FInk: A Financial-Risk Review Assistant for Creator Contracts

Seonuk Kim [1]

Abstract

Webtoon creators may need to review revenue-sharing contracts on a phone before professional advice is available. I present FInk, a local-first assistant that accepts text, photographs, or PDFs; identifies clauses across nine cash-flow risk categories; retrieves supporting passages from a curated Korean corpus; and returns ranked, cited findings with a heuristic 0–100 review-attention score. Only findings backed by score-eligible official sources can change the score. On frozen synthetic fixtures, hybrid detection reached macro-F1 0.92 with a benign false-positive rate of 0.25; retrieval preserved source eligibility; exposure-aware ranking improved point estimates; and the tested runtime made no network calls. The tests verify the implemented pipeline on controlled cases. They do not establish field performance: no real contracts or creator participants were used, and the score remains uncalibrated. FInk is not legal advice. The system is available at https://fink.seonukkim.com.

1. Introduction

Problem definition. This project studies pre-signing financial review for webtoon creators. Given a contract and a limited reading budget, FInk ranks the clauses that may have the largest effect on realized cash flow and shows the source behind each warning. It does not assign a contract-level "safe" or "unsafe" label. The output is a short sequence of questions about the revenue base, permitted deductions, recoupment, settlement timing, minimum guarantees, secondary-rights income, exclusivity, and other terms that change when or how much a creator is paid. The error costs are asymmetric: a missed exposure can affect income after signing, whereas a false alarm mainly costs additional verification time. For example, a 30% share of "net revenue" can pay less than a smaller share of gross receipts when platform and marketing costs are deducted first. FInk therefore treats the revenue base, deductions, recoupment order, and settlement lag as competing claims on a limited review budget.

Motivation. I focus on webtoon creators because their contracts combine contingent revenue sharing with deductions, recoupment, intellectual-property revenue, exclusivity, and production obligations. These terms are financial mechanisms, not peripheral boilerplate: they determine cash inflow, payment delay, downside exposure, and opportunity cost. Yet the first review often happens under time pressure and without a finance or legal specialist present.

The mobile setting shapes the system. A draft may arrive as a message attachment, a PDF, a photographed page, or a clause shown during a meeting. The creator may not be able to take the original document away, and uploading a private agreement to a third-party service can itself be unacceptable. FInk is therefore designed so that a creator can photograph a page or paste text, ask about the relevant clause, and keep the document on the device. Smartphone use is the deployment target; the experiments in this report test a local configuration rather than a production handset.

Professional review remains important, but availability does not guarantee that advice will be used. In a field study of roughly eight thousand retail investors, people who most needed unbiased financial advice were least likely to obtain it, and those who obtained it followed little of it (Bhattacharya et al., 2012). That study concerns investors rather than creators, so I use it only for the broader lesson that lower-friction decision support can matter before professional consultation. FInk helps the creator identify what to ask; it does not replace an adviser or decide whether a contract is lawful or fair.

Approach and scope. FInk segments the input into clauses, detects signals across nine cash-flow categories, retrieves supporting passages from Korean standard contracts, contracting guides, and counseling casebooks, and produces a ranked review. The interface shows a three-level review-effort recommendation, separate counts for official evidence and practice context, and questions tied to each clause. An optional local model answers follow-up questions over the retrieved material, and the review can be saved as a short report.

The evidence gate is the central safeguard. A finding may affect the numerical score only when an A0–A2 official source

[1]Department of Industrial Engineering, UNIST, Ulsan, Korea. Correspondence to: Seonuk Kim <seonuk.kim@unist.ac.kr>.

Final Project Report of IE412 AI for Finance, Spring 2026, UNIST.

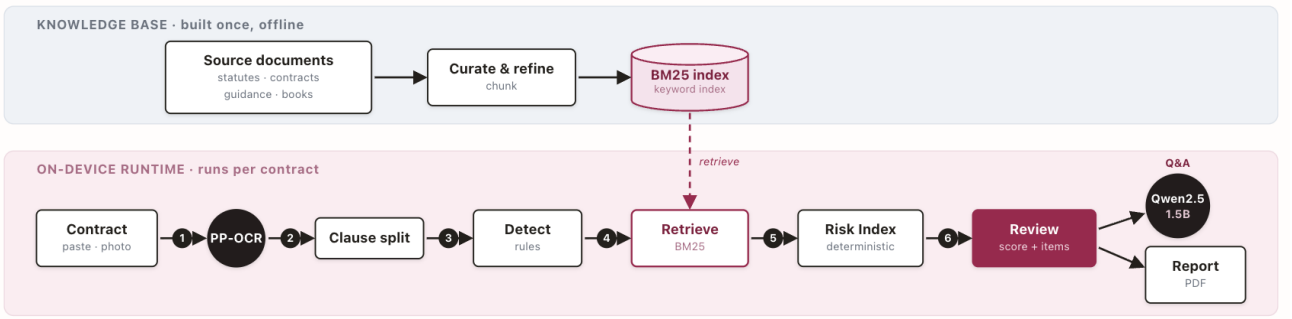

Figure 1. The current FInk pipeline. The knowledge base is curated and indexed once, offline. For each contract, the on-device runtime accepts pasted text or an image, performs OCR when needed, segments the text into clauses, detects financial-risk signals, retrieves supporting material, computes the heuristic Risk Index, and returns a prioritized review. The shipped path uses inspectable rules; the classifier is evaluated separately in Section 3.2. The review supports local Q&A and PDF report export.

supports the corresponding review question. Copyrighted secondary books may inform a non-scoring question, but they cannot raise the score. The gate governs which sources may influence the index; legal validity remains outside the system's scope. The resulting 0–100 value is a bounded aid for ordering attention rather than a probability of loss or an expected-loss estimate.

Contribution and evaluation logic. The main contribution is an implemented local workflow that turns messy creator-contract input into a short, source-backed financial review. The evaluation follows the runtime sequence: detection should find candidate cash-flow exposures without excessive benign alarms; retrieval should preserve source eligibility; ranking should cover more synthetic exposure within a small reading budget; and the runtime should pass scoring, privacy, and execution checks. The contributions are:

- a cash-flow review formulation, nine-category taxonomy, and evidence gate that separates detected concerns from score-eligible official support;
- a smartphone-oriented local prototype connecting text, photo, and PDF input to clause analysis, cited priorities, follow-up questions, and PDF report export; and
- a decision-focused evaluation of detection, source eligibility, ranking, and local execution, with the limits of each result stated explicitly.

2. Related Work

Retrieval for financial documents. Retrieval-augmented generation conditions answers on retrieved evidence rather than parameters alone (Lewis et al., 2020). BM25 (Robertson & Zaragoza, 2009) and dense passage retrieval (Karpukhin et al., 2020) are common alternatives. Financial question answering increasingly treats retrieval as a primary source of error. FinDER studies short, jargon-heavy financial questions (Choi et al., 2025a), and FinAgentBench separates document selection from passage selection (Choi et al., 2025b). FInk uses BM25 over a small, curated Korean corpus because exact legal and financial terms matter and the index must remain inexpensive enough for local execution.

Decision-focused and responsible financial AI. Decision-focused learning evaluates predictions by the downstream decisions they support, not only by predictive error (Elmachtoub & Grigas, 2022). Financial language-model evaluation is also vulnerable to narrative and objective biases: fluent output can sound more conclusive than its evidence warrants (Kong et al., 2026). FInk does not train through an optimizer, but its main evaluation asks whether a ranking covers more financial exposure within a limited reading budget. The interface presents citations and review questions instead of an unqualified verdict.

Local document AI. Small instruction-tuned models such as Qwen2.5 (Qwen Team, 2024) and lightweight OCR such as PP-OCR (Du et al., 2020) make local document processing practical on commodity hardware. In FInk, locality is a privacy requirement as well as a cost and latency choice. It also imposes a capability ceiling: the current report tests one local configuration and does not claim production-level smartphone performance.

3. Method

Figure 1 separates the offline knowledge-base build from the per-contract runtime. At runtime, FInk first identifies candidate cash-flow exposures, then retrieves supporting material, applies the evidence gate, and orders the grounded findings for review.

3.1. Decision task and design criteria

Let $C = \{c_1, \ldots, c_n\}$ be the clauses of one contract. FInk returns an ordering π over C, a source-eligibility state for

Table 1. Nine cash-flow risk categories. A detected signal becomes score-eligible only through the evidence gate.

	Cash-flow risk captured
F1	Settlement transparency and audit rights
F2	Revenue base and deductions
F3	Payment and cash-flow timing
F4	Minimum guarantee and recoupment
F5	Intellectual property and secondary-rights revenue
F6	Term, exclusivity, and opportunity cost
F7	Termination, liability, and penalties
F8	Scope creep and production-cost burden
F9	Other cash-flow risk (residual)

each finding, and a review question tied to the clause. For a reading budget k, the evaluation asks how much synthetic financial exposure appears in the first k positions. The system does not estimate monetary loss directly; it uses category and severity proxies to construct the order.

Five criteria guided the design. The output must support a decision rather than issue a verdict; only official evidence may affect the score; contract text should stay local; the result should fit a short reading budget; and Korean source text should remain canonical, with English used only as a retrieval and reading aid.

3.2. Financial-risk detection

The contract is segmented into clauses, which become the units of detection, retrieval, scoring, and follow-up questions. Each clause is checked against the nine cash-flow categories in Table 1. I evaluate three detector arms: deterministic rules, a constrained local classifier, and the union of their outputs. The public prototype uses rules by default because a creator can inspect why a clause matched. The hybrid arm is a controlled comparison of whether classifier output adds recall without increasing benign false positives; the shipped policy remains rule-based.

3.3. Evidence-gated retrieval

For each detected signal, BM25 retrieves supporting chunks from the knowledge base. A signal becomes score-eligible only when an A0–A2 official source supports the relevant review question. A B/C match may sharpen the question but cannot alter the score. The interface reports official-evidence and practice-basis counts separately. An official citation raises the source standard for a numerical finding. Whether the clause is harmful or invalid remains outside the gate's scope.

3.4. Review-attention score and output

The interface uses the concise label "Risk Index." In this paper I call it a review-attention score to avoid suggesting that it is a calibrated probability or loss estimate. For each category c, grounded severity x_c is mapped to a non-decreasing, saturating contribution $g_c(x_c)$ capped by a weight w_c. The total is

$$R = \sum_c g_c(x_c), \qquad 0 \le R \le 100, \tag{1}$$

with the bound proved in Proposition 1. Saturation prevents repeated signals within one category from increasing the score without limit, consistent with a general risk-assessment design principle (International Organization for Standardization, 2018; International Electrotechnical Commission, 2018). Risk aversion and multi-attribute value theory motivate considering recoverability, uncertainty, and cash-flow dimensions that are not directly comparable (Pratt, 1964; Keeney & Raiffa, 1976); they do not validate the selected weights. The weights, saturation rates, and thresholds are uncalibrated engineering choices. Evidence eligibility and severity remain separate, in the spirit of distinguishing evidence quality from recommendation strength (Guyatt et al., 2008).

The score maps to three review-effort labels: light check, careful check, and professional review recommended. The ranked list uses the same cash-flow dimensions to prefer clauses with larger potential exposure over clauses that merely use severe wording. An optional Qwen2.5-1.5B-Instruct model in a 4-bit build (Qwen Team, 2024) answers questions over retrieved findings. Free-form chat is optional and excluded from the quantitative claims. The browser print path exports a short PDF report without a remote document service.

4. Data and Implementation

4.1. Knowledge-base data and preprocessing

Score-eligible data consist of Korean standard contracts and annotations for cartoons and webtoons (Korea Creative Content Agency, 2025); standard contracts for assignment and licensing of authors' economic rights (Ministry of Culture, Sports and Tourism (Korea) & Korea Copyright Commission, 2018); a fair-contracting guide (Korea Creative Content Agency, 2023); a copyright counseling casebook (Korea Copyright Commission, 2025); and a content-dispute mediation casebook (Content Dispute Resolution Committee (Korea), 2024). I tag these sources A0–A2 according to their role as official rules, standard forms, or public guidance.

Two commercial references, *A Webtoon Creator Gets a Lawyer Friend: What to Know Before Signing a Contract* and *Law and Everyday Life for Koreans*, provide practice-oriented B/C context (Yun et al., n.d.; Ministry of Justice (Korea) & Korea Law-Related Education Center, 2021). Their authority tier reflects source role and redistribution limits rather than usefulness. Because the copyrighted prose cannot be released as a retrieval corpus, I retain only short,

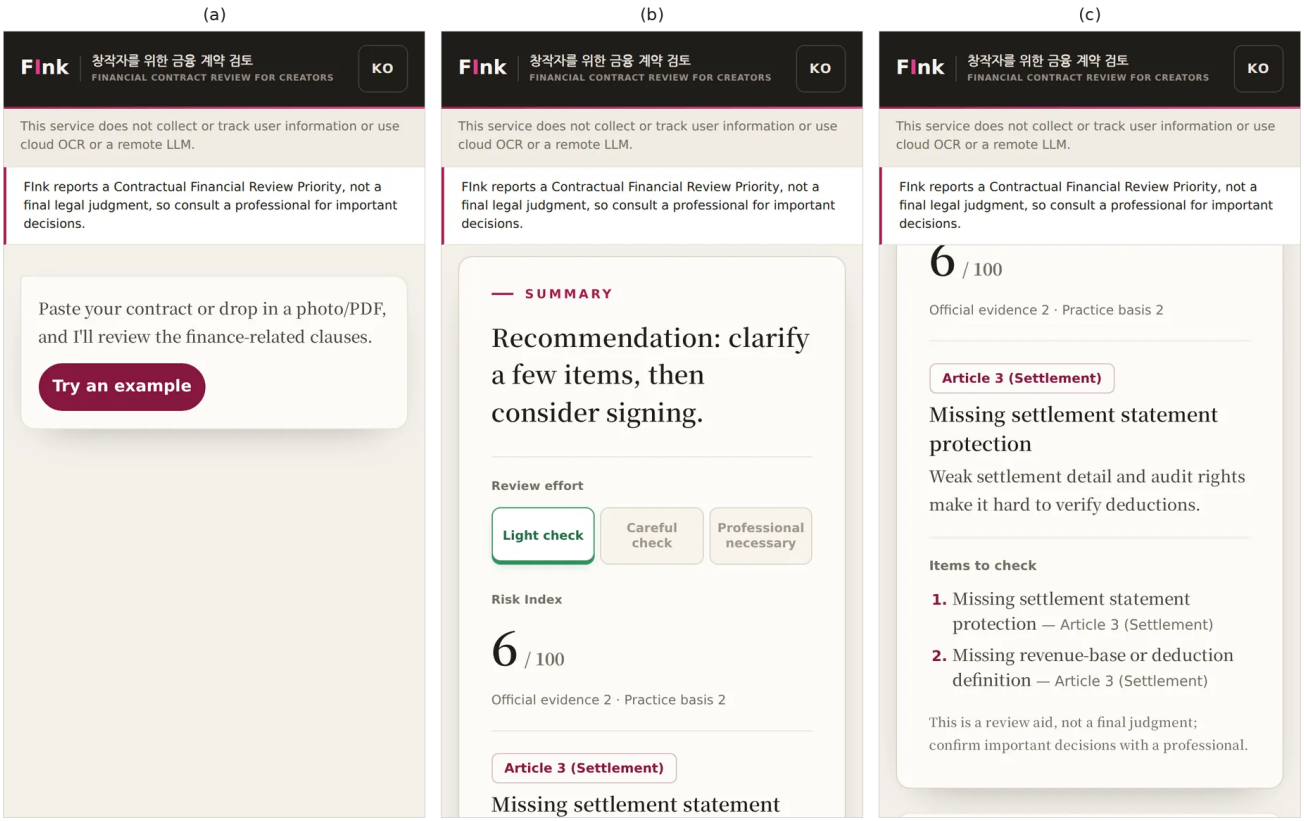

Figure 2. The current smartphone-width English demo. (a) The creator pastes text or uploads a photograph or PDF. (b) The summary shows the review-effort recommendation, the heuristic 0–100 Risk Index, and separate counts for official evidence and practice context. (c) The detailed view names the clause, explains the financial concern, and lists questions to check before signing.

non-verbatim checkpoints. They may suggest a question but cannot raise the score or appear as quoted evidence.

Source PDFs with embedded text were parsed directly, while scanned pages passed through OCR. I normalized article numbers and headings, rejoined clauses split across page boundaries, removed running headers and footers, corrected obvious OCR errors, and converted tables or enumerated provisions into stable text blocks. Chunks follow article or paragraph boundaries; long clauses are split with overlap so that definitions and exceptions are not detached from the main term. Each chunk stores a source identifier, document title, article number or page locator, authority tier, topic, canonical Korean text, and a separate English query aid.

Claude Opus 4.8 produced an initial document structure for the official sources, and ChatGPT 5.5 Pro produced initial non-verbatim checkpoints for the two books. I then compared source identity, numbering, hierarchy, tier labels, and representative chunks with the originals. The BM25 index was frozen only after those checks. Raw commercial-book text is not part of the released evidence corpus. I reviewed additional specialist material, but I did not find a reliable way to transform and redistribute it without risking protected expression; this limits corpus coverage, as discussed in Section 7.

4.2. Evaluation fixtures

Real contracts are private and difficult to redistribute, so evaluation uses frozen synthetic and sanitized fixtures. Tuning used separate development examples. Table 2 states the size and purpose of each fixture. The detection and retrieval sets are too small for population-level claims. The factorial fixture is a controlled ranking test in which hidden oracle exposure weights are used only for evaluation, not by either ranking method.

4.3. Implementation

Runtime stack. The application is served locally with FastAPI. PaddleOCR PP-OCR handles optional photograph and PDF input in its Korean configuration (Du et al., 2020); BM25 retrieves from the curated index; a deterministic module computes the review-attention score; and llama.cpp runs the optional 4-bit GGUF chat model. The model is exposed only after an offline health check. No quantitative result depends on free-form chat. The implemented path connects input, clause analysis, evidence gating, ranking, follow-up

Table 2. Frozen evaluation fixtures.

Component	Fixture size	Purpose
Detection	10 (6 risk, 4 benign)	Category detection; benign FPR
Retrieval	1 bilingual pair	Authority-gate check
Decision	128 (64 risk, 64 benign); 200 resamples	Ranking under a reading budget
Cost trigger	9	Miss cost versus verification effort
Runtime	3	Offline, logging, and network checks

questions, and PDF report export. Real contracts, uploaded files, OCR text, and private source text are excluded from released fixtures and artifacts.

Plan–build–verify loop. I used a repeated plan–build–verify workflow. Claude Code with Opus 4.8 Max converted each feature request into an effort estimate, an acceptance checklist, and ordered tasks. Codex 5.5 at xhigh reasoning effort implemented the tasks; Claude Code then ran the relevant tests and compared the result with the checklist. I reviewed every diff before integration, accepting, modifying, or rejecting changes. A cycle closed only when the implementation and recorded tests matched the checklist. The tools coordinated development; I retained control of the financial taxonomy, authority tiers, scoring policy, acceptance criteria, and release decisions.

Verification gates. Verification combined unit tests for score bounds and saturation, frozen financial scenarios, authority-tier checks, bilingual retrieval checks, privacy checks over logs and exports, and a network-attempt monitor. Each gate maps to a user-visible failure: score overflow, ineligible evidence changing the index, mismatched retrieval, private text leaking into an artifact, or an unexpected outbound call. A failed gate blocks the corresponding release candidate. Additional fixture and artifact details appear in Sections C and D.

5. Results

Each result is tied to a user-facing part of the implemented workflow. Every number comes from a frozen fixture and should be read as a controlled test or descriptive comparison, not as an estimate of field effectiveness.

5.1. Financial-risk detection

Table 3 compares the three detector arms. The hybrid has the highest macro-F1 and the same benign false-positive rate as the rules on this fixture. The model-only arm has twice the benign false-positive rate. Because the fixture contains only ten cases, these values do not estimate performance on unseen contracts.

Table 3. Detection over F1–F9 on ten synthetic cases. Higher macro-F1 is better; lower benign false-positive rate (FPR) is better.

Arm	Macro-F1	Benign FPR
Rule only	0.83	0.25
Model only	0.86	0.50
Hybrid	**0.92**	0.25

5.2. Evidence grounding

On the frozen bilingual query pair, the retrieved source has the intended authority tier and B/C material is not marked score-eligible. Recall@3 and Recall@5 are 1.00, the Korean and English queries resolve to the same evidence, and the cited span overlaps the frozen gold span by 0.85. Because the set contains one bilingual pair, this result verifies the retrieval path only.

5.3. Decision quality

The decision-focused question is whether the ranking covers financial exposure within a small reading budget (Elmachtoub & Grigas, 2022). Oracle exposure coverage at k (OEC@k) is the share of synthetic oracle exposure captured by the top-k clauses. The oracle weights are hidden from both ranking methods and are used only for evaluation.

Exposure-aware ranking has higher point estimates in all four comparisons in Figure 3. With the evidence gate off, OEC@1 rises from 0.69 (95% CI $[0.57, 0.77]$) to 0.81 ($[0.72, 0.90]$), although the intervals overlap. OEC@3 rises from 0.93 ($[0.89, 0.97]$) to 1.00 ($[0.99, 1.00]$), where the intervals are separated. With the gate on, point estimates rise from 0.41 to 0.53 at $k = 1$ and from 0.55 to 0.62 at $k = 3$, but the intervals overlap. Only the ungated OEC@3 comparison shows separated intervals. The remaining differences are directional because their intervals overlap.

Gating lowers absolute coverage because unsupported clauses do not count; the lost coverage is the cost of source control. On the separate nine-case cost-sensitive fixture, a fixed verification trigger recovers 0.67 of consequential cases at a false-trigger rate of 0.33. That result is descriptive and depends on the synthetic miss-cost and verification-effort assumptions.

5.4. Implementation correctness and local runtime

The scoring module passes all frozen formula checks ($13/13$ unit cases and $10/10$ financial-scenario cases), and the score remains in $[0, 100]$ by construction. Across three accepted runtime cases, the harness records zero network attempts and no contract text or file paths across six inspected log and export surfaces. On the text-input path (clause segmentation, rules, BM25 retrieval, scoring, and report assembly; OCR and chat excluded), execution remains below one second with peak process memory in the tens of megabytes in the fixed local environment. These checks cover the tested con-

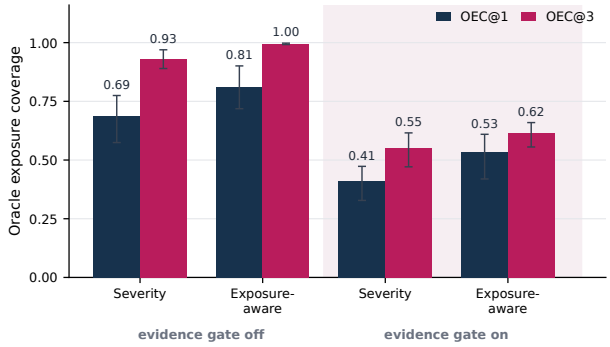

Figure 3. OEC@k on the 128-document synthetic factorial fixture. Bars show point estimates; error bars show 95% bootstrap intervals from 200 resamples. Exposure-aware ranking has higher point estimates in every comparison. The clearest interval separation occurs for OEC@3 with the evidence gate off.

figuration rather than handset latency, battery use, or thermal behavior. Supplementary OCR and extraction results appear in Section B.

6. Discussion

Demonstrated feasibility. The prototype implements the complete path from phone-oriented input to a source-backed review: input becomes clauses, candidate cash-flow exposures are detected, official evidence controls scoring, findings are ranked, and a PDF report is produced without a remote document service. The controlled tests check whether these stages work together on fixed cases. Measuring whether creators make better signing decisions requires a separate user study.

Financial decision objective. Contract review is modeled as an allocation of limited attention. Detection identifies terms that may change cash inflow, deductions, payment timing, downside exposure, or opportunity cost. Retrieval determines whether a finding has an eligible source basis, and OEC@k measures how much exposure the reader encounters before the review budget is exhausted. This links the financial problem to the AI method and evaluation more directly than clause accuracy alone.

Release policy. The evidence gate prevents unsupported scoring but reduces coverage when the corpus lacks an eligible source. I accepted that loss because private or copyrighted material should not determine a public numerical index. The hybrid detector has the best macro-F1 on the small fixture; however, ten cases are insufficient to replace the inspectable rule-based policy, especially when the model-only arm produces more benign alarms.

Privacy and deployment. Local execution reduces disclosure risk in the tested configuration, but it limits model capacity. The 1.5B chat model is therefore optional, and its output does not enter the reported metrics. FInk keeps the creator in the decision by showing sources and questions rather than issuing a ruling, and it directs higher-consequence cases toward professional review.

7. Limitations

Field validation. The current evidence is limited to feasibility tests. I did not analyze real creator contracts or recruit creator participants. The fixtures are small, including ten detection cases and one bilingual retrieval pair. The 128-document factorial set tests a controlled ranking mechanism, but its oracle weights are synthetic. These values cannot estimate unseen-contract error rates or signing outcomes.

Uncalibrated score. The 0–100 scale, category weights, saturation rates, and review thresholds are heuristic engineering choices. The appendix proves boundedness only. The value has no probabilistic or expected-loss interpretation, and equal intervals need not carry equal financial meaning. Calibration would require expert elicitation, creator studies, and observed outcomes; until then, the index should be read only as a relative review order.

Corpus coverage and copyright. The Korean knowledge base focuses on webtoon, copyright, and content contracts. I consulted more specialist material than the live index contains, but I could not identify a safe and reproducible way to transform copyrighted professional texts into retrievable evidence without retaining protected expression. The released corpus is therefore narrower than the material reviewed during development. Non-verbatim B/C checkpoints can also omit legal or commercial nuance, and a missing score-eligible source does not imply that a financial exposure is absent.

Deployment and model limits. Smartphone use is a design target rather than a completed hardware study. I do not report handset latency, battery use, thermal behavior, or cross-device model quality. OCR may fail on difficult images, the local chat model is uncalibrated, and rules may miss paraphrases. FInk remains a review aid; consequential decisions still require a qualified professional.

8. Conclusion

FInk implements the full path from phone-oriented contract input to a source-backed financial review and PDF report while keeping the document local. On frozen fixtures, the hybrid detector has the best measured macro-F1, retrieval preserves source eligibility, exposure-aware ranking has higher point estimates than a severity-only baseline, and the tested runtime makes no network calls. The evidence supports the technical feasibility of the implemented workflow. Next steps are creator evaluation, score calibration, a broader copyright-safe corpus, and direct handset benchmarks.

References

Bhattacharya, U., Hackethal, A., Kaesler, S., Loos, B., and Meyer, S. Is unbiased financial advice to retail investors sufficient? answers from a large field study. *The Review of Financial Studies*, 25(4):975–1032, 2012.

Choi, C., Kwon, J., Ha, J., Choi, H., Kim, C., Lee, Y., Sohn, J.-y., and Lopez-Lira, A. FinDER: Financial dataset for question answering and evaluating retrieval-augmented generation. In *Proceedings of the 6th ACM International Conference on AI in Finance (ICAIF)*, 2025a.

Choi, C., Kwon, J., Lopez-Lira, A., Kim, C., Kim, M., Hwang, J., Ha, J., Choi, H., Yun, S., Kim, Y., and Lee, Y. Finagentbench: A benchmark dataset for agentic retrieval in financial question answering. In *Proceedings of the 6th ACM International Conference on AI in Finance (ICAIF)*, 2025b.

Content Dispute Resolution Committee (Korea). Casebook of content dispute mediation, 2024. Issued in Korean.

Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., Bai, Y., Yu, Z., Yang, Y., Dang, Q., and Wang, H. PP-OCR: A practical ultra lightweight OCR system, 2020. arXiv:2009.09941.

Elmachtoub, A. N. and Grigas, P. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., and Schünemann, H. J. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650):924–926, 2008.

International Electrotechnical Commission. IEC 60812:2018 failure modes and effects analysis (FMEA and FMECA). International Standard, 2018.

International Organization for Standardization. ISO 31000:2018 risk management — guidelines. International Standard, 2018.

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W.-t. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781, 2020.

Keeney, R. L. and Raiffa, H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley, 1976.

Kong, Y., Lee, H., Hwang, Y., Lopez-Lira, A., Levy, B., Mehta, D., Wen, Q., Choi, C., Lee, Y., and Zohren, S. Evaluating llms in finance requires explicit bias consideration. In *Proceedings of the 43rd International Conference on Machine Learning (ICML), Position Paper Track*, 2026.

Korea Copyright Commission. Copyright counseling casebook, 2025. Issued in Korean. Official guidance casebook.

Korea Creative Content Agency. Guide to fair contracting in the cartoon and webtoon sector, 2023. Issued in Korean.

Korea Creative Content Agency. Standard form contracts and annotations for the cartoon and webtoon sector, 2025. Issued in Korean. Score-eligible official source.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Ministry of Culture, Sports and Tourism (Korea) and Korea Copyright Commission. Standard contracts for assignment and licensing of authors' economic rights, 2018. Issued in Korean.

Ministry of Justice (Korea) and Korea Law-Related Education Center. *Law and Everyday Life for Koreans*. Ministry of Justice (Korea) and Korea Law-Related Education Center, 2021. Korean-language legal education reference; non-scoring context; not quoted.

Pratt, J. W. Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136, 1964.

Qwen Team. Qwen2.5 technical report. Technical report, Alibaba Group, 2024. arXiv:2412.15115.

Robertson, S. and Zaragoza, H. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389, 2009.

Yun, Y., Im, A., Kim, S., and Shin, H. *A Webtoon Creator Gets a Lawyer Friend: What to Know Before Signing a Contract*. Bada Publishing, n.d. Korean-language reference; non-scoring context; not quoted.

Appendix

A. Review-Attention Score Bound

Proposition 1 (bounded score). For each cash-flow category c, the implementation enforces $0 \le g_c(x_c) \le w_c$ and $\sum_{c=1}^{9} w_c = 100$. Therefore,

$$0 \le R = \sum_{c=1}^{9} g_c(x_c) \le \sum_{c=1}^{9} w_c = 100.$$

Scope. The inequality proves boundedness only; it neither calibrates the weights or thresholds nor gives equal financial meaning to equal score differences.

B. Supplementary OCR and Extraction Check

On the frozen synthetic camera/OCR fixture, character and word error rates were 0.053 and 0.125. Exact match for money, percentage, date, duration, and clause segmentation was 1.00. These values describe the fixture rather than general OCR or mobile-camera performance.

C. Evaluation and Release Scope

The manifest excludes development examples from the reported fixtures. Release requires all four gate families below to pass:

Gate	Release evidence
Split discipline	Development examples excluded from reported fixtures.
Score and sources	Bound, scenario, authority-tier, and bilingual checks pass.
Privacy surfaces	No contract text or file paths in logs or exports.
Local execution	No outbound attempt in accepted runtime cases.

The public artifact contains only synthetic or sanitized cases and automated checks; it excludes real contracts, private OCR text, raw commercial-book text, and other private data. Any failed gate blocks release.

D. Project Artifact

The repository contains the source code, frozen fixtures, environment notes, and local-demo instructions. It is available at `https://github.com/seonukkim/fink`.