

A Roadmap to Implementing Probabilistic Forecasting Methods

June 2022

ABOUT PERRY WORLD HOUSE

Perry World House is a center for scholarly inquiry, teaching, research, international exchange, policy engagement, and public outreach on pressing global issues.

Perry World House's mission is to bring the academic knowledge of the University of Pennsylvania to bear on the world's most pressing global policy challenges and to foster international policy engagement within and beyond the Penn community.

Located in the heart of campus at 38th Street and Locust Walk, Perry World House draws on the expertise of Penn's 12 schools and numerous globally oriented research centers to educate the Penn community and prepare students to be well-informed, contributing global citizens. At the same time, Perry World House connects Penn with leading policy experts from around the world to develop and advance innovative policy proposals.

Through its rich programming, Perry World House facilitates critical conversations about global policy challenges and fosters interdisciplinary research on these topics. It presents workshops and colloquia, welcomes distinguished visitors, and produces content for global audiences and policy leaders, so that the knowledge developed at Penn can make an immediate impact around the world.

Thanks to the experts named in Appendix: Interviews who provided valuable insight and feedback throughout the process. All errors are the sole responsibility of the authors. It does not represent the positions, policies, or opinions of Penn Global, Perry World House, or the University of Pennsylvania. Thanks to Open Philanthropy, whose support allowed for the drafting of this white paper.

REPORT AUTHORS

Julia Ciocca

Research Fellow, Perry World House

Michael C. Horowitz [on service leave]

Director and Richard Perry

Professor, Perry World House

Jared Rosen

Research Fellow, Perry World House

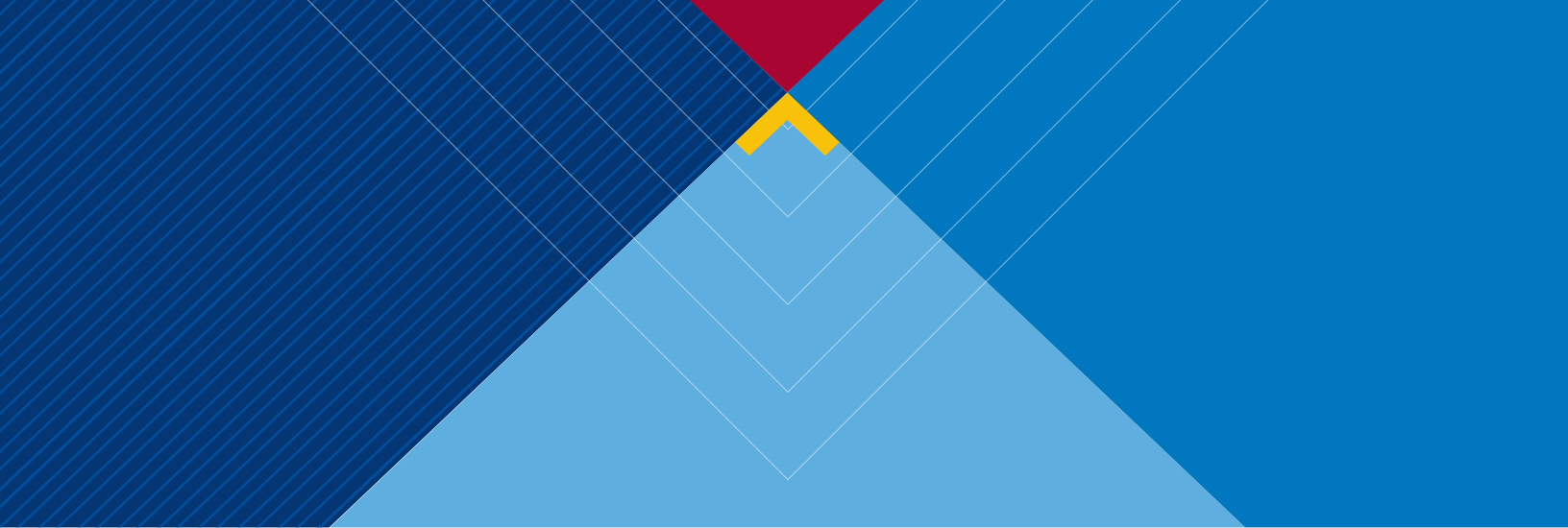
4	EXECUTIVE SUMMARY	
6	INTRODUCTION	
8	HISTORY	
12	PERRY WORLD HOUSE FORECASTING RESEARCH APPROACH	
	PREDICTIVE INTELLIGENCE ASSESSMENT METHODS (PRIAM) WORKING GROUP.....	12
	GLOBAL ORDER COLLOQUIUM 2021: HOW TO SEE THE FUTURE: FORECASTING AND GLOBAL POLICY.....	13
	INTERVIEWS WITH POLICY EXPERTS AND FORMER ANALYSTS.....	13
14	DETERMINING HOW TO MAKE FORECASTING PLATFORMS USEFUL FOR POLICYMAKERS	
15	RESULTS	
18	COSMIC BAZAAR: A MODEL FOR SUCCESS?	
19	INFER: FUTURE OF AMERICAN FORECASTING	
21	CHARTING A ROADMAP: OVERCOMING ADOPTION CHALLENGES & AVOIDING BACKSLIDING	
	BUREAUCRATIC LOCATION	21
	PREDICTION POLLING VS. PREDICTION MARKET.....	21
	ACCESSIBILITY AND USABILITY	22
	QUESTION-WRITING PROCESS	22
	INCLUDE PROBABILITIES AND FORECASTS IN ANALYTIC TRADECRAFT.....	22
23	CONCLUSION	
24	APPENDIX A: INTERVIEW QUESTIONS	
26	APPENDIX B: INTERVIEWS	

> EXECUTIVE SUMMARY

We live in an increasingly data-driven world, where algorithms drive the stock market and shape inventory for all kinds of goods. However, the U.S. intelligence community (IC) still provides analysis to policymakers that looks just as it did over the last several decades. With the announcement that the National Intelligence Council (NIC) will once again be piloting a crowdsourced probabilistic geopolitical forecasting platform, there is a unique opportunity for change. While past efforts to institutionalize crowdsourced forecasting foundered for bureaucratic reasons, the success of the United Kingdom's COSMIC BAZAAR platform illustrates that there is real potential for change.

Crowdsourced geopolitical forecasting is a powerful complement to traditional intelligence analysis, but just creating a platform is not enough. The IC will need to make choices about the platform, its questions, and how it communicates forecasts to IC leaders and policymakers. To inform those choices, this report by Perry World House draws on interviews with policymakers and analysts from the national security community, both those with no experience with geopolitical forecasting and those involved in developing forecasting platforms.





The countries that choose to invest in modernizing their intelligence processes through crowdsourced forecasting methods may gain a competitive advantage in their ability to make better decisions and strategies. Specifically, the results from our research and interviews suggest that, when implementing a new platform, the NIC and those involved in adoption should consider doing the following:

- 1. Use explicit percentages to convey statements of likelihood and probability.**
- 2. Create a classified and unclassified forecasting platform that is potentially open to allies and partner nations.**
- 3. Establish the forecasting platform in an office that is central to the IC, such as the Office of the Director of National Intelligence.**
- 4. Develop a forecast aggregation platform instead of a prediction marketplace.**
- 5. Provide context for aggregated forecasts.**
- 6. Embed the use of probabilities and forecasts into analytic tradecraft guidelines.**

Our interviews, which include many current and former U.S. national security officials, suggest support for these features could make crowdsourced geopolitical forecasting a helpful tool, along with traditional intelligence analysis, in informing U.S. national security decision-making. Given the advantage that crowdsourced forecasting offers, it will be critical for policymakers to think through how forecasting fits into their intelligence processes and how the information derived from forecasting platforms can be most useful to them as well as how features of those platforms can be beneficial.

➤ INTRODUCTION

Probabilistic geopolitical forecasting can be an effective tool for intelligence officers and those who rely on their analysis to better understand the world. Policymakers have an array of options at their disposal when designing geopolitical forecasting programs. These options concern who has access to the program, how they are supposed to interact with provided data, how to extract forecast data and communicate it, and much more. Building on previous research and a unique set of interviews with policymakers and analysts, this report lays out a road map for how governments can make choices to adopt probabilistic geopolitical forecasting methods, especially those involving crowdsourcing, spelling out the benefits to each option and weighing the trade-offs inherent in choosing one program versus another.

The U.S. intelligence community is the most advanced in the world, and while most of their successes are private, some, such as the accurate prediction of Russia's invasion of Ukraine, are more public. However, between the faster-than-expected spread of COVID-19 and the accelerated fall of Kabul from the U.S.' evacuation of Afghanistan, there is no shortage of uncertainty about the world, and more accurate geopolitical assessments would benefit the U.S. government. All policy decisions involve forecasting because crafting policy necessitates having a sense of what the current environment is, the outcome a policymaker desires, and how a policy intervention will influence the likelihood of the outcome.¹

Several organizations, including Perry World House, have supported the creation of a new probabilistic geopolitical forecasting platform within the U.S. intelligence community. The U.S. intelligence community (IC) announced in fall 2021 that the National Intelligence Council (NIC) would move ahead with launching such a platform, using crowdsourced methods.² That announcement follows on other announcements during the Biden administration designed to enhance U.S. government (USG) forecasting capacity. For example, the Biden administration announced the creation of a National Center for Epidemic Forecasting and Outbreak Analytics.³

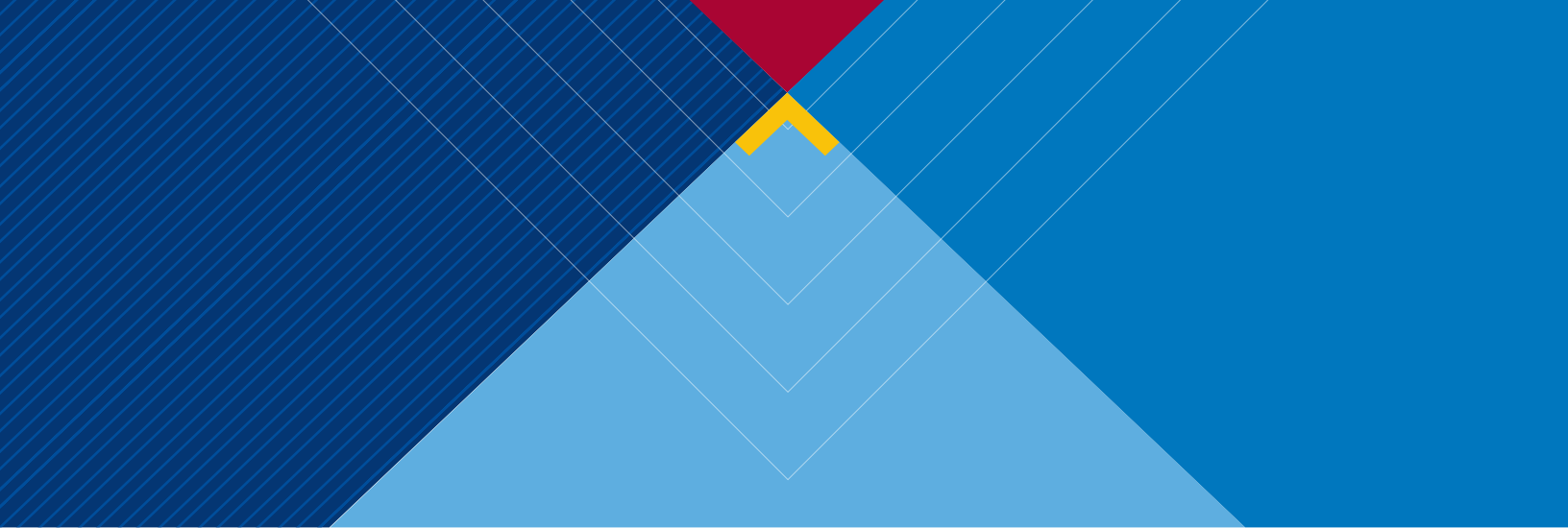
Decades of research into the efficacy of probabilistic forecasting show that methods like crowdsourcing and algorithms can provide information that complements traditional sources of intelligence analysis. These forecasting platforms provide data and insights that can help analysts more accurately assess the world and help policymakers better understand the perceived likelihood of an event. This is accomplished in two ways: quantitative analysis allows analysts to keep score of their correct and incorrect predictions and compels them to avoid vague language.⁴ In aggregate, this results in more accurate forecasts that can feed

1 Beauchamp, Zack. "This Study Tried to Improve Our Ability to Predict Major Geopolitical Events. It Worked." *Vox*, August 21, 2015. <https://www.vox.com/2015/8/20/9179657/tetlock-forecasting>.

2 Perry World House. "How to See the Future: Forecasting and Global Policy." Global Order Colloquium, September 27-28, 2021. <https://global.upenn.edu/perryworldhouse/how-see-future-forecasting-and-global-policy>.

3 "National Security Memorandum on United States Global Leadership to Strengthen the International COVID-19 Response and to Advance Global Health Security and Biological Preparedness." The White House Briefing Room, January 21, 2021. <https://www.whitehouse.gov/briefing-room/statements-releases/2021/01/21/national-security-directive-united-states-global-leadership-to-strengthen-the-international-covid-19-response-and-to-advance-global-health-security-and-biological-preparedness/>.

4 Horowitz, Michael C.; Ciocca, Julia; Kahn, Lauren; and Ruhl, Christian, "Keeping Score: A New Approach to Geopolitical Forecasting." Perry World House, February 2021. <https://global.upenn.edu/sites/default/files/perry-world-house/Keeping%20Score%20Forecasting%20White%20Paper.pdf>.



into intelligence assessments and policymaking. Changes in probabilities on crowdsourced platforms can also serve as early warning indicators that can direct IC and policymaker attention.

With the creation of new U.S. geopolitical forecasting platforms underway, this report will help policymakers better understand the tools at their disposal and different approaches moving forward:

- Where will the platform be housed in the government?⁵
- How will the platform handle classified information?
- Should the platform be a prediction market or forecasting aggregation platform?
- How will the questions be generated?
- How will forecasting fit into current analytical tradecraft methods?
- How will forecasts be communicated to analysts and policymakers?

Growing interest in data analytics across government and industry provides an opportunity to expand the adoption of crowdsourced geopolitical forecasting within the U.S. government, based on the new platforms being launched over the next year. These methods are a complement to traditional intelligence analysis, rather than a replacement. As the analysis below shows, they also have limits. But at a minimum, the early warning function that crowdsourced forecasting can enable for ongoing potential risks, such as a North Korean test of a nuclear weapon, illustrates the potential value for the IC and the policy community.

Keeping score is an essential practice that allows those charged with predicting to improve their skills and have a better understanding of where they were wrong and why. At the same time, replacing conveyances of likelihood with a number—rather than broad terms like “distinctly possible”—helps to clarify the muddied waters of intelligence briefings when the consequences are dire. By making strides toward regularizing the action of keeping score, the U.S. government will improve policymaking.

⁵ As described in this white paper, the National Intelligence Council (NIC) has announced that they will house the forthcoming geopolitical forecasting platform. This report explores other options as well, given the potential for changing bureaucratic conditions.

> HISTORY

The U.S. national security establishment has experimented with a variety of different approaches to probabilistic forecasting since the 1990s. The first prominent modern probabilistic forecasting program in the IC, still ongoing today, was the Political Instability Task Force in 1994.⁶ That program mostly involves the use of time series across countries to make forecasts about political instability around the world. Responding to the need for better interagency information-sharing after the 9/11 attacks, the Defense Advanced Research Projects Agency (DARPA) developed the Futures Markets Applied to Prediction/Policy Analysis Market, though the program was quickly shut down due to controversy about people betting on national security outcomes.⁷

As described in our last report,⁸ the Intelligence Advanced Research Projects Activity (IARPA) has launched a number of different programs over the last few decades. These include the Intelligence Community Prediction Market (ICPM), Forecasting and Understanding from Scientific Exploration (FUSE), Aggregative and Contingent Estimation (ACE), and Forecasting Science and Technology (ForeST).⁹ (See Figure 1.)

These programs featured different types of forecasting platforms and research initiatives, such as statistical models and machine learning, prediction marketplaces, or the use of “super-forecasters” with their independent methods. All represent quantitative forecasting methods that utilize statistical models or machine learning to predict future events.

Crowdsourced approaches have featured heavily in the most successful probabilistic geopolitical forecasting programs to date in the intelligence community. For instance, the ACE program sought to develop and test tools to provide accurate, timely, and continuous probabilistic forecasts and early warning of global events, by aggregating the judgments of many widely dispersed analysts.¹⁰ One team within the ACE program, the Good Judgment Project, outperformed the government’s own classified prediction market by margins of 25-30%.¹¹ Aggregated and crowdsourced methods are not one-size-fits-all approaches.

6 Political Instability Task Force. “Internal Wars and Failures of Governance, 1955-2005.” George Mason University (archived from the original), last updated October 3, 2006. <https://web.archive.org/web/20061208000556/http://globalpolicy.gmu.edu/pitf/>.

7 Hanson, Robin. “The Policy Analysis Market (and FutureMAP) Archive.” George Mason University, 2003. <https://mason.gmu.edu/~rhanson/policyanalysismarket.html>.

8 The previous Perry World House report on geopolitical forecasting, *Keeping Score: A New Approach to Geopolitical Forecasting*, is located here: <https://global.upenn.edu/perryworldhouse/keeping-score-new-approach-geopolitical-forecasting>.

9 Horowitz, et al. “Keeping Score: A New Approach to Geopolitical Forecasting.” 2021. 13.

10 Matheny, Jason. “Aggregated Contingent Estimation.” Proposal Day Template, IARPA Office of Incisive Analysis, May 19, 2010. https://www.iarpa.gov/images/PropersDayPDFs/ACE/ACE_Proposers_Day_Brief.pdf.

11 “Edge Master Class 2015: A Short Course in Superforecasting, Class II.” Interview with Philip Tetlock. August 24, 2015. https://www.edge.org/conversation/philip_tetlock-edge-master-class-2015-a-short-course-in-superforecasting-class-ii.

FIGURE 1: HISTORY OF FORECASTING PLATFORMS

Program Name	Agency	Release Year	Shutdown Year	Platform Details
Political Instability Task Force (PITF)	Central Intelligence Agency	1994	N/A	Initiative housed at George Mason University that curates a dataset to explain the causes of political instability in countries.
Policy Analysis Market (PAM)	DARPA	2003	2003	A marketplace for participants to invest in futures in geopolitical events.
Integrated Conflict Early Warning System (ICEWS)	DARPA	2008	2018	Comprehensive, integrated, automated, generalizable, and validated system to monitor, assess, and forecast national, sub-national, and internal crises.
Intelligence Community Prediction Market (ICPM)	IARPA	2010	N/A	A voluntary marketplace for top-secret cleared government employees and contractors to invest in the outcomes of geopolitical events.
Foresight and Understanding from Scientific Exposition (FUSE)	IARPA	2010	2016	Develops automated methods that aid in the systematic, continuous, and comprehensive assessment of technical emergence using information found in published scientific, technical, and patent literature.
Aggregative Contingent Estimate (ACE)	IARPA	2010	2015	"The ACE Program seeks technical innovations in the following areas: (a) efficient elicitation of probabilistic judgments, including conditional probabilities for contingent events; (b) mathematical aggregation of judgments by many individuals, based on factors that may include: past performance, expertise, cognitive style, metaknowledge, and other attributes predictive of accuracy; and (c) effective representation of aggregated probabilistic forecasts and their distributions." Teams of superforecasters competing. Good Judgment beat everyone.
Open Source Indicators (OSI)	IARPA	2012	2015	"IARPA's Open Source Indicators (OSI) Program aims to fill this gap by developing methods for continuous, automated analysis of publicly available data in order to anticipate and/or detect significant societal events, such as political crises, humanitarian crises, mass violence, riots, mass migrations, disease outbreaks, economic instability, resource shortages, and response to natural disasters. Performers will be evaluated on the basis of warnings that they deliver about real-world events." ¹

1 Sanghani Center for Artificial Intelligence & Data Analytics. "Open Source Indicators (OSI) Grant for Embers." Virginia Tech. 2022. <https://sanghani.cs.vt.edu/grant/iarpa/#:-:text=IARPA's%20Open%20Source%20Indicators%20>.

Figure 1 (cont.)

Program Name	Agency	Release Year	Shutdown Year	Platform Details
Forecasting Science and Technology (ForeST)	IARPA	2013	2015	Led by SciCast at George Mason. Focuses specifically on developments in science and technology, and uses similar techniques. Tracking developments in these fields may help identify advances in weapons systems or emerging technologies in bioterrorism or cyberthreats.
Cyberattack Automated Unconventional Sensor Estimate (CAUSE)	DARPA	2015	2019	An open-source-oriented program with the hope of predicting cyberattacks.
Mercury	IARPA	2015	2019	Researchers wanted to use SIGINT to predict events, similar to how OSI used open-source data, but it was classified. ²
Hybrid Forecasting Competition (HFC)	IARPA	2016	2020	Teams of forecasters in psychology, computer science, and more came together for a multi-year program to develop a program that could predict behaviors and geopolitical events. ³
Forecasting Counterfactuals in Uncontrolled Settings (FOCUS)	IARPA	2017	2021	FOCUS develops and empirically evaluates systematic approaches to counterfactual forecasting. The program, conducted by Johns Hopkins University, produced data that could serve well for future prediction research relying on counterfactuals.
Geopolitical Forecasting Challenge (GFC)	IARPA	2018	2018	A challenge hosted by IARPA that incentivized people to develop forecasting methods for geopolitical events by offering prizes for those with the most accurate predictions.
INFER	National Intelligence Council	2021 -Present	N/A	Currently under development.

Some formats are more suitable for certain departments and jobs than others, but they all share a desire to create falsifiable predictions.

As the prior Perry World House report on geopolitical forecasting described, previous U.S. geopolitical forecasting initiatives, especially those focused on crowdsourcing, have largely stalled or died out for three main reasons. First, some probabilistic forecasting programs moved forward the science of forecasting but

did not have a clear end-user in mind and did not invest enough in communicating with them. Second, geopolitical forecasting represents an advance in technology that is not tangible in the same way that advances in technologies like stealth capabilities or hypersonic gliders are. It is harder for policymakers and the users of forecasting platforms to see the effects of utilizing the program when they are so distanced. Third, as a Government Accountability Office report

2 Rockwell, Mark. "Using SIGINT to detect trends." FCW. February 10, 2015. <https://fcw.com/digital-government/2015/02/using-sigint-to-detect-trends/244427/>.
 3 Dawson, Caitlin. "USC ISI Leads IARPA Contract for developing Hybrid Forecasting Systems." USC Viterbi. October 11, 2017. <https://viterbischool.usc.edu/news/2017/10/usc-isi-leads-iarpa-contract-developing-hybrid-forecasting-systems/>.

referenced, a “chasm” exists between the developed technology and that used by their intended users, “often referred to by department insiders as ‘the valley of death,’ because the acquisition community often requires a higher level of technology maturity than the S&T (science and technology) community is willing to fund and develop.”¹² Ultimately, it seems that this temporary neglect of geopolitical forecasting had little to do with the political saliency of the initiatives and more to do with the valley of death in shifting programs from research and development into more permanent adoption and use.¹³

On the other side, the success of probabilistic forecasting initiatives requires a larger degree of support within the IC. Policymakers have long needed communicable probabilities to better understand the ramifications of policies and what geopolitical events will occur imminently, but this need was not reflected in the analytic tradecraft standards in the IC until recently. Looking at the Intelligence Community Directive (ICD) from 2007, there is no discussion of probabilities in the analytic standards.¹⁴ In fact, IC analytic standards had long been neglected. In a 1991 report by the U.S. House of Representatives Committee of the Armed Services Oversight and Investigations Subcommittee on intelligence successes and failures in Operations Desert Storm and Shield, Gen. Norman Schwarzkopf stated, “I personally feel that there’s a serious need to develop a standardized methodology within the intelligence community for making estimates and predictive analysis. ... The analysis we received was unhelpful.”¹⁵

The 2007 ICD 203 standardized intelligence by codifying and reinstating historical analytical tradecraft standards. Throughout the 20th century, analytic standards within the IC fluctuated based on the desires of policymakers. However, intelligence failures led the IC to institute standards such as objectivity, independence from political beliefs, timeliness, and use of all relevant intelligence sources.¹⁶ Eight years later, in 2015, the Office of the Director of National Intelligence (ODNI) released a new edition of the ICD 203 Analytic Standards, which involved the use of specified estimative language, specifically linking certain words or phrases to ranges of probability.¹⁷ Prior to 2007, analytic standards within the IC were in flux and mandated through the Intelligence Reform and Terrorism Prevention Act. The introduction of ICD 203 systemized analytic standards within the IC. Then the 2015 analytic standards within the IC set probabilities for estimative language.¹⁸ However, these probability ranges are too vague. For instance, the term “likely” is given a range of 55-80% probability, representing a moderate level of confidence.¹⁹ This range spans 25 percentage points. There is a marked difference between there being an 80% chance an event will occur and a 55% chance, but the two probabilities correspond to the same estimative language. Analytic standards allow for consistent, useful intelligence reports open to the integration of new prediction methods, especially geopolitical forecasting.

12 U.S. Government Accountability Office. *Defense Advanced Research Projects Agency: Key Factors Drive Transition of Technologies, But Better Training and Data Dissemination Can Increase Success* (Washington, DC: U.S. Government Accountability Office, November 2015). 3-5. <https://www.gao.gov/assets/680/673746.pdf>.

13 Ibid. See also Niewood, Eliahu H. “Intelligence After Next: Mission-Based Challenges for the Intelligence Community.” *MITRE*, February 19, 2021. <https://www.mitre.org/publications/technical-papers/intelligence-after-next-mission-based-challenges-intelligence-community>.

14 Office of the Director of National Intelligence. “Intelligence Community Directive 203.” June 27, 2007. <https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards%20pdf-unclassified.pdf>.

15 *Intelligence Successes and Failures in Operations Desert Shield/Storm*. Report of the Oversight and Investigations Subcommittee of the Committee on Armed Services, House of Representatives, 103rd Cong., 1st sess. (1993), 30. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a338886.pdf>.

16 Marchio, Jim. “Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis.” *Intelligence and National Security* 29(2): 159-83, March 4, 2014. <https://doi.org/10.1080/02684527.2012.746415>.

17 Office of the Director of National Intelligence, “Intelligence Community Directive 203.” January 2, 2015. <https://irp.fas.org/dni/icd/icd-203.pdf>.

18 Department of the Army, Headquarters. “Intelligence Analysis.” Army Techniques Publication, No.2-33.4. January 10, 2020. <https://irp.fas.org/doddir/army/atp2-33-4.pdf>.

19 Department of the Army, Headquarters. “Intelligence Analysis.” 128; Director of National Intelligence, “Intelligence Community Directive 203.” January 2, 2015. 3.

➤ PERRY WORLD HOUSE FORECASTING RESEARCH APPROACH

To move the debate inside and outside the U.S. government on geopolitical forecasting forward, this report draws on several approaches to explore the key questions outlined above. These include a working group made up of academic experts, policymakers, government officials, and developers; a series of interviews with former and current policymakers and analysts;²⁰ and events and programs such as our Fall 2021 Global Order Colloquium on forecasting and global policy.²¹

Through these projects, Perry World House has compiled input from former policymakers and analysts, as well as academics and practitioners responsible for building and developing forecasting platforms. This aggregation lets us generate a clearer understanding of the goals, concerns, and questions held by both policymakers and analysts as well as potential solutions.

PREDICTIVE INTELLIGENCE ASSESSMENT METHODS (PRIAM) WORKING GROUP

Perry World House's PRIAM Working Group focused on convening policymakers, academics, and experts responsible for designing and implementing forecasting programs. The group met nine times throughout the year, starting in April 2021 and ending in February 2022.²² Each meeting of the working group began with brief presentations from some members regarding their work with forecasting and the insights that they have gained into making forecasting more accessible and effective. These presentations were followed by an hour of conversation among the working group members. Discussions focused on finding solutions to the concerns raised by policymakers or challenges faced by previous iterations of forecasting platforms, so that future systems could be implemented in a better way. They also enabled members to get a better sense of how these issues were being addressed by various fields, firms, and governments, serving as a centralized place in which to share ideas and communicate lessons learned.

²⁰ See Appendix B for a list of policymakers and analysts interviewed for this report.

²¹ Information about the Global Order Colloquium can be found here: <https://global.upenn.edu/perryworldhouse/how-see-future-forecasting-and-global-policy>.

²² The PRIAM Working Group met on April 12, 2021; May 7, 2021; May 11, 2021; May 13, 2021; June 7, 2021; June 29, 2021; July 6, 2021; November 10, 2021; and February 17, 2022.

GLOBAL ORDER COLLOQUIUM 2021: HOW TO SEE THE FUTURE: FORECASTING AND GLOBAL POLICY

Expanding upon the PRIAM Working Group, Perry World House centered its 2021 Global Order Colloquium, titled *How to See the Future: Forecasting and Global Policy*, on the issue of forecasting. The colloquium provided an opportunity to expand the conversation on geopolitical forecasting by hosting three group discussions centered around learning how international economics and global public health experts, in addition to national security experts, use forecasting in their fields.

Perry World House also held three keynote discussions with Deputy Director of National Intelligence Morgan Muir; former President of Liberia Ellen Johnson Sirleaf; and former Deputy National Security Advisor Ben Rhodes. These conversations provided an opportunity to expand on the conversations from the working group that focused on the implementation of forecasting to better understand the areas where forecasting can be useful and what knowledge gaps within these areas forecasting can fill.²³

For instance, climate and security experts speaking at the colloquium had a strong desire for better forecasting data on potential policy initiatives. However, public health experts were more confident in their abilities to make predictions about disease outbreaks. The problem they identified was about communicating those predictions to policymakers and government officials. The colloquium also provided the opportunity for Perry World House to expand the conversation around forecasting, including having colloquium participants publish four pieces on national security forecasting in the journal *Survival*.²⁴

INTERVIEWS WITH POLICY EXPERTS AND FORMER ANALYSTS

Finally, Perry World House conducted twenty-five interviews with policy experts and former analysts to better understand their views on forecasting, how they feel forecasting can be made most effective, and how forecasting systems can overcome the bureaucratic challenges faced by previous attempts at implementation. These interviews provide new information on the elements of probabilistic forecasting, from question design to how to communicate information about the forecasts, that policymakers and analysts find most useful and the questions and challenges that those implementing forecasting face. When policy experts indicated that they had experience using such platforms, we asked them questions about those experiences and how they could be improved. The interview questions are in Appendix A.

Interview candidates were selected based on their experience within the U.S. government and varied in seniority, office, and responsibilities. We interviewed current and former senior government staff while also making sure to interview individuals who had worked as analysts. We did not limit our interviews to only senior staff because in order for a forecasting platform to be successful it would have to cater to the needs of analysts as well as those who receive analysts' reports.

23 Moore, Breanna; Rosen, Jared; and Ruhl, Christian. "How to See the Future: Forecasting and Global Policy." Perry World House, 2021. <https://global.upenn.edu/sites/default/files/perry-world-house/PWH-2021-GO-Colloquium-Report.pdf>

24 The pieces produced by participants of Perry World House's colloquium on forecasting can be found in *Survival: Geopolitics and Strategy*. <https://www.tandfonline.com/journals/tsur20>.

> DETERMINING HOW TO MAKE FORECASTING PLATFORMS USEFUL FOR POLICYMAKERS

Success for a new generation of U.S. geopolitical forecasting initiatives, especially based on crowdsourced methods, requires focusing first and foremost on the needs of end-users, both intelligence analysts initially consuming forecasts and policymakers that will eventually see them. Including policymakers in the discussion at the development stages is crucial to better understanding their approach to and objectives for forecasting. Interviews, like the ones we conducted, provide the opportunity to better understand and engage with the analysts, decision-makers, and communities who will be most likely to engage in forecasting.

Interview questions centered on gauging experience with forecasting, ideas on how to make forecasting most effective, what policymakers would like to see from forecasting, and what concerns might prevent adoption efforts. Additionally, the interview questions also measured participants' views toward the likelihood of adoption.²⁵ The participants varied in their experience and attitudes toward forecasting based on their backgrounds,²⁶ including those from the U.S. Department of Defense, the IC, and the U.S. Department of State. Their responses create new evidence on how to implement probabilistic geopolitical forecasting successfully in the U.S. government and beyond.

²⁵ A full overview of the questions asked of each participant can be found in Appendix A.

²⁶ A full list of survey participants can be found in Appendix B.

> RESULTS

The results show what policymakers believe are valuable features of a geopolitical forecasting platform. This list can help inform those developing a platform on the perspectives of their potential customers. The results below are the outcomes of Perry World House's working groups, interviews, and research of forecasting models and methodologies within the government.

ITEM ONE: INCLUDE EXPLICIT PERCENTAGES

The most immediate desire from policymakers is to present probabilities as explicit percentages as opposed to using terms of likelihood such as “likely,” “moderately likely,” or “possible.” For instance, one policymaker explained they would prefer, “percentage probability, probably rounded to nearest 5, not to oversell precision of the cited methodology.” The policymaker explained that assessments of high, medium, and low likelihoods are all assessments made by analysts rather than consensus from an array of experts. These phrases, effectively detached from numerical values, serve almost no purpose because they correspond to different conceptions of probability to the analyst, briefer, and policymaker. One policy expert interviewed stated plainly: “I like percentages. Descriptor words are only helpful if they’re anchored to precise estimates.” Nowadays, the IC has somewhat adapted to the needs of policymakers by assigning percentages to estimations of likelihood, but as discussed earlier, the probability ranges tied to those terms are far too broad. Furthermore, the use of explicit forecasting platforms ties the probabilities to numerical values calculated from the estimates of fellow users in the IC.

During our interviews, we also asked experts, “What probability would you assign to the phrase ‘moderately likely’?” While most responses varied from 50% to 75%, the average response was about 56%, not too far from the minimum. Furthermore, when experts were asked what their preferred method of communication for likelihood would be, most said that they would prefer percentages. Thus, even with official percentage ranges assigned to estimative language, individual understanding of phrases like “moderately likely” varies significantly and does not necessarily correspond with the officially prescribed percentage range.



ITEM TWO: EXPLAIN DISAGREEMENT BETWEEN ANALYSTS AND THE CROWD

Across nearly all interviews, experts stated that analysts who disagree with forecasts should be required to explain why they disagree with the prediction. Working group members suggested that analysts should be mandated to include platform-generated forecasts in their analysis to ensure adoption into the analysis process. However, policymakers and IC interviewees worried that would create overly rigid guidelines for analysis. Instead, interviewed officials thought that analysts who disagree with the aggregated forecast should have to explain what methods they use to come to their conclusions, discuss the intelligence they include in their memo, and potentially explain why the forecasting platform is incorrect.

Concerning whether analysts should explain their disagreements with forecasted probabilities, one interviewed expert said: “Air the disagreement. You know there’s a tradition for dissenting views in a product. That’s one thing that the NIC has done really well is institutionalize that.” Additionally, subjects believed that analysts who utilize forecasting platforms should explain where the presented number stems from. For instance, if an analyst used a forecasting platform centered in the NIC that received answers from the Central Intelligence Agency, Defense Intelligence Agency, and the Office of Naval Intelligence, then they should detail which offices contributed and how the platform is structured. One interviewed policy expert explained: “I would want to know the evidence behind [the forecast] and the justification for it. And also not just positive but negative. What’s dampening the likelihood?” While analysts should not necessarily be required to utilize forecasts, the ways that forecasting platforms are used should be standardized and regulated to ensure uniformity across agencies and reports.

ITEM THREE: OPEN FORECASTING PLATFORMS TO ALLIES AND PARTNERS

Interviewed policymakers also agreed that intelligence agencies from allied states should also be included in the forecasting platform. While no one specified which countries should or should not be included, policymakers could include multilateral institutions or agreements like NATO, AUKUS, or Five Eyes rather than specifying certain countries. Including allies in forecasting platforms potentially offers more complex and varied perspectives based on different intel gathered and different perspectives or expertise on

geopolitical issues. One policy expert summed it up best: “How forecasting works is based on assumptions and indicators; their assumptions or culture will be necessarily different and therefore important. [It’s] good to see how they are seeing or addressing problems. For more sensitive information and questions, certain intelligence agencies could be specified, and individual forecasts could be shared with allies on a case-by-case basis.”

ITEM FOUR: CREATE CLASSIFIED AND UNCLASSIFIED FORECASTING PLATFORMS

The development of forecasting platforms can be flexible regarding what kind of questions and information is permitted in the forecasting platform. When asked if they would prefer a classified or an unclassified system, policymakers highlighted the benefits of having both available. A classified system would be narrower in scope and involve fewer participants but could enable analysts and policymakers to forecast using more sensitive information. On the other hand, an unclassified platform could be broad enough to include all government employees or be bridged to include different branches while creating fewer security concerns. An unclassified platform may also be easier to access due to fewer security restrictions, promoting more widespread use. Moreover, a broad user base could also allow intelligence agencies to identify superforecasters within other branches of the U.S. government, or outside the government, depending on access rules.

One interviewed policy expert added that having an unclassified geopolitical forecasting platform would also draw analysts away from relying solely on classified information: “People have been conditioned to think that classified info is gospel.” An unclassified platform would also draw analysts away from the spectacle of classified information that may seem groundbreaking but on the whole unreliable. As both classified and unclassified platforms are desired by policymakers, developers may consider creating tiered platforms that can support both unclassified and classified forecasts. Describing the value of having both classified and unclassified platforms, one policy expert said: “A classified version is going to be dominated by ostensible ‘experts.’ We know that experts are the people least likely to see discontinuities. So, you would like to find ways to incorporate people that are good predictors but don’t know much about substance. [But] in the classified version, there was a fair amount of gaming going on. I would like to have both.”

ITEM FIVE: FOCUS ON DELIVERY AND POLICYMAKER INTERACTION

In the interview process, policymakers were also asked about how they would like to receive forecasts. These questions focused on how forecasts would be presented to policymakers, how frequently forecasts would be updated, and what information would be available to policymakers. Throughout the interviews, most policymakers noted concerns about information getting lost or not properly communicated. Thus, they felt that including forecasts in traditional intelligence briefings and reports as an additional data point, especially in instances of analyst disagreement, would help the forecasting process to remain an integrated part of the intelligence process in the long term. One participant echoed the idea that including crowdsourced forecasts in a regular briefing like the President's Daily Briefing (PDB) would be useful saying: "There is the president's daily brief and there is a lot of stuff that comes in from the IC. I think it would be useful to understand a crowdsourced probability in the context of other intel." In addition to including the forecasts in reports, some participants in both the interview process and the PRIAM working group also suggested that policymakers have the opportunity to dig deeper into the data underlying the forecasts if desired.

While including forecasts in more traditional intelligence briefings and reports like the PDB would ensure a smooth integration process and immediate relevance, the delivery of forecasts should not just reflect a passive data point, given the data potentially available over time. In addition to these measures, some participants suggested enabling analysts and policymakers to be able to view forecasting trends over time, in a similar fashion to how market trends are communicated. Others also suggested creating routine times in which forecasts are published, so that policymakers know when to expect updated information. Through these tools, both policymakers and analysts would be able to better understand how geopolitical situations develop and trend over time, leading to the ability to make more informed and complex analyses and decisions in the future.

While there are concerns that providing an overwhelming amount of information would be counterproductive, most felt that allowing opportunities for policymakers to choose when they would like to have deeper engagement with the underlying forecasting

thought process would help to create better decision-making by allowing for more clarity. Policymakers directing analysts on the level of data they would like to see is more efficient than the alternatives and avoids the Goldilocks dilemma of providing overwhelming, and therefore unusable, amounts of information or leaving policymakers unintentionally in the dark on critical data points. Developing forecasting platforms that enable a level of flexibility in this regard can significantly increase the value of such platforms, making them more likely to be adopted.

ITEM SIX: CONTEXTUALIZE FORECASTS

It is also important to figure out how to address the logic behind the forecasts themselves. For example, there is a lack of trust in issues surrounding artificial intelligence (AI) because the outputs from algorithms seem like a black box. Senior IC leaders and policymakers are less likely to trust crowdsourced forecasts if they do not have a sense of why the forecasts land where they do. Incorporating the argumentative logic of the forecasters will help increase trust in the forecasts across the board, making them more usable.

As explained above, experts and policymakers highlighted the importance of giving analysts the context of crowdsourced predictions. Forecasting platforms can meet this demand in two ways. First, they can offer background information as to how and why a question was fielded. Second, they can describe who the participants are in the given question (e.g., participants from the broader IC, participants with top-secret clearance and above, general participation from across government, etc.). Context can help analysts who use forecasts to trust the platform and further understand what trend they are looking at. Discussing the importance of context, one expert said: "I'd want to know how the platform worked. ... Who's being asked? Why were they selected? What was the context of the question being asked (is it in a series of questions)? When was the question posed? Who designed the question?" Answering questions like these can better enable platforms to provide meaningful and usable information to their userbase, increasing the impact and value-add of geopolitical forecasting as well as its likelihood of adoption and longevity. Keeping platforms transparent assuages analysts' initial skepticism and eliminates a barrier to using important data.

> COSMIC BAZAAR: A MODEL FOR SUCCESS?

The United Kingdom currently operates its own probabilistic geopolitical forecasting platform named COSMIC BAZAAR. COSMIC BAZAAR is an aggregative program spanning the whole U.K. government and boasting more than 1,000 participating forecasters and 19,000 forecasts as of November 2021; in April 2021, *The Economist* reported that “more than 10,000 forecasts [had] been made by 1,300 forecasters, from 41 government departments and several allied countries” since the program’s launch in the previous year.²⁷ As forecasters make their predictions, COSMIC BAZAAR ranks them by the accuracy of their predictions. There is no material gain, only the pride of knowing how good (or bad) you are at making accurate predictions about geopolitical events.

COSMIC BAZAAR has demonstrated amazing results in terms of its wide use throughout the U.K. government. But beyond the number of users, COSMIC BAZAAR’s strength lies in its diversity of forecasters and anonymity of forecasters.

While no information about the demographics of COSMIC BAZAAR’s users have been released, the widespread participation in the program suggests that they represent a cognitively diverse group of forecasters as they span not only departments but international borders. Philip Tetlock, an expert on the study of geopolitical forecasting, has identified cognitive diversity as an important part of forecasting.²⁸ COSMIC BAZAAR’s breadth of engagement likely improves the accuracy of its forecasting and creates the ability to identify skilled forecasters across all governmental departments. The United States should look to COSMIC BAZAAR as an example of an open-source forecasting model that has drawn participants from all over the U.K. government. COSMIC BAZAAR shows how such a platform can take predictions from individuals in many departments, how researchers can develop those predictions, and how accurate those predictions tend to be.

27 *The Economist*. “How Spooks Are Turning to Superforecasting in the Cosmic Bazaar.” April 15, 2021. <https://www.economist.com/science-and-technology/2021/04/15/how-spooks-are-turning-to-superforecasting-in-the-cosmic-bazaar>.

28 Tetlock, P.E., and Gardner, D. *Superforecasting: The Art and Science of Prediction*. New York: Crown, 2015. 141.

> INFER: FUTURE OF AMERICAN FORECASTING

At Perry World House’s 2021 Global Order Colloquium, Deputy Director of National Intelligence Morgan Muir announced that the NIC would launch a new geopolitical forecasting capability in 2022.²⁹ Then, in February, the first of three anticipated crowdsourced forecasting platforms was launched under the Integrated Forecasting and Estimates of Risks (INFER) project. Designed and operated by the Applied Research Laboratory for Intelligence and Security (ARLIS) at the University of Maryland and their industry partner, Cultivate Labs, the first platform, called INFER-Public, is designed to capture and aggregate forecast judgments from a wide swath of different forecasters, including researchers, students, and other public communities of interest.

Two more platforms—which will engage participants across the U.S. civilian federal agencies, the U.S. Department of Defense, and the IC—are then expected to launch later in 2022.

INFER’s initial forecasting topics are focusing on emerging science and technology trends in AI, quantum computing, biotechnology, and climate change impacts. ARLIS will share results with U.S. decision-makers to help them develop policies that will improve U.S. science and technology competitiveness. In the future, forecasting topics may expand to geopolitics and economic and societal trends, among other subjects.

INFER-Pub is the evolution of a previous forecasting platform called Foretell from the Center for Security and Emerging Technology at Georgetown University.

²⁹ Morgan Muir’s keynote address can be listened to in its entirety on the Perry World House YouTube channel here: <https://www.youtube.com/watch?v=qmwbXkJA-sA>.

FIGURE 2: COMPARATIVE FEATURES OF FORECASTING PLATFORMS¹

	Intelligence Community Prediction Market (ICPM)	COSMIC BAZAAR	Good Judgment
Open Source or Classified	Classified ²	Open source ³	Open source ⁴
Ally Inclusion	No	Yes ⁵	Public use
Location	United States Intelligence Community ⁶	Broader U.K. government (i.e., civil servants, diplomats, and intelligence analysts) ⁷	Open to the public and began in collaboration with IARPA-ACE ⁸
Program Format (Marketplace vs. Forecast Aggregation)	Marketplace ⁹	Prediction polls ¹⁰	Mixed methods ¹¹
Required vs. Voluntary Analyst Engagement	Voluntary ¹²	Voluntary ¹³	Voluntary ¹⁴

1 For more information on previous forecasting platforms and their development, please refer to Perry World House’s report Keeping Score: A New Approach to Geopolitical Forecasting, which is available here: <https://global.upenn.edu/sites/default/files/perry-world-house/Keeping%20Score%20Forecasting%20White%20Paper.pdf>.

2 Mandel, David R. “Too Soon to Tell if the US Intelligence Community Prediction Market Is More Accurate than Intelligence Reports: Commentary on Stastny and Lehner (2018).” Judgment and Decision Making 14(3), 2019. <http://journal.sjdm.org/19/190417/jdm190417.pdf>.

3 The Economist. “How Spooks Are Turning to Superforecasting in the Cosmic Bazaar.” April 14, 2021. <https://www.economist.com/science-and-technology/2021/04/15/how-spooks-are-turning-to-superforecasting-in-the-cosmic-bazaar>.

4 Good Judgment. “Who Made More Accurate Forecasts—Superforecasters or Intelligence Analysts?” 2022. <https://goodjudgment.com/resources/the-superforecasters-track-record/superforecasters-vs-the-icpm/>.

5 The Economist. “How Spooks Are Turning to Superforecasting in the Cosmic Bazaar.”

6 McHenry, Jonathan. “Three IARPA forecasting efforts: ICPM, HFC, and the Geopolitical Forecasting Challenge.” Federal Foresight Community of Interest. January 26, 2018. <https://www.ffcoi.org/wp-content/uploads/2019/03/Three-IARPA-Forecasting-Efforts-ICPM-HFC-and-the-Geopolitical-Forecasting-Challenge-Jan-2018.pdf>.

7 The Economist. “How Spooks Are Turning to Superforecasting in the Cosmic Bazaar.”

8 Schoemaker, Paul J., and Tetlock, Philip E. “Superforecasting: How to Upgrade Your Company’s Judgment.” Harvard Business Review, May 2016. <https://hbr.org/2016/05/superforecasting-how-to-upgrade-your-companys-judgment>.

9 McHenry. “Three IARPA Forecasting Efforts.”

10 The Economist. “How Spooks Are Turning to Superforecasting in the Cosmic Bazaar.”

11 Goldstein, Seth, et al. “Assessing the Accuracy of Geopolitical Forecasts from the US Intelligence Community’s Prediction Market.” Good Judgment, 2020. <https://goodjudgment.com/wp-content/uploads/2020/11/Goldstein-et-al-GJP-vs-ICPM.pdf>.

12 Mandel. “Too Soon to Tell if the US Intelligence Community Prediction Market Is More Accurate than Intelligence Reports.”

13 The Economist. “How Spooks Are Turning to Superforecasting in the Cosmic Bazaar.”

14 AI Impacts. “Evidence on Good Forecasting Practices from the Good Judgment Project: An Accompanying Blog Post.” <https://aiimpacts.org/evidence-on-good-forecasting-practices-from-the-good-judgment-project-an-accompanying-blog-post/>.

➤ CHARTING A ROADMAP: OVERCOMING ADOPTION CHALLENGES & AVOIDING BACKSLIDING

Bureaucratic support, both from decision-makers and analysts utilizing these systems, will be a critical part of ensuring the successful adoption of any geopolitical forecasting approach. Based on our research, key strategies for building support within the government include location within the government, the platform's design, analyst engagement, and bureaucratic support. While these areas may create difficulties for adoption, they can be overcome by effectively thinking through the purpose and goals of the platform and its ability to improve predictions and decision-making.

BUREAUCRATIC LOCATION

The bureaucratic location of the forecasting platform may play an important role in helping to determine the influence, value, and role of the forecasting platform. For example, as referenced above, the NIC will be hosting a new, classified geopolitical forecasting platform. This makes sense given its overarching role in the IC. For the same reason, ODNI could also host a forecasting platform outside the context of the NIC if it becomes necessary.

The location may influence the scope of the platform because certain agencies might be better at fostering interagency communication. Furthermore, agencies that handle more classified information might be less willing to share intel with non-IC agencies or less intelligence-focused agencies. An institution like the NIC or another part of ODNI should be able to bridge gaps within the IC and get insights from other branches of government.

PREDICTION POLLING VS. PREDICTION MARKET

Crowdsourced geopolitical forecasting can take the form of a marketplace or forecasting aggregation platform. While economists argue that markets are inherently more accurate than forecasting aggregation platforms, recent research building on the experiences of the Good Judgment Project's results in the ACE tournament shows that prediction polling can out-perform prediction marketplaces.³⁰ One reason that aggregation can be more accurate is that marketplaces incentivize savviness, where participants can be rewarded less for the accuracy of predictions and more for the timing of

30 "Edge Master Class 2015: A Short Course in Superforecasting, Class II." Interview with Philip Tetlock. August 24, 2015; Dana, Jason; Atanasov, Pavel; Tetlock, Philip; and Mellers, Barbara. https://www.edge.org/conversation/philip_tetlock-edge-master-class-2015-a-short-course-in-superforecasting-class-ii.

"Are Markets More Accurate Than Polls? The Surprising Informational Value of 'Just Asking,'" *Judgment and Decision Making*, 14(2), March 2019.

<http://journal.sjdm.org/18/18919/jdm18919.html>. See also Atanasov, Pavel; Rescober, Phillip; Stone, Eric; Swift, Samuel A.; Servan-Schreiber, Emile; Tetlock, Philip; Ungar, Lyle; and Mellers, Barbara. "Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls." *Management Science*, 63(3): 691-706, 2016.

them. Therefore, we recommend using a prediction polling forecast aggregation platform rather than a prediction market to maintain accurate forecasts, which are easier to explain to both analysts and policymakers who interact with the forecasts.

ACCESSIBILITY AND USABILITY

Platform design will also be a critical factor in either promoting or hindering the adoption of a forecasting platform. Through the interviews with policymakers and working group meetings of forecasting experts, a key element repeatedly cited as an important feature in promoting forecasting was the accessibility and usability of the platform. Systems that are difficult to use or hard to access will likely see declines in engagement and resistance to using their predictions, even when the platform produces high-quality predictions.

Creating a forecasting platform that offers users an intuitive and accessible design will help make integrating forecasting platforms into existing analysis methods a more seamless process. It is critical for the results to be easy for analysts to understand, and for there to be output options that help them include forecasts in reports that go up to policymakers and senior leaders in the IC. Widgets and other options that enable the rapid export to IC products will help increase the probability of adoption.

Additionally, in the design stages, forecasting platforms should implement designs and parameters that allow for access from multiple devices, search functions, and a clear way in which to filter the topics and questions available on the platform based on primary characteristics such as region or countries involved, topic (e.g., space, AI, insurgency), or time line (e.g., 5 years, 10 years).

QUESTION-WRITING PROCESS

The successful adoption and integration of forecasting platforms will rely heavily on the ability of those platforms to ask questions about the world that policymakers find relevant. The design and scope of questions will play a role in determining how effectively and efficiently forecasting can be used to address these topics. Teams with expertise in question design can help with the process. Additionally, forecasting platforms that enable policymakers to submit questions will allow the forecasting process to remain directly relevant to the decision-makers by targeting the areas in which they feel uncertain or would like additional input about the likelihood of events. This format will also enable the platform and any forecasts made on it to be more relevant to the analysts by giving them a more direct insight into decision-makers' thought processes, the

types of questions they are most interested in, and an ability to track how others are thinking about these issues over time.

With that said, the question-writing process will also need to be monitored to avoid biases, to avoid overwhelming participants with too many questions or too specific questions, and to ensure that forecasters can clearly understand the goal being asked of them. Thus, while encouraging policymakers to have direct input into the questions appearing on the platforms, the forecasting process should likely include an editing process that can more directly and efficiently address these areas of concern.

INCLUDE PROBABILITIES AND FORECASTS IN ANALYTIC TRADECRAFT

As a tool designed to enhance the ability of intelligence analysts and policymakers to better communicate and enable them to be as informed as possible about the likelihood of upcoming events, how analysts utilize forecasting will be critical in ensuring its successful implementation and adoption. With that said, the extent to which analysts interact with forecasting platforms can be managed differently to provide varied levels of investment. Moreover, the ability for participants to discuss questions within the platform can create both intra- and interdepartmental engagement over the prediction questions.

Furthermore, forecasting platforms can be immediately useful to analysts. Alongside potentially promoting greater engagement with others over prediction questions, the ability to track their performance over time can help analysts improve their own abilities to make predictions. Understanding the data and information that others are weighing in their own forecasts, especially superforecasters, may help researchers to develop more complex and accurate approaches to their own analyses. Thus, deep and consistent engagement in forecasting, even when not required, can directly provide a benefit to the analysts using the systems.

Analyst support will be a critical element in determining whether new forecasting platforms succeed. Probabilistic forecasting is a complement to traditional intelligence tools. In an interview, one policymaker explained how to make forecasts more appealing to analysts: "The people who make the forecasts are the wrong person to tell [about how to use the forecasts]. The people who know analysts need to explain it to them. You need good tradecraft." Advocates and entrepreneurs within the government need to clearly communicate this and adopt policies to encourage use with that complementary character in mind.

> CONCLUSION

This report offers insights into how the U.S. government can more effectively adopt probabilistic geopolitical forecasting platforms. It uses the results from working group conversations and interviews to lay out key choices that can make a platform more useful and more likely to succeed. Thinking about policymakers' concerns and potential features listed in this report will be important in helping the platform not only be successfully integrated into the intelligence process but also be maintained long-term as a well-utilized tool.



As the world, including U.S. national security, becomes more data-driven, it is opening the door for the successful adoption of probabilistic geopolitical forecasting initiatives within the U.S. government. Throughout our interview process, each participant was asked to provide their own forecast highlighting their prediction of what they felt was the likelihood of a U.S. government forecasting platform being adopted in the next five to ten years. On average, participants viewed the probability of a geopolitical crowd-sourced forecasting platform being adopted by the U.S. government as 74%. Almost half of participants, 47%, viewed the likelihood as equal to or greater than 70%. These probabilities likely reflect the perceived success of the United Kingdom's COSMIC BAZAAR and the upcoming launch of the INFER platforms in the United States. The increased accuracy and the ability to create an added layer of information to traditional intelligence methods enables countries that can successfully adopt, integrate, and utilize a forecasting system to improve their decision-making. This improved decision-making may ultimately provide an edge in understanding, strategizing, and deciding how to create its policies in a complicated and challenging geopolitical environment.

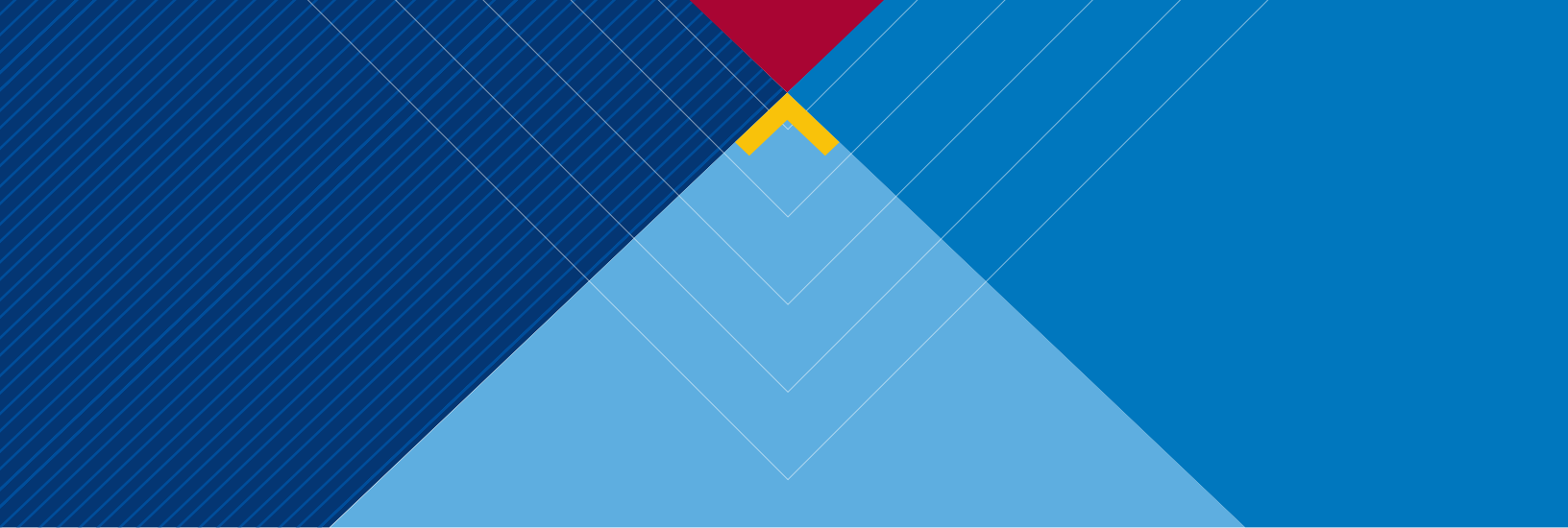
Forecasting platforms have shown the ability to accurately facilitate predictions about different types of geopolitical events, thus enabling better decision-making and more accurate understandings of current geopolitical issues. The work that Perry World House has done on the value of quantitative geopolitical forecasting will hopefully contribute to this better understanding of how to make forecasting effective and long-lasting, especially as the U.S. government works toward creating a new forecasting platform at the NIC. Moving forward, thinking through the ways in which forecasting platforms can be a flexible and integral tool in the pre-existing intelligence process will be critical to ensuring its long-term success.

APPENDIX A: INTERVIEW QUESTIONS

1. First, can we ask a couple of biographical questions?
 - a. Name
 - b. Current affiliation/role
 - c. Former affiliations/roles
 - d. Years of experience in government
2. Before hearing from us or maybe seeing the Perry World House report, had you heard of the concepts we are studying, like crowdsourced forecasting and keeping score?
3. Have you had any experience with any of the probabilistic geopolitical forecasting initiatives in government?
 - a. If so, which ones?
 - b. If so, what did you think of them?
 - c. How many years of experience would you say you have with these initiatives? If you have interacted with probabilistic geopolitical forecasting initiatives what parts of these assessments did/do you find most useful?
 - d. What parts did/do you find least useful?
 - e. Did you end up using the forecast for any output such as a report, analysis, briefing, etc.?
 - f. Why/Why not?
4. Relatedly, how do you think researchers can better communicate the value of these initiatives to policymakers like yourself?
5. When someone hands you a report with predictions, what's the most useful way that they can express the likelihood of an event? For example, is it words of estimative probability (like "somewhat unlikely"), percentages (like "20% probability"), betting odds (like "a 1 in 5 chance"), or maybe some combination of these?
 - a. Why?
6. If you had to assign a percentage to the phrase somewhat likely, what percentage equivalent would you give?
 - a. Would you have given a different percentage if we said moderately likely?

ADDITIONAL QUESTIONS IF THE ANSWER TO QUESTION 2 IS YES

1. One potential goal of a forecasting initiative could be identifying so-called "superforecasters"—basically, high performers with a proven track record in making accurate predictions. Do you think identifying such superforecasters in the U.S. government should be a programmatic goal?
2. Do you think participating in a forecasting platform should be part of analysts' performance evaluation?
 - a. Following on this, do you think analysts' *performance* on these platforms should be part of their job performance evaluation?
3. What incentives or framings do you think would make using forecasts more intrinsically appealing to analysts?
 - a. Do you think that having high support from analysts would increase the resiliency of and trust in forecasting operations?
4. What steps can proponents and developers of forecasting initiatives do to minimize bureaucratic pushback?
5. What should the process for generating and determining forecasting questions look like in order to maximize buy-in from analysts and policymakers?
6. What sort of questions would be most useful? Who should determine which questions to ask? What should be the process for getting input on question formation and design?
7. Are there any questions that should be off-limits?

- 
7. If you saw one of these crowdsourced probabilities in a report, do you think you would need any additional context or information, or would the estimate be enough on its own for you?
 8. Do you think analysts should be *required* to mention these forecasts in their reports, and explain when they disagree with the crowd?
 9. Should a future U.S. government probabilistic geopolitical forecasting initiative include both classified and unclassified platforms?
 10. Should some U.S. allies and partners be allowed to participate in a U.S. government probabilistic geopolitical forecasting platform?
 - a. If so, how could we best facilitate this?
 11. A lot of the efforts to institutionalize forecasting failed to take off and were shut down during the Trump administration, mostly for bureaucratic reasons. What do you think is the best way to make sure new initiatives get past the R&D stage and become more permanent programs?
 12. Thinking about the bureaucratic location of these kinds of initiatives within the U.S. government, *where* do you think they should go?
 13. Finally, we would like to ask you to make your own forecast. What do you think is the likelihood—as a probability or betting odds—that efforts to implement these forecasting methods across the U.S. government will succeed in the next 5–10 years?

APPENDIX B: INTERVIEWS

Derek Chollet, Counselor of the United States Department of State

Charlie Dent, Executive Director and Vice President, Congressional Program, Aspen Institute; Political Commentator, CNN

Kathryn Dura, Doctoral Candidate in Political Science, MIT

Lee Feinstein, Founding Dean and Professor, International Studies, Indiana University Bloomington

Melissa Flagg, Founder, Flagg Consulting LLC; Visiting Fellow, Perry World House

Richard Fontaine, Chief Executive Officer, Center for a New American Security

John Gans, Managing Director of Executive Communications and Strategic Engagement, Rockefeller Foundation; Visiting Fellow, Perry World House

Alice Hunt Friend, Adjunct Professor, American University; Visiting Fellow, Perry World House

Erik Lin-Greenberg, Assistant Professor of Political Science, MIT

Rebecca Lissner, Deputy National Security Advisor to the Vice President, National Security Council

Brian McDowell, Doctoral Candidate, University of Pennsylvania

H.R. McMaster, Fouad and Michelle Ajami Senior Fellow, Hoover Institution, Stanford University

Henri-Paul Normandin, former Ambassador of Canada to Haiti and to the United Nations; Visiting Fellow, Perry World House

Sara Plana, Postdoctoral Fellow, Perry World House

Jonathan Rue, Special Assistant to the Director, Office of the Secretary of Defense Cost Assessment & Program Evaluation (CAPE)

Nadia Schadlow, Senior Fellow, Hudson Institute

Robert Scher, Head of International Affairs, BP America; Visiting Fellow, Perry World House

Randy Schriver, Chairman, Project 2049; Partner, Pacific Solutions LLC; Commissioner, U.S.-China Economic and Security Review Commission

Erin Sikorsky, Director, Center for Climate and Security; Visiting Fellow, Perry World House

Gregory Treverton, Non-Resident Senior Advisor, Center for Strategic and International Studies

Alexander Vershbow, Distinguished Fellow, Atlantic Council; Distinguished Visiting Fellow, Perry World House

Alexander Vindman, Doctoral Candidate, Johns Hopkins University; Military Fellow, Lawfare

Anonymous³¹

³¹ One of our participants requested to remain anonymous.



UNIVERSITY OF PENNSYLVANIA | PERRY WORLD HOUSE
3803 LOCUST WALK, PHILADELPHIA, PA 19104

215.573.5730

@PERRYWORLDHOUSE
FACEBOOK.COM/PERRYWORLDHOUSE

GLOBAL.UPENN.EDU/PERRYWORLDHOUSE