AI for Climate Finance: Agentic Retrieval and Multi-Step Reasoning for Early Warning System Investments

Anonymous ACL submission

Abstract

Tracking financial investments in climate adaptation is a complex and expertise-intensive task, particularly for Early Warning Systems (EWS), which lack standardized financial reporting across multilateral development banks (MDBs) and funds. To address this challenge, we introduce an LLM-based agentic AI system that integrates contextual retrieval, fine-tuning, and multi-step reasoning to extract relevant financial data, classify investments, and ensure compliance with funding guidelines. Our study focuses on a real-world application: tracking EWS investments in the Climate Risk and Early Warning Systems (CREWS) Fund. We analyze 25 MDB project documents and evaluate multiple AI-driven classification methods, including zero-shot and few-shot learning, fine-tuned transformer-based classifiers, chain-of-thought (CoT) prompting, and an agent-based retrieval-augmented generation (RAG) approach. Our results show that the agent-based RAG approach significantly outperforms other methods, achieving 87% accuracy, 89% precision, and 83% recall. Additionally, we contribute a benchmark dataset and expert-annotated corpus, providing a valuable resource for future research in AI-driven financial tracking and climate finance transparency.[1]

1 Introduction

Recent advances in Large Language Models (LLMs) have transformed investment tracking, financial reporting, and compliance monitoring in climate finance. However, tracking financial flows and categorizing investments in Early Warning Systems (EWS) remains challenging due to the lack of standardized structures and terminologies in financial reports from Multilateral Development Banks (MDBs) and climate funds.

Motivation. Early Warning Systems (EWS) are essential for disaster risk reduction and climate resilience. The United Nations (UN) has prioritized universal EWS access by 2027 through its Early Warnings for All (EW4All) initiative, emphasizing that timely warnings reduce economic losses and save lives. Studies show that 24 hours of advance warning can reduce damages by 30%, while every dollar invested in early warning systems saves up to ten dollars in avoided losses[2]. Despite their importance, EWS investments lack financial transparency, as MDB reports often fail to systematically classify and track funding allocations. This study addresses this gap by developing an AI-driven system to automate investment tracking in the Climate Risk and Early Warning Systems (CREWS) Fund. Traditional NLP methods struggle with the inconsistencies and variability in financial reporting, making manual tracking impractical.

Context. EW4All underscores the need for financial transparency in climate adaptation. However, MDB financial reports lack standardized categorization, contain both structured and unstructured data, and use inconsistent terminology across institutions. Existing NLP models fail to generalize across diverse reporting formats and require extensive labeled data. Addressing these challenges necessitates advanced AI techniques capable of reasoning over heterogeneous financial documents.

Contribution. We introduce the EW4All Financial Tracking AI-Assistant, a system designed to automate EWS investment classification in MDB reports. *a)* It employs multi-modal processing to extract financial information from text, tables, and graphs, improving classification accuracy across diverse document formats. *b)* It handles heterogeneous reporting structures, adapting to inconsistencies in MDB financial disclosures with AI-driven categorization techniques. *c)* It integrates multi-

[1] We will open-source all code, LLM generations, and human annotations.

[2] See Appendix A for more on EWS.

step reasoning and retrieval, leveraging retrieval-augmented generation (RAG) and chain-of-thought (CoT) prompting for enhanced explainability and expert validation.

Our system significantly outperforms existing methods, achieving 87% accuracy, 89% precision, and 83% recall—representing a 23% improvement over traditional NLP approaches. The agent-based RAG method surpasses zero-shot, few-shot, and fine-tuned transformer baselines, demonstrating the effectiveness of AI-driven reasoning for structured financial tracking.

Implications. By improving climate finance transparency, this AI-driven approach provides structured, evidence-based insights into MDB investments. The integration of retrieval-augmented generation and agentic AI enhances decision-making, financial accountability, and policy formulation in global climate investment tracking. This work contributes to broader AI applications in climate finance, supporting international initiatives that seek to optimize resource allocation for climate resilience.

2 Related Literature

RAG improves knowledge-intensive tasks by integrating external retrieval with LLM generation (Lewis et al., 2020), yet traditional RAG remains limited by static retrieval pipelines. Agentic RAG enhances adaptability by incorporating iterative retrieval and decision-making, improving factual accuracy and multi-step reasoning (Xi et al., 2023; Yao et al., 2023; Guo et al., 2024). Multi-agent frameworks extend this by refining retrieval for applications such as code generation and verification (Guo et al., 2024; Liu et al., 2024), advancing explainability and human-AI collaboration.

In-Context Learning (ICL) allows LLMs to generalize from few-shot demonstrations without fine-tuning (Brown et al., 2020), but its effectiveness hinges on example selection. Retrieval-based ICL improves prompt efficiency, and reward models further refine in-context retrieval (Wang et al., 2024). CoT prompting facilitates step-by-step reasoning, significantly boosting performance in arithmetic and commonsense tasks (Wei et al., 2022; Kojima et al., 2022). Self-consistency decoding enhances CoT by aggregating multiple reasoning paths (**?**), while example-based prompting strengthens complex question-answering capabilities (Diao et al., 2024).

3 Methodology

Our methodology comprises four main steps: ❶ PDF parsing and chunking, ❷ context augmentation, ❸ storage and retrieval from a vector database, and ❹ classification and budget allocation using multiple methods. The final and fifth steps ❺ are verification by an expert group and updating the database (see Figure 1).

3.1 PDF Parsing and Chunking

For each PDF document d in our dataset $\mathcal{D}$, we begin by extracting its raw text T_d using the *LlamaParse* (LlamaIndex, 2024) PDF parser:

$$T_d = \text{LlamaParse}(d).$$

Subsequently, the extracted text is partitioned into two distinct types of content:

– **Table Chunks:** Tables within the document are automatically detected and extracted as separate chunks.

– **Text Chunks:** The remaining textual content is segmented based on markdown-style headers. Each resulting text chunk comprises a header (title) and the paragraphs that follow.

The overall set of chunks is defined as $\mathcal{C} = \mathcal{C}_{\text{text}} \cup \mathcal{C}_{\text{table}}$, where $\mathcal{C}_{\text{text}}$ and $\mathcal{C}_{\text{table}}$ denote the sets of text and table chunks, respectively. This separation allows us to treat structured data (tables) and unstructured text differently in subsequent processing stages.

3.2 Context Augmentation

To enhance the context of each chunk, we augment it with a concise summary that situates it within the full document (Anthropic, 2024). Given a chunk $c \in \mathcal{C}$ and the full document text T_d, The prompt $P_{\text{Context}}(c, T_d)$ is used to generate a two-sentence context summary using an LLM, $\text{ctx}(c) = \text{LLM}(P_{\text{Context}}(c, T_d))$. The augmented chunk c' is then formed by concatenating the original chunk with its contextual summary:

$$c' = c \oplus \text{ctx}(c),$$

where $\oplus$ denotes concatenation. This augmentation improves the disambiguation of the content during later retrieval and classification stages.

3.3 Storage and Retrieval in a Vector Database

Each augmented chunk c' is stored in a vector database (vdb) along with relevant metadata, including a unique file identifier f derived from the

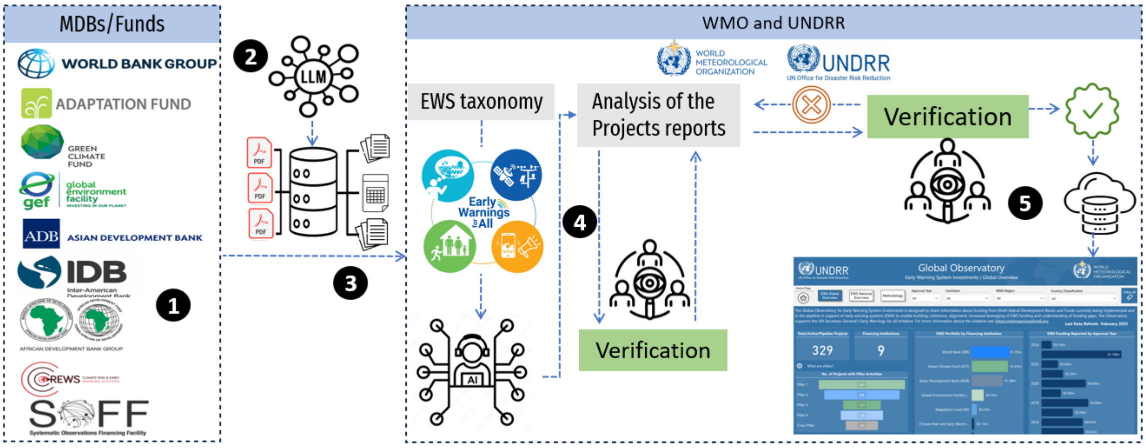

Figure 1: AI-driven financial tracking pipeline for EWS investments, integrating MDB data, LLM-based classification, and expert verification by WMO and UNDRR. The different steps are: ❶ PDF parsing and chunking, ❷ context augmentation, ❸ storage and retrieval from a vector database, ❹ classification and budget allocation, ❺ verification and updating the database.

PDF file name: $\text{meta}(c') = \{\texttt{file\_name} : f\}$. The storage operation is performed as:

$$\text{VDB\_store}(c', \text{meta}(c')).$$

For downstream processing, we query the vdb using a tailored query q in conjunction with the file identifier f to retrieve a fixed number of relevant chunks (specifically, five per document) that are most likely to contain information on pillars and budget allocations:

$$\mathcal{R}(f) = \text{VDB\_query}(q, f) \quad \text{with} \quad |\mathcal{R}(f)| = 5.$$

3.3.1 Hybrid Retrieval via Rank Fusion

In addition to the above procedure, we employ a hybrid search strategy that combines dense vector search with BM25F-based keyword search (Robertson and Zaragoza, 2009) to leverage both semantic similarity and exact lexical matching. Let $\mathcal{R}_v(q, f)$ denote the set of candidate chunks retrieved via dense vector search, and let $\mathcal{R}_k(q, f)$ denote the candidate chunks obtained via BM25F keyword search. To fuse these two retrieval sets, we use Reciprocal Rank Fusion (RRF) (Cormack et al., 2009). For each candidate chunk $c \in \mathcal{R}_v(q, f) \cup \mathcal{R}_k(q, f)$ we compute an RRF score as:

$$\text{RRF}(c) = \sum_{i \in \{v,k\}} \frac{1}{\text{rank}_i(c) + K},$$

where $\text{rank}_i(c)$ is the rank of c in retrieval system i (with lower ranks corresponding to higher relevance) and K is a smoothing constant (typically set to 60). The final set of retrieved chunks is then given by selecting the top five candidates according to their RRF scores:

$$\mathcal{R}(f) = \text{Top5}\Big(\mathcal{R}_v(q, f) \cup \mathcal{R}_k(q, f),\ \text{RRF}(c)\Big).$$

This hybrid method harnesses the semantic sensitivity of dense vector retrieval alongside the precise lexical matching of BM25F, thereby enhancing the overall disambiguation and retrieval performance during downstream processing.

3.4 Classification and Budget Allocation

For each retrieved chunk $c' \in \mathcal{R}(f)$, we apply the following four methods to classify the chunk (i.e., assign it a class y from the five pillars) and to allocate an associated budget B.

3.4.1 Zero-Shot and Few-Shot Classification

In this approach, we construct a prompt $P_{\text{Class+Budget}}(c')$ that includes the content of the augmented chunk and (in the few-shot setting) several annotated examples. The LLM is then queried to simultaneously produce an outcome classification y and an associated budget B:

$$\{y, B\} = \text{LLM}(P_{\text{Class+Budget}}(c')).$$

This method leverages the pre-trained knowledge of the LLM, with few-shot prompting guiding its responses.

3.4.2 Fine-Tuned Transformer-Based Classifier

In another approach, we fine-tune a transformer-based classifier M_{ft} on a labeled dataset $\{(c'_i, y_i)\}_{i=1}^N$. The model is used to classify each augmented chunk $y = M_{\text{ft}}(c')$. Subsequently, an LLM is used to determine the budget allocation of each class. The prompt $P_{\text{Budget}}(c', y)$ is constructed using the the chunk and its classification.

$$B = \text{LLM}(P_{\text{Budget}}(c', y)).$$

The final result for each chunk is the tuple $\{y, B\}$.

3.4.3 Few-Shot-V2: Chain-of-Thought (CoT)

This approach employs a three-step COT strategy, resulting in a tuple $\{y, B\}$:

1 **Reformatting:** If c' represents a table, it is reformatted into a clean markdown table:

$$c'' = \text{LLM}(P_{\text{reformat}}(c')).$$

Otherwise, we set $c'' = c'$.

2 **Classification:** A classification prompt is used to classify the (reformatted) chunk:

$$y = \text{LLM}(P_{\text{Class}}(c'')).$$

3 **Budget Allocation:** A subsequent prompt allocates the budget $B = \text{LLM}(P_{\text{Budget}}(c'', y))$.

3.4.4 Agent-Based Approach

This method uses an agent that follows a sequence of instructions and performs RAG queries:

1. **Instruction Generation:** The agent, primed with examples of annotated PDFs and the desired output format, generates a list of sub-task instructions $I = \{i_1, i_2, \ldots, i_k\}$ to complete the classification and budget allocation task. It also generates a list of queries $Q = \{q_1, q_2, \ldots, q_l\}$ to use if the sub-tasks require querying the vdb.
2. **Sub-Task and Query Mapping:** The agent maps instructions I to queries Q.
3. **Sub-Task Execution:** For each instruction i_j, is the sub-task requires querying the vdb, a retrieval is performed to extract relevant chunks:

$$c'_{i_j} = \text{VDB\_query}(q_{i_j}, f)$$

4. **Sub-Task Validation:** The agent performs a self-healing step to validate that the retrieved chunks c'_{i_j} are sufficient. If not, a new query $q_{i_j}^{\text{new}}$ is generated and the retrieval is repeated:

$$c'_{i_j}{}^{\text{final}} = \begin{cases} \text{VDB\_query}(q_{i_j}^{\text{new}}, f), & \text{if } c'_{i_j} \text{ is insufficient,} \\ c'_{i_j}, & \text{otherwise.} \end{cases}$$

5. **Final Formatting:** After finishing all the sub-tasks, the final step formats the output as JSON:

$$\{y, B\} = \text{LLM}(P_{\text{Format}}(\{\text{result}_I\}))$$

4 Results

We evaluated our methodology on an evaluation set comprising a collection of PDF documents from the CREWS Fund. Our evaluation focuses on how accurately the budget is distributed across the EWS Pillars for each document. To this end, we assess three key metrics: accuracy, precision, and recall. Table 1 summarizes the performance of each method, where the metrics for the agent-based approach are highlighted in bold due to its superior performance.

Method	Accuracy	Precision	Recall
Zero-Shot	0.41	0.40	0.61
Few-Shot	0.42	0.45	0.64
Transformer	0.41	0.64	0.32
Few-Shot-CoT	0.51	0.63	0.71
Agent	**0.87**	**0.89**	**0.83**

Table 1: Evaluation metrics for budget distribution across the EWS Pillars.

The results indicate that the agent-based approach significantly outperforms the other methods, achieving higher accuracy, precision, and recall. This suggests that the integration of retrieval-augmented generation and dynamic sub-task execution in the agent method greatly enhances the effectiveness of budget allocation across the pillars.

5 Conclusion

Automating financial tracking of EWS investments is crucial for improving climate finance transparency and accountability. In this study, we introduced the EWS4All Financial Tracking AI-Assistant, a novel system that integrates multi-modal processing, hierarchical reasoning, and RAG for document classification and budget allocation. Our experiments on 25 project documents from the CREWS Fund demonstrated that an agent-based approach significantly outperforms traditional NLP methods, achieving 87% accuracy, 89% precision, and 83% recall. The system effectively addresses challenges related to document heterogeneity, structured and unstructured data integration, and cross-organizational inconsistencies. Beyond improving financial tracking, our work contributes a benchmark dataset for future AI research in climate finance. By combining AI-driven classification, retrieval, and reasoning, this approach enhances decision-making processes in MDBs and supports evidence-based climate investment policies. Future work will focus on extending the system to handle a broader range of MDB financial documents, improving model generalization, and integrating real-time updates for dynamic financial tracking.

Limitations

While our approach demonstrates significant improvements in automating financial tracking for EWS investments, several limitations remain. First, our system relies on existing financial reports from MDBs, in this case CREWS, which are often heterogeneous and may contain incomplete or ambiguous financial allocations. In cases where funding details are missing or inconsistently reported, even advanced retrieval-augmented generation (RAG) and multi-step reasoning approaches may struggle to provide accurate classifications. Second, the classification system is influenced by the training data used in fine-tuning and prompt engineering. Despite expert annotations, the model may still exhibit biases in investment classification, particularly when encountering novel financial structures or terminology not well-represented in the dataset. Third, while our agent-based RAG system achieves state-of-the-art performance on structured and unstructured financial data, its generalizability to other climate finance applications outside EWS has not been fully explored. Future work should assess model robustness across different sustainability reporting frameworks and financial instruments. Finally, our system assumes that financial tracking can be improved through AI-assisted reasoning; however, its real-world effectiveness depends on institutional adoption, policy integration, and alignment with evolving financial disclosure regulations.

Ethics Statement

Human Annotation: This study relies on annotations provided by domain experts from the World Meteorological Organization (WMO), who possess extensive knowledge of Early Warning Systems (EWS). These experts played a pivotal role in the design and conceptualization of the study. Their deep understanding of both the contextual and practical aspects of the collected data ensures the accuracy and relevance of the annotations. The use of expert annotations minimizes the risk of misclassification and enhances the reliability of the model's outputs.

Responsible AI Use. This tool is intended as an assistive system to enhance transparency and efficiency in financial tracking, not as a replacement for human analysts. Expert oversight remains crucial in interpreting financial classifications, addressing edge cases, and ensuring compliance with policy frameworks. By open-sourcing our dataset and model, we encourage responsible use and further validation to refine the system's applicability in real-world climate finance decision-making.

Data Privacy and Bias: This study does not involve any personally identifiable or sensitive financial data. All data used in this research originates from publicly available sources under a Creative Commons license, ensuring compliance with data privacy regulations. While we find no evidence of demographic biases in the dataset, we acknowledge that financial reporting by multilateral development banks (MDBs) may reflect institutional biases in investment classification. Our model operates as a decision-support tool and should not replace human judgment in financial tracking and policy decisions.

Reproducibility Statement: To ensure full reproducibility, we will release all PDFs, codes, EWS-taxonomy, and expert-annotated data used in this study. Our approach aligns with best practices in AI transparency and responsible research dissemination. However, we encourage users of this dataset and model to consider ethical implications when applying automated financial tracking systems in real-world decision-making contexts. For vector database storage and retrieval, we utilized Weaviate, an open-source, scalable vector search engine that efficiently indexes high-dimensional embeddings. Additionally, for reasoning and large language model (LLM) interactions, we integrated OpenAI's o1 API, leveraging its advanced capabilities to process, analyze, and infer patterns from financial document data.

References

Anthropic. 2024. Introducing contextual retrieval. https://www.anthropic.com/news/contextual-retrieval.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:1877–1901.

Gordon V. Cormack, Charles L.A. Clarke, and Stephan Buettcher. 2009. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 758–759. ACM.

Shizhe Diao, Pengcheng Wang, Yong Lin, Rui Pan, Xiang Liu, and Tong Zhang. 2024. Active prompting with chain-of-thought for large language models.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large language model based multi-agents: A survey of progress and challenges.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Junwei Liu, Kaixin Wang, Yixuan Chen, Xin Peng, Zhenpeng Chen, Lingming Zhang, and Yiling Lou. 2024. Large language model-based agents for software engineering: A survey.

LlamaIndex. 2024. Llama parse: A genai-native document parsing api. Accessed via LlamaIndex Cloud API.

Gianluca Pescaroli, Sarah Dryhurst, and Georgios Marios Karagiannis. 2025. Bridging gaps in research and practice for early warning systems: new datasets for public response. *Frontiers in Communication*, 10:1451800.

Stephen Robertson and Hugo Zaragoza. 2009. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389.

Andrew C Tupper and Carina J Fearnley. 2023. Mind the gaps in disaster early-warning systems—and fix them. *Nature*, 623:479.

Omar Velazquez, Gianluca Pescaroli, Gemma Cremen, and Carmine Galasso. 2020. A review of the technical and socio-organizational components of earthquake early warning systems. *Frontiers in Earth Science*, 8:533498.

Jie Wang, Alexandros Karatzoglou, Ioannis Arapakis, and Joemon M Jose. 2024. Reinforcement learning-based recommender systems with large language models for state reward and action modeling. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 375–385.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. 2023. The rise and potential of large language model based agents: A survey.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*.

A Early Warning Systems (EWS)

A.1 Definition and Purpose

Early Warning Systems (EWS) are integrated frameworks designed to detect imminent hazards and alert authorities and communities before disasters strike. In essence, an EWS combines hazard monitoring, risk analysis, communication, and preparedness planning to enable timely, preventive actions. Early warnings are a cornerstone of disaster risk reduction (DRR) – they save lives and reduce economic losses by giving people time to evacuate, protect assets, and secure critical infrastructure[3]. By empowering those at risk to act ahead of a hazard, EWS help build climate resilience: they are proven to safeguard lives, livelihoods, and ecosystems amid increasing climate-related threats[4]. In summary, an effective EWS ensures that impending dangers are rapidly identified, warnings reach the impacted population, and appropriate protective measures are taken in advance.

A.2 EWS Taxonomy

A robust EWS involves several fundamental components that work together seamlessly. The United Nations identify four interrelated pillars necessary for an effective people-centered EWS (Pescaroli et al., 2025). This taxonomy serves as a structured framework to categorize EWS components and activities, facilitating a consistent approach to analyzing early warning systems across various domains. Our approach in this paper is based on these four fundamental pillars of EWS and one

[3] See `https://www.unisdr.org/files/608_10340.pdf`.

[4] See, `https://www.unep.org/topics/climate-action/climate-transparency/climate-information-and-early-warning-systems`.

cross-pillar, ensuring a comprehensive understanding of risk knowledge, detection, communication, and preparedness.

Early Warning System (EWS) Taxonomy Prompt

An Early Warning System (EWS) is an integrated system of hazard monitoring, forecasting, and prediction, disaster risk assessment, communication, and preparedness activities that enables individuals, communities, governments, businesses, and others to take timely action to reduce disaster risks before hazardous events occur.
When analyzing a text, it is essential to determine whether it falls under EWS components and activities, which vary across multiple sectors and require coordination and financing from various actors.
The taxonomy is based on the Four Pillars of Early Warning Systems and one cross-pillar:

Pillar 1: Disaster Risk Knowledge and Management (Led by UNDRR)

This pillar focuses on understanding disaster risks and enhancing the knowledge of communities by collecting and utilizing comprehensive information on hazards, exposure, vulnerability, and capacity.
Illustrative examples:
– Inclusive risk knowledge: Incorporating local, traditional, and scientific risk knowledge.
– Production of risk knowledge: Establishing a systematic recording of disaster loss data.
– Risk-informed planning: Ensuring decision-makers can access and use updated risk information.
– Data rescue: Digitizing and preserving historical disaster data.
Keywords: Risk mapping, vulnerability mapping, disaster risk reduction (DRR), climate information.

Pillar 2: Detection, Observation, Monitoring, Analysis, and Forecasting (Led by WMO)

This pillar enhances the capability to detect and monitor hazards, providing timely and accurate forecasting.
Illustrative examples:
– Observing networks enhancement: Strengthening real-time monitoring systems.
– Hazard-specific observations: Improving monitoring of high-impact hazards.
– Impact-based forecasting: Developing quantitative triggers for anticipatory action.
Keywords: Forecasting, seasonal predictions, multi-model projections, climate services.

Pillar 3: Warning Dissemination and Communication (Led by ITU)

Effective communication ensures that early warnings are received by those at risk, enabling them to take timely action.
Illustrative examples:
– Multichannel alert systems: Use of SMS, satellite, sirens, and social media.
– Standardized warnings: Implementation of the Common Alerting Protocol (CAP).
– Feedback mechanisms: Enabling community input on warning effectiveness.
Keywords: Communication systems, multichannel dissemination, emergency broadcast systems.

Pillar 4: Preparedness and Response Capabilities (Led by IFRC)

Timely preparedness and response measures translate early warnings into life-saving actions.
Illustrative examples:
– Emergency preparedness planning: Developing anticipatory action frameworks.
– Public awareness campaigns: Educating communities on disaster response.
– Emergency shelters: Construction of cyclone shelters, evacuation centers.
Keywords: Preparedness planning, emer-

gency drills, public education on disaster response.

Cross-Pillar: Foundational Elements for Effective EWS

Cross-cutting elements critical to the sustainability and effectiveness of EWS include governance, inclusion, institutional arrangements, and financial planning.
Illustrative examples:

- Governance and institutional frameworks: Defining roles of agencies and stakeholders.
- Financial sustainability: Mobilizing and tracking finance for early warning systems.
- Regulatory support: Developing and enforcing data-sharing legislation.

Keywords: Institutional frameworks, governance, financial sustainability, data management.

Each of these components is vital. Only when risk knowledge, monitoring, communication, and preparedness work in unison can an early warning system effectively protect lives and properties. Gaps in any one element (for example, if warnings don't reach the vulnerable, or if communities don't know how to respond) will weaken the whole system. Thus, successful EWS are people-centered and end-to-end, linking high-tech hazard detection with on-the-ground community action.

A.3 Importance for climate finance

EWS are widely recognized as a high-impact, cost-effective investment for climate resilience. By providing advance notice of floods, storms, heatwaves and other climate-related hazards, EWS significantly reduce disaster losses. Studies indicate that every $1 spent on early warnings can save up to $10 by preventing damages and losses.[5] For example, just 24 hours' warning of an extreme event can cut ensuing damage by about 30%, and an estimated USD $800 million investment in early warning infrastructure in developing countries could avert $3–16 billion in losses every year[6]. These economic benefits underscore why EWS are considered "no-regret" adaptation measures, i.e., they pay for themselves many times over by protecting lives, assets, and development gains.

Given their proven value, EWS have become a priority in climate change adaptation and disaster risk reduction funding. International climate finance mechanisms, such as the Green Climate Fund, Climate Risk and Early Warning Systems (CREWS) Fund, and Adaptation Fund along with development banks, are channeling resources into EWS projects, from modernizing meteorological services and hazard monitoring networks to community training and alert communication systems. Strengthening EWS is also central to global initiatives like the United Nations' Early Warnings for All (EW4All), which calls for expanding early warning coverage to 100% of the global population by 2027. Achieving this goal requires substantial financial support to build new warning systems in climate-vulnerable countries and to maintain and upgrade existing ones. Climate finance is therefore being directed to help develop, implement, and sustain EWS, ensuring that countries can operate these systems (e.g. funding for equipment, data systems, and personnel) over the long term. In summary, investing in EWS is essential for climate resilience. It not only reduces humanitarian and economic impacts from extreme weather, but also yields high returns on investment. Financial support for EWS, whether through dedicated climate funds, loans and grants, or public budgets, underpins their development and sustainability, making it possible to deploy cutting-edge technology and foster prepared communities. By mitigating the worst effects of climate disasters, EWS help safeguard development progress, which is why they feature prominently in climate adaptation financing and strategies.

Hence, investing in EWS is essential for climate resilience. It not only reduces humanitarian and economic impacts from extreme weather, but also yields high returns on investment. Financial support for EWS, whether through dedicated climate funds, loans and grants, or public budgets, underpins their development and sustainability, making it possible to deploy cutting-edge technology and foster prepared communities. By mitigating the worst effects of climate disasters, EWS help safeguard development progress, which is why they feature prominently in climate adaptation financing and strategies.

---

[5] See, https://wmo.int/news/media-centre/early-warnings-all-advances-new-challenges-emerge.

[6] See, https://www.unep.org/topics/climate-action/climate-transparency/climate-information-and-early-warning-systems.

A.4 Current challenges

Despite their clear benefits, there are several challenges in financing and implementing EWS effectively. Key issues include:

Data Inconsistencies and Lack of Standardization: EWS rely on data from multiple sources (weather observations, risk databases, etc.), but often this data is inconsistent, incomplete, or not shared effectively across systems. Differences in how hazards are monitored and reported can lead to gaps or delays in warnings. Likewise, there is a lack of standardization in early warning protocols and data formats between agencies and countries (Velazquez et al., 2020; Pescaroli et al., 2025). Incompatible data systems and inconsistent methodologies (for example, different trigger criteria for warnings or varying risk assessment methods) make it difficult to integrate information. This fragmentation hinders the creation of a "common operating picture" of risk. Data harmonization and common standards (for data collection, forecasting models, and warning communication) are needed to ensure EWS components work together seamlessly.

Institutional and Cross-Organizational Barriers: An effective EWS cuts across many organizations, national meteorological services, disaster management agencies, local governments, international partners, and communities. Coordinating these actors remains a challenge. In many cases, efforts are siloed: meteorological offices may issue technical warnings that don't fully reach or engage local authorities or the public. There are gaps in governance, clarity of roles, and inter-agency communication that can weaken the warning chain. Improving EWS often requires overcoming bureaucratic boundaries and fostering cooperation between different sectors (e.g., linking climate scientists with emergency planners). Interoperability issues, i.e.,ensuring different organizations' technologies and procedures align, are also a hurdle (Tupper and Fearnley, 2023). As the World Meteorological Organization (WMO) states, connecting all relevant actors (from international agencies down to community groups) and adapting plans to real-world local conditions is complex[7]. Sustained commitment, clear protocols, and partnerships are required to break down these barriers so that EWS operate as a cohesive, cross-sector system.

Financing Gaps and Sustainability: While funding for EWS is rising, it still lags behind what is needed for global coverage and maintenance. Many high-risk developing countries lack the resources to install or upgrade EWS infrastructure (radar, sensors, communication tools) and to train personnel. Fragmented financing is a problem. Support comes from various donors and programs without a unified strategy, leading to potential overlaps in some areas and stark gaps in others. For instance, recent analyses show that a large share of EWS funding is concentrated in a few countries, while Small Island Developing States (SIDS) and Least Developed Countries (LDCs) remain underfunded despite being highly vulnerable[8]. Even when initial capital is provided to set up an EWS, securing long-term funding for operations and maintenance (software updates, staffing, equipment calibration) is difficult. Without sustainable financing, systems can degrade over time. Ensuring financial sustainability, co-financing arrangements, and political commitment is critical so that EWS are not one-off projects but enduring services.

In addition to the above, there are challenges in technological adoption and last-mile delivery: for example, reaching remote or marginalized populations with warnings (issues of language, literacy, and reliable communication channels) and building trust so that people heed warnings. Climate change is also introducing new complexities – hazards are becoming more unpredictable or intense, testing the limits of existing early warning capabilities. Overall, addressing data and standardization issues, improving institutional coordination, and closing funding gaps are priority challenges to fully realize the life-saving potential of EWS.

A.5 Relevance to this study

Our work is focused on the financial tracking and classification of investments in climate resilience, and EWS represent a prime example of such investments. Early warning projects often cut across sectors and funding sources – they might include components of infrastructure, technology, capacity building, and community outreach. Because of this cross-cutting nature, tracking where and how money is spent on EWS can be difficult without a clear classification system. Different organizations may label EWS-related activities in various ways (e.g. "hydromet modernization", "dis-

[7] See, https://wmo.int/news/media-centre/early-warnings-all-advances-new-challenges-emerge.

[8] See, https://wmo.int/media/news/tracking-funding-life-saving-early-warning-systems.

aster preparedness", "climate services"), leading to inconsistencies in investment data. By establishing a standardized framework to define and categorize EWS investments, the study helps create a "big-picture view" of early warning financing. This enables analysts and policymakers to identify overlaps, gaps, and trends that were previously obscured by fragmented data.

Moreover, improving the classification of EWS funding directly supports broader resilience initiatives. For instance, the newly launched Global Observatory for Early Warning System Investments is already working to tag and track EWS-related expenditures across major financial institutions. Such efforts mirror the goals of this study by highlighting the need for consistent tracking, transparency, and coordination in climate resilience finance. Better classification of investments means stakeholders can pinpoint where resources are going and where additional support is needed to meet global targets like the "Early Warnings for All by 2027" pledge. In short, EWS feature in this study as a critical category of climate resilience investment that must be clearly identified and monitored.

By including EWS in its financial tracking framework, the study provides valuable insights for decision-makers. It helps determine how much funding is allocated to early warnings, from which sources, and for what components (equipment, training, maintenance, etc.). This information is crucial for evidence-based decisions on scaling up EWS: for example, spotting a shortfall in community-level preparedness funding, or recognizing successful investment patterns that could be replicated. Ultimately, linking EWS to the study's financial tracking reinforces the message that climate resilience investments can be better managed when we know their size, scope, and impact area. By classifying EWS expenditures systematically, the study contributes to stronger accountability and strategic planning in building climate resilience, ensuring that early warning systems – and the communities they protect – get the support they urgently need.

B Dataset Construction

In this study, we analyze financial information extracted from PDFs containing both structured and unstructured data. Unlike conventional benchmark datasets, these documents exhibit high heterogeneity in their formats—some tables are well-structured, while others embed financial figures within free-text paragraphs or are scattered across multiple rows and columns. Additionally, many numerical values correspond to multiple rows within the same column, creating challenges in extraction, alignment, and interpretation.

The annotated data, provided by experts in CSV format, along with the corresponding PDFs, can be found in the supplementary materials of this paper.

The dataset consists of 298 rows of expert annotations and contains the following 9 columns: *Fund, Project ID, Component, Outcome/Expected-Outcome/Objectives, Output/Sub-component, Activity/Output Indicator, Page Number, Amount,* and *Label*.

The total amount of Early Warning Systems (EWS) is computed as the sum of all *Amount* values for a given project.

The annotated dataset (CSV file and PDFs) consists of financial reports and investment documents sourced from publicly available institutional records, which are intended for public information and research and transparency purposes. The dataset is used strictly within its intended scope—analyzing financial tracking in climate investments—and adheres to the original access conditions. Additionally, for the artifacts we create, including benchmark datasets and classification models, we specify their intended use for research and evaluation in automated financial tracking and ensure they remain compliant with ethical research guidelines.