CENTRE FOR ECONOMIC POLICY RESEARCH

CEPR

December 2024

Harnessing AI: How to develop and integrate automated prediction systems for humanitarian anticipatory action[1]

Humanitarian Forecast Working Group[2]

1 INTRODUCTION

Despite unprecedented access to data, resources, and wealth, the world faces an escalating wave of humanitarian crises. Armed conflict, climate-induced disasters, and political instability are displacing millions and devastating communities. Nearly one in every five children are living in or fleeing conflict zones (OCHA, 2024). Often the impacts of conflict and climatic hazards – such as droughts and flood – exacerbate each other, leading to even greater suffering. As crises unfold and escalate, the need for timely and effective humanitarian action becomes paramount.

Sophisticated systems for forecasting and monitoring natural and man-made hazards have emerged as critical tools to help inform and prompt action. Traditional forecasting of hydro-meteorological events has recently been supplemented by advanced data analytics and machine learning approaches that have the potential to greatly improve performance and increase lead times ahead of an event unfolding. These automated forecasting systems aim to predict trends and anomalies, including in novel domains such as conflict emergence and internal displacement, enabling proactive and informed decision making. The full potential for automated forecasting systems to inform anticipatory action (AA) is immense but is still to be realised (Altay and Narayanan, 2020; Wagner and Jaime, 2020). By providing early warnings and predictive insights, these systems could help organisations allocate resources more efficiently, plan interventions more effectively, and ultimately save lives and prevent or reduce humanitarian impact.

The idea of AA, triggered by early warning, is summarised in Figure 1.[3] Whereas the traditional approach delivers assistance and unlocks funding after a shock has occurred, AA requires funding to be unlocked and assistance to be provided before a

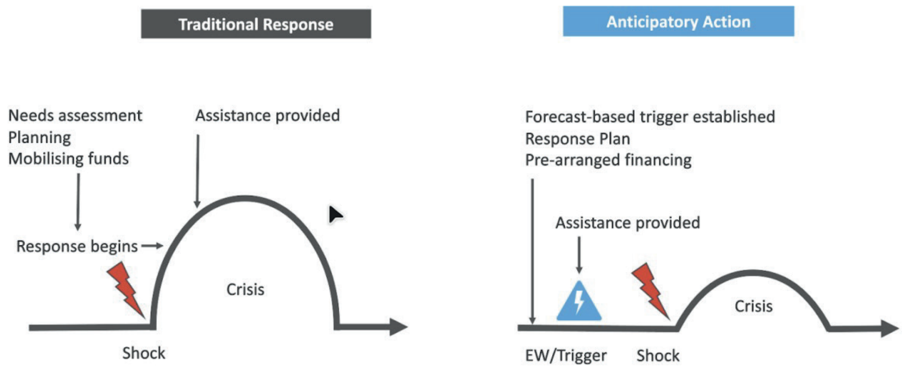
1 This Policy Insight is the result of a workshop hosted at the Institut d'Anàlisi Econòmica (IAE-CSIC) in Barcelona in May 2024 under the title "Growing Together: Prediction, Prevention and Preparedness". It brought together academics as well as representatives from foreign offices, UN agencies, the International Federation of Red Cross and Red Crescent Societies and INGOs that are at the forefront of augmenting existing systems for anticipatory action with the effective use of machine learning. The focus of the workshop was on using forecasts to improve humanitarian assistance allocation and planning as well as taking preventive action. This follow-up paper seeks to capture key debates when it comes to the use of conflict forecasts and other early warning systems to improve decision making in humanitarian anticipatory action or the adjacent fields of development work, foreign aid and diplomacy. Finance by the German Federal Foreign Office (GFFO) grant number AA38230003 is gratefully acknowledged. Mayoral and Mueller also acknowledge support by the RECIPE program funded by the Foreign, Commonwealth & Development Office (FCDO). Finally, we want to thank Marcel Meyer (UNDP), Nadia Noumri (UNICEF), Mohammed Harun Rashid (UNDP) and Gary Milante (World Bank) for their contributions to this brief.

2 Full list of authors at end of document. Please cite this document as: Humanitarian Forecast Working Group (2024), "Harnessing AI: How to develop and integrate automated prediction systems for humanitarian anticipatory action", CEPR Policy Insight No. 135.

3 See Clark and Dercon (2016) for more background.

shock occurs, triggered by agreed-upon monitoring thresholds or forecasts. Effective implementation of anticipatory action ideally therefore requires three elements: a pre-agreed trigger, pre-agreed activities, and pre-arranged financing.[4]

Figure 1 The key idea of anticipatory action



Source: UNICEF.

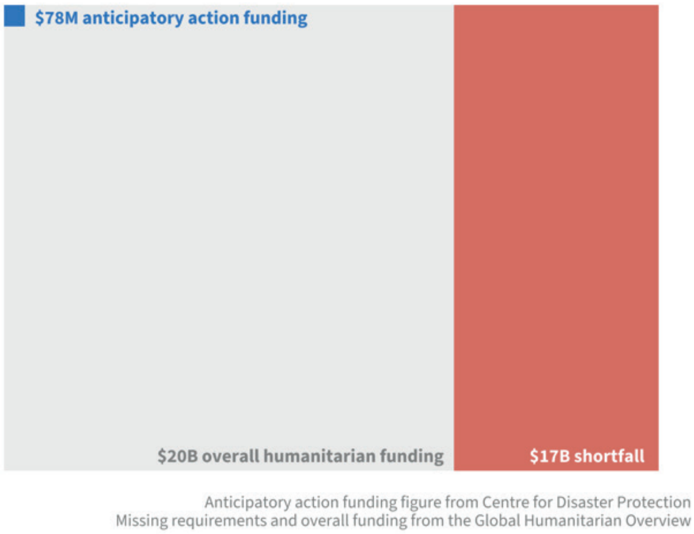
Automated systems have begun to be successfully implemented for AA and response to disasters triggered by natural hazards (extreme weather events in particular), but their potential for AA in the context of armed conflict seems underused. Figure 2 represents the humanitarian funding situation in 2021 graphically, with $20 billion in overall funding contrasting with $78 million in AA funding, and a $17 billion funding shortfall. This means that AA made up less than 0.4% of total humanitarian funding in 2021. While there are targets within organisations to increase funding shares to AA, these are not common.[5] Where there is AA, it is typically focused on pre-arranging action plans and corresponding financing, for example in the case of droughts, flooding, or the outbreak of disease. AA may also include forecasting and decision-making frameworks. With all this in place, action can be taken ahead of a shock and can be faster and more effective, thus reducing the impact of the crisis. AA can also extend beyond natural hazards: steps to prevent armed conflict early on may also yield high returns on investment (e.g. Mueller et al., 2024). However, there are significant technical, ethical, and organisational difficulties that need to be taken into account for quantitative forecasting systems to be systematically used in AA.

In this Policy Insight, we provide an account of these difficulties and the current solutions in place. We start by describing the kinds of decisions faced by agents in the humanitarian sector and exploring the current landscape of conflict and other hazard forecasting technologies, examining their capabilities and limitations to monitor and predict armed conflict and the onset and shifting levels of humanitarian needs. We delve into the serious constraints posed by ethical concerns and how predictions can be brought into the decision-making process required for AA.

We then propose ways forward for both academic researchers and practitioners. We discuss the possibilities and difficulties of the integration of quantitative forecasting systems for decision making in the humanitarian sector. But this Policy Insight is also a call for academic research to develop new types of systems that would allow practitioners in this sector to improve anticipatory action. We furthermore propose a shared vocabulary (see the glossary of key terms in the appendix), which will be indispensable for meaningful communication as the use of forecasting systems becomes more commonplace in the humanitarian sector.

4 See https://www.unocha.org/anticipatory-action
5 The only donor with an explicit target is the German Federal Foreign Office (GFFO). The other donors are more vaguely committed (as part of G7) to "*strive to significantly increase our financial support in anticipatory action programming*".

Figure 2 Anticipatory action funding relative to overall humanitarian funding, 2021



Source: OCHA presentation.

2 DECISION MAKING IN THE HUMANITARIAN SECTOR

This section highlights who the main actors in the humanitarian sector are, what decisions they are faced with in practice, how these decisions are made, and the extent to which forecasts could be useful. We also touch on the extent to which some of these elements differ for actors more focused on development or diplomacy rather than humanitarian responses.

The main decisions of interest in this Policy Insight are where to intervene, when, followed by how to intervene with which tools. Intervention can be further broken down into AA and response. Decision making in humanitarian response is characterised by urgency, heuristics, and uncertainty. Incident command systems aimed at reducing friction and confusion during crisis management focus on response, not anticipation. Meanwhile, AA seeks to plan and intervene ahead of new crises emerging or existing crises deepening. There is also a common distinction between slow- versus sudden-onset hazards. AA cannot prevent all impacts caused by fast-onset acute hazards such as flash floods or earthquakes so acute crises will always need a response element, but there is still scope for more analytic approaches before the crisis emerges as well as when making allocation decisions during the response. In the case of slow-onset crises, usually relating to the impacts of climatic change such as drought, there is more time to monitor early warning signals and ideally engage in AA. However, as repeated examples from the Horn of Africa alone show, such warning signs have often been ignored and intervention delayed until a situation reaches crisis point (Dempsey and Hillier, 2012, Buchanan-Smith and Davies, 1995; Hammon and Maxwell, 2002; Hillbruner and Moloney 2012; Maxwell et al., 2023).

Decisions require us to form beliefs about the status quo and the future, namely, whether there is a crisis and how likely it is that a crisis will emerge or continue. These beliefs can be formed through either a more intuitive or a more analytical approach, as shown in Table 1. The tension between intuitive and analytical decision making is a key thread running throughout this Policy Insight. Sometimes it becomes a tension within an organisation between which approach to take and how knowledge can be most effectively exchanged between analytical and intuitive parts of the organisation. At the intuitive end of the spectrum, there may only be broad, subjective opinions about what constitutes a crisis and information processing is quick, subjective and highly context-specific. An analytical approach works with an intersubjectively operationalised definition of crisis. In the latter case, crises are defined on the basis of a set of outcome variables reaching pre-agreed thresholds, with data processed on the basis of similarly pre-agreed rules. An example is the Integrated Food Security and Nutrition Conceptual Framework by the Integrated Food Security Phase Classification,[6] which underpins the IPC's classification system for identifying the current and projected number of people living in different phases or severity levels of acute food insecurity, chronic food insecurity, and acute malnutrition. While some organisations already have forecasts to support their decision making, this does not necessarily mean that they are effectively used for an analytically informed approach to a decision. We discuss how systems that are expert-based (and thus more on the intuitive end) by default might effectively make use of forecasts in Section 4.

Table 1 Two modes of crisis early warning for decision making

Intuitive mode	Analytic mode
Broad, subjective opinions of what constitutes a *crisis*	Crisis is operationalised as (a set of) concretely defined outcome variable(s)
Interpretability and timely availability of data is key	Quantitative structure(-ability) of data is key
Information processing is quick, subjective and context-specific • Importance of information often assessed based on experience • Causal relationships are assumed in a way that (usually) aligns with intuition • Prone to cognitive biases	Information processing is based on (intersubjectively agreed upon) rules, e.g., quantitative algorithms • Importance of information derived based on quantitative models; sometimes non-intuitive results • Models rarely allow to draw conclusions on causal relations • Prone to biases inherent in the data

Source: GFFO presentation.

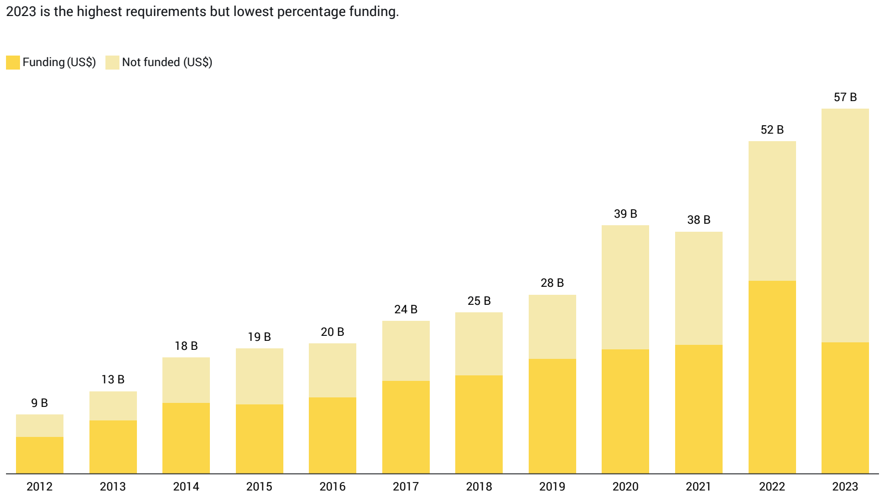
In the humanitarian sector, decisions are made against the backdrop of a severe funding shortage. The rise in funding appeals shown in Figure 3, taken from the Global Humanitarian Overview (GHO) 2024, reflects the increasing trends in internal armed conflict- and climate-related crises over the last decade.[7] Funding has increased but has not kept up with the rise in funding requirements. Since the charts in Figure 3 were produced for the GHO, progress against UN appeals has increased to 43% (US$24 billion). This still makes it the lowest coverage relative to declared need since 2012. As discussed previously, the vast majority of funding is allocated to crisis response rather than to AA in the humanitarian sector.

The mostly ad hoc element of emergency needs assessments and response funding has been described as a 'begging bowl' approach that comes with inequities and inefficiencies (Clarke and Dercon, 2016, Chapter 2). Benefactors have to choose where to allocate resources, while those asking for money are faced with uncertainty over how much they will get. AA can reduce these uncertainties if benefactors set out ex-ante response plans, detailing when and how resources will be allocated in case (or before) a disaster strikes. There has been an increased interest among donors in trialling the adoption of parametric insurance practices, as exemplified by the UK's Foreign, Commonwealth Development Office (FCDO) Risk Pools Programme (RPP).[8] This development has been supported by improved access to high-quality and frequently updated monitoring data, such as wind strengths in case of a hurricane, that might determine pay-outs.

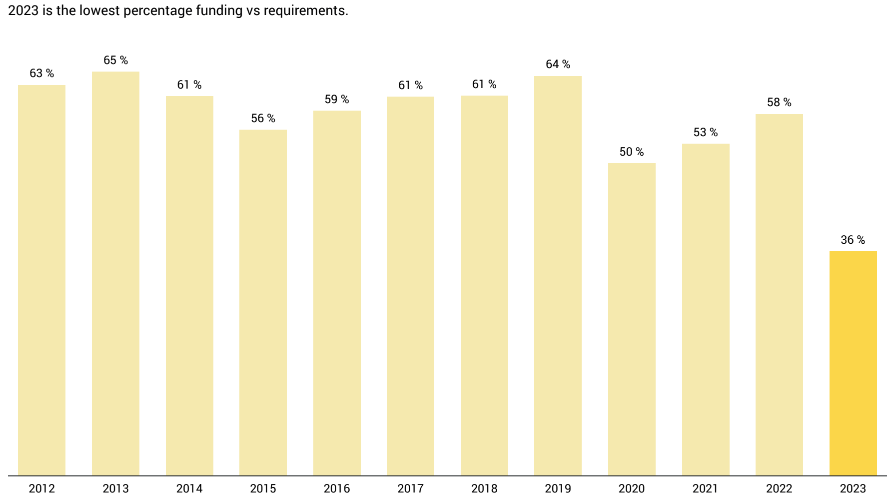
6 See https://www.ipcinfo.org/ipc-manual-interactive/overview/17-the-ipc-conceptual-framework/en/
7 It should be noted that comparing figures of requirements and funding over time is challenging as the methodology of calculating humanitarian requirements has not remained consistent since 2012: the Consolidated Appeals Process (CAP) was replaced by the Enhanced Humanitarian Programme Cycle (HPC), in the Joint Intersectoral Analysis Framework (JIAF) was later embedded.
8 See https://devtracker.fcdo.gov.uk/programme/GB-1-203469/summary

Figure 3 Humanitarian funding appeals and funding coverage

Appeal funding vs requirements, 2012 - 2023 (as of 4 December 2023)



Appeal funding coverage, 2012 - 2023 (as of 4 December 2023)



Note: All requirements and funding between 2012 and 2022 are as of end-year. Requirements and funding for 2023 as of 4 December 2023.

Source: HumanitarianAction.

2.1 Actors and funding decisions in the humanitarian sector

Over 5,000 organisations are involved in providing humanitarian aid to millions of people each year.[9] The United Nations (UN) alone has several agencies related to humanitarian provision, supported by the Inter-Agency Standing Committee (IASC) as the longest-standing and highest-level humanitarian coordination forum of the UN system.[10] So-called clusters (and the four areas of responsibilities within Protection) bring together UN and non-UN partners around 15 technical areas of humanitarian action, with the aim of coordinating preparedness and response activities.[11] Often funds are pooled through UN agencies, such as the Central Emergency Response Fund (CERF) or the Country-Based Pooled Funds (CBPF) established by the UN Office for the Coordination of Humanitarian Affairs (OCHA). Membership organisations and networks in the humanitarian sector may similarly pool funds, for example in the Disaster Response Emergency Fund (DREF) of the International Federation of Red Cross and Red Crescent Societies (IFRC) or the Start Network, which pools funds from different donors accessible to over 90 NGO member organisations. When humanitarian need arises, there is also a large community of actors that play a role in coordinating and implementing support. National governments and local

9 See https://humanitarianoutcomes.org/projects/gdho/graphics
10 See https://interagencystandingcommittee.org/the-inter-agency-standing-committee
11 See https://reliefweb.int/topics/cluster-coordination

humanitarian actors are usually the first to respond when disasters strike and can often access areas that international actors cannot. The focus in this Policy Insight is primarily on funding decisions by large international actors of the kind present at the workshop that inspired this paper. They leverage the largest amount of funds and their approach to AA can act as a model for others in the field.

Additionally, there are 'enabling' agencies focused on supporting data and analysis needs such as providing hydro-meteorological modelling, insights from satellite images, or data on fatalities – as collected by the Uppsala Conflict Data Program (UCDP) or Armed Conflict Location and Event Data (ACLED) – and crisis analysis.[12] There are also knowledge hubs or multi-stakeholder forums that bring different data sources together. For example, INFORM is a collaboration of the Inter-Agency Standing Committee Reference Group on Risk, Early Warning and Preparedness and the European Commission. As part of this collaboration, sine 2014 the European Commission's Disaster Risk Management Knowledge Centre (DRMKC) has published the INFORM Risk Index, an open-source risk assessment for humanitarian crises and disasters updated twice a year. INFORM has since developed additional products such as INFORM Severity (2020) and INFORM Climate Change (2022), with INFORM Warning under development. The effective collaboration between large international actors and these enabling organisations is critical to inform AA in line with the best available information.

The most important decision faced by funders and organisations on the ground is how to allocate scarce resources while upholding humanitarian principles. For sudden-onset emergencies like earthquakes, the UN's CERF, the IFRC's DREF and the Start Network's Start Fund are the major international funding mechanisms. They share characteristics but also have specific differences in terms of scale, speed, and mandate. At the international level, the UN Office for the Coordination of Humanitarian Affairs (OCHA) plays a central role in the allocation of funds through the UN, leading both the CERF and CBPFs. For the Red Cross and Red Crescent, the funding for anticipatory action is largely centralised through the DREF, which pioneered the use of pre-agreed triggers and protocols, and which has continued to innovate through the use of complementary insurance mechanisms. The Start Network has a global fund and several country-level funds that disburse funding to its members.

Aside from pooled funding mechanisms, the UN and IFRC run specific emergency appeals for a range of situations such as epidemics, conflict, and population displacement. These funding mechanisms do not tend to use anticipatory models since they are at a greater scale than is possible within the lead-time available from a forecast. Nevertheless, there is an untapped potential for models to determine estimated populations affected and the geographical severity of needs. To date, affected population figures have been estimated in an ad-hoc and unsystematic manner by the UN and IFRC.

To determine levels of funding requirements, the humanitarian sector has developed sophisticated guidance for needs assessment. These take an analytic approach to decision making as much as possible, relying on sources such as the IASC INFORM products, People in Need (PiN), and other suitable sources for information as well as on frameworks to structure the process. In the case of cluster coordination, humanitarian needs and response analysis is a collective process through which a Humanitarian Country Team (HCT) identifies the humanitarian needs of affected populations and develops plans and mobilises resources to respond to the most pressing ones on the ground.[13] For example, a Multi-cluster/sector Initial Rapid Needs Assessment (MIRA) is an inter-agency needs assessment and analysis process that

12 For example organisations, see the H2H Network (https://h2hnetwork.org/).
13 See https://interagencystandingcommittee.org/humanitarian-planning

is supposed to be carried out within the first three days of a disaster,[14] although this is rarely (if ever) the case. The Joint and Intersectoral Analysis Framework (JIAF) similarly seeks to improve the way humanitarian actors jointly plan and respond to (often protracted) crises, through a rigorous, evidence-based, and comprehensive joint and intersectoral analysis system.[15] An example of the Humanitarian Needs and Response Plan (HNRP) for Sudan is shown in Figure 4. In the implementation of response plans, more decisions need to be made about where exactly to allocate resources and through which channels. This is where the live collection of systematic as well as ad-hoc qualitative and quantitative data aids organisations in carrying out triage and reaching those most in need.

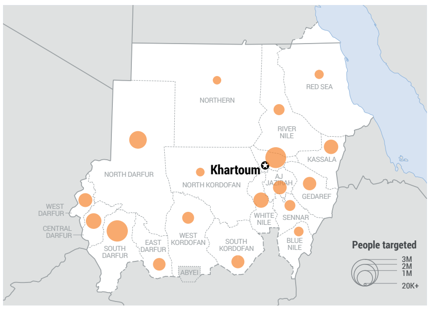
Figure 4 Example from Sudanese Humanitarian Needs and Response Plan (HNRP) for 2024

People in need and people targeted by sex, age and disability

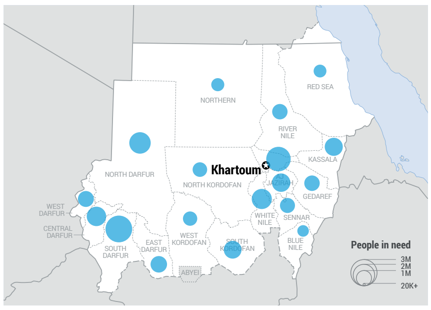
		WOMEN	CHILDREN	ADULT	OLDER PEOPLE	WITH DISABILITY	REQUIREMENTS
PEOPLE IN NEED	24.8M	26%	48%	47%	5%	15%	$2.7B
PEOPLE TARGETED	14.7M	26%	48%	47%	5%	15%	

M: Million / B: Billion

People in need by state



People targeted by state



People in need and people targeted by cluster

CLUSTER	REQUIREMENTS	PEOPLE IN NEED	PEOPLE TARGETED	% OF PEOPLE TARGETED
Education	$131.0M	9.1M	4.2M	46%
Shelter and Non-Food Items	$212.4M	8.6M	2.1M	25%
Food Security and Livelihoods	$581.2M	19.3M	11.4M	59%
Health	$178.6M	14.7M	4.9M	33%
Nutrition	$350.1M	4.7M	1.9M	42%
Child Protection	$81.0M	6.2M	3.0M	48%
Gender-based violence	$62.8M	6.7M	1.8M	27%
General Protection	$64.1M	7.7M	3.6M	46%
Mine Action	$15.3M	6.0M	3.8M	64%
Site Management	$14.8M	3.2M	1.6M	50%
Water, Sanitation and Hygiene	$230.9M	18.9M	8.9M	47%
Refugee Response	$631.2M	963K	963K	100%
Coordination & Common Services	$25.0M	-	-	-
Emergency Telecommunications	$6.3M	-	-	-
Logistics	$111.0M	-	-	-

Note: The figure details for each cluster the number of people in need, how many of them are targeted and the relevant funding requirements, the cluster objectives and response for the education cluster.

Source: Humanitarian Action

14 See https://emergency.unhcr.org/coordination-and-communication/cluster-system/multi-cluster-sector-initial-rapid-needs-assessment-mira

15 See https://www.jiaf.info/

While HNRPs are clearly indispensable, they do not strictly fall under AA.[16] There are underlying triggers for when a situation is deemed critical enough to warrant a plan or new crisis funding appeal, but these plans are developed in response to crisis rather than being predetermined plans that are acted upon in response to pre-agreed thresholds. The rest of this Policy Insight focuses on AA that can be taken ahead of the peak of a looming crisis rather than the details of response operations once a crisis has moved beyond the stage of AA. As a report published by the World Food Programme (WFP) points out, "[t]here is a growing body of evidence pointing towards the positive impact of AA, yet it is often fragmented, incomplete in scope, and in need of methodological improvements" (Weingärtner et al., 2020).

2.2 Intersections with the development sector and foreign offices

Historically, there has been a separation between humanitarian and development funding streams at the international donor level, resulting in diverging interests. Yet there has been an increasing recognition of the need for so-called 'nexus' approaches to bridge this gap. In 2016, the Humanitarian–Development–Peace (HDP) nexus emerged out of the 2016 World Humanitarian Summit with the aim to "ensure strong cooperation, collaboration and coordination between humanitarian, development and peacebuilding efforts at the national level to ensure collective outcomes on the basis of joined-up, coherent, complementary and risk-informed analysis, planning and action".[17] In the case of humanitarian AA, triggers, plans and funding may change depending on the levels of development an area has received. Upfront development and investment to increase resilience to disasters has implications for potential humanitarian needs.

UNDP development strategies and approaches have taken a more holistic perspective in the past ten years, reinforcing the importance of the HDP nexus and resilience-based peace and development. As a development organisation, UNDP's role in the nexus includes driving system-wide coherence on nexus approaches by acting as a bridge between humanitarian, development and peace actors, and addressing root causes through the scaling up of integrated development and peace programming. Specifically, for UNDP this means acting at the global/headquarters level and at the country office/field levels. At the global and strategic level, the administrator of the UNDP is a Vice-Chair of the Joint Steering Committee to Advance Humanitarian and Development Collaboration (JSC) at the UN. The JSC was established in 2017 as a key mechanism to promote greater coherence of humanitarian and development action in crises through closer collaboration, integrated approaches, and joint missions in the field. At the country office or field level, UNDP's work encompasses both anticipation and prevention of crises, as well as crisis response and recovery.

Coordinating AA across these domains comes with challenges in practice. Development, whether supported by UN agencies or countries' foreign offices, often operates over long time horizons, with incremental progress. This is a different pace from crises that require urgent humanitarian assistance and protection. Furthermore, just like humanitarian action, sustainable development suffers from an immense funding shortfall. Most recently, the UN estimated the development financing gap to be a $4.2 trillion annually, up from $2.5 trillion before the COVID-19 pandemic (United Nations, 2024). Nevertheless, more effective collaboration between the humanitarian sector and development efforts is needed. Climate change and cascading effects of crises show that sustainable development is crucial to building long-term resilience.

Diplomacy, peacebuilding, and prevention of armed conflict are also directly relevant to humanitarian crises and wider development efforts. It is now widely accepted that armed conflict is a key contributing factor to slow economic development, lack

16 Risk analysis and AA do play an increasingly prominent role in some HNRPs, such as the 2024 plan for Somalia (see https://humanitarianaction.info/plan/1180).
17 See https://www.undp.org/crisis/humanitarian-development-and-peace-nexus

of state capacity, and humanitarian crisis. However, the integration of conflict, economic development, and humanitarian action within organisations is challenging. This is further complicated by moves in several Western countries that aim to make development aid an arm of foreign policy.

In contrast to humanitarian crises due to disasters, the impact of which can (at best) be mitigated but the hazard not stopped, one goal of conflict early warning is to enable organisations and actors to agree on measures of AA that weaken or completely prevent the hazard from happening. However, and also in contrast to AA for crises caused by natural hazards, the measures that can be taken are much less clearly defined and also more context-specific for the case of armed conflicts. This brings us back to the difference and tension between *intuitive* and *analytical* approaches. For short-term AA, diplomacy – and, in extreme cases, military action – are instruments, but they need highly context-specific information that can usually not be provided from an early warning system but instead has to come from experts on the ground. Thus, for the case of conflict early warning, quantitative early warning systems in organisations such as foreign ministries are usually only one building block of a broader early warning system that includes field experts and other actors. In such a context, conflict early warning systems complement and sometimes also compete with more traditional (*intuitive*) ways of information processing.

Early warning systems can also help at a more strategic level, for example in the programming of development programmes that serve several goals at once but increasingly need to take account of conflict aspects. There is often no clear distinction between policies for conflict prevention and development efforts. But given academic research on the role of specific types of aid in triggering conflict (Nunn and Quian, 2014; Sonno, 2023; Rohner, 2024), development actors need to take care to not equate the two. Within organisations, the allocation of conflict experts matters: often they are only allocated to situations of already violent conflict, which could hinder preventative approaches. Ideally, early warning systems would help to justify paying attention to development earlier on.

Thus, in designing an early warning system, the organisational setting in which it operates has to be taken into account. In some settings, conflict early warning systems have to be designed in such a way as to be explainable and understandable to non-technical experts and decision makers; explainability of the results of quantitative models can sometimes be even more important for building trust and becoming impactful for decision makers than highly accurate predictions. In other circumstances – for example, in the allocation of scarce resources across countries and regions – early warning systems can serve as a quantitative benchmark for strategic decision-making processes. This can make performance aspects relatively more important.

2.3 Weighing up the use of forecasts for anticipatory action

Technically, the definition of AA used does not require forecasts; it is possible for AA plans to solely respond to threshold levels reached in monitoring systems that report observations on the current state of the world. However, looking ahead can be crucial: rather than only releasing funds for flood support, for example, when levels of rainfall reach a certain level, funds could be released much sooner when taking into account rainfall forecasts.

Broadly speaking, forecasts use past data to predict the future. The fast progress in machine learning methods and increasing data availability over the last decade has made machine learning-based forecasts much more accurate. The use of forecasts falls at the analytic end of the decision -making spectrum. Forecasts also are typically more suited to supporting AA rather than response work as they provide forward-looking information updates. However, we need to distinguish protracted crises from new crises. Quantitative forecasting systems are most useful before the start of new crises, whereas keeping track of crisis dynamics is most useful in protracted contexts. However, forecasting new phenomena is particularly difficult. In practice, some organisations are already making use of forecasts, which will be highlighted in Section 4.5. Here we briefly outline the potential challenges and benefits of their use in decision making.

The key to thinking about how quantitative forecasts can be built into decision making is the choice of a target variable for training the algorithm. Section 3 will delve deeper into this matter, but Table 2 already provides an overview of examples. When forecasting events that imply humanitarian need, the algorithm is trained with data on relevant past events, such as droughts. The drought forecast can then be used to forecast impact or need in a second step. When forecasting impact or need directly, the causal chain between drivers and impact to need does not need to be known, and instead outcomes are directly forecasted. Predicting these different layers has very different technical requirements and the extent of endogeneity of the target plays a key role in whether they can be prevented with a decision or whether the decision can only build on the forecast.

Table 2 Illustrating machine learning targets for forecasting systems

	Classifier example	Regression example
Forecasting events	Predicting the likelihood of an armed conflict outbreak in the next 12 months	Prediction water levels of a river two weeks in advance
Forecasting impacts	Predicting that refugees will leave a country and flee to another in two months from now	Predicting the number of displaced individuals leaving a municipality in the following year
Needs	Predicting that food will become scarce without outside help within the next two weeks in a region	Predicting the number of additional medical kits required in a hospital next year

The remainder of this section outlines five potential benefits as well as six important challenges to the use of forecasts in the context of AA.

Benefit 1: An extra 'opinion'. Making decisions in complex situations is challenging. Possession of expert knowledge does not make technocrats immune to bias or groupthink. With a forecast that provides a methodologically objective assessment, decision makers can triangulate between that algorithmic prediction and the predictions or insights of expert advisors. When both agree, then action can be taken with greater confidence. When they disagree, it may prompt important further debate. One way to think of quantitative forecasts is as improvements over naive base rates, which previous research has shown are one of the most important factors for good judgement (Tetlock and Gardner, 2016, p. 116). In terms of complementarity, this brings together the ability of models to weigh a large number of multidimensional factors and the superior ability of analysts to consider idiosyncratic factors that would be difficult to capture with quantitative data.

Benefit 2: Improved comparability. When making decisions across different countries and contexts, it can be challenging to make meaningful comparisons. Thus, a numerical prediction that takes on comparable values across locations and time can be a highly useful addition to decision making as it provides a relative benchmark – for example, related to the difference in risk of violent conflict in one location versus another, or of long- and short-term trends over time. When accessed by local parts of the organisation, quantitative forecasts can even strengthen their case within the hierarchy of that organisation. Protracted conflicts often stop making headlines in news media once the initial 'shock factor' wears off. Comparable forecasts help to ensure these regions remain on the radar of policymakers and practitioners, and that (relatively) small but devastating conflicts catch the attention of the international community.

Benefit 3: Greater accountability. In some domains there is bias towards inaction. With a numerical forecast, a threshold value can be decided at which a recorded decision must be taken to act or not to act. Usually, only the decision to act is recorded and it may thus be harder to take. In this scenario, however, the decision to inaction is also recorded, making it less of a quiet default option. It can also be a counterweight to idiosyncratic factors like office politics and differences in the pull of different country desks. Furthermore, if the forecasts are open and accessible to the public, they could help empower local populations to promote their own safety by holding decision makers accountable for acting on the evidence at hand and spearheading locally owned peacebuilding initiatives.

Benefit 4: New insights and performance. The use of forecasts can bring obscure conflict patterns to light that would not have been visible to humans otherwise. These include nonlinear relationships and interactions between variables. So-called black box models are particularly effective at using these patterns to make predictions. With growing availability of quantitative data, these approaches are becoming increasingly better in their predictions and will outperform human experts eventually.

Benefit 5: Capacity-building effects. Building the capacity to trigger anticipatory action through forecasts can generate long-term benefits beyond reducing the immediate risk or impact of a crisis. By mobilising resources and strategic planning ahead of potential emergencies, AA systems enhance the preparedness and resilience of local actors, equipping them to respond more effectively to future, even unforeseen, crises. Prepositioned resources, such as food or supplies, can remain valuable for subsequent emergencies. Additionally, AA often strengthens coordination and planning frameworks within affected areas, fostering greater local ownership and readiness. These capacity-building effects can justify some automation, as they contribute to more sustainable and proactive crisis management.

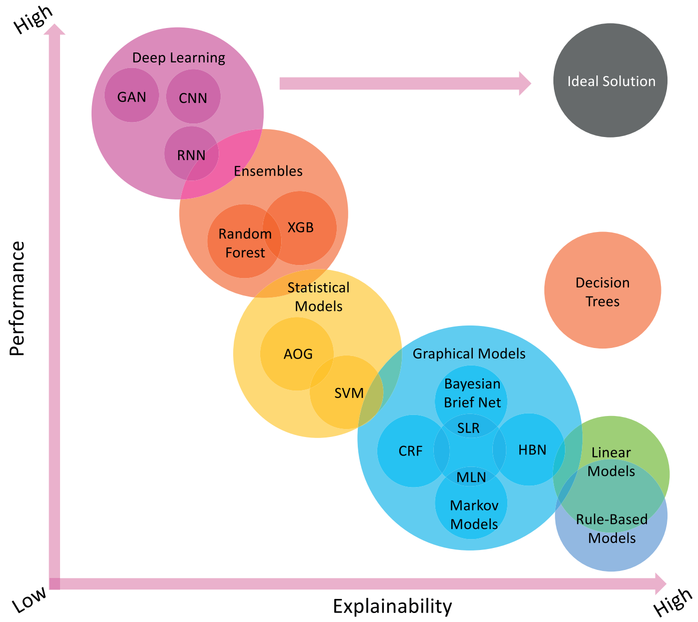
Challenge 1: Data availability. Algorithms learn from data, so where data are not available, are limited, or of poor quality, algorithms will not be able to generate helpful predictions, or risk generating harmful predictions. For example, the improving availability of high-quality meteorological and natural hazard data (see Section 3.3) means that the use of forecasts in humanitarian AA related to floods and droughts is quite advanced in some organisations. Similarly, global conflict fatalities data have been systematically recorded (see Section 3.2), whereas political event data sets have often been discontinued or have only been generated for a smaller number of countries.

Challenge 2: Time horizons in different domains. Technically, forecasts can be made on any time horizon. However, in order for an algorithm to learn to predict five years into the future, it needs to have data available that include many examples of what happened five years ahead, which is often not obtainable. In UN development agencies, there is currently a strong emphasis on using social sciences approaches from strategic foresight, scenario building and futures studies to improve institutional mandates. These qualitative strategic approaches that focus on long timescales of five to ten years may therefore find it hard to link their planning with quantitative forecasting (e.g., of armed conflict) that might cover timescales of six months to three years. The use of forecasts in development may therefore require more nuance in interpretation to derive actionable insights or longer time series for quantitative forecasts. There may also be a mismatch between the time frames of some existing forecasts and the timeframe for strategy development and funding cycles among donors. This makes collaboration – for example, Conflict Forecast recently working with the FCDO to make forecasts on a three rather than one year horizon – indispensable.

Challenge 3: Trust in algorithms. There is still a significant amount of unease when it comes to using algorithms to make decisions. Algorithms are able to outperform humans significantly in some domains where humans might be more prone to bias, but algorithms often perform worse in more complex situations (e.g., Sunstein, 2023). Even where algorithms outperform humans, so-called algorithm aversion means that humans judge algorithms more harshly on their mistakes (Dietvorst et al., 2014). This mistrust in machines can make it harder for algorithms to enter decision-making processes, even when there will still be a ‘human in the loop’ before final decisions are made. There may be worries about heading down a slippery slope into full automation, or the potential for function creep whereby the algorithmic approach is expanded to additional areas (Levy et al., 2021). Full automation might be of particular concern regarding responsibility: algorithms cannot be held to account, and it is not clear that we want machine learning engineers to be solely responsible instead of humans with more domain knowledge of the field in which predictions are made.

Challenge 4: Explainability versus performance. One way of mitigating mistrust in algorithms is ensuring their predictions or recommendations are easy to understand. However, while simpler rules-based or linear algorithms are relatively easy to understand, they are less good at making predictions. Figure 5 visualises this trade-off. Deep learning models are especially suitable for large quantities of data, including satellite or other images. Tree-based ensembles like Random Forest have been particularly effective and popular for making predictions in social science-related fields that tend to have fewer data points. As mentioned above, less explainable black box models are especially effective at exploiting nonlinear relations and interactions, which typically allow them to perform much better. Efforts in explainable AI have sought to reduce this trade-off, aiming to highlight how a model came to its conclusions. For example, feature importance values give insight into which of the included variables are particularly relevant when making the predictions. SHapley Additive exPlanations (SHAP) values go a step further and highlight both the size and the direction that a feature contributes to an individual prediction’s value, i.e., whether the feature is associated with increase or decrease in the prediction value. The language used here is important, for we cannot speak of a feature “leading to” a specific value in a way that implies a known causal relationship.

Figure 5 Explainability and performance in quantitative forecasting



Source: Yang et al. (2021)

Challenge 5: Impacts of forecasts. In Section 4.1 we expand on the concern surrounding self-fulfilling and self-defeating forecasts. Here we briefly illustrate these issues with two examples of highly simplified scenarios. Predicting armed conflict could become self-fulfilling, for example because foreign investors might become aware of this forecast and withdraw investment or close borders. This could in turn worsen the country’s economic outlook and reduce the opportunity cost for individuals to pick up arms. This points to further ethical concerns about who a given forecast should be available to. Self-defeating forecasts might occur when a decision maker becomes very good at mitigating the risk that the forecast predicts. Using the example of conflict again, it could be that mitigation of violence becomes so effective that the areas with high risk never end up erupting in armed conflict. This would render recent data somewhat useless for further training any predictive models, for what the algorithm would see are all the features commonly associated with violence, but no armed conflict emerging. This is already a problem to some extent in the data that we train on, for there are always some efforts somewhere to mitigate conflict, which the model is unaware of. The only way in which this could be resolved is the development of a comprehensive data set of all conflict mitigation activities carried out over time.

Challenge 6: Endogeneity. Decision makers need to be aware of whether a variable they forecast is endogenous to their decisions. Humanitarian needs at a location can, for example, be amplified or prevented by the decision maker. A famous example from economics is the Lucas critique,[18] which is a criticism of econometric policy evaluation procedures that fail to recognise that the reaction of an economy varies systematically with changes in policy. When predicting needs and impacts, this is particularly important. This will be discussed in more detail in Section 4.

3 FORECASTING HAZARDS: REQUIREMENTS AND BEST PRACTICE

This section highlights the frontier of efforts in forecasting conflict or natural hazards, or directly forecasting the impact of events. The first section additionally outlines the importance of target choice (and its availability). While there has also been an increased interest in forecasting political events such as coup d'etats, this Policy Insight omits such efforts as their humanitarian impact is more ambiguous and overall mitigation and intervention strategies typically fall outside of the humanitarian sphere. We additionally omit systems that are not strictly quantitative forecasts based on machine learning methods. This includes the very valuable work done by the International Crisis Group's CrisisWatch[19] and its Environmental Early Action & Risk Tracking Hub.[20] Finally, we omit other monitoring efforts that are not strongly about the future – for example, the use of satellite images to identify conflict-related activities such as troop movement, damage to infrastructure (Mueller et al., 2021) or graves (Howarth et al., 2024).

3.1 The importance of target choice

The most fundamental requirement for any forecasting model is a reliable, regularly updated dataset that captures the varying intensity of the hazard of interest. This dataset collection/provision exercise can be defined as *monitoring*. Most effort typically goes into monitoring and it sometimes involves sophisticated AI-driven systems. The data generated this way can then be used as targets for forecasting systems. This section focuses on *forecasting*, whereby these historical data inform models that make predictions about the future.

The starting point for developing a forecasting system is a clear, quantitative definition of the hazard to be forecasted – the target variable. This is directly derived from the selected dataset, highlighting the importance of long-term funding for high-quality data collection projects that enable the continued delivery of forecasts. Furthermore, the data availability will influence every aspect of any early warning system: the statistical methods that can be used, its spatial/temporal coverage, and the performance that can potentially be reached.

Taking this one step further, the choice of target will be the most important factor in determining its potential usefulness. In general, this comes down to choosing the following:

18 See https://research.hhs.se/esploro/outputs/bookChapter/Lucas-Critique/991001480353806056
19 See https://www.crisisgroup.org/crisiswatch
20 See https://icg-prod.s3.amazonaws.com/s3fs-public/2023-12/EEARTH-2-pager.pdf

CEPR

1. **A prediction task**: Is the decision maker interested in the likelihood of an event occurring or not (classification), or the magnitude/intensity of the event (regression)? Are they interested in making forecasts of the events themselves, of their impact or the resulting need for support?
2. **A forecasting horizon**: Is the decision maker interested in informing immediate humanitarian responses (short-term forecasts) or development initiatives/strategic planning (long-term forecasts)? Is there a time lag between the availability of a forecast and the possibilities to act on it?
3. **A spatial unit**: Is the decision maker interested in the local or national level?
4. **Update frequency:** How often does the decision maker need or want new forecast values? Is annual too infrequent? Is there capacity and desire to absorb monthly updates?

Again, the choice of dataset dictates what is possible. If the data record a binary indicator of a crisis at the country/year level, then a forecast of the magnitude of the crisis at the local/month level will be impossible. To elucidate this further, the following subsections outline frequently used datasets in the respective domains, as well as the typical target variable definitions.

3.2 Conflict data and forecasting

Early warning system practitioners in the conflict landscape usually rely on one of two datasets: the Armed Conflict Location & Event Data Project (ACLED) and the Uppsala Conflict Data Program (UCDP).[21] These differ their spatial/temporal coverage, update frequency, and data validation processes. Both are suitable for developing forecasting systems, yet their event definitions would lead to different target definitions.

- **UCDP event definition:** "An incident where armed force was used by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date".[22] The UCDP distinguishes between three types of violence: state-based, one-sided and non-state events.
- **ACLED event definition:** "Involvement of designated actors – e.g. a named rebel group, a militia, or state forces (with the sole exception of Unidentified Armed Group and generic categories including Rioters, Protesters, and Civilians). They occur at a specific named location (identified by name and geographic coordinates) and on a specific day."[23]

A key difference is that in the ACLED definition a fatality is not a necessary condition for an event. This means they it has a more granular coding of event types such as battles, violence against civilians and riots, but also non-violent ones such as protests and strategic developments. This opens up the possibility of defining target variables for conflict not related to fatalities, but instead with respect to events. Slightly adjacent to general conflict data and prediction is the field of predicting mass killings. While this section does not elaborate on these systems, two noteworthy projects are the Early Warning Project (a joint initiative of the United States Holocaust Memorial Museum and Dartmouth College) as well as the Atrocity Forecasting Project by the Australian National University.

The following provides an overview of prominent early warning systems that rely on target variables based on fatalities data from UCDP.[24] The Violence & Impacts Early-Warning System (VIEWS) focuses on predicting the number of fatalities per country

21 Clearly, organisations may collect their own data to inform interventions (e.g. UNDP, UN OCHA, UNHCR) or collaborate with local partners to gather on-the-ground information that is critical for humanitarian responses (e.g., IFRC, Welthungerhilfe, Start Network). Other monitoring efforts include automated event detection (e.g., POLECAT (Halterman et al., 2023) or GDELT) and event-specific tracking (e.g., Oryx in Ukraine). Whilst acknowledging this, we focus on ACLED and UCDP since they are publicly available, regularly updated, and have wide temporal/spatial coverage.
22 Source: UCDP codebook (https://ucdp.uu.se/downloads/ged/ucdp-ged-50-codebook.pdf).
23 Source: ACLED codebook (https://acleddata.com/knowledge-base/codebook/).
24 For a comprehensive overview, the reader is directed to Rod et al. (2024).

worldwide for 1-36 months into the future. This is also complemented by detailed geographic forecasts at the Peace Research Institute Oslo (PRIO) grid-cell level (0.5° resolution, ~55x55km at the Equator) for Africa and the Middle East. One clear benefit of this approach is that it permits the generation of uncertainty intervals.[25] In practical terms, this allows for making a statement such as "our forecast estimates country X will experience between 50 and 100 fatalities in Y months' time" as opposed to "our forecast estimates country X will experience 75 fatalities in Y months' time". Further, it enables an evaluation of 'tail risk'. This moves the prediction task beyond capturing just the most likely outcome and towards extracting information about the likelihood of rare, but possibly extremely damaging, outcomes. This could be a crucial step towards making quantitative forecasts of armed conflict relevant for the humanitarian sector. Conflict Forecast forecasts intensity but also utilises threshold-based definitions as a target variable. It predicts the likelihood that monthly fatalities will either exceed 0 or 0.5 fatalities per one million inhabitants in the next three or 12 months. The use of a per-capita measure in threshold-based approaches is critical to ensure the forecast is comparable across countries.[26] Conflict Forecast also predicts the likelihood that monthly fatalities will exceed zero in the next 12 months at the PRIO grid-cell level worldwide. Finally, the Patterns of Conflict Emergence (PaCE) takes a different approach to conflict forecasting using the UCDP data. PaCE aims to uncover recurring patterns and temporal sequences leading up to conflicts by analysing a wide range of data, including financial markets, news articles, diplomatic documents, and satellite imagery. Through machine learning methods, PaCE identifies motifs and sequences that precede armed conflict events, providing early warning and preventive action tools that deepen the understanding of conflict dynamics.

By contrast, the Conflict Alert System (CAST) relies on ACLED data to forecast event counts rather than conflict fatalities or conflict risk. The Water Peace and Security (WPS) partnership also uses ACLED data to make conflict predictions, but only for the African continent, West Asia, South and Southeast Asia. They also specifically learn from and predict ACLED events associated with fatalities, predicting two and twelve months into the future. Meanwhile CAST produces monthly forecasts of the number of battles, explosions/remote violence, and violence against civilians events for six months into the future. Forecasts are made at the subnational (admin1) level, and the model includes a host of conflict features including neighbouring violence, non-violent events, seasonality, and the characteristics of the most active actors in the area. This is especially valuable for policymakers who want to focus on a wider scope of events and do not wish to limit their definition of conflict to only battle-related deaths. For example, Section 2 raised the distinction between slow and rapid onsets. In the case of slow onsets, monitoring of protests and other non-fatal events might deliver the best forecast for fatal violence types.

It should be noted that the 'best' definition of a target variable remains an open debate across the conflict prevention and preparedness community.[27] Bridging the gap between quantitative forecasters and policymakers requires a two-way dialogue, and the choice of target should be driven by the decision makers' use case. If the objective is to identify possible escalations/de-escalations in ongoing armed conflicts, then a direct prediction of fatalities (accompanied with uncertainty estimates) over shorter forecasting horizons is perhaps most relevant. Yet, if the aim is to inform strategic planning, development initiatives and/or soft power interventions, then longer-term forecasts that estimate the likelihood of exceeding a predefined level of violence might be more valuable.

25 A competition to generate uncertainty estimates of fatalities was launched in 2023 by the VIEWS team, with funding from the German Ministry of Foreign Affairs and PREVIEW. The final forecasts from the 12 participating teams are now available on the challenge's data dashboard (https://predcomp.viewsforecasting.org), where a running evaluation will be presented upon each monthly release of UCDP conflict data for the true future (July 2024 to June 2025).
26 For example, the probability of exceeding 25 deaths per year will have a widely different meaning for very populous countries versus small island states.
27 This was one of the motivations for the "Growing Together: Prediction, Prevention and Preparedness" workshop at the IAE (see https://econai.iae-csic.org).

CEPR

3.3 Natural hazard data and forecasting

Hazards related to the natural world have a much broader scope; alongside weather forecasts, early warning systems exist for hazards such as floods, droughts, earthquakes, landslides, tsunamis, volcanic eruptions, and others. Therefore, the number of forecasting systems is vast and a comprehensive overview is beyond the scope of this Policy Insight. Instead, we note a number of key features of the natural hazard early warning landscape:

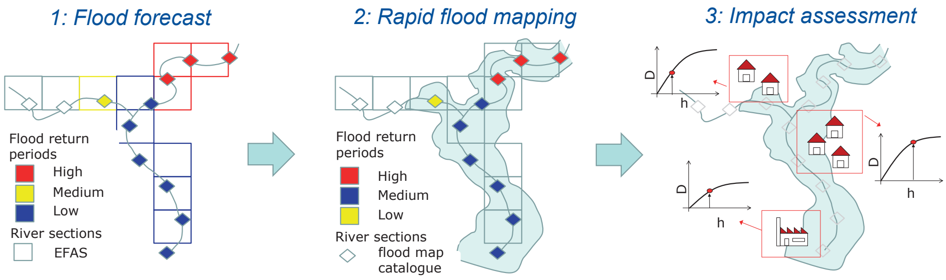
1. **Data availability**: The most frequently used data for natural hazards relate to meteorological variables such as temperature, rainfall, and wind speeds. These are widely available with high spatial granularity and long time horizons. In the case of fast-onset crises, accurate monitoring can be sufficient to enable 'analytical' decision making.[28] However, two clear gaps exist. The first is a comprehensive database that records historic and existing mitigation practices. It is simple for a model to learn that precipitation over and above the historical average is a strong proxy for a possible flood. It is much harder, without the aforementioned data, for it to understand that this flood will never materialise due to good policy (e.g., irrigation systems). The second is with respect to 'out-of-sample' forecasts, which are required to train and evaluate machine learning forecasting systems.[29] Access to these data is a critical barrier to developing forecasting systems for slow-onset crises.
2. **Disparate systems**: Even for a singular hazard (say, droughts), there exists a multitude of systems covering different geographies and types of drought. This not only highlights the need for a general consensus on the definition of a crisis for specific hazards, but also makes it challenging to make cross-geography comparisons. The Global Drought Information System (GDIS) is a rare effort to "pull together the best non-prescriptive drought information from local providers and provide an "apples to apples" comparison of drought conditions around the world".[30] However, the GDIS's consolidated database of drought forecasts starkly demonstrates that sophisticated technologies are only possible for high-income countries.
3. **Multi-hazard early warning systems** (**MHEWS**): The sector has recently redoubled its efforts to build MHEWS in response to the UN Secretary General's Executive Action Plan of the Early Warnings for All (EW4All) initiative. Specifically, these aim to cover multiple hazardous events to "minimize inefficiencies, maintenance costs, and duplication, and maximize investments in awareness, education, and preparedness" (UNDRR, 2023). UNDRR highlights that 101 countries now have multi-hazard early warning systems, but with significant regional/income disparities and only one-third of member states having multi-hazard monitoring and forecasting systems. Whilst not a forecasting system, the Global Disaster Alert and Coordination System (GDACS) is a prominent example of monitoring across multiple hazards.[31] It provides a global map of disaster alerts in the last four days, their expected impact on the affected population, as well as an integrated platform to coordinate AA and crisis response on the ground.

28 See Section 4.5 for best practice implementations of anticipatory action in the natural hazard domain.
29 For example, easily accessible data records actual/observed measures of temperature. Yet, forecasting at any point in time necessarily implies uncertainty about the future. Say we are standing in July 2024 and want to forecast the probability of a drought in the next three months. We need forecasts of temperature in the next three months, since we do not have access to future observed values. Training a good machine learning model requires iteratively rolling through time, 'pretending' that we do not know the future, making a prediction and evaluating the performance. Therefore, a long history of 'out-of-sample' meteorological forecasts is required.
30 See https://gdis-noaa.hub.arcgis.com/
31 See https://gdacs.org

3.4 Impact data and forecasting

Impact-based forecasting implies moving beyond early warning systems for hazards (whether that be conflict-related or natural hazards) towards directly forecasting their potential impact. In the case of natural hazards, forecasting impact (e.g., population in need) is relatively commonplace. For example, the Global Flood Awareness System (GloFAS) from Copernicus complements its flood forecasts with extent mapping and exposure information to assess regional impacts (see Figure 6). Considered exposure information includes population, infrastructure (health, education, airport facilities), and land cover. Another example is FEWS NET, which uses a scenario-based methodology to make eight-month-ahead projections of acutely food insecure populations. Similarly, the Anomaly hot Spots of Agricultural Production (ASAP) project aims to provide timely monitoring of possible crop production anomalies through frequent updates.

Figure 6 Global Flood Awareness System Rapid Risk Assessment by the earth observation component of the European Union's Space programme Copernicus



Source: Dottori et al. (2017).

By contrast, the conflict forecasting community generally lacks similar solutions. The main difference between the two is that *natural hazards are exogenous* to dynamics in human society, whereas armed conflict is a direct result of these dynamics, i.e., *armed conflict is endogenous*.[32] This not only complicates conflict forecasting but it even makes the study of conflict impacts difficult – for example, has conflict reduced GDP or has the agricultural crisis that triggered the conflict reduced GDP?

Given the endogeneity of conflict, predicting armed conflict and acting on forecasts has markedly different consequences compared to natural hazards. We discuss the ethical concerns related to decision making using forecasts in Section 4.1; here we discuss the practical implications. First, we cannot control the weather, so mitigation is the primary intervention instrument in the natural hazard domain – for example, excessive rainfall will lead to a flood irrespective of short-term responses but, given an accurate prediction, decision makers can work to minimise the impact on the affected area. Yet, effective policy in the conflict space can actually affect whether violence materialises or not. Table 3 summarises the difference.

32 This is not to deny huge endogenous aspects of natural hazards. Human decisions, including where and how we build infrastructure, manage resources, address socioeconomic factors like poverty and inequality, make land use and urban planning choices, and contribute to climate change, play a critical role in shaping the vulnerability and exposure of populations to natural hazards by influencing both their frequency and severity.

Table 3 Characterisation of armed conflict and natural hazard for forecasting and response

Aspect	Natural hazards	Armed conflict
Origin	Exogenous, triggered by natural processes.	Endogenous, driven by human actions and decisions.
Predictability	Often unpredictable, but patterns can be studied to anticipate occurrences.	Complex and dynamic, influenced by socio-political factors.
Response	Requires scientific and technical solutions, as well as community preparedness.	Involves political negotiation, peacekeeping, and humanitarian aid.

This leads to a host of political ramifications. For example, there is concern that public forecasts could be misused by violent actors/regimes or actually trigger conflict by amplifying the causes. Further, acting on forecasts requires accountability and ownership of the outcomes. This may point to the development of in-house, private systems, but organisations should leverage and collaborate with academics to leverage the significant progress made in this domain. Cross-country comparisons also open up political and economic sensitivities. A model from an international organisation that states risks are higher in country X versus country Y, and driven by factor Z, may damage relationships with respective governments. It also has subsequent impacts on financial markets due to signalling effects.

Hence, moving to directly forecasting impacts of armed conflict might help to bridge the gap between conflict forecasting and decision making. This may support reduced politicisation of the issue by directly quantifying humanitarian need. It may also more closely align with the processes of decision makers. Long-term policies and immediate crisis responses are often targeted and calibrated based on the affected population, rather than the expected number of deaths.

One example, and perhaps the most severe impact of armed conflict, is forced displacement. This refers to both cross-border refugees and internally displaced people (IDP). State-of-the-art forecasting systems for displacement are currently provided by the UN, which provides monthly updated nowcasts of country-level refugee stocks, and the Danish Refugee Council, which provides bi-annual updates to forecasts of total displacement for 26 countries in the next three years.[33] It is notable how forecasting systems of IDPs are largely out of the public domain, although the Internal Displacement Monitoring Centre (IDMC)[34] and International Organization for Migration (IOM)[35] are pioneering efforts to improve historical data and forecasting models.

Other possible metrics which may be useful for informing resource allocation for interventions, as well as contextualising predictions, could be forecasts of expected economic (e.g., destruction of infrastructure, loss of human capital), environmental (e.g. loss of biodiversity, crop land), health and social (e.g., child health (Tapsoba, 2023) or sexual violence against women) impacts. In fact, significant progress has been made in this area, typically relying on the use of satellite imagery. For example, Mueller et al. (2021) demonstrate the possibility of automating detection of war-related building destruction in the Syrian civil war. UNOSAT is perhaps the leader in this field, with open-source products tracking, for example, road network damage in

33 See https://pro.drc.ngo/what-we-do/innovation-and-climate-action/predictive-analysis/foresight-displacement-forecasts/. A worldwide refugee forecasting system developed by the EconAI group, based on public UNHCR, UCDP, news text and Google Search data is operational at the end of 2024.
34 See https://www.internal-displacement.org/internal-displacement-updates/
35 See https://www.arcgis.com/home/webmap/viewer.html?webmap=7ba4279889c74b2d926b9b5202c46c4c

Gaza resulting from the most recent conflict[36] and flooding extent in Bangladesh.[37] However, none of these potential outcome variables currently has the global coverage that could inform a generalised forecasting system.

Additional prominent examples of impact-based forecasting include Montandon and GO from the International Federation of the Red Cross (IFRC). Montandon, a global crisis database, collects information on observed and forecasted hazards, their impacts and the response (by whom, where and to what effect). The data feed into decision-making processes through the GO platform, which helps National Societies prepare, respond and learn from crises. VIEWS is also moving the direction of impact forecasting[38] through a series of research projects that seek to extend its conflict prediction system with a number of impact components in the coming years (e.g., Societies at Risk, ANTICIPATE, and VIEWS-PIN).[39] These forecasts are also set to incorporate estimates of forecast uncertainty.

4 INTEGRATING FORECASTING INTO DECISION MAKING

The end of Section 2 already highlighted potential benefits and concerns regarding the use of forecasts in the humanitarian sector. This section expands on some of the ethical and practical issues in developing suitable forecasts and integrating them into decision making. This includes a consideration of the potential audiences for forecasts, as well as what this implies for target choice, model evaluation, and the interpretation of forecasts in humanitarian AA. The section closes with examples of existing practice.

4.1 Ethical considerations

The issue here is not terminator-style autonomous AI systems running wild or predictions at the individual level that come with huge ethical concerns. The machine learning systems deployed in armed conflict and displacement forecasting typically deploy simple methods like random forest algorithms or simple neural network architectures, which essentially produce sophisticated summaries of the available data. Nonetheless, serious ethical concerns remain when integrating quantitative forecasting systems into decision making.

According to the Fundamental Principles of the International Red Cross and Red Crescent Movement, proclaimed in Vienna in 1965, humanitarian action must adhere to the principles of humanity, impartiality, neutrality, independence, voluntary service, unity, and universality. IASC (2024) summarises the principle of humanity as follows:

> *Human suffering must be addressed wherever it is found. The purpose of humanitarian action is to protect life and health and ensure respect for human beings.*

This has direct consequences for the use of forecasting systems. For example, the principle prohibits organisations working in this area from making a forecast public if this would have a significant risk of negative humanitarian consequences. This is a key constraint to keep in mind in the design of decision-making systems that include quantitative displacement forecasts.

There are obviously legal limits to the use of AI, but they are unlikely to bind for development. The EU AI Act, for example, categorises AI systems into four risk groups – unacceptable, high, limited, and minimal risk – and imposes stricter requirements on higher-risk systems. Two key areas that overlap with displacement forecasting are

36 See https://unosat.org/products/3883
37 See https://unosat.org/products/3897
38 See https://viewsforecasting.org/news/views-awarded-crafd-funding-for-cutting-edge-innovation-to-forecast-the-need-for-humanitarian-assistance/
39 See https://viewsforecasting.org/research/sar-anticipate/; https://viewsforecasting.org/research/sar-anticipate/ and https://viewsforecasting.org/research/views-pin/

migration, asylum and border control management (such as automated examination of visa applications), and the administration of justice and democratic processes (such as AI solutions for searching court rulings). These applications could lead to the misuse of quantitative displacement forecasts, necessitating the controlled dissemination of forecasts. This justifies the regulation of access to such forecasts, either through in-house production or careful screening of external users. According to the EU AI Act, high-risk AI systems must be transparent about their operation and data usage, aligning with the need to communicate forecast limitations and methodologies clearly. The regulation emphasises human oversight to ensure decisions are not solely automated, and it incorporates ethical principles like respect for human dignity, non-discrimination, and fairness.

In what follows we will assume that quantitative forecasts can play a pivotal role, i.e., that they have the credibility to influence decisions of some decision maker (local or foreign) that is exposed to the forecast (directly or indirectly). But how does information become pivotal? It matters most where the recipient of the forecast does not have any prior information or is willing to trust the forecast more than other information. Putting it differently, quantitative forecasts can play the role of levelling the playing field in which the least informed benefit the most. However, the communication and accessibility of the forecast plays an important role. Depending on whether forecasts are open or kept from the public, they will trigger different information updating.

The key distinction to be made when deciding on communication is whether forecasts are self-reinforcing or self-defeating. A self-fulfilling forecast is a prediction or expectation that causes itself to become true due to the behaviour it inspires. In other words, the forecast influences actions that lead to its fulfilment. A self-defeating forecast is a prediction that prevents itself from becoming true due to the behaviour it inspires. Self-defeating forecasts lead to actions that undermine the predicted outcome. When predicting negative outcomes like armed conflict or displacement, the general rule is that self-fulfilling predictions need to be disseminated with care, whereas there is a moral imperative for self-defeating forecasts to be communicated. However, there are exceptions to this rule. A self-fulfilling forecast of displacement could be worth communicating if it prevents civilian fatalities; a self-defeating forecast of displacement should not be shared if the reason is that actors prevent the local population from fleeing.

Two case studies illustrate how differential information updating can trigger self-fulfilling prophecies. In an attempt to prevent armed conflict, a well-meaning outside actor might announce that they will monitor violence in a set of villages through satellite imagery and disseminate the images through an easily accessible webpage. This can trigger more attacks on the monitored villages if local armed actors are motivated in their violence by an attempt to displace local populations or the attention they gain from outsiders. Note that in the example it is the fact that the information of the violence is disseminated through the webpage, visible to everyone, that triggers an increase in violence. Similarly, forecasts of violence can trigger more violence. If, for example, forecasts are only conducted in hidden systems and sold to outside investors who have an incentive to protect their investments, then this withdrawal of resources can make existing conflicts worse.

Bias is a serious issue for any forecasting system. However, which biases matter most is often misunderstood. By far the most serious issue is biased training data. If forecasting systems are trained on labels that give more weight to some types of conflict and ignore other conflicts, then this bias will be reflected back to the user of the forecast. If some type of displacement is not reported by officials for political reasons, then it is nearly impossible to forecast. Data quality for the outcome to be predicted and a good definition of this outcome are absolutely key when getting into quantitative forecasting, and using biased training data can have serious consequences. Biased data for predictors is also a concern, but this bias will typically show up in performance measures of the system if the target data are correct. It is possible to explicitly test for biases in the data on predictors. In summary, biased predictors (features) will be flagged to the user by the system, biased outcomes (labels) will bias the entire forecasting system and can remain undetected. The two main data providers in this area – ACLED and UCDP – are under constant scrutiny regarding their coding methods and reporting biases. It is crucial that the academic community is able to uphold this scrutiny.

Existing armed violence forecasting systems like CAST, Conflict Forecast, PACE and VIEWS are all public, but some monitor access for intense users. The implicit belief is that levelling the informational playing field will trigger more self-defeating prophecies than self-fulfilling ones and that making information available is therefore an ethical imperative. However, it is not clear whether this would hold for impact-focused forecasting models that predict impacts such as displacement.

4.2 Forecast audience

Integrating forecasts into decision making requires a good understanding of who will use them. In practice, there are often substantial inter-disciplinary and individual differences in the target audience for forecasts. Such differences include the perception of what a 'good' forecast is and what it can be expected to deliver, as well as the prioritisation and willingness to act upon forecasts. There may also be differences in what determines the level of trust a decision maker places in quantitative modelling outputs like forecasts. Some may be particularly interested in the selection of input data or a model's explainability, while others will place more importance on how well a model performs on different evaluation metrics. Sometimes what matters even more is the personal relationship between the person communicating the forecasts and the receiving decision maker, their alignment of world views and ways of understanding reality (e.g., *analytical* versus *intuitive*), or their ability to bridge gaps in these views. In some cases, efforts to build trust in and use forecasts are hampered by a broader institutional culture, predefined procedures, and communication norms. Depending on the audience, it may also be especially critical that the final data product comes with high-quality and appealing visualisations that are perceived to be intuitive, with an appropriate level of complexity or customisability of how results are displayed.

Different audiences are interested in and able to engage with different levels of complexity of forecast outputs. Data-savvy people might want more nuance and higher-quality forecasts, while others may want a simple map or a single number with straightforward interpretability to make quick decisions or to rank countries. Especially among the latter, data literacy may be limited, which makes it harder to build trust in forecasting systems due to subjective perceptions of how well a forecast is performing. A probabilistic forecast's performance is most meaningfully evaluated in aggregate, considering many events and predictions over time. However, it is more intuitive for many to focus only on the now, evaluating forecasting performance on the basis of key single events (for example, by asking whether the Sudan crisis was predicted). There is not even a single correct answer to this question, for it depends on what threshold we set for deeming a forecast to indicate an 'event' (for example, in terms of what probability of conflict onset or how many fatalities were predicted).

4.3 Trade-offs in choosing targets and evaluating forecasts

Before a forecast can be integrated into decision-making processes, it needs to be decided what is being forecast and how the forecast is evaluated. It may be that a single forecast cannot meet all the needs of a decision maker. Thus, the trade-offs outlined in this section can also be read as an invitation to work with more than one target definition (although this in turn increases the amount of information that needs to be processed by decision makers).

Sections 2 and 3 already touched on the importance of target choice. Here the focus is on the effect the target choice has on the ease with which results might be interpreted. The first decision is about opting for a regression or a classification task. A benefit of the classification task is the bounded output of a probability between 0 and 1, which is quite unambiguously comparable across countries. A regression target is usually defined on a per capita basis to make it comparable, although this makes it somewhat susceptible to inaccuracies in population data. Furthermore, in absolute terms, one fatality per 100,000 implies more fatalities for a larger country. For example, for Chad, with a population of less than 20 million, one fatality per 100,000 inhabitants implies fewer than 200 deaths, while for Nigeria, with a population of around 220 million, it implies 2,200 deaths. In practice, classification and regression values might be forecast alongside each other. For example, Conflict Forecast highlights the risk of outbreak of conflict in countries that have not currently crossed the threshold definition of being in armed conflict, but then displays intensity (i.e., deaths per capita) forecast values for countries already deemed to be in armed conflict. Such thresholds for being in armed conflict or civil war have been a matter of debate and will always have an element of arbitrariness. However, this choice cannot be avoided – forecasting continuous values or continuous log values will also direct attention of the algorithm to specific parts of the distribution.

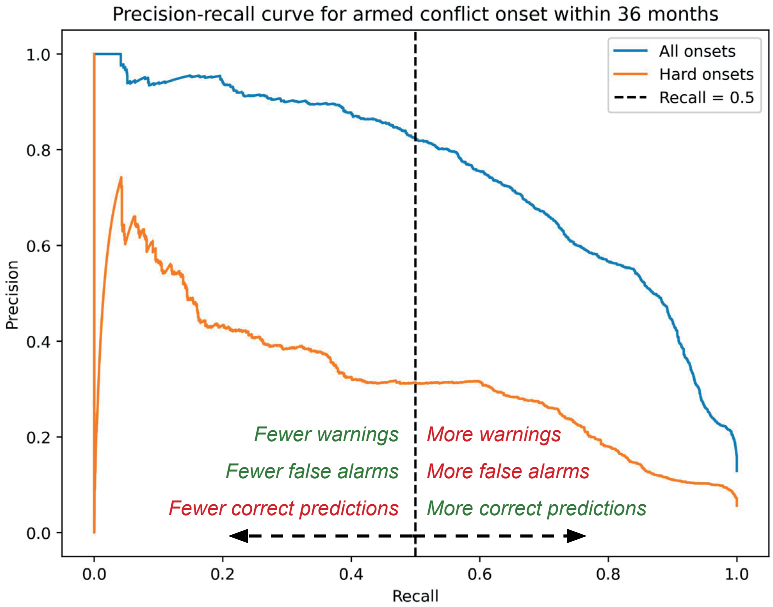
If one chooses to work with a classification target, there are further decisions to be made. Again taking the example of conflict, there are options to predict 'incidence' (i.e., the occurrence of conflict) or 'onset' (i.e., the first occurrence of armed conflict in a series of conflict-periods in a given location). Arguably, onsets are more relevant for decision makers, not least because incidences are easier to predict due to the conflict trap: once a country is in conflict it is likely to remain in conflict for some time. Alternatively, there could be an 'escalation target' where an event is defined on the basis of an increase in fatalities. Afterall, a conflict may have started with relatively low levels of deaths, but decision makers may still be interested in receiving an early warning if the conflict has a high risk of becoming more violent; it would still need to be decided *how much* more violent is deemed a suitable escalation threshold.

The trade-offs regarding spatial resolution and temporal horizons are quite straightforward and thus not discussed at length here. The question of spatial resolution is typically answered by data availability. Even if the target variable is available at a subnational level, it might be that the most relevant predictor variables are only available at the national level. In that case, there would not be enough variability in the predictors for an algorithm to learn relevant correspondences. Then a simpler prediction tool such as autoregressive models might be more appropriate to extrapolate from a single time series into the future. Choice of spatial resolution (as well as data availability) also depends on whether a decision maker is more interested in highly localised predictions because they are operating in a single country or region, or whether they are making allocation decisions at a global level with a need to compare countries. The most suitable forecast horizon will also depend on the use case as mentioned in Section 3. There are still different ways of generating a value of, for example, fatalities within the next three months. One option would be to simply train the model on a transformed target that is "deaths within the next three months". Another could be to train a model on the exact number of deaths in one, two, and three months and to then sum the predictions.

A final trade-off relates to what evaluation metric to optimise for. An important step in generating forecasts is so-called hyperparameter tuning. A hyperparameter is a model argument whose value is set before the learning process begins. What hyperparameters are available for tuning depends on the algorithm chosen – for example, in tree-based models there is an option to set the maximum depth of a decision tree. Hyperparameter tuning involves creating predictions with different hyperparameter configurations to determine which combination leads to the best performance. However, this requires a decision on what makes one set of predictions better or worse than another.

Are false alarms preferable to missed alerts? This question captures the trade-off between precision and recall as two popular evaluation metrics for event detection classification tasks (although there are many more, including combinations of these two). The trade-off is visualised in the precision recall curve shown in Figure 7. The y-axis shows *precision*, which is the measure of accuracy concerning positive predictions, specifically reflecting the proportion of cases predicted as positive that are indeed correct. High precision indicates that, when the system identifies a case as positive, it is likely to be accurate. The x-axis shows *recall*, which measures the system's ability to identify all true positive cases, reflecting its completeness. High recall signifies that the system is effective at capturing all actual positives, even if it may include some incorrect positives among them.

Figure 7 Illustration of precision recall trade-off from Conflict Forecast



In summary, precision emphasises *accuracy* of positive predictions, while recall focuses on the *coverage* of actual positive cases. If one is particularly keen to predict as many relevant events as possible, i.e., not missing alerts, then recall will be the preferred metric to optimise. Precision is more suitable if one wants to minimise the chances of false alarms. Which of these is preferable will depend on the decision at hand. In some contexts, false alarms could be costly (for example, due to resources being irretrievably allocated). However, missing relevant events is often even more costly and a few false alarms (i.e., a slightly lower precision) may be an acceptable price to pay in return for higher recall, which is usually the approach taken in medical prediction contexts.

In practice, the appetite for false positives (i.e., low precision) in policy circles is often low. The reason is that spending on false positives is hard to justify organisationally or to the donor community. This means that there is a tendency to shift towards the left in Figure 7, i.e., to become more conservative and act less often on early warning signals. This trade-off is amplified when predicting hard onsets after long periods of peace,[40] as illustrated by the orange curve in Figure 7.

For regression tasks, there is also a host of evaluation metrics to choose from, including variations of the mean squared error (MSE) or mean absolute percentage error (MAPE). Choices will depend on factors such as whether large deviations between prediction and target should be punished and whether the error should be in the same unit as the

40 See Mueller and Rauh (2022), who discuss that outbreaks after long periods of peace are hard marchine learning problems.

target. A unique feature of regression tasks is that they can be reconfigured to predict distributions rather than making mere point predictions. This additionally captures how uncertain a model is about the prediction. It also expands the types of evaluation metrics that can be calculated, including the continuous ranked probability score (CRPS) or mean interval scores (MIS). The scoring committee for the 2020 VIEWS prediction competition, which tasked participating teams with predicting changes in fatalities from state-based armed conflict, used a suite of different metrics when evaluating the submissions. The organisers also came up with Targeted Absolute Distance with Direction Augmentation (TADDA), a new metric specifically developed for evaluating forecasted change in fatalities (Vesco et al., 2022).

Finding good, intuitive benchmarks is often key. A common benchmark in conflict forecasting is the 'no-change' model. A no-change model is a one that simply predicts the same value for the next time period as was observed in the current or most recent time period. Benchmarks like these illustrate situations in which quantitative forecasting models need to beat more simple heuristics, and often these models are surprisingly hard to beat.

4.4 Interpreting and using forecasting results

A key question is how forecasts can smoothly integrate into decision making and how the additional information can meaningfully interact with existing expertise and insights. This also relates directly to how results are communicated – between people or from a screen to a human. While it lies beyond the scope of this Policy Insight and the expertise of the authors to comprehensively refer to the literature on how to effectively embed forecasts into decision making, this section will draw on selected findings alongside the experience of those who attended the workshop. It should be stressed that we do not see humans and algorithms as being in competition with each other. As mentioned in Section 3, the aim is for numerical predictions to act as an additional source for the triangulation of information in decision making.

When interpreting and using forecasts it is critical that they are not seen as uncovering causal relationships. This is a common error when sharing forecast values with a wider audience. Even simple linear regression models risk being misinterpreted this way, although the catchphrase of "correlation is not causation" may still be familiar in that context. Section 3 already touched on the issue of better-performing models being less explainable and the attempts to mitigate this through feature importance or SHAP value analysis. Such explainability tools can be misleading if they are meant to serve as proxies but are taken at face value. The values provided also only speak to their relevance of predictor variables in the model, not the real-world scenario. Afterall, there are likely to be factors in the real world that relate (some even causally) to the target variable that are not captured in a dataset for the model to take into consideration. Exposing causal relationships in non-experimental data from outside lab settings is hard and is at the core of what academics in economics, political science, and other disciplines spend their careers doing. In fact, the state of the art in causal modelling, by focusing on correctly estimating causal effects (even if just for one variable), is orthogonal to predictive performance (Ward et al., 2010; Shmueli, 2010; Pearl, 2009).

Policymakers are often vocal about the need for forecast outputs to be simple. Some advocate for a single text line or number, others for the use of linguistic probability categories such as "very likely" or "high risk". What is suitable in practice will strongly depend on who is digesting the forecast for which purpose. When using forecasts in practice there is sometimes a step in which a single forecast value reaching a critical threshold leads to an alarm and a series of actions are triggered. For such a use case, a single number is useful. But the process of building trust between the intuitive and analytical members of an organisation requires additional information available to those who want to see beyond the simple single number output. Sometimes there may be a skills gap between what a decision maker could ideally take into account and what they are currently able to do when using a forecast. Thus, it is worth considering whether there is a medium-term need for training to improve data literacy rather than seeking to further simplify forecasting outputs. A particular example is a forecast's uncertainty, which could be key to the weight it is given in decision making. However, forecasts with uncertainty outputs are also somewhat harder to interpret and might warrant additional training for decision makers.

The use of linguistic probabilities in place of numeric probabilities has been questioned in research on the communication of probabilities. In the context of the communication of intelligence, Dhami and Mandel (2021) summarise their finding as follows: "Numeric probabilities are not without drawbacks (e.g., they are more difficult to elicit and may be misunderstood by receivers with poor numeracy). However, these drawbacks can be ameliorated with training and practice, whereas the pitfalls of linguistic probabilities are endemic to the approach." For example, people's interpretations of probability terms may be vague, different from person to person and even for one individual over time or in different circumstances. This could be addressed through a so-called lexicon that matches probability intervals with probability terms, although they would ideally undergo extensive testing to fine-tune the matches (at least for one language and context). However, even then issues remain. It is difficult to combine the interactions of probabilities if they are no longer numerical and the terms are coarse. Furthermore, there is a problem of directionality in communication if more positive ("some chance") or more negative ("doubtful") terms are used for the same low probability, which might be interpreted as an implicit policy recommendation.

Once a decision maker sees a numerical forecast, this is likely to have an anchoring effect. This raises the question of whether a human should be explicitly asked to 'commit' to a numerical forecast of their own so they can compare their own and the algorithm's perspectives independently. In practice, this may be challenging to implement for a variety of reasons, such as a reluctance of humans to make such predictions and commit them to a record, and it may be impractical to keep the algorithmic prediction a secret until then. If the aim is to make the humans better at forecasting, this can be achieved with target training (e.g., Mellers et al., 2015). It might then be interesting to combine the forecasts by machines and humans to see if we can get better results, especially if we think that humans and machines err in specific ways and thus neatly complement each other. Benjamin et al. (2023) conducted related lab experiments that included conditions in which human forecasters had access to different types of information when making their own forecasts. Results show that "skilled forecasters who had access to machine-generated forecasts outperformed those who only viewed historical data". Thus perhaps the issue of anchoring is not a hurdle and the focus instead needs to be on how much information the decision maker(s) at hand can meaningfully make use of. It should be noted that Benjamin et al.'s paper was specifically about crowd forecasting, which is somewhat different from the small-team or individual decision makers who might digest forecasts in a policy context. Furthermore, the algorithms used were fairly simple univariate time series models, where humans are more likely able to produce a reasonable forecast. The usually multivariate forecasting systems at the core of the current Policy Insight may have different effects on human predictions. Buy-in and reluctance towards using data and model-based forecasts was also less of an issue in Benjamin et al. (2023) due to the artificial lab conditions, but it is a significant issue in real conditions, as workshop participants reported from their own experiences.

For the effective use of a forecast in decision making, the numerical values may need to be contextualised according to potential consequences. This relates back to the use of forecasts to inform scenarios as well as to the discussion from Section 3.4 on forecasting the hazard or the impact. Jean et al. (2023) find in the context of integrating probabilistic flood forecasts into a decision-making process that contextualising forecasts with their potential impact implications is key. They further recommend that there should ideally be ongoing communication between forecasters and decision makers and that decision makers should receive training on the correct interpretation of (in their case, hydrological) forecasts. Such communication channels may be difficult to maintain in practice if the forecasts are not generated in-house. However, discussions about how to show potential harm alongside forecasts (or directly forecasting impact) and how to effectively equip decision makers with relevant skills are still highly relevant. There are many ways in which a forecast can be contextualised, one being to draw parallels between situations. This lies at the core of the conflict forecasting approach of PaCE (see Section 3.2), whereby the trajectories of countries are compared and used for prediction. This can in itself be a form of communicating context as it gives decision makers past situations to look at that the model has identified as being similar to the current one (Chadefaux, 2021).

A final question is how forecasts are shared and displayed and by whom – with a focus on practicality rather than the ethical concerns already covered in Section 4.1. Some forecasts are publicly available and can only be accessed through dashboards without having access to the underlying data or being able to easily download historical forecasts, while others may provide an application programming interface (API) to allow users to access the data with ease and incorporate them into internal monitoring or forecasting systems. Which method is preferred will depend on the user and how much time or technical expertise they have in their team. Offering both could be ideal, but maintaining dashboards in addition to generating the underlying forecasts costs money and may be a burden to academics and others providing the data. Furthermore, designing such dashboards and interfaces well requires a different kind of expertise, and there may not be sufficient budget for deliberation over aspects of user interaction and experience (UIX), even though it matters. In an experiment regarding the visual communication of tornado risks, Sutton and Fisher (2021) find that the "use of colour, properties of text presentation, and contents of messages affect attention allocation". Thus, the effective interpretation of forecasts and their integration into decision making may also depend on the subtle ways in which these forecasts are visually presented.

Integrating forecasts into decision making is a challenging task and there is no 'one size fits all' approach. Especially in large organisations, there is a great variety of potential audiences for a forecast. In contexts like UN agencies, there are still many questions to address regarding the systematic integration of insights from forecasts with intelligence from people on the ground in Country Offices. However, there are many efforts to make forecasts readily available wherever it is ethically sound, to come up with meaningful metrics, and to make the outputs of black box models more interpretable.

4.5 Examples of existing practice

The use of forecasts in humanitarian and development decision making is no longer hypothetical. This section highlights the experiences of six organisations: the IFRC, OCHA, Start Network, UNDP, UNHCR, UNICEF, and Welthungerhilfe.[41] The focus in this context is on how data and forecasting models are used and how they inform decision making and the distribution of funding. A catalogue of predictive models in the humanitarian sector has already been curated by the OCHA's Centre for Humanitarian Data[42] and several other reviews of forecasting and early warning systems exist. We first highlight systems that have already integrated quantitative early warning into decision making. We then discuss systems currently under development.

41 Most of these organisations were present at the workshop.
42 See https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/

4.5.1 IFRC

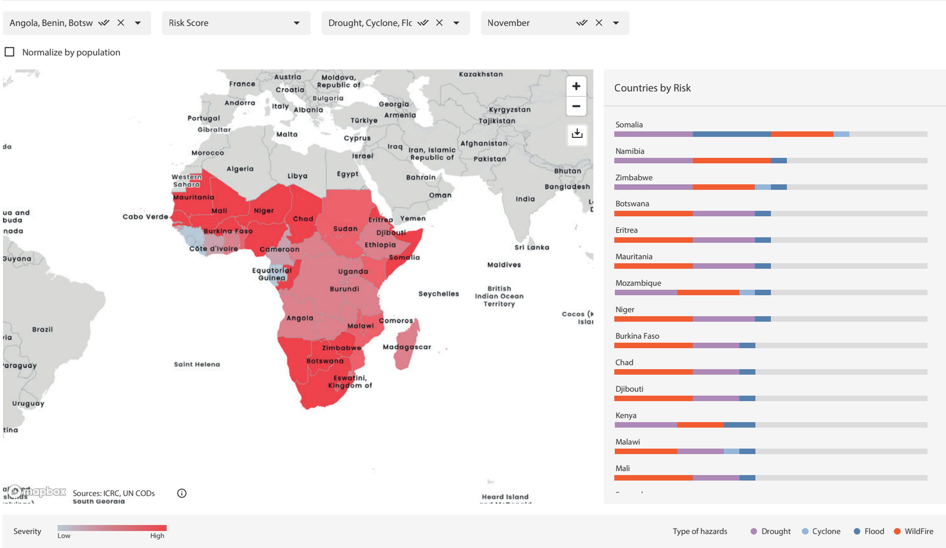
Anticipation funds and systems: The IFRC has two main methods to fund emergencies: the Disaster Relief Emergency Fund (DREF) and Emergency Appeals. Anticipatory action is largely funded through the DREF. The Anticipatory Pillar of the DREF is a fast, reliable and efficient way of getting money to the Red Cross and Red Crescent societies in anticipation of disasters. It helps them save lives and reduce, or even prevent, the damage and losses caused by disasters in communities. It uses a forecast-based financing approach: based on meteorological forecasts and risk analysis, the IFRC provides funding to national societies for early action in advance of a predicted hazard. Money is then released automatically when pre-defined forecast thresholds, or 'triggers', are met.

The key vehicle to establish the automatic release of funding are early action protocols (EAPs). These are formal plans produced by national societies. They outline the early actions that will be taken when a specific hazard is forecasted to impact communities. The EAP provides pre-approved funding for up to five years and includes (i) pre-positioning of the stock needed to enable early actions; (ii) annual readiness activities so the national society is prepared and on standby to respond; and (iii) pre-agreed early action activities designed to save lives and livelihoods once a hazard is forecasted.

Simplified early action protocols (sEAPs) are a lighter approach available to Red Cross and Red Crescent societies, also drawing on the principles of forecast-based financing. These are simplified plans with a lower budget and a shorter life span than a full EAP. A national society can have up to three sEAPs, each focusing on one hazard.

Data used: The IFRC's Risk Watch was launched in 2020 to monitor hazard information and disseminate relevant warnings to the 191 Red Cross/Red Crescent national societies. The evidence and analytical tools that form part of this risk module are hosted on the IFRC's GO platform (see Figure 8). They are public goods and use the best-available open-source, open-access data from IFRC, national societies, as well as partners from governments, academia, the UN, the private sector and other international organisations. The list of data sources for hazards and impacts is long and continues to expand as more sources are added to the GO platform. The platform also hosts information about responses, anticipatory action measures, and their respective impacts over time.

Figure 8 IFRC Go Africa regional page on 28 November 2024, showing risk across October, November and December



Note: This is the "risk score" view of seasonal risks. Alternatives are "people exposed" and "people at risk of displacement". Risk Watch also collates data to show imminent events, such as floods, storms, earthquakes, droughts or wildfires.

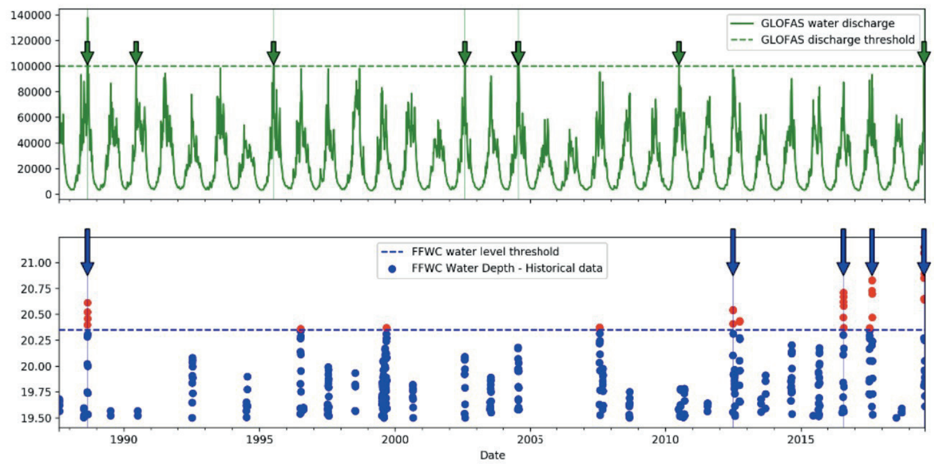
Overall seasons risk scores per country shown are taken from the INFORM Risk Index but additionally accounting for seasonality, while population exposure is based on UNDRR data, expected displacement figures from the IDMC, and food insecurity risk from the Integrated Food Security Phase Classification (IPC). For this, the IFRC draws on data from the Pacific Disaster Center (PDC), the WFP's Advanced Disaster Analysis and Mapping (ADAM), and the Global Disaster Alert and Coordination System (GDACS).

Examples of actions taken: Eight EAPs were activated in 2023 (compared with seven activations in in 2022), consisting of six full EAPs (Zambia Floods, Ecuador Floods due to El Niño, Uganda Floods, Kenya Floods, Zimbabwe Drought, Honduras Drought) and the first two activations of EAPs (Djibouti Floods and Kazakhstan Cold Wave). The rate of growth of EAPs is doubling each year.

4.5.2. OCHA

Anticipation funds and systems: OCHA's pilot framework for collective anticipatory action in Bangladesh, initiated in 2021 and updated in 2023, targets severe monsoon flooding of the Jamuna River (Figure 9).[43] It integrates a two-stage trigger system using the GloFAS and Flood Forecasting & Warning Centre (FFWC) models to predict flood risk and activate funds. The pilot aims to reach up to 440,000 people in five vulnerable districts. CERF funding, alongside collaboration with the Red Cross and government, facilitates pre-arranged, multi-sectoral interventions.

Figure 9 Historical analysis of triggers in the GloFAS and FFWC models for the anticipatory humanitarian action pilot for the 2020 monsoon floods in Bangladesh



Data used: The model employs forecasts from GloFAS and FFWC with a two-step trigger process. At the heart sits a comparison of current water discharge and water levels with historical data, shown in Figure 9. The team then produce estimates for the appropriate thresholds for the activation of the anticipatory action framework, so that the activation would only occur in events with a severe impact. Teams from OCHA, IFRC, and BDRCS discuss the findings of the analysis and revise thresholds in the framework based on the Centre's analysis and recommendations.

Examples of actions taken: In 2020, the system allocated $5.2 million in anticipatory funds to aid 200,000 people before flooding. The 2023 pilot further expanded, targeting over 400,000 beneficiaries with cash transfers, hygiene kits, and early warning messages in collaboration with NGOs and government agencies.

43 See https://centre.humdata.org/anticipatory-action-in-bangladesh-before-peak-monsoon-flooding/

4.5.3 Start Network

Anticipation funds and systems

Start Network has two funds available to the almost 100 Start Network member organisations to access funds for anticipated crises:

1. The global Start Fund, established in 2014, disburses money within 72 hours of a crisis alert, making it one of the quickest humanitarian financing mechanisms available. It targets three specific areas: underfunded small- to medium-scale crises, forecasts of impending crises, and spikes in chronic humanitarian crises. Start Network members can utilise event forecasts and impact predictions to raise a Start Fund anticipation alert. The Start Fund accepts anticipation alerts for any risk that could have humanitarian impacts, including flooding, heatwaves, cold waves, displacement due to conflict, electoral tensions, disease outbreaks, drought, volcanic activity, and tropical cyclones.

Figure 10 Start Network alert cycle

Note: Figure shows the process of the Start Fund alerts.

2. Start Ready, launched in 2022, pre-positions financing for recurring and predictable crises such as floods, droughts, and heatwaves. Start Network member networks collaborate at a national level to design and establish a disaster risk financing system, which involves analysing risks and developing contingency plans. Funds are rapidly disbursed when pre-agreed thresholds (return period events) are reached in risk models and are used by Start Network members to implement pre-agreed plans for anticipatory and early response activities.

Data used: In the case of the Start Fund, member organisations can choose which forecasts of events and predictions about their impact to use when raising an anticipation alert. In the case of Start Ready, NGO members that are contributing to the pool jointly agree upon which forecast and monitoring systems to use and what the critical set of thresholds are for funds to be released. When the crisis anticipation function of the Start Fund opened in 2016, the Global FOREWARN[44] community was created alongside it to support members to act ahead of the occurrence of hazardous events and their impacts. At a global and national level, FOREWARN aims to support Start Network members to identify the high-quality hazard and risk information

44 Forecast-based Warning, Analysis, and Response Network.

shared and produced by scientists and experts, and to operationalise this information to improve the quality of anticipation work. It also provides expert support to decision making for both the Start Fund and Start Ready funding mechanisms.

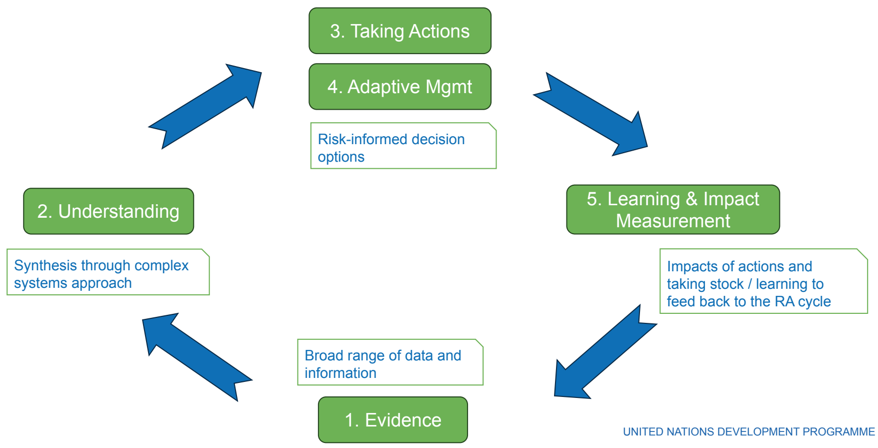
Examples of actions taken: In 2023, over 20% of alerts raised to the Start Fund were in anticipation of crises. Start Fund anticipation alerts for conflicts and violence have typically addressed well-defined situations with specific dates, such as elections or announced military actions. In the initial Start Ready risk pool for 2022-2023, 590,019 people across eight countries were protected from ten climate risks. Over 12 months, Start Ready was activated eight times, providing pre-agreed funding to NGOs for early response to drought in Zimbabwe, Senegal, and Somalia; heatwaves in Pakistan; a cyclone in Madagascar; and flooding in the Democratic Republic of the Congo. A total of £2,854,861 was distributed to 15 agencies, funding 19 projects and directly benefiting approximately 283,275 people before, during, and after the crises.

4.5.4 UNDP

Anticipation funds and systems: The recent (November 2023) update of UNDP Crisis Bureau's Standard Operating Procedures (SOPs) for Crisis Response and Recovery places crisis anticipation and prevention at their core. The new SOPs include, for the first time, a well-defined approach for crisis anticipation and prevention as part of risk-informed development, including institutional processes for dedicating resources (UNDP corporate assistance) based on anticipatory risk analyses. The recent update of the SOPs provides new opportunities for the use of conflict forecasting in risk-informed development.

The UNDP is moving away from a binary (i.e., right or wrong) approach to action towards a more cyclical approach shown in Figure 11. This allows for the gradual integration of new insights.

Figure 11 How UNDP makes sense of data, including forecasts, in order to inform action and build an evidence base for improved future action



Predictive data and forecasts generally fall into the first step of this cycle. However, once this information has been absorbed it can have impact at three different levels in the stages of taking action and adaptive management:

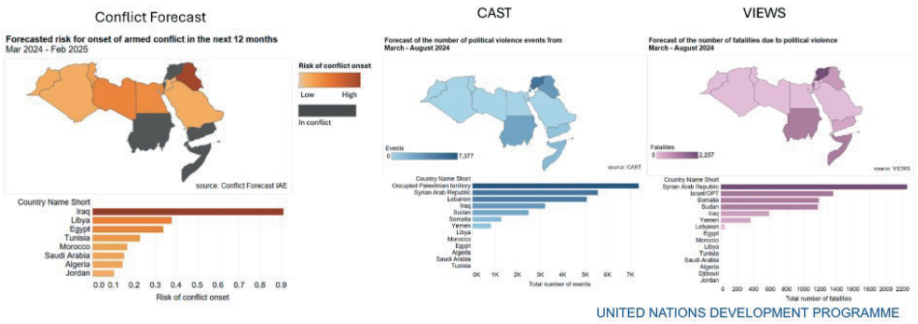
1. Preventive programming and funding: it can alter what is funded where.
2. Programmatic and operational capacity enhancement: it can shift the focus from post-crisis actions to preventive actions.
3. Policy advocacy to partners: it can affect messaging to internal and external partners and stakeholders.

Data used: The recently established Risk Anticipation Hub (RAH) in UNDP's Crisis Bureau plays a key role in realising the SOPs. It integrates risk analyses from different teams, providing data-driven analyses and data visualisations to country offices, and it strengthens forward-looking risk analyses based on predictive analytics and conflict forecasting.

The link to decision-making is currently under development. The following steps are currently taken:

1. Automated ingestion of a selection of publicly available conflict forecasts into an internal data warehouse that feeds a set of (interactive) data visualisation products to allow for use and integration of conflict forecasting data with other data sources around risk anticipation. Figure 12 illustrates the juxtaposition of several conflict forecasting systems in a Crisis Risk Dashboard for the MENA region.

Figure 12 Dashboard under development at UNDP for integrating conflict risk into decision making



2. Proof-of-concept for using conflict forecasting data as an additional (experimental) data layer for UNDP internal Regional Horizon Scanning processes to identify priority countries in all UNDP regions globally.
3. Ongoing conceptualisation for how to best include and aggregate different existing conflict forecasting products into one newly developed INFORM Warning risk index.
4. Evaluation of feasibility and usefulness of VIEWS forecasts to inform conflict analyses in an East African Country Office, with an ongoing innovation workstream around the use of (what-if/counterfactual) conflict forecasting runs for scenario-based analyses.
5. Integration of conflict forecast data in regional data visualisation dashboard.
6. Facilitation of internal discussion around advantages and disadvantages of using predictive analytics in decision making as part of continuous institutional learning for how to improve operational risk anticipation.

4.5.5 UNHCR

Anticipation funds and systems: UNHCR works to protect and improve the futures of refugees, whether they are on the move, in long-term camps, or recently displaced by conflict or other emergencies. UNHCR has a dedicated data science team working on early warning models and supporting the organisation in using data-driven methods to operationalise findings.

Data used: UNHCR holds a lot of its own data, including data generated from its IT case management tool (the Profile Global Registration System, or ProGres) and the Population Registration and Identity Management EcoSystem (PRIMES). UNHCR also produces its own nowcasts of monthly refugee and asylum seeker stocks and flows. These data flow into the forward-looking predictive models used by UNHCR, of which there are four:

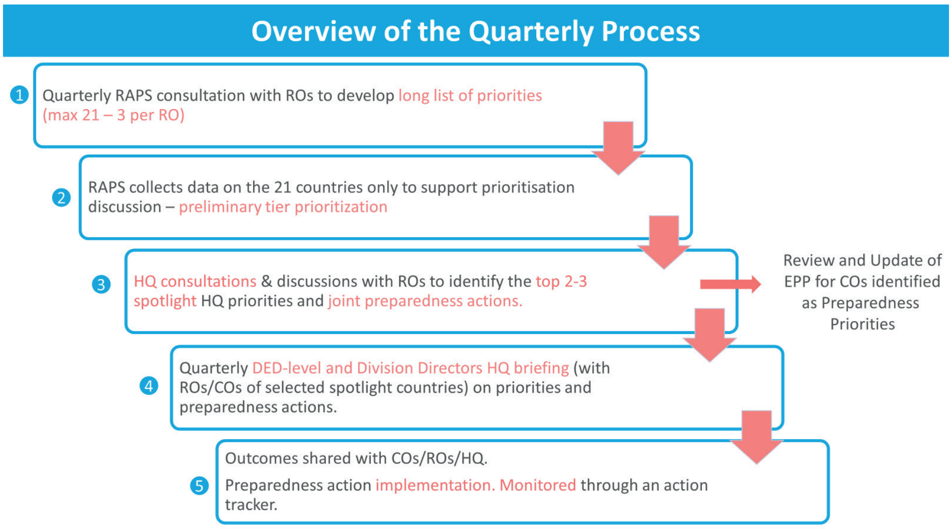
1. The SIMUGRAV model produces forecasts of bilateral forced displacement stock and flows population figures (refugees and asylum seekers) three years ahead based on a gravity framework of push and pull factors affecting displacement. The model was first developed in 2022 and is updated yearly.
2. The Early Warning Risk Index generates a monthly risk index for displacement one, three, and six months ahead. It estimates the risk of refugee flows above a certain threshold (e.g. 2,000, 5,000, 10,000, and 25,000).
3. The Climate Induced Forced Displacement Early Warning (CLIFDEW-GRID) model creates predictions for Eastern, Central, and Western Africa at the subnational 0.1° grid-cell level. It is funded through the Complex Risk Analytics Fund (CRAF'd) and places a particular focus on climate change.
4. The unaccompanied and separated children (UASC) projection model predicts the number of arrivals of UASCs, usually by month but sometimes by week. The model is under development with the University of East Anglia and is to be updated each year. For now, the focus is on South Sudan, Democratic Republic of the Congo, and Somalia.

4.5.6 UNICEF

Anticipation funds and systems: As emergencies become more frequent and resources invested in humanitarian action grow, expectations for UNICEF to deliver on the core commitments for children (CCCs) in humanitarian action in a timely and effective manner increase. Emergency preparedness at UNICEF is being increasingly mainstreamed as everyone's responsibility, beyond specific functions or typical emergency contexts. Key efforts include the expansion of the UNICEF Horizon Scan tool, a process that identifies emerging risks and links them to financial and technical preparedness support for UNICEF country offices.

Figure 13 visualises the quarterly internal UNICEF Horizon Scan process. The process starts with the Risk Analysis and Preparedness Section, within the Office of Emergency (EMOPS), consulting with Regional Offices (ROs) to develop a long list of priorities. Data are then collected on these countries to support prioritisation discussions, to inform a preliminary prioritisation. Subsequent HQ consultations, notably with the other teams in EMOPS, the Programme Group (PG), the Supply Division (SD), the Division of People and Culture (DPC), the Global Communication and Advocacy (GCA) and Public Partnership Division (PPD), focus on identifying the top two or three HQ priorities, along with joint preparedness actions. These priorities and actions are then reviewed during quarterly high-level briefings, chaired by the Deputy Executive Director for Humanitarian Action with other HQ Division Directors. These preparedness priorities are also fed back to the CO emergency preparedness plans via the Emergency Preparedness Platform (EPP). Outside of this quarterly cycle there may be additional prioritisations in response to requests by Regional Emergency Advisers (REAs).

Figure 13 UNICEF Horizon Scan process



Data used: The data insights of the Horizon Scan rely on UNICEF's Global Risk Tool. The two main data pillars of risk and capacity. The risk pillar draws on a variety of data providers and forecasts in domains such as armed conflict, climate and economic risk, as well as public health vulnerabilities, potential impact and overall risk. Sources include INFORM, ACLED, Conflict Forecast, the International Research Institute for Climate and Society (IRI) of the Colombia Climate School, FEWSNET and WHO.

Examples of actions taken: Through the Horizon Scanning process, 17 Country Offices were supported with Co-Funding allocations in 2024, amounting to a total of $ 4.85 million in preparedness funds. Five simulation exercises to strengthen preparedness were also organised. GIS mapping and scenario support were provided to more than 20 countries.

4.5.7 Welthungerhilfe

Anticipation funds and systems: Welthungerhilfe has been supporting the global drive for a humanitarian paradigm shift towards more anticipatory thinking and action. Spearheaded by its multi-year project, the Welthungerhilfe Anticipatory Humanitarian Action Facility (WAHAFA), Welthungerhilfe aims to identify and analyse disaster risks and their impacts, support the development of mechanisms for locally led anticipatory action, and secure funding to implement these mechanisms. In addition to Welthungerhilfe's own country teams and partners, WAHAFA offers German NGOs and their local humanitarian partners the opportunity to actively shape and drive this change by facilitating access to the necessary pillars of anticipatory action: (i) *capacity* - organisations will be invited to take part in knowledge exchange and trainings; (ii) *build* – participating German NGOs and their local humanitarian partners receive conceptual and financial support for the development of trigger models and localised Anticipatory Action Plans; and (iii) *fuel* – this also includes secured financing for the operationalisation of these plans and implementation of pre-agreed measures once there is an early warning

Data used: Welthungerhilfe relies on hazard and impact data and existing forecasting models to develop its Anticipatory Action Plans (AAPs). Historical information on hazard events and their impacts and in-depth vulnerability and capacity analysis of potentially at-risk communities are combined to identify areas of highest hazard risk. Use of this information with hazard forecasts, which include regional seasonal climate outlooks, precipitation forecasts, alongside observational data such as community-managed river gauges and indices such as WRSI, allows Welthungerhilfe to perform impact-based forecasting. This enables identification of a threshold (trigger) that, once crossed, releases pre-arranged funding and initiates the implementation of pre-agreed anticipatory humanitarian actions. The use of 'soft triggers' (expert consensus) combined with 'hard triggers' listed previously is one approach to managing the limitations of data and forecast uncertainty within the AAPs.

Examples of actions taken: Since 2019, Welthungerhilfe has piloted Anticipatory Humanitarian Action projects to mitigate disaster impacts before they fully unfold, starting with Madagascar's first Anticipatory Action Plan for drought. This initial plan, developed in close collaboration with Madagascar's National Disaster Risk Management Authority and meteorological services, supported nearly 10,000 people with cash to help families avoid negative coping strategies such as skipping meals or selling livestock due to predicated crop failure, while leading to the adoption of the WRSI by the government. Building on this success, Welthungerhilfe expanded to Zimbabwe and Kenya and, since 2023, has scaled up under the WAHAFA programme to nine countries in sub-Saharan Africa in collaboration with seven German NGOs and 16 local partners, focusing on tropical cyclones, droughts, floods, and epidemics. In 2024, Welthungerhilfe and its partner the Farm Community Trust of Zimbabwe implemented anticipatory actions in Zimbabwe to address predicated drought impacts, providing livestock support and water access. In South Sudan, Welthungerhilfe and its partner Hope Restoration South Sudan prepared for anticipated floods by distributing cash and flood relief kits, protecting vulnerable communities before the onset of the flooding.

5 AGENDA FOR RESEARCHERS AND HUMANITARIANS

A lot of work on all levels remains to be done. Importantly, many of the developments in this area can only be tackled if development is coordinated across academic research teams, data providers, and humanitarians. Reliable funding is another critical component for many of the things we discuss below. An initiative like CRAF'd seems particularly well-placed as a coordination and funding device, but a public, strategic plan is required for successful implementation. In this section, we lay out some of the most urgent areas for development.

The key component for quantitative forecasts is the **data landscape**. Researchers and humanitarians need to pool data effectively and data sharing needs to be incentivised by the donor community. We have seen the development of large efforts around the continued provision of armed conflict data by UCDP and ACLED – this clearly demonstrates the effectiveness of building data environments. High-quality and spatially disaggregated armed conflict data are a key pillar also in future; the efforts here need to be maintained. The provision of good-quality displacement data should be a second priority. UNHCR has recently made an important step towards data sharing at the aggregate, yearly level, but there is a need for fine-grained internal displacement data as these are one of the key drivers of humanitarian need. An example here is PRIMARI by the IOM, which is currently funded by CRAF'd. The project aims to forge a composite dataset spanning up to 60 countries, drawing on the rich data resources of the Displacement Tracking Matrix (DTM). Efforts like these need to be coordinated with both academics and humanitarians working in this area and continued if deemed useful by the community.

One aspect where humanitarians are not forthcoming enough is in providing data on their own policies to the academic community. The reasons for this vary, but there is generally no systematic attempt to track implemented policies at the spatial and temporal granularity needed to conduct impact studies or to cost policy responses appropriately. Policy data are essential to understand how preventive approaches can be priced appropriately, but they are also crucial to understand the effectiveness of the policies themselves.

There is some disagreement regarding how the data should be made available and likely no 'one size fits all' solution. As our previous discussion showed, there are serious ethical concerns to be kept in mind and the answers will need to be developed in exchanges between academics, data providers, and analysis providers. This is because the trade-offs can only be evaluated in discussions involving all stakeholders. One aspect of availability is whether data should be made available through dashboards or APIs. The general preference at the workshop was for APIs, but their suitability depends on how many users possess the capacity to develop and maintain their own, tailor-made analysis pipeline that access APIs. The need to continuously maintain dashboards produces extra costs for the organisations providing the data, but the maintenance cost for dashboards is typically small when compared to the cost of generating the data.

A critical aspect alongside new data provision is data continuity. Discontinued data collection efforts lead to all prediction efforts grinding to a halt until a comparable data source is set up. If organisations can signal that they will maintain data pipelines, this generates strong incentives for 'downstream' organisations to build analysis around these inputs; this is also true for data access methods. For end-users who start to rely on dashboards or APIs, this generates the risk that they will disappear.

Another important aspect of data provision is the availability of more granular data (both temporal and spatial), since this is the only way granular forecasts can be made. Different parts of the same organisation often have different information and data needs, thus calling for forecasts at two levels. Macro- or country-level forecasts can help in making strategic decisions, such as strategic planning, tracking hotspots, and funding allocations, while micro- or administrative-level forecasts are needed for quick adjustments to operational implementation and programming. Many organisations have significant field presence that would benefit from more frequent forecasts at lower administrative levels.

The second main pillar next to improving the data landscape is **analysis and forecasting**. This is where the data get ingested to generate decision-critical summaries. Forecasts are one possibility, but good monitoring is also important. Academic research can play a crucial role here. Many of the methods needed to solve the issues in practical applications are not solved methodologically. Examples include:

- **Machine learning panel methods.** Most hazards can be better predicted with panel data as otherwise it is impossible to predict hazards that have not taken place at a location before. At the same time, data-rich environments like armed conflict prediction call for machine learning methods. This requires improving integration between econometric panel data methods and machine learning.
- **Forecasts of distributions in time and space**. This would allow forecast outputs to react flexibly to policy requirements. The current machine learning toolkit requires specific targets for training and this interrupts the communication between quantitative providers and intuitive users. If forecasting systems predicted joint distributions of violence outcomes in time and space, then these forecasts could be used flexibly to answer an infinite number of conditional questions.
- **Dual systems that estimate policy effects while forecasting**. This would allow forecast systems to address a broader set of challenges effectively. One reason that quantitative forecasting systems are used more for climate hazards is that these forecasts are not 'political', i.e. the outcomes do not react endogenously to the forecast. In conflict or displacement prediction, the problem of self-fulfilling and self-defeating prophecies currently leads to trade-offs in which either ethical standards or forecast performance can be maintained, but not both.
- **Subnational forecasts for armed conflict, displacement and need.** Need forecasting is currently out of reach due to lack of data. Approaches range from forecasting conflict and then using associations to costs to attempts

to forecast displacement, which is one step closer to needs. However, policymakers complain that even current subnational conflict forecasts are not up-to-date and precise enough to provide effective decision aid.

True out-of-sample competitions should be an essential part of the collective learning process. The VIEWS team is a global leader and has provided the public good of posing challenges and evaluating teams in a transparent way. Humanitarians should propose challenges on outputs that they find most useful. However, funding tools for this are often too short-term – running a forecasting challenge takes more than one year and most short-term funding vehicles are therefore not adequate.

More comprehensive performance evaluations and uncertainty estimates could help with analysis and interpretation. Practitioners understand that forecasts are not always correct, but often do not have a good sense of how accurate the forecasts are. Easily understandable and intuitive performance evaluations and uncertainty estimates, beyond the standard machine learning performance metrics, could help practitioners develop a better sense of probability and understand the accuracy of a model or forecast. This in turn would help practitioners in making measured, informed decisions.

In the policy-practitioner community, it is important to work on interpreting forecasts and effective storytelling that lead to informed decision making. In addition to forming a sense of probability and magnitude in terms of forecasts, combining quantitative predictions with on-the-ground qualitative information can help to interpret forecasts from a holistic standpoint. Local knowledge can be used to corroborate and calibrate forecasts to form a better 'sense' of conflict contexts. Leveraging field presence by systematising both quantitative and qualitative data collection from country/field offices can add to this effort. Equally important is working on storytelling, often through visualisations, and creating appropriate messaging and narratives from data. Practitioners, particularly decision makers, are more likely to remember and act upon a coherent and powerful narrative.

Pre-positioned **finance** that can be triggered by agreed, objective triggers is desirable as it allows for a much faster response. The example shown in Figure 9 is a particularly strong one here as it is maximally transparent, but this sort of system with transparent triggers would generate critical ethical problems in armed conflict because violence might react to the offer of financing. Strategic intransparency is important but needs to be organised carefully. Humanitarian organisations need to think carefully about the trade-offs. Donors also play an important role as the provision of pre-positioned finance ultimately depends on them.

Finally, a key insight from the exchange that sparked this Policy Insight is the fact that the **exchange across communities** is key to make academic research most useful for the policy community. Moving closer to predicting humanitarian demand, for example, is crucial for organisations, but this requires a collaboration between humanitarians who understand this demand and experts in building forecasting systems. The negative repercussions of publishing a forecast are often not known to outsiders and so, also here, regular exchanges are crucial.

6 GLOSSARY OF KEY TERMS

GLOSSARY OF TERMS

Artificial intelligence (AI): The simulation of human intelligence by machines that are programmed to perform complex cognitive functions.[45] It is a set of technologies that enable computers to perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyse data, make recommendations, and more.[46]

Anticipatory action: Defined as acting ahead of predicted hazards to prevent or reduce acute humanitarian impacts before they fully unfold. This requires pre-agreed plans that identify partners and activities, reliable early warning information, and pre-agreed financing, released predictably and rapidly when an agreed trigger-point is reached.[47] Note that the terminology here has evolved dramatically over the years from EWEA/ Advanced preparedness, to Forecast-based financing, to Anticipatory Action, to Impact forecasting, to EW4All, but the concept largely remains the same.[48]

Damages/ impact analysis: Estimating the potential harm and losses that could result from a conflict.

Disaster risk reduction: Aimed at preventing new and reducing existing disaster risk and managing residual risk, all of which contribute to strengthening resilience and therefore to the achievement of sustainable development.[49]

Development: A multidimensional undertaking to achieve a higher quality of life for all people. Economic development, social development and environmental protection are interdependent and mutually reinforcing components of sustainable development.[50]

(Multi-hazard) early warning systems (MH)EWS: An integrated system of hazard monitoring, forecasting and prediction, disaster risk assessment, communication and preparedness activities systems and processes that enable individuals, communities, governments, businesses and others to take timely action to reduce disaster risks in advance of hazardous events. EWS include the following four key and interrelated components:

- Disaster risk knowledge based on systematic collection of data and disaster risk assessments.
- Detection, monitoring, analysing and forecasting of hazards and possible consequences.
- Dissemination and communication, by an official source, of authoritative, timely, accurate and actionable warnings, and associated information on likelihood and impact.
- Preparedness at all levels to respond to the warnings received

Multi-hazard early warning systems address several hazards and/or impacts of similar or different type in contexts where hazardous events may occur alone, simultaneously, cascading or cumulatively over time, and taking into account the potential interrelated effects.[51]

Evaluation metrics:

- **Recall:** The fraction of (actual) conflicts that a model was able to correctly forecast. This evaluation metric takes missed predictions (false negatives) into account, but not false positive predictions.
- **Precision:** The fraction of all cases predicted to be a conflict that actually are conflicts. Alternatively, the probability that a case forecasted to experience conflict will actually experience conflict. This evaluation metric takes false positive predictions into account, but not missed predictions (false negatives).

Finance: Funds released to prevent or mitigate hazards or to address the needs that have already emerged as a result.

Disaster risk finance (DRF): The strategies and instruments employed to manage the financial impact of disasters on governments, businesses, and individuals. It encompasses a range of financial tools and mechanisms designed to ensure that sufficient funds are available for disaster response, recovery, and reconstruction, thereby enhancing resilience and reducing the economic burden of disasters.[52]

45 See https://www.unocha.org/publications/report/world/briefing-note-artificial-intelligence-and-humanitarian-sector
46 See https://cloud.google.com/learn/what-is-artificial-intelligence
47 Source: GGFO (https://www.auswaertiges-amt.de/en/newsroom/news/g7-anticipatory-action/2531236); see also OCHA resources with video (https://www.unocha.org/anticipatory-action).
48 See also REAP (2022).
49 See https://www.undrr.org/terminology/disaster-risk-reduction
50 See https://research.un.org/en/docs/dev
51 See https://www.undrr.org/words-into-action/guide-multi-hazard-early-warning
52 See https://www.financialprotectionforum.org/what-is-disaster-risk-finance-drf

OR a term covering financial mechanisms, arranged in advance of disasters, for use in disaster risk management activities

Pre-arranged finance (PAF): A specific form of disaster risk financing that has been approved in advance of a crisis and that is guaranteed to be released to a specific implementer when a specific pre-identified trigger condition is met (REAP, 2022).

Fragility: The opposite of robustness. In the development context, a fragile country is one that exhibits significant vulnerability due to weak institutions, poor governance, and limited capacity to manage internal and external stresses, which can lead to instability and hinder sustainable development (OECD, 2022).

Forecasting: The effort to predict a future event, for example through an algorithm that learns from past data. The output could be a probability between 0 and 1 of a defined event occurring (such as a river rising above a given level) or a point estimate/interval for a specific outcome (such as 50, or between 0 and 100, people displaced).

Foresight: The systematic exploration of multiple plausible futures to inform present decisions, emphasizing flexibility and strategic planning to navigate uncertainties effectively (IMF, 2021). Foresight differs from forecasting in that forecasting uses quantitative data to predict future trends, while foresight involves a broader, qualitative understanding of potential future scenarios.

Hazard: A process, phenomenon or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption or environmental degradation.[53]

Humanitarian action: refers to activities aimed at saving lives, alleviating suffering, and maintaining human dignity during and after crises, including natural disasters and conflicts. It encompasses the provision of essential services such as food, water, shelter, and medical care, and is guided by principles of humanity, neutrality, impartiality, and independence.[54]

INFORM Risk: A global, open-source risk assessment for humanitarian crises and disasters. It can support decisions about prevention, preparedness and response.

Machine learning (ML): A subset of artificial intelligence that enables systems to learn from data and improve over time without explicit programming.[55] Supervised machine learning involves using labelled data to train an algorithm, which then makes predictions or classifications based on input data, and evaluates its accuracy through an error function.

Monitoring: The practice of systematically observing and collecting data to tackle the need for a description of the current or previous situations.

Parametric insurance: An approach that ensures rapid payouts based on pre-agreed triggers, such as rainfall levels, helping to mitigate the impact of disasters promptly and effectively.

Preparedness: The knowledge and capacities developed by governments, professional response and recovery organisations, communities and individuals to effectively anticipate, respond to, and recover from, the impacts of likely, imminent or current hazard events or conditions (IASC, 2013).

Prevention: Long-term strategies to address underlying causes and prevent conflict from occurring

Rapid-onset/sudden-onset events: May be climate-related (e.g., floods, cyclones, landslides, tornadoes, wildfires), geologic-related (earthquakes, tsunamis, or volcanic eruptions) or not (e.g., chemical explosion or critical infrastructure failure).[56]

Resilience: The ability of a system, community or society exposed to hazards to resist, absorb, accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner.[57] Definitions vary, however. The dictionary definition is the quality of being able to return quickly to a previous good condition after problems. This is different from not being affected by a shock, i.e. robustness.

Risk: Defined differently depending on the context. Risk is sometimes the expected impact of a conflict, calculated as its predicted probability of occurring times its expected cost. In the context of machine learning, risk is also used as a synonym for just the predicted probability of a negative event.

53 See https://www.undrr.org/terminology/hazard
54 See https://www.ghdinitiative.org/ghd/gns/principles-good-practice-of-ghd/principles-good-practice-ghd.html
55 See https://www.ibm.com/es-es/topics/machine-learning
56 See https://inee.org/eie-glossary/sudden-onset-disaster
57 See https://www.undrr.org/terminology/resilience

UNDRR defines risk as the combination of the probability of an event and its negative consequences. This definition emphasises the multifaceted nature of risk, which includes:

- **Hazard:** Natural or human-induced physical events or phenomena that can cause harm.
- **Exposure:** The presence of people, property, systems, or other elements in hazard-prone areas.
- **Vulnerability:** The susceptibility of the exposed elements to suffer harm.
- **Capacity:** The strengths and resources available to manage and reduce risk

Compound risk: Combined impact of multiple interacting risks that occur simultaneously or in succession, resulting in more severe consequences than when these risks occur independently.

Systemic risk: Latent risks within a system not typically tracked as hazards but understood to have the potential to impact overall system performance when certain conditions change; these risks are often emergent, not immediately apparent, and may not fit traditional disaster classifications (UNDRR, 2019).

Conflict onset: The initial outbreak or beginning of organized, violent conflict within or between states or groups. It typically marks the point at which tensions escalate into active violence, surpassing a threshold of casualties or incidents.

Slow-onset events: Climate crisis events including sea level rise, increasing temperatures, ocean acidification, glacial retreat and related impacts, salinization, land and forest degradation, loss of biodiversity and desertification.[58]

58 See https://unfccc.int/resource/docs/2010/cop16/eng/07a01.pdf

REFERENCES

Altay, N. and A. Narayanan (2022), "Forecasting in humanitarian operations: Literature review and research needs", *International Journal of Forecasting* 38(3): 1234-1244.

Behar, A. and S. Hlatshwayo 2021), "How to implement strategic foresight (and why)", IMF Guidance Note.

Benjamin, D.M., F. Morstatter, A.E. Abbas et al. (2023), "Hybrid forecasting of geopolitical events", *AI Magazine* 44(1): 112-128.

Buchanan-Smith, M. and S. Davies (eds) (1995), *Famine Early Warning and Response*, Practical Action Publishing.

Chadefaux, T. (2021), "A shape-based approach to conflict forecasting", *International Interactions* 48(4): 633–648.

Clarke, D.J. and S. Dercon (2016), "Defining the Problem: Begging bowls and benefactors", in D.J. Clarke and S. Dercon (eds), *Dull Disasters? How planning ahead will make a difference*, Oxford University Press.

Dempsey, B. and D. Hillier (2012), "A Dangerous Delay. The cost of late response to early warnings in the 2011 drought in the Horn of Africa", Joint Agency Briefing Paper, Oxfam International and Save the Children UK.

Dottori, F., M. Kalas, P. Salamon, A. Bianchi, L., Alfieri and L. Feyen (2017), "An operational procedure for rapid flood risk assessment in Europe", *Natural Hazards and Earth System Sciences* 17: 1111–1126.

Dhami, M.K. and D.R. Mandel (2021), "Words or numbers? Communicating probability in intelligence analysis", *American Psychologist* 76(3): 549–560.

Dietvorst, B., J.P Simmons and C. Massey (2014), "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err", *Journal of Experimental Psychology General* 144(1).

Halterman, A., B.E. Bagozzi, A. Beger, P. Schrodt and G. Scraborough (2023), "PLOVER and POLECAT: A New Political Event Ontology and Dataset", SocArVix.

Hammond, L. and D. Maxwell (2002), "The Ethiopian Crisis of 1999–2000: Lessons Learned, Questions Unanswered", *Disasters* 26(3): 262-279.

Hillbruner, C. and G. Moloney (2012), "When early warning is not enough—Lessons learned from the 2011 Somalia Famine", *Global Food Security* 1(1): 20-28.

Howarth, C.N., K. Khoshnood, N.A. Raymond et al. (2024), *Special Report: Corroboration of Widespread Body Disposal and Significant Increase in Gravesite Activity from RSF Attacks in El-Gezira*, Humanitarian Research Lab at Yale School of Public Health.

Jean, V., M.A. Boucher, A. Frini and D. Roussel (2023), "Fully integrating probabilistic flood forecasts into the decision-making process across southern Quebec, Canada: some factors to consider", *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques* 49(2): 153–170.

IASC – Inter-Agency Standing Committee (2013), *Common Framework for Preparedness*, IASC Reference Group on Risk, Early Warning and Preparedness.

IASC (2024), "Why Humanitarian Leadership Matters", Chapter A in *Leadership in Humanitarian Action: Handbook for Humanitarian Coordinators*.

Levy, K., K. Chasalow and S. Riley (2021), "Algorithms and Decision-Making in the Public Sector", *Annual Review of Law and Social Science* 17.

Maxwell, D., P. Howe and M. Fitzpatrick (2023), *Famine Prevention: A Landscape Report*, Feinstein International Center, Tufts University. /

Mellers, B., E. Stone, T. Murray et al. (2015), "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions", *Perspectives on Psychological Science* 10(3): 267-281.

Mueller, H. and C. Rauh (2022), "The Hard Problem of Prediction for Conflict Prevention", *Journal of the European Economic Association* 20(6): 2440–2467.

Mueller, H., A. Groeger, J. Hersh, A. Matranga, J. Serrat (2021), "Monitoring war destruction from space using machine learning", *Proceedings of the National Academy of Sciences of the USA* 118(23): e2025400118.

Mueller, H., C. Rauh, A. Ruggieri and B. Seimon (2024), "'If You Only Have a Hammer': Optimal Dynamic Prevention Policy", BSE Working Paper 1465.

Nunn, N. and N. Qian (2014), "US Food Aid and Civil Conflict", American Economic Review 104(6): 1630–66. https://www.aeaweb.org/articles?id=10.1257/aer.104.6.1630

OCHA – United Nations Office for the Coordination of Humanitarian Affairs (2024), *Global Humanitarian Overview 2024*.

OECD – Organisation for Economic Co-operation and Development (2022), *States of Fragility 2022*.

Pearl, J. (2009), "Causal inference in statistics: An overview", *Statistics Surveys* 3: 96-146.

REAP – Risk-informed Early Action Partnership (2022), *Glossary of Early Action Terms, 2022 Edition*.

Rød, E.G. T. Gåsste and H. Hegre (2024), "A review and comparison of conflict early warning systems", *International Journal of Forecasting* 40(1): 96-112.

Rohner, D. (2024), "Mediation, Military, and Money: The Promises and Pitfalls of Outside Interventions to End Armed Conflicts", *Journal of Economic Literature* 62(1): 155-195.

Shmueli, G. (2010), "To explain or to predict?", *Statistical Science* 25(3): 289-310.

Sonno, T. (2023), "Globalization and Conflicts: the Good, the Bad, and the Ugly of Corporations in Africa", working paper.

Sunstein, C.R. (2023), "The use of algorithms in society", forthcoming in *The Review of Austrian Economics*.

Sutton, J. and L.M. Fischer (2021), "Understanding Visual Risk Communication Messages: An Analysis of Visual Attention Allocation and Think-Aloud Responses to Tornado Graphics", *Weather, Climate and Society* 13(1): 173-188.

Tapsoba, A. (2023), "The cost of fear: Impact of violence risk on child health during conflict", *Journal of Development Economics* 160, 102975.

Tetlock, P. and D. Gardner (2016), *Superforecasting: The Art and Science of Prediction*, Random House.

United Nations (2024), "New UN report calls for trillions more in development investment to rescue Sustainable Development Goals", Press Release, 9 April.

UNDRR –UN Office for Disaster Risk Reduction (2019), *Global Assessment Report on Disaster Risk Reduction*.

CEPR

UNDRR (2023), *Global Status of Mult-Hazard Early Warning Systems*.

Vesco, P., H. Hegre, M. Colaresi, R.B. Jansen, A. Lo, G. Reisch and N.B. Weidmann (2022), “United they stand: Findings from an escalation prediction competition”, *International Interactions* 48(4): 860–896.

Wagner, M. and C. Jaime (2020), “An Agenda for Expanding Forecast-based Action to Situations of Conflict”, Global Public Policy Institute.

Ward, M.D., B.D. Greenhill and K.M. Bakke (2010), “The perils of policy by p-value: Predicting civil conflicts”, *Journal of Peace Research* 47(4): 363-375.

Weingärtner, L., T. Pforr and E. Wilkinson (2020), *The Evidence Base on Anticipatory Action*, World Food Programme.

Yang, G., Q. Ye and J. Xia (2022), “Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond”, *Information Fusion* 77: 29-52.

LIST OF AUTHORS

Margherita Philipp, FEA and EconAI
Matthias Amling, Welthungerhilfe
Andreas Beger, Basil Analytics
Seth Caldwell, UN OCHA
Luke Caley, IFRC (Red Cross & Red Crescent)
Thomas Chadefaux, Trinity College Dublin
Michael Colaresi, University of Pittsburgh & PRIO
Marcus Elten, UNDRR
Anna Farina, Start Network
Ella Gerry, Start Network
Justin Ginnetti, IFRC
Håvard Hegre, VIEWS, PRIO
Johannes Hengelbrock, German Federal Foreign Office
Katayoun Kishi, ACLED
Sun-joo Lee, UNDP
Angelica Lindqvist-McGowan, VIEWS, Uppsala University
Alexandra Málaga, FEA & EconAI
Laura Mayoral, IAE-CSIC, Barcelona School of Economics
Hannes Mueller, IAE-CSIC, Barcelona School of Economics, CEPR
Christopher Rauh, IAE-CSIC, Barcelona School of Economics, University of Cambridge, PRIO and CEPR
Jo Rixom, FCDO
Ben Seimon, FEA and EconAI
Ramon Talvi-Robledo, FEA & EconAI
Renato Vassallo, FEA and EconAI
Kevin Wyjad, UNICEF

THE CENTRE FOR ECONOMIC POLICY RESEARCH

The Centre for Economic Policy Research (CEPR) is a network of over 1,700 research economists based mostly in European universities. The Centre's goal is twofold: to promote world-class research, and to get the policy-relevant results into the hands of key decision-makers. CEPR's guiding principle is 'Research excellence with policy relevance'. It was founded in the UK in 1983, where it is a Charity, and in November 2019 CEPR initiated the creation of an Association under French law, in order to provide a vehicle for an expansion in France. The members of the Conseil d'Administration of the Association are identical to the UK Board of Trustees.

CEPR is independent of all public and private interest groups. It takes no institutional stand on economic policy matters and its core funding comes from its Institutional Members, projects that it runs and sales of publications. Because it draws on such a large network of researchers, its output reflects a broad spectrum of individual viewpoints as well as perspectives drawn from civil society. CEPR research may include views on policy, but the Trustees/members of the Conseil d'Administration of the Association do not give prior review to its publications. The opinions expressed in this report are those of the authors and not those of CEPR.