

From Algorithm to Action: Conflict Forecasting and Public Opinion on Preventive Humanitarian Intervention

Jungmin Han*

Thomas Chadeaux†

October 12, 2026

This manuscript presents a pre-analysis plan prepared for preregistration on the Open Science Framework (OSF).

Abstract

This project examines whether computational conflict forecasting can increase public support for preventive humanitarian intervention. While conflict-forecasting research has primarily focused on improving predictions of political violence, and humanitarian-intervention research has largely examined public attitudes after civilian suffering becomes visible, less is known about whether predictive information can generate support for action before large-scale violence occurs. We argue that computational conflict forecasts can strengthen support for preventive intervention by increasing perceived certainty about conflict escalation, reflected in citizens' assessments of the likelihood of future violence and their confidence in the available information. We further examine whether these effects depend on the forecasting system's reported predictive performance and citizens' preexisting perceptions of AI's functional capabilities and benefits. To test these expectations, we design a randomized vignette experiment involving approximately 1,500 respondents in the United States. Respondents evaluate a hypothetical humanitarian crisis in an African country and are randomly assigned to one of five conditions that vary the presence, informational source, and reported predictive performance of early-warning assessments. Following the vignette, respondents assess the likelihood of conflict escalation, their confidence in the available information, and their support for preventive military intervention, alongside evaluations of moral responsibility, intervention costs, and expected effectiveness. These measures allow us to examine whether computational forecasts influence intervention support through perceived certainty about future violence and whether their effects vary with prior perceptions of AI. The study thus shifts attention from whether political violence can be forecast accurately to whether predictive information can generate public support for preventive action before large-scale violence occurs.

*Postdoctoral Researcher, Department of Political Science, Trinity College Dublin

†Professor, Department of Political Science, Trinity College Dublin.

1. Introduction

Preventive humanitarian intervention poses a distinctive political dilemma. Governments may have strong reasons to act before large-scale violence occurs, since early intervention can prevent crises from becoming more severe, costly, and difficult to contain (George, 2000; Karreth & Tir, 2013; Beardsley et al., 2017). Yet prevention also requires action under uncertainty: before violence has fully materialized, the humanitarian consequences that intervention seeks to avert remain prospective, while the political, financial, and human costs of intervention are immediate and visible. The 1994 genocide in Rwanda starkly illustrates the consequences of delayed action. Reflecting on the international response four years later, President Bill Clinton acknowledged that the international community “did not act quickly enough after the killing began,” while subsequent U.S. policy emphasized the need to identify warning signs and respond to atrocities at an earlier stage (Scheffer, 1998; U.S. Department of State, 1999). At the same time, acting early can expose leaders to domestic criticism if military deployments appear unnecessary, costly, or excessively risky, particularly when the feared catastrophe has not yet occurred. This tension was evident during the 2011 Libya crisis, when President Barack Obama warned of the “potential for mass murder of innocent civilians” while simultaneously ruling out the deployment of U.S. ground troops (Obama, 2011a). In subsequently explaining the intervention, Obama emphasized the “costs and risks of intervention” and stressed that action could be undertaken without putting American troops on the ground (Obama, 2011b). Preventive humanitarian action therefore confronts democratic leaders with a central political challenge: how to justify certain and immediately observable costs in response to harms that remain uncertain and not yet fully visible.

Recent advances in conflict forecasting offer one possible way to address this informational challenge. Improvements in data availability, computational methods, and Artificial Intelligence (AI) increasingly allow researchers and policy organizations to estimate where and when violent conflict is likely to occur (Hegre et al., 2017; Cederman & Weidmann, 2017; Murphy et al., 2024). By combining diverse indicators into systematic, forward-looking assessments of escalation risk, forecasting systems can provide information that is difficult to derive from fragmented observations alone. These tools are increasingly used in humanitarian planning, displacement forecasting, peacekeeping, and early-warning systems (Duursma & Karlsrud, 2019; Beduschi, 2022; Kjærøum & Madsen, 2025). Their political significance may therefore extend beyond prediction itself. By making the prospective consequences of inaction more concrete and credible, conflict forecasts may help governments justify preventive action before a crisis reaches its most destructive stage.

Better warning, however, does not necessarily produce preventive action. Research on early warning has long identified a warning–response gap in which even timely and credible information may fail to generate a political response because decision-makers confront uncertainty about appropriate policy options, concerns about false alarms, and the costs of intervention (George & Holl, 1997; Austin, 2004; Meyer et al., 2020). One potentially important part of this gap is public opinion. Existing research shows that citizens evaluate humanitarian intervention through moral, instrumental, and international-legitimacy considerations (Kreps & Maxey, 2018; Dill & Schubiger, 2021; Wallace, 2019; Mazepus et al., 2024), but most studies examine crises in which civilian suffering is already observable. Much less is known about whether predictive information can generate public support for acting before such suffering has occurred, or whether its political effects depend on the reliability of the forecast.

This study examines whether computational conflict forecasting can help bridge the gap between prediction and preventive action. We design a randomized vignette experiment among

adults in the United States, in which respondents evaluate a hypothetical humanitarian crisis before large-scale fighting has begun. The experiment varies the presence, informational source, and reported predictive performance of early-warning assessments, comparing AI-based conflict forecasts with the absence of an explicit warning and qualitative assessments by local experts. We examine whether AI-based forecasts increase perceived certainty about conflict escalation and public support for preventive humanitarian intervention, whether stronger reported predictive performance amplifies these effects, and whether citizens' preexisting perceptions of AI's functional capabilities and benefits condition their responses. We further investigate whether perceived certainty about escalation, reflected in assessments of the likelihood of future violence and confidence in the available information, mediates the effects of computational forecasting on intervention support. We also examine how early-warning information shapes citizens' evaluations of the moral justification, cost acceptability, and expected effectiveness of preventive intervention. In doing so, the study shifts attention from whether conflict can be forecast accurately to whether predictive information can become politically actionable before large-scale violence occurs.

2. Humanitarian Crises, Uncertainty, and Public Opinion on Preventive Intervention

Humanitarian intervention occupies a distinctive position in public attitudes toward the use of force abroad. Compared with interventions pursued for objectives such as regime change or other forms of internal political change, humanitarian missions generally receive relatively high levels of public support ([Jentleson, 1992](#); [Jentleson & Britton, 1998](#); [Eichenberg, 2005](#); [Brownlee, 2020](#); [Kiratli, 2024](#); [Levin & Musgrave, 2025](#)). A central explanation is their moral purpose: because humanitarian interventions aim to protect civilians from large-scale violence, citizens may view them as fulfilling a moral obligation rather than simply advancing strategic interests ([Kreps & Maxey, 2018](#); [Tomz et al., 2020](#); [Chu & Lee, 2024](#)). At the same time, public support also depends on instrumental and procedural considerations, including the expected effectiveness and costs of intervention and whether the operation is internationally authorized ([Dill & Schubiger, 2021](#); [Mazepus et al., 2024](#); [Wallace, 2019](#)). Citizens may therefore endorse the objective of civilian protection while remaining reluctant to support an operation that appears costly, risky, ineffective, or insufficiently legitimate ([Davies & Johns, 2016](#)).

Most of this literature, however, examines crises in which civilian suffering is already visible. Preventive humanitarian action poses a different political problem because it must be undertaken before large-scale violence has fully materialized.¹ This timing can be consequential: once conflict escalates, intervention may become more difficult and costly, while opportunities for lower-cost preventive measures narrow ([George, 2000](#); [Karreth & Tir, 2013](#); [Beardsley et al., 2017](#)). Yet acting earlier also means acting under greater uncertainty. Citizens may be asked to support diplomatic initiatives, humanitarian assistance, financial commitments, or military deployments when it is still unclear whether violence will escalate, how severe it will become, or whether intervention will ultimately prove necessary ([George & Holl, 1997](#); [George, 2000](#); [Muggah & Whitlock, 2022](#)). Preventive intervention therefore creates an asymmetry be-

¹A partial exception is [Levin & Musgrave \(2025\)](#), who examine support for interventions intended to prevent prospective crimes against humanity, although their focus is on electoral intervention rather than preventive military action.

tween uncertain future benefits and immediate, visible costs: the civilians who may be protected have not yet suffered the anticipated harm, while the burdens of intervention are incurred in the present.

This uncertainty creates a political constraint for both national and international actors. Democratic governments may find it difficult to justify costly preventive measures when the threat remains probabilistic, while international organizations must decide whether uncertain risks warrant scarce diplomatic attention, personnel, and resources. Existing research on UN conflict prevention shows that preventive engagement is more likely when escalation risks become visible through signals such as previous violence or regional contagion, while less immediately observable predictors of future conflict attract less consistent action (White et al., 2018; Cunningham et al., 2026). These institutional decisions are also embedded in a broader political environment shaped by member-state support and public legitimacy (Dellmuth & Tallberg, 2021; Sommerer & Tallberg, 2026). Public attitudes may therefore matter not only for whether national governments undertake preventive action, but also for the political feasibility of international preventive efforts.

Historical cases illustrate this problem. Prior to and during the 1994 genocide in Rwanda, international actors received substantial warning of organized mass violence, yet those warnings failed to generate an adequate preventive response (George & Holl, 1997; Meyer et al., 2020). In Bosnia, policymakers likewise confronted uncertainty about the trajectory of the conflict alongside concerns about the costs and casualties associated with military involvement, contributing to delayed and limited intervention as civilian violence intensified (George, 2000). These cases suggest that failures of prevention cannot be understood solely as failures to detect danger. They also reflect the difficulty of justifying immediate action against consequences that remain uncertain at the moment when prevention is still possible.

This leaves an important gap in the literature on public opinion and humanitarian intervention. Existing studies tell us a great deal about how citizens evaluate intervention once humanitarian suffering is apparent, but much less about how they respond when the case for action rests on expectations about future violence. If uncertainty makes preventive action harder to justify politically, then the availability of credible information about escalation risk may be consequential for public support.

3. Conflict Forecasting and the Politics of Preventive Action

Conflict forecasting offers one possible way to reduce this informational constraint. Advances in computational methods and data availability have improved researchers' ability to estimate where and when violence is likely to occur, moving beyond slow-moving structural indicators toward more dynamic information on conflict processes (Chadefaux, 2014; Hegre et al., 2017; Cederman & Weidmann, 2017; Murphy et al., 2024). Computational approaches with machine learning and Artificial Intelligence can synthesize these signals into forward-looking assessments of escalation risk, potentially giving decision-makers a more systematic basis for acting before violence fully materializes (Esberg & Mikulaschek, 2021; Beduschi, 2022; Mueller & Rauh, 2022).

Forecasting tools are increasingly being incorporated into practical efforts to support anticipatory action. In the humanitarian sector, predictive systems have been used to anticipate displacement, food insecurity, and humanitarian needs so that resources can be mobilized before crises reach their most severe stage (Beduschi, 2022; Kjærøum & Madsen, 2025). In Ukraine,

forward-looking estimates of displacement and return patterns have informed humanitarian and social-protection planning under rapidly changing wartime conditions (Krelinova et al., 2025). Predictive analysis has also been proposed within UN peace operations to identify areas at heightened risk of violence and inform decisions about troop deployment, mediation, civilian protection, and other preventive activities (Duursma & Karlsrud, 2019; Murphy et al., 2024). These applications demonstrate that forecasting can expand the informational basis for acting before escalation closes the preventive window.

Yet improvements in forecasting do not automatically produce improvements in prevention. A longstanding literature on early warning identifies a persistent warning–response gap: timely and credible warnings may still fail to generate action (George & Holl, 1997; Austin, 2004; Meyer et al., 2020). Forecast production is only one stage in a broader political process in which information must be communicated, interpreted, accepted, and connected to feasible responses. Policymakers may discount warnings because crises compete for limited attention, the effectiveness of intervention is uncertain, false alarms carry political and material costs, or action appears premature before violence becomes visible (George & Holl, 1997; Austin, 2004; Muggah & Whitlock, 2022). Indeed, technological advances may generate more numerous or more precise warnings without altering the political incentives that determine whether actors respond (Hirblinger et al., 2024; Meyer et al., 2020). The central challenge is therefore not simply to predict violence more accurately, but to make predictive information politically actionable.

Public opinion is one potential part of this translation process. Preventive foreign policy often requires governments to justify present costs on the basis of anticipated rather than observed harm. Where citizens remain uncertain that violence will escalate, they may regard preventive measures as unnecessary, premature, or insufficiently justified. Conflict forecasts may therefore have political effects beyond improving elite information: by making the risk of future violence more concrete and credible, they may also affect whether citizens regard preventive action as warranted. In this sense, forecasting may help narrow the warning–response gap not only by informing decision-makers, but also by shaping the public environment in which preventive policies must be justified.

Taken together, existing research on conflict forecasting has focused primarily on improving predictive performance and producing more reliable warnings (Hegre et al., 2017; Cederman & Weidmann, 2017; Murphy et al., 2024). Much less is known about how such warnings are received by citizens or whether they can increase the political feasibility of preventive action. This study addresses that gap by asking whether conflict-forecasting information increases public support for preventive humanitarian intervention, whether its effect depends on the reported accuracy of the forecast, and under what conditions predictive warnings are most persuasive.

4. Theory and Hypotheses

This section develops theoretical expectations about how the source and reported predictive performance of conflict-related information shape public support for preventive humanitarian intervention. The central argument is that preventive intervention presents citizens with an informational problem: the costs of action are immediate, whereas the humanitarian consequences that intervention is intended to prevent remain uncertain. Conflict forecasting can alter this calculation by providing systematic information about whether an emerging crisis is likely to escalate. We therefore focus primarily on an uncertainty-reduction mechanism and subsequently examine whether the influence of forecasting depends on the humanitarian stakes of the crisis

and the material costs of intervention.

Average Effects of Conflict Forecasting

Citizens evaluating preventive humanitarian intervention face a distinctive informational challenge: they must decide whether costly military action is warranted before the anticipated humanitarian harm has materialized. Unlike responses to ongoing atrocities, preventive intervention requires citizens to assess the likelihood of future violence under conditions of uncertainty. When it remains unclear whether political tensions will escalate into major violence, citizens may be reluctant to support preventive action because the prospective humanitarian benefits of intervention remain uncertain (George & Holl, 1997; George, 2000). Early-warning assessments may alleviate this informational constraint by providing forward-looking information about the risk of conflict escalation.

Contemporary conflict-forecasting systems increasingly employ statistical and machine-learning models to analyze political, economic, and security conditions and generate systematic predictions of future violence (Hegre et al., 2017; Mueller & Rauh, 2022; Murphy et al., 2024). By integrating multiple indicators and identifying patterns across previous episodes of conflict, these systems can translate fragmented information about ongoing political tensions into explicit assessments of escalation risk. Rather than merely describing current conditions, computational forecasts provide an inference about an outcome that has not yet occurred. Such information may help citizens form clearer expectations about the likelihood of future violence and evaluate whether preventive intervention is warranted.

Research on advice-taking and algorithm appreciation provides an additional theoretical basis for expecting computational forecasts to influence judgments under uncertainty. The Judge–Advisor framework describes how individuals incorporate external information into their judgments when evaluating uncertain outcomes (Sniezek & Buckley, 1995; Yaniv & Kleinberger, 2000; Soll & Mannes, 2011). Although individuals may initially hold incomplete or uncertain beliefs, advice can influence their subsequent assessments depending on the weight they assign to the information received. Importantly, individuals do not necessarily respond equally to advice generated by different sources. Logg et al. (2019) demonstrate that individuals sometimes rely more heavily on algorithmic advice than on human judgments, including in geopolitical forecasting tasks involving lay participants. Similarly, Bogert et al. (2022) find that algorithmic advice can increase reliance on external information and confidence in subsequent judgments, even when advice quality is held constant. These findings suggest that computational forecasts may carry particular evidentiary weight when citizens evaluate uncertain future events.

The informational influence of computational forecasting may therefore differ from that of qualitative warnings provided by local experts. Local NGOs, community leaders, and regional specialists possess valuable contextual knowledge and can identify developments that signal potential escalation. Computational forecasting systems, however, systematically integrate multiple indicators and compare current conditions with patterns observed across a broader range of cases to generate forward-looking assessments of conflict risk (Hegre et al., 2017; Mueller & Rauh, 2022). Research on humanitarian anticipatory action likewise highlights the potential of computational models to synthesize diverse sources of information and identify emerging threats before large-scale harm occurs (Beduschi, 2022). Although both computational and qualitative assessments can warn of impending violence, citizens may perceive computational forecasts as providing a more systematic and reliable basis for anticipating whether political tensions will escalate.

We conceptualize this informational mechanism as perceived certainty about conflict es-

calation, encompassing two related dimensions: perceived escalation likelihood and epistemic confidence. The former concerns citizens' subjective assessments of how likely major violence is to occur, whereas the latter concerns their confidence in the information supporting those assessments. Computational forecasts may strengthen both dimensions by providing explicit predictions about future escalation and by presenting those predictions as the products of systematic, data-driven analysis. Compared with receiving no explicit warning, citizens exposed to a computational forecast may therefore perceive escalation as more likely and regard the available information as more reliable. Likewise, if citizens assign greater evidentiary weight to computational than qualitative assessments, AI-based forecasts should produce greater perceived certainty about the prospect of future violence.

These expectations lead to the following hypotheses:

H1a: Citizens who receive an AI-based conflict forecast will exhibit greater perceived certainty about the likelihood of conflict escalation than those who receive no warning about the risk of future violence.

H1b: Citizens who receive an AI-based conflict forecast will exhibit greater perceived certainty about the likelihood of conflict escalation than those who receive an early warning based on qualitative assessments by local experts.

Greater perceived certainty about future escalation should, in turn, strengthen public support for preventive humanitarian intervention. Warning scholarship emphasizes that information about impending threats does not automatically generate political action; recipients must regard warnings as sufficiently credible and decision-relevant to justify a response (Meyer et al., 2020). This requirement is particularly consequential for preventive intervention because the anticipated humanitarian consequences have not yet materialized, while the financial and military costs of intervention are comparatively immediate. Citizens who perceive escalation as unlikely or remain doubtful about the available information may therefore consider preventive intervention premature or unjustified.

By contrast, when citizens regard major violence as likely and the supporting information as reliable, the anticipated humanitarian consequences of inaction may become more compelling. Computational forecasts may consequently increase support for preventive intervention by strengthening citizens' expectations that major violence will occur without preventive action. This influence should be particularly pronounced if citizens regard systematic computational assessments as more informative than qualitative warnings. Accordingly, AI-based conflict forecasts should generate greater support for preventive humanitarian intervention than either the absence of an explicit warning or a comparable warning based on qualitative assessments by local experts.

H2a: Public support for preventive humanitarian intervention will be higher among citizens who receive a computational conflict forecast than among citizens who receive no warning about the risk of future violence.

H2b: Public support for preventive humanitarian intervention will be higher among citizens who receive a computational conflict forecast than among citizens who receive a qualitative warning from local sources.

The preceding arguments suggest that perceived certainty about conflict escalation constitutes a central mechanism linking computational forecasts to public support for preventive intervention. By providing systematic assessments of future violence, AI-based forecasts may

increase perceived escalation likelihood and confidence in the available information. These changes may, in turn, strengthen the perceived justification for undertaking costly preventive action before major violence occurs. Thus, the influence of computational forecasts on public support for humanitarian intervention should operate, at least in part, through increased perceived certainty about the likelihood of conflict escalation.

H3a: The positive effect of AI-based conflict forecasts on public support for preventive humanitarian intervention, relative to receiving no warning, will be mediated by increased perceived certainty about the likelihood of conflict escalation.

H3b: The positive effect of AI-based conflict forecasts on public support for preventive humanitarian intervention, relative to qualitative warnings from local experts, will be mediated by increased perceived certainty about the likelihood of conflict escalation.

Predictive Performance of Conflict Forecasting

The effects of computational conflict forecasts should depend not only on whether predictive information is available, but also on the reported predictive performance of the forecasting system. Preventive humanitarian intervention requires citizens to evaluate potential threats before anticipated violence has fully materialized, when uncertainty about whether escalation will occur is especially high (George & Holl, 1997; George, 2000). Conflict forecasts remain probabilistic rather than definitive statements about future events (Hegre et al., 2017). Consequently, the credibility and informational value of an early warning may depend on the forecasting system's historical ability to anticipate political violence.

A forecasting system with stronger reported predictive performance should provide citizens with greater confidence in the available information about future escalation. When past warnings have frequently been followed by major violence, citizens may regard a new warning as a more reliable indication of an impending threat. Conversely, a weaker record of predictive performance leaves greater doubt about whether the warning signals genuine danger or constitutes a false alarm, a problem long emphasized in the early-warning literature (George & Holl, 1997; Muggah & Whitlock, 2022). Because most citizens are unlikely to evaluate the technical properties of computational forecasting systems directly, information about historical predictive performance may serve as an accessible cue to the reliability of the warning. Higher reported performance should therefore strengthen perceived certainty about conflict escalation by increasing confidence in the available information and reinforcing beliefs that major violence is likely to occur.

H4a: Citizens who receive a computational conflict forecast with higher reported predictive performance will exhibit greater perceived certainty about the likelihood of conflict escalation than those who receive an otherwise identical forecast with lower reported predictive performance.

Greater perceived certainty about future escalation should, in turn, increase support for preventive humanitarian intervention. Decisions about preventive action require citizens to weigh the potential consequences of failing to intervene against the financial, military, and political costs of acting before violence has occurred. Decision-theoretic approaches to conflict prediction emphasize the importance of balancing the costs of false warnings against those of failing

to respond to genuine threats (Mueller & Rauh, 2022; Kjærum & Madsen, 2025). When citizens regard an early warning as more reliable, the anticipated consequences of inaction may become more compelling, making preventive intervention appear more justified despite its associated costs. Accordingly, stronger reported predictive performance should increase public support for preventive humanitarian intervention. Whereas a warning with a weaker performance record may leave citizens reluctant to endorse costly preventive action, a warning supported by a stronger historical record may provide a more persuasive informational basis for intervention. This expectation is consistent with arguments that the political influence of early warnings depends on whether their recipients regard them as credible and sufficiently persuasive to justify a response (Meyer et al., 2020).

H4b: Public support for preventive humanitarian intervention will be higher among citizens who receive a computational conflict forecast with higher reported predictive performance than among those who receive an otherwise identical forecast with lower reported predictive performance.

The preceding arguments suggest that perceived certainty about conflict escalation constitutes a potential mechanism linking the reported predictive performance of computational forecasts to public support for preventive intervention. Stronger historical performance may increase confidence in the available information and reinforce expectations of impending violence. These informational perceptions may subsequently make the anticipated threat more compelling and strengthen citizens' willingness to support preventive military action. Thus, the effect of reported predictive performance on intervention support should operate, at least in part, through increased perceived certainty about conflict escalation.

H4c: The positive effect of higher reported predictive performance on public support for preventive humanitarian intervention will be mediated by increased perceived certainty about the likelihood of conflict escalation.

Individual Predispositions and the Effects of Conflict Forecasting

The effects of computational conflict forecasting may depend partly on citizens' preexisting perceptions of the functional benefits and capabilities of Artificial Intelligence (AI). Individuals who regard AI as useful, capable, and beneficial may be more inclined to perceive computational forecasts as credible and informative assessments of future conflict risks. Such predispositions may increase their willingness to incorporate AI-generated predictions into their judgments about whether political tensions are likely to escalate into major violence. By contrast, citizens who are less convinced of AI's functional benefits and capabilities may be more inclined to discount computational forecasts, even when these forecasts provide information about an otherwise uncertain future threat.

These differences in attitudes toward AI should condition how computational forecasts influence perceived certainty about conflict escalation. Citizens with more favorable perceptions of AI may place greater confidence in computational assessments and consequently develop stronger expectations that major violence will occur. This tendency may be especially pronounced when comparing computational forecasts with qualitative warnings based on local expertise. Although both types of early warning provide assessments of future conflict risks,

individuals with favorable attitudes toward AI may assign greater evidentiary weight to computational forecasts because they regard AI-based analysis as particularly capable of identifying patterns and anticipating future developments. Accordingly, the effects of computational forecasts on perceived certainty about escalation should be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

H5a: The positive effect of receiving an AI-based conflict forecast, relative to receiving no early warning, on perceived certainty about conflict escalation will be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

H5b: The positive effect of receiving an AI-based conflict forecast, relative to a qualitative warning from local experts, on perceived certainty about conflict escalation will be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

Individual predispositions toward AI may also condition how citizens respond to information about the predictive performance of computational forecasting systems. Citizens who already regard AI as capable and useful may be particularly receptive to evidence that a forecasting system has demonstrated strong predictive performance. Higher reported precision may reinforce their confidence in the reliability of computational assessments and strengthen their expectations about future escalation. Conversely, citizens with less favorable attitudes toward AI may remain skeptical of computational forecasts even when presented with evidence of relatively high predictive performance. Thus, the effect of reported predictive performance on perceived certainty about conflict escalation may depend on citizens' prior evaluations of AI's capabilities.

H5c: The positive effect of higher reported predictive performance on perceived certainty about conflict escalation will be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

Beyond their influence on perceived certainty, preexisting attitudes toward AI may also condition how computational forecasts shape public support for preventive humanitarian intervention. Citizens who regard AI-based predictions as credible and informative may consider them a stronger evidentiary basis for undertaking preventive action before major violence occurs. When computational forecasts indicate a high likelihood of escalation, individuals with favorable attitudes toward AI may therefore be more willing to support preventive intervention despite its associated financial and military costs. By contrast, individuals who are skeptical of AI's capabilities may be less willing to rely on computational forecasts when evaluating whether preventive action is warranted. The positive influence of computational conflict forecasting on intervention support should consequently be greater among citizens with more favorable perceptions of AI's functional capabilities and benefits.

H5d: The positive effect of receiving an AI-based conflict forecast, relative to receiving no early warning, on public support for preventive humanitarian intervention will be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

H5e: The positive effect of receiving an AI-based conflict forecast, relative to a qualitative warning from local experts, on public support for preventive humanitarian intervention will be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

Finally, citizens' prior attitudes toward AI may shape how information about predictive performance translates into support for preventive humanitarian intervention. For individuals who already regard AI as capable and reliable, stronger reported predictive performance may reinforce the credibility of a warning and increase their willingness to endorse preventive action. Conversely, individuals who are less favorable toward AI may be less responsive to differences in reported predictive precision, limiting the influence of such information on their policy preferences. Consequently, the effect of stronger predictive performance on intervention support should be more pronounced among citizens with favorable attitudes toward AI.

H5f: The positive effect of higher reported predictive performance on public support for preventive humanitarian intervention will be stronger among citizens with more favorable perceptions of AI's functional capabilities and benefits.

5. Research Design

We test the theoretical expectations using a randomized vignette experiment, building on previous survey-experimental research on public attitudes toward humanitarian intervention and foreign involvement (Kreps & Maxey, 2018; Kiratli, 2024; Chu & Lee, 2024; Han & Park, 2025). Respondents evaluate a hypothetical political crisis in which the U.S. government considers preventive military intervention before large-scale violence has begun. The experiment employs a five-condition between-subjects design that varies whether an early warning is provided, the informational source underlying the warning, and the reported historical predictive performance of an AI-based forecasting system.

The study will be conducted among approximately 1,500 voting-age adults in the United States recruited through Prolific. The United States provides a substantively relevant setting because public opinion can constrain government decisions concerning overseas military operations, including humanitarian intervention (Hildebrandt et al., 2013; Maxey, 2020; Tomz et al., 2020). Its historical and contemporary involvement in international military and humanitarian operations also makes the U.S. public particularly relevant for examining political support for preventive action (Jentleson & Britton, 1998; Kreps & Maxey, 2018; Mitkov, 2023; Lai & Reiter, 2005; Chu & Lee, 2024; Lupu & Wallace, 2024).

Before encountering the experimental vignette, respondents complete a pretreatment questionnaire collecting demographic and political characteristics, including gender, age, race, education, household income, geographic location, party identification, and political ideology. These variables provide descriptive information about the sample and allow us to assess heterogeneity in treatment effects across demographic and political groups.

Of particular importance to our theoretical argument, the pretreatment questionnaire measures general attitudes toward artificial intelligence using five items adapted from the General Attitudes toward Artificial Intelligence Scale (Schepman & Rodway, 2020). These items capture favorable perceptions of AI's functional benefits, capabilities, and potential advantages relative to human decision-making. We use them to examine whether individuals with more favorable prior attitudes toward AI respond differently to AI-based early warnings and information concerning forecasting performance.

The pretreatment questionnaire also includes a basic numerical-comprehension question asking respondents to compare the predictive accuracy of two hypothetical weather forecasting systems. This measure allows us to examine whether understanding percentage-based fore-

casting information conditions responses to the numerical performance information presented in the experiment. All respondents will be retained in the primary analyses regardless of their numeracy responses.

Hypothetical Humanitarian Crisis

We situate the hypothetical crisis in an unnamed African country. This choice provides a plausible context for preventive humanitarian intervention while limiting the influence of respondents' prior knowledge or attitudes toward a specific country or ongoing conflict. Previous survey experiments have similarly employed hypothetical African settings to investigate public support for foreign intervention (Recchia & Chu, 2021; Mitkov, 2023). Nevertheless, we recognize that the African setting may activate racial, geographic, or paternalistic associations that affect public evaluations of humanitarian intervention (Chu & Lee, 2024). Because the geographic setting remains constant across experimental conditions, such associations do not themselves constitute differences in the randomized information, although they may shape how respondents interpret the warnings.

All respondents receive the same introductory scenario describing an emerging civil war between the government and an armed opposition group. Large-scale fighting has not yet begun, but escalation is expected to result in substantial civilian casualties. This scenario establishes a common humanitarian context in which the potential consequences of escalation are serious but have not yet materialized. Holding the anticipated civilian consequences constant allows us to focus on how information about future escalation, rather than variation in the expected humanitarian costs of violence, shapes public support for preventive intervention. The vignette information is delivered in both written and audio form. The audio recordings were generated using GPT-5.6 and reproduce the substantive content of the written scenarios. Providing the information through both formats is intended to facilitate comprehension of the experimental materials. Respondents are told:

Imagine that a civil war may be developing in a foreign country in Africa. Tensions have existed between the government and an armed opposition group, but large-scale fighting has not yet begun.

If the situation escalates into major violence, at least several hundred people are expected to be killed during its initial stages. Most of the victims would be civilians living in the affected areas, including people who are not directly involved in the fighting. Civilians may also be deliberately targeted as the conflict intensifies.

Following the introductory scenario, respondents complete two comprehension checks concerning whether large-scale fighting has already begun and the expected consequences of escalation. Respondents who fail these checks do not proceed to the randomized warning treatments.

Experimental Treatments: Early-Warning Assessments

Following the common introductory scenario, respondents are randomly assigned to one of five conditions. Each condition presents an announcement by a U.S. government official regarding the developing crisis and a proposed preventive military intervention. Across all conditions,

the government is considering deploying approximately 5,000 U.S. military personnel at an estimated cost of around \$1 billion to prevent major violence. Holding the proposed military commitment constant allows us to isolate the effects of early-warning information on public evaluations of an otherwise identical intervention proposal.

The five conditions differ in the presence of an explicit early warning, the source of the assessment, and the information provided about its historical predictive performance.

(1) No-Warning Baseline. In the baseline condition, respondents are informed that the U.S. government is closely monitoring the ongoing tensions but that large-scale violence has not yet occurred. The government is considering preventive military intervention, but no explicit forecast concerning the likelihood of escalation is provided. This condition establishes the baseline level of public support for preventive intervention in the absence of an explicit early-warning assessment.

(2) AI-Based Early Warning (w/o Performance Information). In the second condition, respondents receive an explicit warning that major violence is very likely to occur in the near future. The announcement states that the warning is based on an early-warning assessment conducted using artificial intelligence. Respondents are informed that the AI-based system analyzed political, economic, security, and conflict-related information and concluded that the likelihood of escalation into major violence within the next few months is extremely high. No information is provided about the system's historical predictive performance. Comparing this condition with the no-warning baseline identifies the effect of introducing an explicit AI-based early warning, relative to government monitoring without a specific forecast. Because the two conditions differ in both the presence of a warning and its AI-based source, this comparison captures their combined effect rather than the independent effect of AI attribution.

(3) AI-Based Early Warning (w/ 60% Precision). The third condition provides the same AI-based warning but additionally communicates the forecasting system's historical predictive performance. Respondents are informed that the system has a precision of approximately 60%. The announcement explains that when the system issued comparable warnings in the past, major violence followed in approximately six out of ten similar cases. This condition introduces explicit numerical information about the observed reliability of AI-generated warnings.

(4) AI-Based Early Warning (w/ 90% Precision). The fourth condition is identical to the preceding condition except that the reported precision is approximately 90%. Respondents are informed that major violence followed approximately nine out of ten comparable warnings in the past. The randomized comparison between the 60% and 90% conditions identifies the causal effect of communicating stronger historical predictive performance while holding the forecasting source and the qualitative escalation warning constant. The numerical information refers to the frequency with which major violence followed comparable positive warnings rather than the forecasting system's overall classification accuracy. Accordingly, we interpret the contrast as a manipulation of reported positive predictive performance, not necessarily of the unconditional probability that violence will occur in the hypothetical case.

(5) Qualitative Human-based Early Warning. The fifth condition provides an explicit warning that major violence is very likely to occur, but attributes the assessment to regional conflict

experts using field reports and contextual analysis. Respondents are informed that the experts reviewed information from local NGOs, community leaders, and other sources familiar with conditions on the ground and concluded that the likelihood of escalation into major violence within the next few months is extremely high. Comparing this condition with the AI-based warning without numerical performance information allows us to examine whether substantively similar warnings have different effects when grounded in AI-based computational analysis rather than qualitative assessments by regional experts. Because the treatments differ in their descriptions of the underlying analytical processes, this contrast captures the effect of the experimentally presented information source and its associated description, rather than a completely isolated AI label.

Experimental Comparisons. The five-condition design permits four principal comparisons. These comparisons allow us to distinguish the influence of warning provision, forecasting source, and reported predictive performance on public support for preventive intervention.

- First, comparing the AI-based warning without performance information with the no-warning baseline assesses the combined effect of receiving an explicit AI-based escalation warning.
- Second, comparing the AI-based warning without performance information with the qualitative expert-based warning assesses whether public responses depend on the analytical source underlying otherwise similar warnings.
- Third, comparing the 90% and 60% AI-warning conditions identifies the effect of stronger reported historical predictive performance.
- Fourth, comparing each numerical-performance condition with the AI-based warning without performance information assesses how explicitly communicating historical predictive performance changes responses relative to a warning that does not disclose such information.

Outcome and Mechanism Measures

Following the experimental vignette, respondents answer six questions concerning their assessments of escalation risk, the available information, and the proposed military intervention. To examine the informational mechanisms underlying responses to early warnings, we measure two conceptually distinct intermediate outcomes.

- First, perceived escalation likelihood is measured by asking respondents how likely they believe the situation would be to escalate into major violence if no foreign intervention were undertaken. Responses range from very unlikely to very likely on a seven-point scale. This measure captures respondents' subjective assessments of the prospective threat.
- Second, confidence in the available information is measured by asking respondents how confident they are that the information provides an accurate indication of whether major violence will occur. Responses range from not at all confident to very confident on a seven-point scale. This measure captures perceived informational reliability and episodic confidence in the assessment of future violence.

Although related, the two measures represent different components of the proposed informational mechanism. Perceived escalation likelihood concerns the estimated probability of violence, whereas epistemic confidence concerns respondents' confidence in the informational basis of that assessment. A warning may increase perceived escalation risk without necessarily increasing confidence, or increase confidence without substantially changing estimated risk. We therefore analyze the measures separately rather than combining them into a single uncertainty-reduction index.

The primary dependent variable is support for preventive military intervention. Respondents indicate the extent to which they support or oppose the U.S. government carrying out the proposed operation, using a seven-point scale ranging from strongly oppose to strongly support. This measure directly captures the principal political outcome of interest.

The survey also includes three additional measures of respondents' evaluations of preventive military intervention.

- First, perceived moral responsibility is measured by asking whether carrying out the proposed operation would help fulfill the country's moral responsibility to protect people at risk. This measure captures the perceived humanitarian justification for military action.
- Second, cost acceptability is measured by asking how acceptable respondents consider the financial and military costs of the proposed intervention, given the risk of major violence. Because the proposed operation is identical across treatment conditions, this measure allows us to examine whether differences in early-warning information influence willingness to accept the costs of preventive action.
- Third, perceived intervention effectiveness is measured by asking respondents how successful they believe the proposed operation would be in preventing violence. This measure captures expectations about the efficacy of preventive military action and provides an opportunity to examine an alternative explanation for changes in intervention support.

Together, these measures allow us to investigate whether early-warning information affects public support through changes in perceived escalation risk, confidence in the available information, and broader evaluations of the proposed intervention.

Treatment Comprehension and Information Leakage

Following the outcome questions, respondents complete two factual treatment-comprehension questions. The first asks which type of information was used to assess escalation risk, distinguishing AI-based assessment, qualitative expert analysis, and the absence of a specified early-warning method. The second asks whether respondents were informed that major violence followed approximately 60% or 90% of comparable warnings, or whether no such historical-performance information was provided. These questions assess respondents' recognition of the randomized warning information. They are treated as measures of treatment comprehension rather than as direct measures of the theoretical mechanisms. Finally, respondents are asked whether the hypothetical scenario brought to mind a particular African country. This open-ended question allows us to explore whether respondents associated the unnamed setting with a specific real-world case.

6. Analysis Plan

Effects on Perceived Risk and Confidence

The theoretical argument proposes that early-warning information influences support for preventive intervention partly by changing how individuals evaluate the likelihood of future violence and the reliability of the available information. We therefore estimate the effects of the randomized warning treatments on two intermediate outcomes: perceived escalation likelihood and confidence in the available information. For each outcome, we estimate:

$$M_{ki} = \alpha_k + \beta_{1k}AI_i + \beta_{2k}AI_i^{60} + \beta_{3k}AI_i^{90} + \beta_{4k}QUAL_i + \varepsilon_{ki},$$

where M_{ki} denotes intermediate outcome k for respondent i . The treatment contrasts mirror those used in the primary intervention-support analysis. This allows us to determine whether the same experimental differences that affect intervention support also alter respondents' perceptions of escalation risk and confidence in the information.

We do not assume that perceived escalation risk and epistemic confidence are interchangeable. A higher perceived probability of violence reflects a stronger assessment of the threat, whereas greater confidence indicates stronger belief in the reliability of the available information. Accordingly, the two outcomes will be analyzed separately. Evidence that an early-warning treatment increases both perceived risk and confidence would be consistent with the proposed informational mechanism, but would not by itself establish that either variable causally mediates the treatment effect on intervention support.

Treatment Effects on Public Attitudes toward Preventive Actions

Because each respondent evaluates a single randomly assigned vignette, the respondent is the unit of analysis. The primary analyses include eligible respondents who consented to participate and passed the comprehension checks administered before randomization.

The main dependent variable is support for the proposed preventive military intervention, measured on a seven-point scale. We estimate average treatment effects using ordinary least squares (OLS) regression, treating the seven-point response as approximately continuous. We report heteroskedasticity-robust standard errors and 95% confidence intervals. The principal model is specified as:

$$Y_i = \alpha + \beta_1AI_i + \beta_2AI_i^{60} + \beta_3AI_i^{90} + \beta_4QUAL_i + \varepsilon_i,$$

where Y_i denotes respondent i 's support for preventive military intervention. The treatment indicators represent the AI-based warning without disclosed performance (AI_i), the AI-based warning with 60% precision (AI_{60_i}), the AI-based warning with 90% precision (AI_{90_i}), and the qualitative expert-based warning ($QUAL_i$). The no-warning condition serves as the omitted reference category. Under random assignment, the treatment coefficients estimate average differences in intervention support relative to the no-warning baseline. We will present estimated means for each experimental condition together with the relevant treatment contrasts and their confidence intervals.

The principal comparisons correspond to the theoretical expectations concerning early-warning information.

- First, the comparison between the AI-based warning without disclosed performance and the no-warning baseline estimates the combined effect of receiving an explicit AI-based early warning. Under the reference-category coding above, this effect is represented by β_1 .
- Second, the comparison between the AI-based warning without disclosed performance and the qualitative expert-based warning evaluates whether the informational source of the warning affects public support. The corresponding contrast is:

$$\Delta_{\text{AI-QUAL}} = \beta_1 - \beta_4.$$

- Third, the comparison between the AI-warning conditions reporting 90% and 60% precision identifies the effect of stronger reported predictive performance:

$$\Delta_{90-60} = \beta_3 - \beta_2.$$

- Fourth, we examine the effects of communicating historical predictive performance by comparing the 60% and 90% conditions separately with the AI-based warning that does not disclose performance information:

$$\Delta_{60-\text{AI}} = \beta_2 - \beta_1,$$

$$\Delta_{90-\text{AI}} = \beta_3 - \beta_1.$$

These contrasts distinguish the effects of introducing an early warning, changing its informational source, and communicating numerical evidence about the historical performance of an AI-based forecasting system. To assess robustness, we will additionally estimate covariate-adjusted models incorporating prespecified pretreatment demographic and political characteristics. Because assignment to the warning conditions is randomized, these covariates are not required for identification of the average treatment effects, but may improve precision.

We additionally estimate treatment effects on three outcomes: perceived moral responsibility, acceptability of intervention costs, and expected intervention effectiveness. These analyses investigate whether early-warning information affects broader evaluations of the proposed military intervention. The moral-responsibility measure assesses whether the warning changes the perceived humanitarian justification for preventive action. The cost-acceptability measure evaluates whether information suggesting a greater risk of violence makes respondents more willing to accept the fixed financial and military costs of intervention. The expected-effectiveness measure examines whether warning information also influences beliefs about the capacity of the proposed operation to prevent major violence. This provides an alternative potential pathway through which forecasting information could influence intervention support. Because these outcomes are measured after treatment, we first analyze them as dependent variables in their own right, estimating the causal effect of each randomized warning condition on the corresponding evaluation. We will report the results separately rather than combine the measures into a single index, given their distinct theoretical meanings.

Mediation Analysis

To investigate the proposed informational mechanisms, we conduct a causal mediation analysis focusing on perceived escalation likelihood and confidence in the available information.

The early-warning treatments are randomly assigned, whereas the proposed mediators are measured after treatment and are not independently randomized. Consequently, random treatment assignment identifies the effects of warning information on the mediators but does not, by itself, identify the causal effects of the mediators on intervention support. We therefore distinguish experimentally identified treatment effects from mediation estimates that depend on additional identification assumptions.

Following the potential-outcomes approach to causal mediation analysis, we estimate indirect effects under the sequential ignorability assumptions, including the absence of unmeasured mediator-outcome confounding conditional on treatment assignment and appropriate pretreatment covariates. The analysis further requires assumptions concerning overlap and the absence of problematic treatment-induced mediator-outcome confounding. These assumptions cannot be established solely through the experimental randomization. For each principal treatment contrast, we estimate a mediator model:

$$M_i = \alpha_M + \beta_M T_i + \boldsymbol{\delta}_M^\top \mathbf{Z}_i + \varepsilon_{Mi},$$

and an outcome model:

$$Y_i = \alpha_Y + \beta_Y T_i + \theta M_i + \boldsymbol{\delta}_Y^\top \mathbf{Z}_i + \varepsilon_{Yi},$$

where T_i denotes the relevant randomized treatment contrast, M_i represents the proposed mediator, Y_i denotes intervention support, and $\mathbf{Z}_i$ contains prespecified pretreatment covariates.

The models provide the basis for estimating average causal mediation effects and average direct effects under the relevant assumptions. Where theoretically appropriate, specifications allowing treatment-mediator interactions will also be considered. The primary mediation analyses focus on two comparisons: (1) the AI-based warning without disclosed performance versus the no-warning condition and (2) the 90% versus 60% AI-warning conditions. The AI-versus-qualitative-warning comparison will also be examined to assess whether differences in informational source are consistent with the proposed mechanisms. We analyze perceived escalation likelihood and confidence separately, rather than treating them as components of a single scale. This distinction allows us to examine whether warning information operates primarily by increasing perceptions of the prospective threat or by increasing confidence in the information used to assess that threat.

Because the mediators may be causally related, the separate mediation estimates will not be interpreted as mutually exclusive components of the total treatment effect. Nor will their sum be interpreted as the total mediated effect without additional assumptions. To assess the sensitivity of the mediation results to potential unmeasured mediator-outcome confounding, we will conduct sensitivity analyses examining how violations of the identifying assumptions affect the estimated indirect effects. Given the observational nature of the mediator-outcome relationship, the mediation results will be interpreted cautiously as assumption-dependent evidence concerning the proposed mechanisms rather than definitive identification of the causal pathways.

Furthermore, we conduct analyses examining whether moral responsibility, cost acceptability, and perceived intervention effectiveness are associated with the effects of warning information on intervention support. The first two measures capture evaluations through which heightened concern about future violence might translate into greater willingness to undertake costly preventive action. Perceived intervention effectiveness represents a distinct explanation: respondents may support intervention because they believe the proposed operation is likely to succeed. We will first report the experimentally identified effects of warning assignment on each of these outcomes. We may then conduct mediation analyses to assess whether the observed patterns

are consistent with these proposed pathways. Because the measures are collected after treatment and are not independently randomized, we will not interpret simple associations among the intermediate variables as evidence establishing their causal ordering. In particular, any exploratory sequential mediation analysis linking early-warning information to perceived escalation risk, moral or cost evaluations, and intervention support will be presented as strongly assumption-dependent.

Moderating Effects of Prior AI Attitudes

An additional theoretical expectation concerns whether individuals' prior attitudes toward AI condition their responses to AI-based conflict forecasts. Using the five pretreatment AI-attitude items, we construct a composite measure of favorable attitudes toward the functional benefits and capabilities of AI. Responses will be coded so that higher values indicate more favorable attitudes. We will assess the internal consistency and dimensionality of the selected items before constructing the scale. To estimate heterogeneous treatment effects, we interact the prespecified treatment contrasts with the standardized AI-attitude measure. For a given binary treatment comparison, the model is:

$$Y_i = \alpha + \beta_1 T_i + \beta_2 A_i + \beta_3 (T_i \times A_i) + \boldsymbol{\delta}^\top \mathbf{Z}_i + \varepsilon_i,$$

where A_i denotes the pretreatment AI-attitude measure. The coefficient β_3 tests whether the effect of the randomized warning treatment varies with respondents' preexisting attitudes toward AI. Our principal moderation analyses focus on (1) the AI-based versus no warning comparison, (2) the AI-based versus qualitative expert-based warning contrast and (2) the 90% versus 60% AI-warning contrast. These comparisons are particularly relevant for assessing whether favorable attitudes toward AI condition responses to computational forecasting and information about its performance.

We will also examine whether prior AI attitudes moderate treatment effects on perceived escalation likelihood and epistemic confidence. These analyses can help establish whether favorable attitudes toward AI condition not only intervention preferences but also the interpretation of AI-based warning information. Because prior AI attitudes are measured rather than randomized, the interaction results will be interpreted as heterogeneity in the causal effects of warning assignment across observed attitude levels, rather than as causal effects of AI attitudes themselves.

Numerical Comprehension and Robustness Checks

The survey includes a pretreatment numerical-comprehension question concerning the relative accuracy of two weather forecasting systems. We will conduct supplementary analyses examining whether responses to the numerical-performance treatments differ between respondents who correctly and incorrectly answer this question. In particular, we will test whether numerical comprehension moderates the treatment contrast between the 90% and 60% precision conditions. Because numerical comprehension is measured before treatment, these analyses can assess whether the effects of reported predictive performance depend on respondents' ability to interpret percentage-based information. The primary analyses will retain all randomized

respondents in the analysis sample regardless of their responses to the numeracy question or post-treatment comprehension questions.

As descriptive robustness checks, we will also examine treatment effects among respondents who correctly identify the experimentally assigned warning source or precision information. However, because treatment comprehension is measured after randomization and may itself be affected by treatment assignment, these conditional analyses will not be interpreted as preserving randomized causal comparisons.

Finally, we will conduct exploratory analyses of the open-ended information-leakage question to determine whether respondents associated the hypothetical crisis with particular African countries. These responses will be used to describe the associations generated by the scenario rather than to define the primary analysis sample. The primary conclusions will be based on the randomized treatment contrasts in the full prespecified analysis sample, with the mechanism, moderation, and robustness analyses providing additional evidence concerning the interpretation and scope of the treatment effects.

References

- Austin, A. (2004). Early warning and the field: a cargo cult science? In *Transforming ethnopolitical conflict: The berghof handbook* (pp. 129–150). Springer.
- Beardsley, K., Cunningham, D. E., & White, P. B. (2017). Resolving civil wars before they start: The un security council and conflict prevention in self-determination disputes. *British Journal of Political Science*, 47(3), 675–697.
- Beduschi, A. (2022). Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks. *International Review of the Red Cross*, 104(919), 1149–1169.
- Bogert, E., Lauharatanahirun, N., & Schecter, A. (2022). Human preferences toward algorithmic advice in a word association task. *Scientific reports*, 12(1), 14501.
- Brownlee, J. (2020). Cognitive shortcuts and public support for intervention. *Journal of Conflict Resolution*, 64(2-3), 261–289.
- Cederman, L.-E., & Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectations? *Science*, 355(6324), 474–476.
- Chadefaux, T. (2014). Early warning signals for war in the news. *Journal of Peace Research*, 51(1), 5–18.
- Chu, J. A., & Lee, C. A. (2024). Race, religion, and american support for humanitarian intervention. *Journal of Conflict Resolution*, 68(10), 2076–2100.
- Cunningham, D. E., et al. (2026). Under what conditions does preventive mediation occur in self-determination disputes? *Journal of Peace Research*. (Advance online publication) doi: 10.1093/jopres/xjaf033
- Davies, G. A., & Johns, R. (2016). R2p from below: Does the british public view humanitarian interventions as ethical and effective? *International Politics*, 53(1), 118–137.
- Dellmuth, L. M., & Tallberg, J. (2021). Elite communication and the popular legitimacy of international organizations. *British Journal of Political Science*, 51(3), 1292–1313.
- Dill, J., & Schubiger, L. I. (2021). Attitudes toward the use of force: Instrumental imperatives, moral principles, and international law. *American Journal of Political Science*, 65(3), 612–633.
- Duursma, A., & Karlsrud, J. (2019). Predictive peacekeeping: Strengthening predictive analysis in un peace operations. *Stability: International Journal of Security and Development*, 8(1).
- Eichenberg, R. C. (2005). Victory has many friends: Us public opinion and the use of military force, 1981–2005. *International security*, 30(1), 140–177.
- Esberg, J., & Mikulaschek, C. (2021). Digital technologies, peace and security: Challenges and opportunities for united nations peace operations. *Peace and Security: Challenges and Opportunities for United Nations Peace Operations* (August 24, 2021).
- George, A. L. (2000). Strategies for preventive diplomacy and conflict resolution: Scholarship for policymaking. *PS: Political Science & Politics*, 33(1), 15–20.

- George, A. L., & Holl, J. E. (1997). *The warning-response problem and missed opportunities in preventive diplomacy*. Carnegie Commission on Preventing Deadly Conflict Washington, DC.
- Han, J., & Park, H. (2025). State atrocities, rebel governance, and public opinion on humanitarian intervention. *OSF Preprint*. Retrieved from https://osf.io/preprints/socarxiv/q7n5j_v3
- Hegre, H., Metternich, N. W., Nygård, H. M., & Wucherpfennig, J. (2017). *Introduction: Forecasting in peace research* (Vol. 54) (No. 2). Sage Publications Sage UK: London, England.
- Hildebrandt, T., Hillebrecht, C., Holm, P. M., & Pevehouse, J. (2013). The domestic politics of humanitarian intervention: Public opinion, partisanship, and ideology. *Foreign Policy Analysis*, 9(3), 243–266.
- Hirblinger, A. T., Wählich, M., Keator, K., McNaboe, C., Duursma, A., Karlsrud, J., ... others (2024). Making peace with un-certainty: Reflections on the role of digital technology in peace processes beyond the data hype. *International Studies Perspectives*, 25(2), 185–225.
- Jentleson, B. W. (1992). The pretty prudent public: Post post-vietnam american opinion on the use of military force. *International studies quarterly*, 36(1), 49–74.
- Jentleson, B. W., & Britton, R. L. (1998). Still pretty prudent: Post-cold war american public opinion on the use of military force. *Journal of Conflict Resolution*, 42(4), 395–417.
- Karreth, J., & Tir, J. (2013). International institutions and civil war prevention. *The Journal of Politics*, 75(1), 96–109.
- Kiratli, O. S. (2024). Policy objective of military intervention and public attitudes: A conjoint experiment from us and turkey. *Political Behavior*, 46(2), 1257–1279.
- Kjærum, A., & Madsen, B. S. (2025). Pushing the boundaries of anticipatory action using machine learning. *Data & Policy*, 7, e8.
- Krelinova, K., Loktieva, I., & Jusselme, D. (2025). Forecasting displacement and solutions for decision-making in volatile contexts: A case study from ukraine. *Data & Policy*, 7. doi: 10.1017/dap.2024.46
- Kreps, S., & Maxey, S. (2018). Mechanisms of morality: Sources of support for humanitarian intervention. *Journal of Conflict Resolution*, 62(8), 1814–1842.
- Lai, B., & Reiter, D. (2005). Rally ‘round the union jack? public opinion and the use of force in the united kingdom, 1948–2001. *International studies quarterly*, 49(2), 255–272.
- Levin, D. H., & Musgrave, P. (2025). To what end? policy objectives and us public support for political warfare. *Foreign Policy Analysis*, 21(2), oraf008.
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational behavior and human decision processes*, 151, 90–103.
- Lupu, Y., & Wallace, G. P. (2024). The laws of war and public support for foreign combatants. *International Organization*, 78(4), 823–852.

- Maxey, S. (2020). The power of humanitarian narratives: a domestic coalition theory of justifications for military action. *Political Research Quarterly*, 73(3), 680–695.
- Mazepus, H., Aloyo, E., & Kubalová, N. (2024). What drives support for armed humanitarian intervention? experimental evidence from dutch citizens on international law and probability of success. *Research & Politics*, 11(3), 20531680241272613.
- Meyer, C. O., De Franco, C., & Otto, F. (2020). *Warning about war: Conflict, persuasion and foreign policy*. Cambridge: Cambridge University Press. doi: 10.1017/9781108644006
- Mitkov, Z. (2023). Determining support for humanitarian interventions: Prospect theory versus cues. *Foreign Policy Analysis*, 19(1), orac035.
- Mueller, H., & Rauh, C. (2022). The hard problem of prediction for conflict prevention. *Journal of the European Economic Association*, 20(6), 2440–2467. doi: 10.1093/jeaa/jvac025
- Muggah, R., & Whitlock, M. (2022). Reflections on the evolution of conflict early warning. *Stability: International Journal of Security and Development*, 10(1).
- Murphy, M., Sharpe, E., & Huang, K. (2024). The promise of machine learning in violent conflict forecasting. *Data & Policy*, 6, e35. doi: 10.1017/dap.2024.27
- Obama, B. (2011a, March 18). *The president on libya: “our goal is focused, our cause is just, and our coalition is strong”*. <https://obamawhitehouse.archives.gov/blog/2011/03/18/president-libya-our-goal-focused-our-cause-just-and-our-coalition-strong>. The White House.
- Obama, B. (2011b, March 28). *Remarks by the president in address to the nation on libya*. <https://obamawhitehouse.archives.gov/the-press-office/2011/03/28/remarks-president-address-nation-libya>. The White House.
- Recchia, S., & Chu, J. (2021). Validating threat: Io approval and public support for joining military counterterrorism coalitions. *International studies quarterly*, 65(4), 919–928.
- Scheffer, D. J. (1998, December 10). *Genocide and atrocity crimes*. https://1997-2001.state.gov/policy_remarks/1998/981210_scheffer_genocide.html. (Remarks by Ambassador-at-Large for War Crimes Issues)
- Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in human behavior reports*, 1, 100014.
- Snizek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational behavior and human decision processes*, 62(2), 159–174.
- Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, 27(1), 81–102.
- Sommerer, T., & Tallberg, J. (2026). Effects of popular legitimacy on international organizations: An elite survey experiment. *The Journal of Politics*, 88(2), 732–745.
- Tomz, M., Weeks, J. L., & Yarhi-Milo, K. (2020). Public opinion and decisions about military force in democracies. *International organization*, 74(1), 119–143.

- U.S. Department of State. (1999, October 29). *The international community and genocide in rwanda*. https://1997-2001.state.gov/policy_remarks/1999/991029_scheffer_rwanda.html. (Remarks by Ambassador-at-Large for War Crimes Issues David J. Scheffer)
- Wallace, G. P. (2019). Supplying protection: The united nations and public support for humanitarian intervention. *Conflict Management and Peace Science*, 36(3), 248–269.
- White, P. B., Cunningham, D. E., & Beardsley, K. (2018). Where, when, and how does the UN work to prevent civil war in self-determination disputes? *Journal of Peace Research*, 55(3), 380–394. doi: 10.1177/0022343317744826
- Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, 83(2), 260–281.

Survey Questionnaires

Consent Check

This research study aims to enhance our understanding of personal attitudes towards political matters. To accomplish this, we will administer a brief questionnaire to individuals living in the United States. The survey will take approximately 5–10 minutes. The questionnaire is completely anonymous: your answers cannot be traced back to you and will be kept confidential. There are no risks associated with participating in this survey. You may leave the survey at any point if you feel uncomfortable with any question. If you choose to exit without finishing, your answers will not be saved. Once you submit, you cannot retract your responses because the survey is anonymous. Please note that if you do not answer the consent or attention-check question correctly, the survey may conclude at that point without proceeding further.

Q1. Are you eligible for this survey, and do you agree to participate?

- Yes (1); No (2)

Pre-treatment Survey Questions

Q2. What is your gender?

- Male (1); Female (2); Non-binary / third gender (3); Prefer not to say (4)

Q3. What is your age?

- 18–29 (1); 30–39 (2); 40–49 (3); 50–59 (4); 60 or above (5)

Q4-1. (US) Choose one or more races that you consider yourself to be.

- White or Caucasian (1); Black or African American (2); American Indian/Native American or Alaska Native (3); Asian (4); Native Hawaiian or Other Pacific Islander (5); Other (6); Prefer not to say (7)

Q5-1. (US) What is the highest level of education you have completed?

- Some high school or less (1); High school diploma or GED (2); Some college, but no degree (3); Associates or technical degree (4); Bachelor's degree (5); Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS, etc.) (6); Prefer not to say (7)

Q6-1. (US) What was your total household income before taxes during the past 12 months?

- Less than \$25,000 (1); \$25,000–\$49,999 (2); \$50,000–\$74,999 (3); \$75,000–\$99,999 (4); \$100,000–\$149,999 (5); \$150,000 or more (6); Prefer not to say (7)

Q7-1. (US) In which state do you currently reside?

Dropdown: Alabama (1) ... I do not reside in the United States (53).

Q8-1. (US) Generally speaking, do you think of yourself as a Republican, Democrat, or Independent?

- Republican (1); Democrat (2); Independent (3); Other party (4); No preference (5); Don't know (6)

Q9. [Pre-treatment Attention Check] Please choose two options that include the word "Somewhat." This question is designed to check how much a respondent pays attention to this survey.

- Agree (1); Somewhat Agree (2); Neither Agree nor Disagree (3); Somewhat Disagree (4); Disagree (5)

Q10. In general, how would you describe your political views?

- Conservative (1); Somewhat Conservative (2); Neither Conservative nor Liberal (3); Somewhat Liberal (4); Liberal (5)

Q11. Two weather forecasting systems predict whether it will rain. System A's predictions are correct 50% of the time, while System B's predictions are correct 80% of the time. Which system makes more accurate predictions?

- System A (1); System B (2); Both systems have the same percentage of correct predictions (3); Not sure (4)

Q13. Please indicate how much you agree or disagree with each statement.

[Responses: Strongly disagree (1) – Strongly agree (7)]

- There are many beneficial applications of Artificial Intelligence.
- Artificially intelligent systems can perform better than humans.
- Much of society will benefit from a future full of Artificial Intelligence.
- An artificially intelligent agent would be better than an employee in many routine jobs.
- Some complex decisions are best left to artificially intelligent systems.

Treatment Vignette

We will present a scenario regarding a political event in a foreign country. Some aspects of the scenario may seem significant to you, while others may not. Please adjust your speakers or headphones to a comfortable volume. The following content will be presented in both written and audio form.

Q14. Do you agree to read the details very carefully, and then give your most thoughtful answers? [Yes, No]

Humanitarian Crisis Scenario:

Imagine that a civil war may be developing in a foreign country in Africa. Tensions have existed between the government and an armed opposition group, but large-scale fighting has not yet begun.

If the situation escalates into major violence, at least several hundred people are expected to be killed during its initial stages. Most of the victims would be civilians living in the affected areas, including people who are not directly involved in the fighting. Civilians may also be deliberately targeted as the conflict intensifies.

[Pre-treatment Attention Check] Before continuing, we need to make sure you read this information carefully.

Q15. What is the current situation?

- (1) Large-scale fighting between the government and a rebel group has already begun.
- (2) Large-scale fighting between the government and a rebel group has not yet begun.
- (3) The conflict has already ended

Q16. According to the scenario, what would be the main consequence if the situation escalated into major violence?

- (1) Civilian casualties.
- (2) Economic collapse.
- (3) Collapse of the government.

Treatment: Warning condition

The survey makes a transition into a government statement regarding humanitarian intervention into the situation.

Regarding the situation, a U.S. government official made the following announcement.

(1) Baseline without early-warning:

We are closely monitoring ongoing tensions between the government and an armed opposition group in the country. Although tensions remain high, large-scale violence has not yet occurred.

Given the ongoing tensions in the country, the U.S. government is considering deploying military forces to prevent major violence. We estimate that such an operation would require approximately 5,000 U.S. military personnel and cost around \$1 billion.

(2) Early Warning with AI-based Forecasting (precision undisclosed)

We are closely monitoring ongoing tensions between the government and an armed opposition group in the country. Although tensions remain high, large-scale violence has not yet occurred.

However, we have recently received a warning that major violence is very likely to occur in the near future. This warning is based on an early-warning assessment conducted using artificial intelligence. The AI-based system analyzed a wide range of political, economic, security, and conflict-related information from the country and concluded that the likelihood of escalation into major violence within the next few months is extremely high.

Given this early warning, the U.S. government is considering deploying military forces to prevent major violence. We estimate that such an operation would require approximately 5,000 U.S. military personnel and cost around \$1 billion.

(3) Early Warning with AI-based Forecasting (60% precision)

We are closely monitoring ongoing tensions between the government and an armed opposition group in the country. Although tensions remain high, large-scale violence has not yet occurred.

However, we have recently received a warning that major violence is very likely to occur in the near future. This warning is based on an early-warning assessment conducted using artificial intelligence. The AI-based system analyzed a wide range of political, economic, security, and conflict-related information from the country and concluded that the likelihood of escalation into major violence within the next few months is extremely high. The system has a precision of about 60%. In other words, when it issued this type of warning in the past, major violence occurred in about 6 out of 10 similar cases.

Given this early warning, the U.S. government is considering deploying military forces to prevent major violence. We estimate that such an operation would require approximately 5,000 U.S. military personnel and cost around \$1 billion.

(4) Early Warning with AI-based Forecasting (90% precision)

We are closely monitoring ongoing tensions between the government and an armed opposition group in the country. Although tensions remain high, large-scale violence has not yet occurred.

However, we have recently received a warning that major violence is very likely

to occur in the near future. This warning is based on an early-warning assessment conducted using artificial intelligence. The AI-based system analyzed a wide range of political, economic, security, and conflict-related information from the country and concluded that the likelihood of escalation into major violence within the next few months is extremely high. The system has a precision of about 90%. In other words, when it issued this type of warning in the past, major violence occurred in about 9 out of 10 similar cases.

Given this early warning, the U.S. government is considering deploying military forces to prevent major violence. We estimate that such an operation would require approximately 5,000 U.S. military personnel and cost around \$1 billion.

(5) Early warning with Qualitative Evidence

We are closely monitoring ongoing tensions between the government and an armed opposition group in the country. Although tensions remain high, large-scale violence has not yet occurred.

However, we have recently received a warning that major violence is very likely to occur in the near future. This warning is based on an early-warning assessment conducted by regional conflict experts using field reports and contextual analysis. The experts reviewed reports from local NGOs, community leaders, and other sources familiar with conditions on the ground and concluded that the likelihood of escalation into major violence within the next few months is extremely high.

Given this early warning, the U.S. government is considering deploying military forces to prevent major violence. We estimate that such an operation would require approximately 5,000 U.S. military personnel and cost around \$1 billion.

Post-treatment Measurement

- If no foreign intervention were undertaken, how likely do you think it is that the situation would escalate into major violence? [Very unlikely (1) – Very likely (7)]
- How confident are you that the available information provides an accurate indication of whether major violence will occur? [Not at all confident (1) – Very confident (7)]
- To what extent do you think carrying out the proposed intervention would help fulfill your country's moral responsibility to protect people at risk in this situation? [Not at all (1) – To a very great extent (7)]
- How acceptable do you think the financial and military costs of the proposed intervention are, given the risk of major violence? [Not at all acceptable (1) – Completely acceptable (7)]
- To what extent would you support or oppose your government carrying out this operation? [Strongly oppose (1) – Strongly support (7)]

- How successful do you think the proposed action would be in preventing violence? [Very unsuccessful (1) – Very successful (7)]
- According to the scenario, what type of information was used to assess the risk of major violence?
 - (1) No specific early-warning method was described.
 - (2) An AI-based early-warning system.
 - (3) Analyses by local experts, NGOs, and community leaders.
 - (4) Not sure
- According to the scenario, what information was provided about how often major violence followed similar warnings?
 - (1) About 60%
 - (2) About 90%
 - (3) No information about previous warning performance was provided
 - (4) Not sure
- Is there a specific foreign country in Africa that comes to mind as you read the scenario above? If so, please write the name of the country. If not, please type “NA.”