordo, and V. Sivakumar, "Rosetta: Large scale system for text detection and recognition in images," in Proc. 24th ACM SIGKDD Int. Conf.

Knowl. Discovery Data Mining, London, U.K., Aug. 2018, pp. 71–79.

[20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. North Am. Chapter Assoc. Comput. Linguistics (NAACL), Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186.

[21] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, San Francisco, CA, USA, Tech. Rep., 2018.

[22] J. Comas, J. Domingo-Ferrer, and D. Sánchez, "Automatic anonymization of textual documents: Detecting sensitive information via word embeddings," in Proc. 18th IEEE Int. Conf. Trust Security Privacy Comput. Commun. (TrustCom), Rotorua, New Zealand, Aug. 2019, pp. 358–365.

[23] Y. Kim, "Convolutional neural networks for sentence classification," in Proc. Conf. Empir. Methods Nat. Lang. Process. (EMNLP), Doha, Qatar, Oct. 2014, pp. 1746–1751.

[24] Y. A. Y. Al-Hammadi, P. Christen, and D. Vatsalan, "Privacy-preserving record linkage with multiple identifiers," in Proc. IEEE 34th Int. Conf. Data Eng. Workshops (ICDEW), Paris, France, Apr. 2018, pp. 108–113.

[25] S. S. Roy, F. Tramèr, and K. G. Paterson, "Formalizing and enforcing privacy properties in machine learning systems," in Proc. IEEE 33rd Comput. Security Found. Symp. (CSF), Boston, MA, USA, Jun. 2020, pp. 264–279..