Critical Reading Exam

> **Examination:** PhD Critical Reading Exam (Doctoral Commission, June 2026)
>
> **Candidate:** Stergios Konstantinidis • *Department of Information Systems (DESI), HEC Lausanne, University of Lausanne (UNIL)*
>
> **Topic:** *Large Language Models for Historical Document Intelligence: Retrieval-Augmented Generation, Automated Evaluation, and Agentic Architectures*
>
> **Repositories:** GitHub CRE Repository • Overleaf Exam Workspace
>
> **Primary Examination Documents:** **Exam Paper (PDF, 8 Pages)** • **Presentation Deck (PDF, 13 Slides)** • **Presentation Deck (PPTX, 11 MB)**

1. The 5 Literature Review Articles (Critical Reading Exam)

The exam selection systematically examines five cornerstone publications spanning four interconnected theoretical pillars in historical document intelligence:

Article	Primary Contribution	Addressed Architectural Vulnerability	Archival & Methodological Relevance
Tran et al. (JCDL 2024) *RAG for Historical Newspapers*	First end-to-end RAG system over multilingual historical newspapers (NewsEye); proposed hybrid dense retrieval and NER-enhanced re-ranking with ground-truth-free LLM metrics.	Degraded OCR query drift; severe annotation scarcity in historical corpora.	Establishes the problem setting: multilingual, OCR-corrupted archival text requires hybrid dense retrieval and robust reference-free evaluation.

Article	Primary Contribution	Addressed Architectural Vulnerability	Archival & Methodological Relevance
Guan et al. (AAAI 2024) *Synthetic OCR Degradation & Robust Retrieval*	Principled synthetic OCR degradation generation and test-time adaptation targeting proper nouns and out-of-vocabulary entities.	Severe training data scarcity for historical periods; catastrophic forgetting of rare historical entities.	Provides methodology for synthetic degradation modeling and test-time safeguard mechanisms to prevent hallucinated modernisation of archaic nomenclature.
Gu et al. (The Innovation 2026) *A Survey on LLM-as-a-Judge*	Systematic taxonomy of LLM judges: position bias, verbosity bias, self-enhancement bias, and calibration protocols.	Unreliable evaluation in ground-truth-free RAG evaluation pipelines.	Establishes the theoretical and methodological foundation for validating automated evaluation metrics in ground-truth-free archival pipelines.
Sun et al. (EMNLP 2025) *DocAgent: Multi-Modal Long-Context Framework*	Agentic architecture featuring selective retrieval, multimodal document inspection tools, and iterative answer verification agents.	Context window limits; inability of pure text LLMs to resolve visual layout artifacts.	Demonstrates how agentic decomposition, multimodal tool use, and multi-turn verification overcome the limitations of flat passage retrieval.
Lewis et al. (NeurIPS 2020) *Foundational Retrieval-Augmented Generation*	The foundational dual-encoder dense retrieval and seq2seq generator architecture trained end-to-end with latent documents.	Hallucination in closed-book parametric memory; inability to update knowledge.	The theoretical origin of RAG; reveals how modern assumptions (clean English Wikipedia) break down when deployed over centuries of degraded archives.

2. Theoretical & Architectural Synthesis

The Foundational Tension in Historical Document Intelligence

Across all reviewed articles, a persistent architectural tension emerges: **the fundamental assumptions that make LLM-based document intelligence effective are systematically violated by the historical archives that most urgently need it:**

1. **Assumption of Orthographic Cleanliness:** Standard RAG models assume high-quality input text. In historical newspapers, character error rates (CER) frequently exceed 15–25% due to font fading, broken typefaces, ink bleed, and complex multi-column typography. Dense vector spaces must therefore be explicitly regularized against OCR degradation.
2. **Assumption of Monolingual Modernity:** Foundational retrievers are optimized for contemporary English. Archival collections encompass archaic dialects, historical spelling variants, and evolving syntactic conventions across multiple centuries. Multilingual embeddings combined with robust spatial indexing are essential to bridge temporal linguistic drift.
3. **Assumption of Annotation Abundance:** Modern benchmarks rely on millions of human QA annotations. Archival collections possess virtually zero labeled question-answering pairs. Reconciling this gap requires robust ground-truth-free evaluation frameworks carefully calibrated against prompt bias, verbosity artifacts, and positional skew.
4. **Assumption of Flat Screen Interfaces:** Traditional document retrieval presents flat lists of snippets. Navigating interrelated historical narratives demands agentic reasoning, multimodal layout inspection, and interactive exploration that turn passive retrieval into active sensemaking.

---

3. Research Milestone & Examination Context

- **Doctoral Commission:** Faculty of Business and Economics (HEC Lausanne), University of Lausanne.
- **Examination Date:** June 2026.
- **Outcome:** Passed

Revision #13
Created 2026-09-09 16:20:30 UTC by Stergios
Updated 2026-09-17 12:24:48 UTC by Stergios