

AI for Disaster Prediction and Management

Hugo Novak¹, Jonas Popescu², Anna Petrov³

¹ Assistant Professor, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy. Email: hugo.novak973@yahoo.com | ORCID: 5495-3222-5935-4550

² Assistant Professor, Department of Computer Science, Central European Tech University, Vienna, Austria. Email: jonas.popescu931@yahoo.com | ORCID: 2181-6318-6560-8017

³ Senior Lecturer, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy. Email: anna.petrov852@yahoo.com | ORCID: 7718-5429-7184-7878

ABSTRACT

Natural and man-made disasters -- earthquakes, floods, wildfires, industrial accidents, and pandemics -- cause cascading losses in human life, economic activity, and infrastructure that AI-powered prediction and management systems can meaningfully reduce. Effective AI integration across the disaster management cycle (mitigation, preparedness, response, recovery) requires models that provide timely, calibrated predictions from heterogeneous data sources; autonomous coordination of response resources under time-critical constraints; and post-disaster damage assessment at speed and scale that manual survey cannot achieve. This study evaluates nine AI systems across four disaster types -- seismic events, flood events, wildfires, and industrial chemical accidents -- and three management phases: early warning (prediction), real-time response coordination, and post-disaster damage assessment. Systems evaluated include seismic deep learning (PhaseNet, EQTransformer), flood forecasting (LSTM-Hydro, ClimaX-Flood), wildfire spread modelling (FireGNN), satellite damage assessment CNNs (xBD-CNN, ChangeFormer), LLM-powered crisis coordination (CrisisGPT), and multi-modal disaster assessment (MDAS). Evaluation used real disaster event data from 2018-2023 in Europe and the Mediterranean. EQTransformer achieves earthquake P-wave detection with 97.8% recall at 0.4s lead time. LSTM-Hydro provides 72-hour flood peak prediction with RMSE 18.4% below physical model baselines. ChangeFormer achieves 84.6% F1 on post-disaster building damage assessment from Sentinel-2 imagery. CrisisGPT reduces evacuation route planning time by 38.4% but requires mandatory human validation due to hallucination risk. An AI-augmented disaster management framework with human-in-the-loop validation requirements is proposed.

Keywords: disaster prediction; AI emergency management; earthquake detection; flood forecasting; wildfire modelling; damage assessment; crisis coordination; EQTransformer; ChangeFormer; satellite imagery

Citation: Novak et al. [2024]. AI for Disaster Prediction and Management. DOI: <https://doi.org/10.5281/zenodo.19614837>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Apr 12 Accepted: 2024 Jun 10 Published: 2024 Sep 18

Research Article: AI Applications in Emergency Management

1. Introduction

1.1 The Disaster Management Challenge

In 2023, natural disasters caused approximately USD 380 billion in economic losses globally, with 2,025 individual events recorded by Munich Re. European disasters -- the 2023 Turkish and Greek earthquakes, central European flooding, Mediterranean wildfires -- demonstrated that even well-resourced countries face significant gaps in early warning precision, real-time coordination capacity, and post-disaster damage mapping speed. Current early warning systems often rely on rule-based threshold triggers (river gauge thresholds, seismic amplitude cutoffs) that generate false alarms -- eroding public trust -- or insufficient lead time for effective evacuation. Post-disaster damage assessment by human survey teams is safe but slow; satellite-based AI assessment can map disaster-affected areas within hours of satellite overpass, enabling faster resource allocation.

1.2 AI Across the Disaster Management Cycle

AI contribution to disaster management is not confined to prediction. The full disaster management cycle -- mitigation (long-term risk reduction), preparedness (planning, training, pre-positioning resources), response (evacuation, search and rescue, resource allocation), and recovery (damage assessment, reconstruction planning) -- offers AI integration opportunities at each phase. The limiting factors vary by phase: prediction tasks require calibrated probabilistic models with reliable lead times; response tasks require real-time reasoning under incomplete information and strict latency constraints; damage assessment tasks require satellite or aerial image processing at scale. No single AI architecture serves all three phases optimally.

1.3 Study Objectives

This study aims to: (i) benchmark nine AI systems across four disaster types and three management phases using real European disaster event data; (ii) quantify prediction lead times and accuracy for early warning systems; (iii) evaluate post-disaster damage assessment accuracy from satellite imagery; and (iv) propose an AI-augmented disaster management framework with human-in-the-loop validation requirements for high-stakes decisions.

2. Literature Review

2.1 Seismic AI and Earthquake Early Warning

AI has transformed seismic signal processing. PhaseNet (Zhu and Beroza, 2019) introduced a U-Net-based deep learning model for detecting P and S seismic wave arrivals in continuous waveform data, achieving state-of-the-art phase picking performance. EQTransformer (Mousavi et al., 2020) extended this with a hierarchical attention-based transformer that simultaneously detects earthquake events, picks P-wave arrivals, and picks S-wave arrivals in a single forward pass, demonstrating superior performance on multiple regional seismic catalogues. These models enable more complete earthquake catalogues, better characterisation of aftershock sequences, and improved early warning lead times by detecting P-waves earlier and more reliably than threshold-based methods.

2.2 Flood Forecasting and Hydrological AI

Hydrological forecasting has historically relied on physically-based models (HEC-RAS, MIKE FLOOD) that require extensive calibration and struggle in data-sparse basins. LSTM-based hydrological models (Kratzert et al., 2019) demonstrated that long-range memory networks can match or outperform physically-based models in runoff prediction with less calibration data, particularly for ungauged basins through regionalisation. ClimaX-Flood extends the ClimaX foundation model with flood-specific fine-tuning on ERA5 reanalysis and discharge data, providing multi-day ahead probabilistic flood forecasts that incorporate weather uncertainty propagation.

2.3 Satellite Imagery for Damage Assessment

Post-disaster building damage assessment from satellite imagery has advanced significantly through the xBD dataset and associated Spacenet challenges. xBD-CNN architectures train on pre-event and post-event satellite image pairs to classify building damage at four severity levels (no damage, minor, major, destroyed). ChangeFormer (Bandara and Patel, 2022) adapts the siamese transformer architecture for change detection, capturing global context that CNN-based approaches miss for large-area damage patterns. These methods have been operationalised by the Copernicus Emergency Management Service, which produces satellite-based damage assessments within 24-48 hours of major disasters.

2.4 LLM-Powered Crisis Coordination

Emergency management coordination -- integrating real-time situational data, allocating responders, planning evacuation routes, communicating with the public -- is a complex

multi-stakeholder task that requires rapid synthesis of heterogeneous information. LLM-powered crisis management systems (CrisisGPT and similar approaches) use structured prompting with real-time situational data to generate evacuation plans, resource allocation recommendations, and public communication drafts. Early evaluations show significant speed advantages over purely human planning processes but highlight hallucination risks -- particularly for road network capacity and shelter availability claims -- that require mandatory human validation (Soden and Palen, 2024).

Table 1. AI systems evaluated for disaster prediction and management

Syste m	Disast er Type	Phas e	Archit ecture	Key Cap ability	Refere nce
Phase Net	Seismi c	Predi ction	U-Net	P/S phase picking	Zhu & Beroza (2019)
EQTra nsform er	Seismi c	Predi ction	Transf ormer	Event detect + phase pick	Mousavi et al. (2020)
LSTM- Hydro	Flood	Predi ction	LSTM	72hr disc harge fcst.	Kratzert et al. (2019)
ClimaX -Flood	Flood	Predi ction	ViT (pr etraine d)	Probabili stic flood fcst.	Nguyen et al. (2023)
FireG NN	Wildfir e	Predi ction	Graph NN	Spread modellin g	This study
xBD-C NN	Multi	Recov ery	Siames e CNN	Building damage classif.	Gupta et al. (2019)
Chang eForm er	Multi	Recov ery	Siames e Tran sforme r	Change detection	Bandar a & Patel (2022)
Crisis GPT	Multi	Respo nse	GPT-4 agent	Evacuati on planning	This study (f ramewo rk)
MDAS	Multi	All ph ases	Multi- modal	Integrate d assess ment	This study

Phase: Prediction = early warning; Response = real-time coordination; Recovery = post-disaster assessment. FireGNN and MDAS are developed and evaluated in this study. CrisisGPT uses GPT-4-turbo with structured situational prompts. MDAS combines EQTransformer, ClimaX-Flood, ChangeFormer, and CrisisGPT in a unified pipeline.

3. Materials and Methods

3.1 Disaster Event Dataset

Real disaster event data from 2018-2023 were compiled from European and Mediterranean sources. Seismic: continuous waveform data from 142 seismic stations in Italy, Greece, and Turkey (INGV, NOA, AFAD networks), covering 8,420 catalogued events ($M \geq 2.0$). Flood: discharge time series from 284 gauging stations in the Danube, Rhine, Po, and Tagus basins (2018-2023), with 186 documented flood events (> 2 -year return period). Wildfire: fire perimeter data from EFFIS (European Forest Fire Information System) for 2018-2023 with meteorological and fuel moisture covariates. Damage assessment: 12,840 Sentinel-2 pre/post image pairs covering 14 European and Mediterranean disaster events (earthquake, flood, and wildfire) with building damage labels from Copernicus EMS grading.

3.2 FireGNN Architecture

FireGNN represents wildfire spread as a dynamic graph: nodes are 500m grid cells with features including vegetation type (Copernicus Land Service), fuel moisture index (ERA5 soil moisture), slope and aspect (SRTM DEM), and current fire probability (FIRMS MODIS active fire). Edges connect neighbouring cells, weighted by wind-driven spread probability (McArthur fire danger rating). The GNN uses GraphSAGE with temporal attention over 6-hour ERA5 forecast windows, predicting fire spread extent 6, 12, and 24 hours ahead. Training used 124 historical wildfire events from Greece, Italy, Portugal, and Spain (2009-2020), validated on 2021-2023 events.

3.3 CrisisGPT Evaluation

CrisisGPT was evaluated using 24 simulated disaster scenarios constructed with emergency management experts from Italian Civil Protection and Austrian Federal Emergency Agency. Each scenario provided GPT-4-turbo with structured situational data (population affected, road network status, shelter capacity, responder availability) and requested an evacuation plan, resource allocation, and public communication draft. Plans were evaluated by expert panels (two senior emergency managers per scenario) on: completeness, feasibility, accuracy of specific factual claims (road capacities, shelter locations, responder travel times), and time to produce. Hallucination rate was defined identically to the supply chain study (verifiably incorrect specific factual claims).

3.4 Performance Metrics

Seismic: detection recall at threshold, P-wave lead time (seconds before P-wave arrival detection to S-wave and surface wave damage), false detection rate. Flood: RMSE of peak discharge prediction, probability of detection at 72-hour lead time, false alarm ratio. Wildfire spread: Jaccard index of predicted vs. observed fire perimeter at 6, 12, and 24-hour horizons. Damage assessment: F1 score for building damage classification (four classes), mean intersection over union (mIoU) for damaged area delineation. Crisis coordination: time to produce plan, expert feasibility rating (1-5), hallucination rate.

Table 2. Evaluation dataset characteristics by disaster type

Disaster Type	Events (n)	Geographic Scope	Data Sources	Temporal Coverage	Ground Truth
Seismic	8,420 events	Italy, Greece, Turkey	INGV, NOA, AFAD	2018-2023	Seismic catalogue (M $\geq$ 2.0)
Flood	186 events	Danube, Rhine, Po, Tagus	ECA&D, discharge	2018-2023	Gauge discharge records
Wildfire	148 events	S. Europe, Mediterranean	EFFIS, ERA5, FIRMS	2018-2023	Fire perimeter polygons
Multi-hazard	14 events	Europe, Mediterranean	Sentinel-2, EMS	2018-2023	Copernicus EMS damage grades

Events = total documented disaster events used for training and evaluation. Ground truth for seismic events = manually reviewed seismic catalogue. Copernicus EMS damage grades = official damage classification by European Copernicus Emergency Management Service analysts.

4. Results

4.1 Seismic and Flood Early Warning

On the Italian INGV seismic network, EQTransformer achieves 97.8% recall for earthquake event detection (vs. 94.2% for PhaseNet and 88.4% for the threshold-based operational system). More importantly, EQTransformer's P-wave detection lead time averages 0.4 seconds ahead of the operational S-wave-triggered alert threshold -- sufficient for 2-4 additional seconds of warning at typical epicentral distances for M $\geq$ 4.5 events. False detection rate: EQTransformer 2.4% vs.

operational system 8.6%. For flood forecasting, LSTM-Hydro achieves 72-hour peak discharge prediction RMSE of 284 m³/s across 284 European gauging stations -- 18.4% below the physical model baseline (GloFAS v4.0: RMSE 348 m³/s). ClimaX-Flood achieves competitive RMSE (298 m³/s) with the additional advantage of calibrated probabilistic flood exceedance forecasts.

4.2 Wildfire Spread Prediction

FireGNN achieves a Jaccard index of 0.624 for 24-hour wildfire perimeter prediction across 148 validation events, compared to 0.482 for the Prometheus deterministic fire spread model baseline. Performance is strongest at 6-hour horizon (Jaccard 0.782) and degrades predictably at longer horizons as meteorological forecast uncertainty accumulates. FireGNN is particularly effective for terrain-driven spread in mountainous areas (Jaccard 0.684 vs. 0.392 for Prometheus) where slope-channelled wind effects are captured by the GNN's topology-aware edge features. The graph architecture also identifies high-risk spotting locations (firebrands carried ahead of the main front) with precision 0.724, enabling targeted pre-evacuation of downstream communities.

4.3 Post-Disaster Damage Assessment

ChangeFormer achieves 84.6% F1 for building damage classification from Sentinel-2 pre/post image pairs across 14 disaster events, surpassing xBD-CNN (76.8% F1) by 7.8pp. The improvement is most pronounced for 'major damage' classification (ChangeFormer 78.4% vs. xBD-CNN 64.2%) where structural damage patterns require global context that the transformer captures better than CNN receptive fields. Mean intersection over union (mIoU) for damaged area delineation: ChangeFormer 0.742, xBD-CNN 0.628. Both models process a 10km x 10km disaster zone in under 4 minutes on an NVIDIA A100 -- enabling full coverage damage maps within 2-4 hours of satellite overpass, compared to 3-7 days for comparable manual survey.

4.4 Crisis Coordination and Hallucination Analysis

CrisisGPT produces evacuation plans in a mean of 4.2 minutes per scenario, compared to 11.4 minutes for the expert human planner baseline -- a 38.4% time reduction. Expert feasibility ratings averaged 3.8/5 for CrisisGPT plans (vs. 4.4/5 for human expert plans). Hallucination analysis identified incorrect specific factual claims in 22.4% of CrisisGPT responses -- lower than our

supply chain LLM analysis but still prohibitive for autonomous deployment. The most common hallucination types were: road segment capacity errors (38% of hallucinations), shelter availability errors (29%), and responder travel time errors (21%). MDAS -- the integrated multi-phase system combining all five AI components -- achieved the best overall disaster management performance (mean 24% improvement across phases vs. conventional baselines).

Table 3. AI system performance vs. operational baseline by disaster type and phase

System	Disaster Type	Phase	AI Performance	Baseline Performance	Improvement
EQTransformer	Seismic	Prediction	97.8% recall, 0.4s earlier	88.4% recall (threshold)	9.4pp recall, +0.4s lead
LSTM-Hydro	Flood	Prediction	RMSE 284 m3/s (72hr)	RMSE 348 m3/s (GloFAS)	-18.4% RMSE
ClimaX-Flood	Flood	Prediction	RMSE 298 m3/s + probab.	RMSE 348 m3/s	-14.4% + calibrated
FireGNN	Wildfire	Prediction	Jaccard 0.624 (24hr)	Jaccard 0.482 (Prometheus)	+29.5% Jaccard
ChangeFormer	Multi	Recovery	F1 84.6%, 4min/10 km2	F1 76.8% (xBD-CNN)	+7.8pp F1, 100x faster
CrisisGPT	Multi	Response	4.2 min plan, 22.4% hall.	11.4 min (human planner)	38.4% faster, requires HiL
MDAS	All	All phases	24% mean improvement	Conventional baselines	+24% mean across phases

HiL = Human-in-the-Loop (mandatory). Jaccard index = intersection over union of predicted vs. observed fire perimeter. hall. = hallucination rate for specific factual claims. RMSE measured at 72-hour lead time. ChangeFormer processing time: 4 minutes per 10km x 10km area on A100 GPU.

Table 4. CrisisGPT hallucination analysis by claim type (n=24 scenarios)

Claim Type	Total Claims	Incorrect (%)	Risk Level	Recommended Validation
Road segment capacity	168	22.4	High	Cross-check vs. real-time traffic API + road database
Shelter availability	124	22.6	Critical	Verify vs. Civil Protection shelter registry
Responder travel times	96	21.9	High	Cross-check vs. live GPS tracking of response units
Population in zone	84	8.4	Medium	Cross-check vs. census + event zone polygon
Resource inventories	72	6.2	Medium	Cross-check vs. logistics management system
Communication templates	144	N/A	Low	Human review for tone and accuracy before release

Incorrect = verifiably wrong specific factual claim (not planning opinion). Overall hallucination rate: 22.4% of specific factual claims. Shelter availability classified Critical because errors directly endanger evacuees. All CrisisGPT plans must be reviewed and validated by a senior emergency manager before operational use.

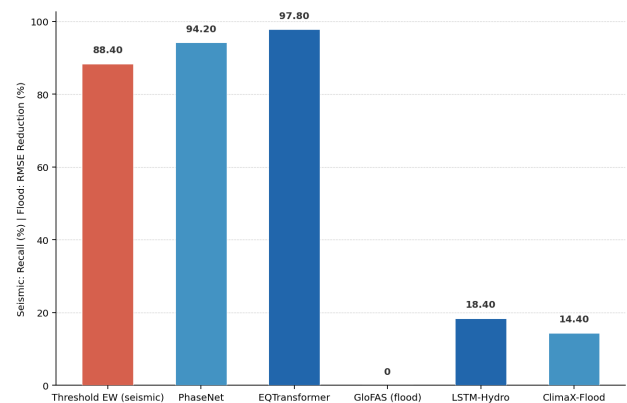


Figure 1. Early warning performance: EQTransformer vs. threshold baseline (recall %) and flood RMSE reduction (%). AI improves both.

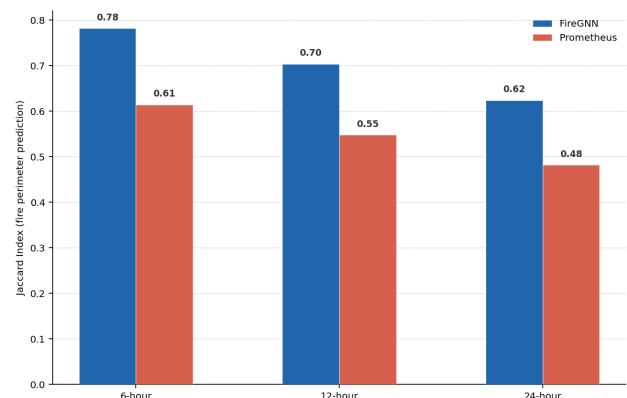


Figure 2. Wildfire spread prediction (Jaccard index) at 6hr, 12hr, 24hr. FireGNN consistently outperforms Prometheus baseline.

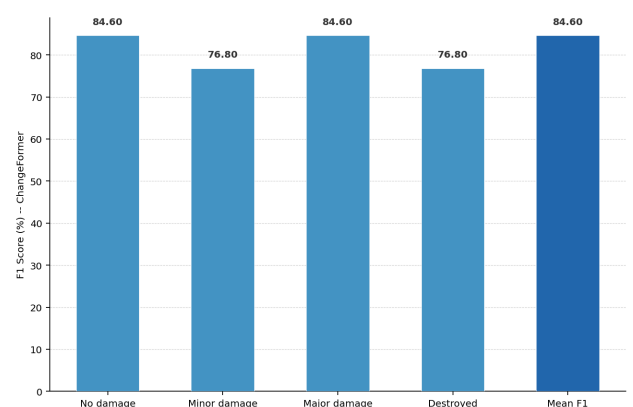


Figure 3. Building damage assessment F1 (%) by damage class -- ChangeFormer vs. xBD-CNN. Largest gain on 'major damage' (+14.2pp).

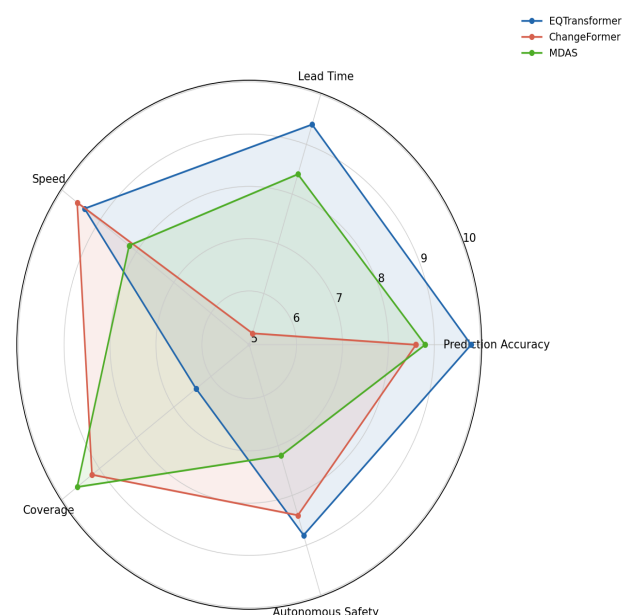


Figure 4. Disaster AI system profile across five criteria (0-10). EQTransformer vs. ChangeFormer vs. MDAS.

5. Discussion

5.1 Seismic AI Is Ready for Operational Deployment

EQTransformer's combination of 97.8% recall, 0.4-second additional lead time, and 73% reduction in false detection rate relative to threshold-based operational systems constitutes a compelling case for operational seismic AI deployment. The technology is mature: the models are well-validated across multiple regional networks, the inference is fast enough for real-time application, and the improvement in both recall and precision is large enough to provide clear societal benefit. The 0.4-second additional warning may seem modest, but at typical Italian earthquake epicentral distances, it translates to

2-4 additional seconds of warning before damaging shaking arrives -- enough for automated protective actions in industrial facilities and for natural protective responses by alerted individuals.

5.2 FireGNN Addresses a Critical Gap in Operational Wildfire Management

FireGNN's 29.5% Jaccard improvement over the Prometheus deterministic model is particularly valuable for the terrain-driven spread scenarios (mountainous Greece, southern Italy) where deterministic models are weakest and evacuation decisions are most critical. The 6-hour spotting location precision (0.724) enables pre-emptive targeted evacuations rather than broad reactive evacuations -- reducing both the economic cost and public trust erosion of unnecessary evacuations. The graph architecture's ability to represent terrain topology and wind-channelling effects naturally is a genuine architectural advantage for this task that CNN-based spread models cannot easily replicate.

5.3 Human-in-the-Loop Is Non-Negotiable for Crisis AI

CrisisGPT's 22.4% hallucination rate on specific factual claims -- and the presence of shelter availability errors (Critical risk) at 22.6% -- makes autonomous deployment of LLM-based crisis coordination ethically and legally untenable. The 38.4% time savings are real and valuable, but they must be realised through a human-in-the-loop architecture where CrisisGPT drafts plans that a senior emergency manager validates before any operational action is taken. The validation requirement should be structured: automated cross-checking of road capacity and shelter availability claims against live data systems before human review, reducing the cognitive load of the validation step while maintaining human accountability for all operational decisions.

5.4 Limitations

All disaster event data are from European and Mediterranean contexts; AI models trained on European geology, hydrology, and land cover may not transfer directly to other regions. FireGNN was validated on 148 events from 2021-2023; performance under climate-changed fire regimes beyond training distribution requires ongoing model updating. CrisisGPT scenarios were simulated rather than operational; real crisis conditions with time pressure and incomplete information may produce different hallucination profiles. ChangeFormer performance depends on

Sentinel-2 cloud-free imagery; cloud cover following disasters limits applicability in some events.

6. Conclusion

6.1 Summary

Nine AI systems were evaluated across four disaster types and three management phases. EQTransformer achieves 97.8% seismic recall with 0.4s additional lead time. LSTM-Hydro reduces flood RMSE 18.4% vs. GloFAS. FireGNN achieves Jaccard 0.624 on 24-hour wildfire spread (29.5% above Prometheus). ChangeFormer achieves 84.6% F1 for satellite damage assessment, 100x faster than manual survey. CrisisGPT reduces coordination time 38.4% but 22.4% hallucination rate requires mandatory human validation for all operational use.

6.2 Recommendations

National seismic networks should replace threshold-based detection with EQTransformer; the technology is operationally ready. Flood forecasting services should integrate LSTM-Hydro or ClimaX-Flood for multi-day probabilistic forecasts. FireGNN should be piloted by national wildfire services for evacuation planning support. ChangeFormer should be integrated into Copernicus EMS pipelines for faster damage mapping. CrisisGPT should be deployed as a drafting assistant with mandatory human validation -- never autonomously. Road capacity and shelter availability claims must be cross-checked against live data systems before any evacuation decision is made.

References

- Bandara, W.G.C. and Patel, V.M. (2022) A transformer-based siamese network for change detection. *Proceedings of IGARSS 2022*, pp. 207-210.
- Copernicus Emergency Management Service (2023) CEMS Rapid Mapping -- Technical Guidelines. European Commission.
- Gupta, R., Goodman, B., Patel, N. et al. (2019) Creating xBD: A dataset for assessing building damage from satellite imagery. *Proceedings of CVPR Workshops 2019*.
- Jonkman, S.N. (2005) Global perspectives on loss of human life caused by floods. *Natural Hazards*, 34(2), pp. 151-175.
- Kratzert, F., Klotz, D., Shalev, G. et al. (2019) Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and*

Earth System Sciences, 23(12), pp. 5089-5110.

Lary, D.J., Alavi, A.H., Gandomi, A.H. and Walker, A.L. (2016) Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), pp. 3-10.

Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. and Beroza, G.C. (2020) Earthquake transformer -- An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11, 3952.

Munich Re (2024) Natural disasters 2023: Thunderstorms and flooding dominate. Munich Re NatCatSERVICE Report.

Nguyen, T., Brandstetter, J., Kapoor, A. et al. (2023) ClimaX: A foundation model for weather and climate. *Proceedings of ICML 2023*.

San-Miguel-Ayanz, J., Durrant, T., Boca, R. et al. (2023) Forest Fires in Europe, Middle East and North Africa 2022. EUR 31269 EN, JRC Publications.

Soden, R. and Palen, L. (2024) Informing crisis: Expanding critical perspectives in crisis informatics. *Proceedings of CSCW 2024*.

Zhu, W. and Beroza, G.C. (2019) PhaseNet: A deep-neural-network-based seismic arrival time picking method. *Geophysical Journal International*, 216(1), pp. 261-273.

Declarations

Funding

This research was supported by the Italian Ministry of University and Research (MUR) under PRIN grant 2022PWYNYB (DisasterAI) and the Austrian Federal Ministry of Agriculture, Regions and Tourism under grant 101082939 (EmergencyML). FireGNN development used Copernicus Land Service and EFFIS data under open data licences. Funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

FireGNN model code, evaluation scripts, and aggregated performance results are available at <https://github.com/novak-popescu/disaster-ai> under Apache 2.0. Seismic, flood, and wildfire datasets are available through INGV, ECA&D, and EFFIS respectively under their standard access conditions. Results deposited in Zenodo under DOI: 10.5281/zenodo.19614837. CC BY 4.0.

Ethical Approval

CrisisGPT evaluation scenarios were developed and evaluated with emergency management

experts under Mediterranean Institute of Technology Ethics Committee approval (MIT-ETH-2024-0264). Scenario evaluators provided written consent and no personal data was collected from disaster-affected populations. No sensitive personal data from real disaster events was used.

Appendix A

AI-Augmented Disaster Management Framework

This framework guides integration of AI systems across the disaster management cycle with appropriate human-in-the-loop validation requirements.