

DECEMBER 2021

---

# Wisdom of the Crowd as Arbiter of Expert Disagreement

Case Study: Future of the DOD-Silicon Valley  
Relationship

CSET Issue Brief



AUTHOR  
Michael Page

## Executive Summary

Can probabilistic forecasting tools such as the wisdom of the crowd help *stakeholders*—the experts and policymakers at the table—reduce uncertainty and disagreement on big questions? Although studies have shown that crowd forecasting is an effective tool for generating accurate forecasts on well-defined questions, it has yet to find a regular home in the policymaker tool kit. Policymakers are often interested in big questions that are not directly forecastable, and quantitative forecasts on their own lack the context in which to understand their importance.

To address these obstacles, the Center for Security and Emerging Technology (CSET) trialed a method for incorporating crowd forecasting into a stakeholder debate on a big question: What is the future of the relationship between the U.S. Department of Defense (DOD) and Silicon Valley? Based on stakeholder interviews, we broke the big question down into forecastable components relating to the state of the relationship and *factors* that affect the relationship. We then elicited forecasts on those components from stakeholders and the crowd, using Foretell, the crowd forecasting platform piloted by CSET. The stakeholders' forecasts provide the context—such as areas of disagreement or uncertainty—in which to understand the wisdom of the crowd. In general, we expect the crowd's forecasts will be most useful where stakeholder disagreement on the big question can be traced to disagreement on specific forecasts. Under these circumstances, the crowd is akin to a reliable third party that can be invoked to arbitrate a disagreement.

After interviewing and surveying 17 stakeholders, we divided them into two cohorts based on their overall expectations for the DOD-Silicon Valley relationship. The *improve* cohort expects the relationship to improve over the next five years, and the *same/worse* cohort expects it to stay the same or worsen. We then looked for differences between the cohorts' forecasts and views on how a factor affects the DOD-Silicon Valley relationship. For forecast disagreements, we compared the cohorts' forecasts with the views of the crowd—comprising more than 2,700 forecasts

from nearly 400 forecasters—to see whether the crowd’s view is more in line with the *improve* cohort or the *same/worse* cohort.

The process succeeded in identifying forecast and factor disagreements that might help explain why the stakeholders have different expectations for the DOD-Silicon Valley relationship. The stakeholders’ forecasts generally correlate with their expectations for the relationship, and the crowd’s views are closer to the *same/worse* cohort more often than the *improve* cohort. The correlations between forecasts and overall expectations are strongest—and the difference between the cohorts’ forecasts statistically significant—for three metrics:

- **Forecast disagreement on percentage of Americans with low confidence in the military.** The *same/worse* cohort forecasts lower confidence in the military. The crowd agrees with the *same/worse* cohort.
- **Forecast disagreement on “big tech” revenue.** The *improve* cohort forecasts higher revenue. The crowd agrees with the *same/worse* cohort.
- **Forecast disagreement on value of “big tech” DOD contracts.** The *improve* cohort forecasts a higher value of “big tech” DOD contracts. The crowd agrees with the *same/worse* cohort.

The stakeholders generally agree on how changes in a factor would affect the DOD-Silicon Valley relationship. The one exception was also where we saw the most significant disagreement between the cohorts:

- **Factor disagreement on *strength of U.S. tech sector*.** Nearly half of the *improve* cohort believes a stronger U.S. tech sector would improve the DOD-Silicon Valley relationship. None of the *same/worse* cohort agrees.

This study was a proof of concept for how probabilistic forecasting tools could be incorporated into policy debates on a big question. We hope this effort—or efforts like it—can achieve the following:

1. A more-focused debate that zeroes in on the underlying points of disagreement.
2. Adjustments in stakeholders' expectations that reduce overall levels of uncertainty and disagreement.
3. A record of what the stakeholders and crowd expected to happen that makes it more likely surprising developments will be noticed and accorded their due weight.

Visit [www.cset-foretell.com/issue-campaigns-dod-sv](http://www.cset-foretell.com/issue-campaigns-dod-sv) to view the forecast questions and see the most up-to-date data.

## Table of Contents

Executive Summary .....	2
Introduction .....	6
Methodology .....	8
Step 1: Stakeholder Interviews .....	8
Step 2: Stakeholder Survey .....	10
Step 3: Crowd Forecasts .....	12
Step 4: Forecast Comparison.....	12
Findings.....	13
Issue Breakdown .....	14
Factor Forecasts.....	17
Factor 1: U.S. Distrust of Military/Government .....	19
Factor 2: U.S. Political Polarization .....	24
Factor 3: Silicon Valley Protests .....	25
Factor 4: Strength of U.S. Tech Sector .....	26
Factor 5: China Military Aggression .....	30
Factor 6: China tech capabilities.....	35
Outcome Measure Forecasts .....	38
Outcome Measure 1: Transactions .....	40
Outcome Measure 2: Personnel.....	43
What should stakeholders do with this information? .....	45
Conclusion .....	47
Author .....	49
Acknowledgments .....	49
Appendix.....	50
Participating Stakeholders.....	50
Forecast Summary Tables .....	52
Endnotes .....	56

## Introduction

Policymaking involves predictions on big issues, such as the future of the United States-China relationship or the effect a foreign policy will have on global stability.<sup>1</sup> Policy debates on these matters are often mired in conflicting—and to others, black box—intuitions of experts and policymakers. Making these intuitions more interpretable and measurable could help dissolve otherwise intractable disagreement and improve policymaking.

Consider the role probabilistic forecasting could play in policy debates. For any well-defined question that can be answered with a numeric probability, we can run trials to identify the most accurate forecasting method. The Intelligence Advanced Research Projects Activity (IARPA) did exactly this. From 2010 to 2014, it ran a tournament to identify the most effective methods for estimating the likelihood of geopolitical events. The Good Judgment Project, a team led by University of Pennsylvania professors Philip Tetlock and Barbara Mellers, won the tournament every year using a version of the wisdom of the crowd: the collective judgment of a large, diverse group.<sup>2</sup>

Incorporating probabilistic forecasting methods such as crowd forecasting into policy debates has proven challenging. For one, the questions of interest to policymakers are too big to directly forecast. As noted by the founders of the Good Judgment Project, Philip Tetlock and Barbara Mellers, along with J. Peter Scoblic, “the specificity required to make questions rigorously resolvable precludes asking ‘big’ questions.”<sup>3</sup> Big questions can be broken down into components that are forecastable, but if the breakdown does not track how policymakers think about the big question, the resulting forecasts are not useful. A second challenge is that probabilistic forecasts by themselves—such as a 7 percent chance of conflict in the South China Sea in the next six months—lack the context in which to understand their importance. Should a policymaker conclude the chance of conflict is higher or lower than they previously believed? What are the broader implications of either conclusion?

To address these challenges, we trialed a method in which the stakeholders effectively chose *what* to forecast, and the crowd provided the actual forecasts.<sup>4</sup> To provide context for the crowd's forecasts, we also elicited forecasts from the stakeholders and used those forecasts to identify sources of uncertainty and disagreement where the crowd's forecasts can add particular value.

We began with a big question on which stakeholders have conflicting intuitions: What is the future of the DOD-Silicon Valley relationship? DOD has prioritized strengthening its relationship with Silicon Valley to ensure it has access to cutting-edge technology and talent. In 2015, it launched the Defense Innovation Unit to accelerate military adoption of commercial technology. The following year saw the creation of the Defense Innovation Board, chaired by former Google CEO Eric Schmidt, with a mandate to bring the best of Silicon Valley to the U.S. military. An employee protest of Google's involvement with DOD in 2018 led to headlines of a culture clash between DOD and Silicon Valley. Today, stakeholders' views vary widely on the extent of the rift, its causes, and whether the relationship is trending in a positive or negative direction.<sup>5</sup>

The remainder of this brief discusses our effort to pierce the veil of this debate, including our methodology, findings, and practical takeaways.

Crowd forecasting is a member of a growing family of probabilistic forecasting methods including quantitative modeling, human judgment, and combinations of the two. While our application of the methodology described here employs crowd forecasting using Foretell, it could accommodate any probabilistic forecasting tool. The broader goal motivating our work is to better integrate state-of-the-art analytical tools into high-stakes policymaking.

## Methodology

This section summarizes the four steps of our study:

1. **Stakeholder interviews.** Selected and interviewed stakeholders and, based on those interviews, broke the big question down into forecastable components.
2. **Stakeholder survey.** Surveyed the stakeholders to understand their overall expectations for the DOD-Silicon Valley relationship; how they believe specific factors affect the relationship; and their forecasts on 19 identified metrics.
3. **Crowd forecasts.** Published metrics on Foretell to elicit ongoing forecasts from a community of more than 2,000 registered forecasters.
4. **Forecast comparison.** Compared the stakeholders' and crowd's forecasts, focusing specifically on points of disagreement.

### *Step 1: Stakeholder Interviews*

**Stakeholder selection.** We interviewed 17 *stakeholders*: individuals currently or formerly associated with DOD, industry, or academia who have a professional interest in the future of the DOD-Silicon Valley relationship. Our goal in selecting stakeholders was to have a group of people with informed and divergent views on the topic to mimic the discussions that might occur in a policy debate. To identify potential stakeholders, we created an initial list comprising individuals we believed to have the above properties, and then used Elicit,<sup>6</sup> an AI research assistant, to find other individuals sharing similar properties based on their online presence. See the Appendix for a list of participating stakeholders.

**Stakeholder interviews.** The objective of the interviews was to identify measures of the state of the DOD-Silicon Valley relationship (*outcome measures*)—such as contracts between DOD and Silicon Valley—and geopolitical, cultural, and economic *factors*

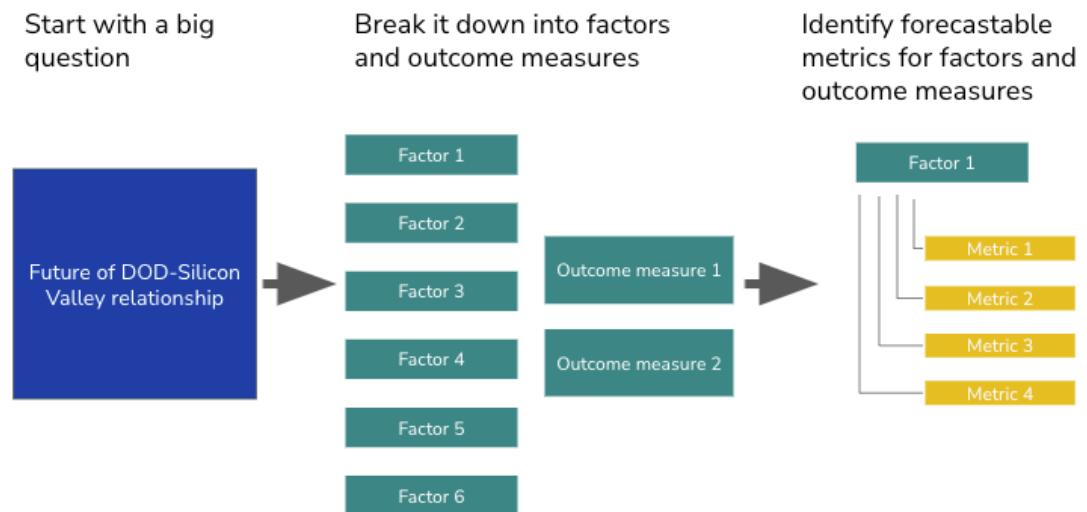


stakeholders believe affect the relationship—such as *China military aggression* or *strength of the U.S. tech sector*. To identify candidate outcome measures, we asked stakeholders to describe the relationship today, what would constitute a better or worse relationship, and where they expected the relationship to be in five years. To identify candidate factors, we asked the stakeholders to imagine scenarios in which the relationship is better or worse than their expectations in five years and provide the most likely reasons.

### Selection of factors, outcome measures, and related metrics.

Factors and outcome measures are too general to forecast directly. To forecast them, we selected 19 specific events or trends (*metrics*) that partially capture the factors and outcome measures. For example, for the *China military aggression* factor, we selected four metrics, including the chance China shoots at another country's vessel in the South China Sea in the next six months, and the number of times Japan responds to Chinese threats to its airspace over the next three years. Our selection of metrics was informed by the stakeholder interviews. Figure 1 illustrates the steps in breaking down a big question into forecastable metrics.

Figure 1. Process of breaking down big question



Source: CSET.

## ***Step 2: Stakeholder Survey***

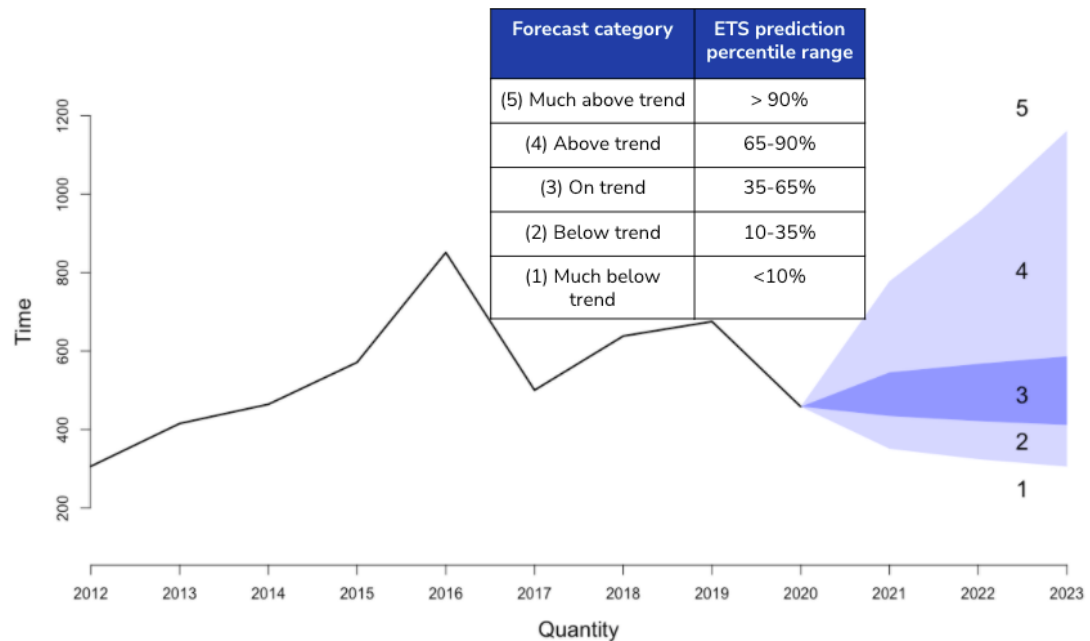
Following the interviews, we sent the stakeholders a three-part survey:<sup>7</sup>

**Part I: Overall expectations.** Part I of the survey asked the stakeholders how they expect the DOD-Silicon Valley relationship will change over the next five years.<sup>8</sup> We divided stakeholders into two roughly equal-sized cohorts based on their responses to this question.

**Part II: Factor significance.** Part II of the survey asked the stakeholders to assess the effect of increases and decreases in the factors on the DOD-Silicon Valley relationship. Based on these responses, we identified the factors for which the stakeholder consensus is that an increase would improve the DOD-Silicon Valley relationship, and the factors for which the opposite is true. Where the stakeholders disagreed about the effect of a factor on the relationship, we looked at whether views differ between the two cohorts (*factor disagreements*).

**Part III: Stakeholder forecasts.** Part III of the survey asked the stakeholders to forecast the 19 metrics informing the factors and outcome measures.<sup>9</sup> We asked the stakeholders to forecast two types of metrics: time series and binary. *Time series* metrics are supported by historical data and track quantity over time. An example is the number of times Japan responds to Chinese threats to its airspace over the next three years, as shown in Figure 2.

Figure 2. Forecast categories for time series metrics



Source: CSET.

For time series metrics, we used a simpler forecast elicitation process for the stakeholders than for the crowd. The crowd provided precise forecasts for each time period. For the stakeholders, we divided the next three years into five forecast categories and asked stakeholders to select one.<sup>10</sup> *Forecast categories* are bands of possible outcomes that correspond to prediction intervals generated by an exponential triple smoothing (ETS) projection from the historical data. For example, the *on trend* forecast category includes the 35th to 65th ETS prediction percentile interval, and the *above trend* forecast category includes the 65th to 90th ETS prediction percentile interval. The average stakeholder forecast shown in Tables 1 and 2 in the Appendix is an average of the midpoint of each selected forecast category. For example, the average of an *on trend* (50th percentile midpoint) forecast and *above trend* (78th percentile midpoint) forecast is the 64th ETS prediction percentile.

*Binary* metrics track the probability of an event occurring during a set time period. An example is the chance that China shoots at another country's vessel in the South China Sea in the next six months. Unlike for time series metrics, forecasts on binary metrics cannot be described in terms of forecast categories. Because of this, we asked stakeholders to provide a precise (0 to 100 percent) probability for the event.

### ***Step 3: Crowd Forecasts***

To elicit the wisdom of the crowd, we used Foretell, CSET's crowd forecasting platform. Foretell has nearly 2,000 registered forecasters, consisting in large part of graduate students and professionals with an interest in technology and security policy. Within this group are 50 pro forecasters, those whose forecasts during the previous year were most accurate. The 19 metrics were initially published on Foretell at some point between August 17 and September 8, 2021, and remain live at the time of this brief's publication.<sup>11</sup> Unlike the stakeholders' forecasts, which were captured at a single point in time, the crowd forecasts are ongoing and can be updated in response to new information. Unlike the stakeholders, for time series metrics the crowd provides precise forecasts for each year or half-year of the forecast period. The consensus crowd forecast provided below is the average of all forecasts. The data below is current as of November 4, 2021, at which point the questions had received more than 2,700 forecasts from 382 forecasters.

### ***Step 4: Forecast Comparison***

Lastly, we compared the stakeholders' forecasts with the crowd's forecasts. Our goal was to highlight where and how the crowd's forecast might be most useful for the stakeholders. Toward that end, we looked for correlations between the stakeholders' overall expectations for the relationship and their forecasts. For metrics with notable correlations, we broke the stakeholders' forecasts down by cohort, allowing us to compare the average forecasts of the two cohorts. Differences between the cohorts' forecasts

provide potential explanations for why the cohorts have different overall expectations for the DOD-Silicon Valley relationship. For example, if the stakeholders agree that increases in *China military aggression* would improve the relationship, but disagree about whether the relationship will improve, a possible explanation is they disagree about whether, or by how much, *China military aggression* will increase. We can evaluate the merits of that explanation by looking for differences between the cohorts' forecasts on the *China military aggression* metrics.

We also looked at whether a correlation between the stakeholders' forecasts and their overall expectations for the relationship is in the expected direction. We based this assessment on the stakeholders' consensus view on how a metric's associated factor or outcome measure relates to the DOD-Silicon Valley relationship. Returning to the example above, the stakeholder consensus is that increases in *China military aggression* would improve the relationship. Accordingly, we would expect that if the cohorts' forecasts are different for the *China military aggression* factors, the *improve* cohort's forecasts would be higher.

## Findings

We divided the stakeholders into two cohorts based on their overall expectations for the DOD-Silicon Valley relationship. The *improve* cohort comprises the 59 percent (10) of stakeholders who expect the relationship to improve over the next five years. The *same/worse* cohort comprises the 41 percent (7) of stakeholders who expect the relationship to remain the same (5) or worsen (2). What accounts for the cohorts' different expectations, and what does the wisdom of the crowd have to say about them?

Below is a three-part analysis of these findings. The first part summarizes the stakeholders' views on the factors, including where the cohorts disagree about the effect of a factor on the overall relationship (*factor disagreements*). The second and third parts compare the stakeholders' forecasts and the crowd's forecasts on the factor and outcome metrics, respectively. They

also highlight where the cohorts' forecasts diverge (*forecast disagreements*) and with which cohort the crowd agrees, if any.

### ***Issue Breakdown***

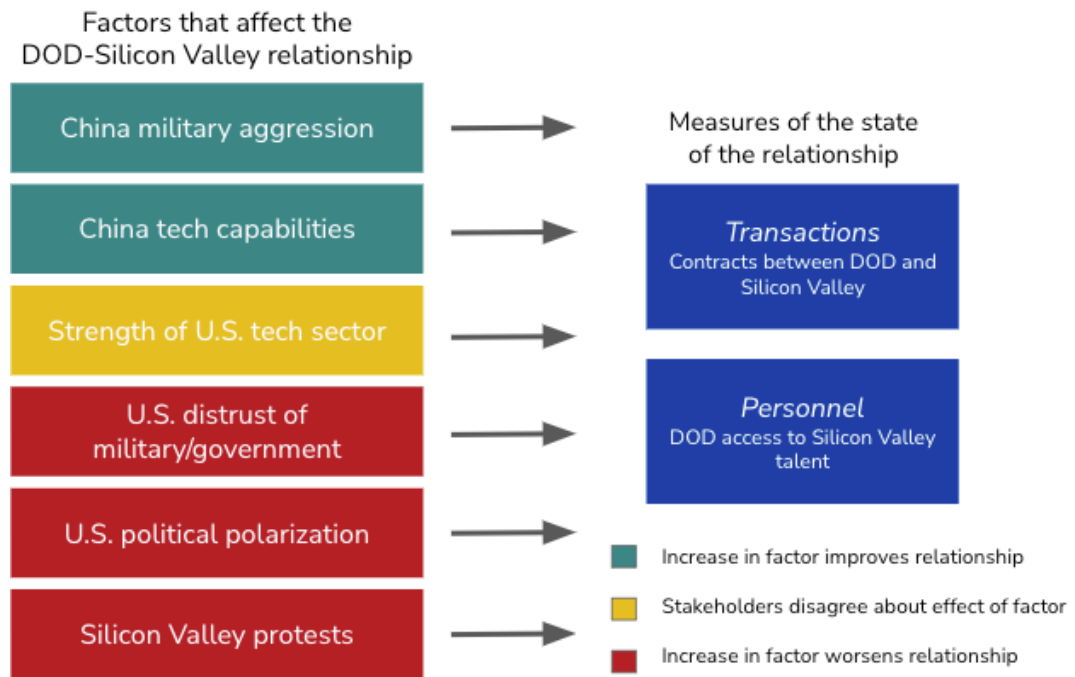
As shown in Figure 3, we identified two measures of the state of the DOD-Silicon Valley relationship (*outcome measures*) based on the stakeholder interviews:

1. *Transactions*: Contracts between DOD and Silicon Valley; and
2. *Personnel*: DOD access to tech sector talent.

We then identified six geopolitical, cultural, and economic *factors* stakeholders think affect the relationship:

1. China military aggression,
2. China tech capabilities,
3. U.S. tech sector strength,
4. U.S. political polarization,
5. U.S. distrust of military / government, and
6. Silicon Valley protests.

Figure 3. Stakeholder breakdown of future of DOD-Silicon Valley relationship



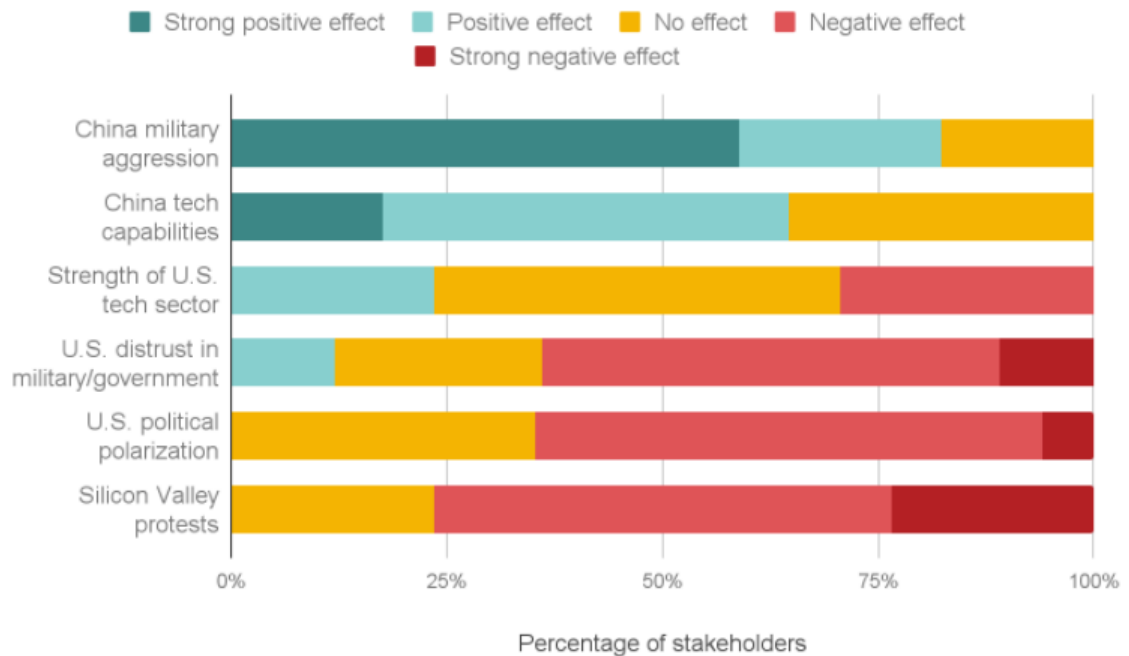
Source: CSET stakeholder survey.

The stakeholders' views on how and whether each factor affects the DOD-Silicon Valley relationship vary. To understand this variance, we surveyed the stakeholders on the effect changes in each factor would have on the relationship.<sup>12</sup>

As shown in Figure 4, for five of the six factors, the stakeholders generally agree on the effect, if any, an increase in that factor would have on the DOD-Silicon Valley relationship.<sup>13</sup> For example, the stakeholders generally agree that an increase in *China military aggression* would have a positive impact on the relationship (82 percent expect it would improve the relationship), but that an increase in *Silicon Valley protests* would have a negative impact on the relationship (76 percent expect it would worsen the relationship). The importance stakeholders place on these factors varied by cohort. The *improve* cohort believes an increase in *China military aggression* would have a strong positive effect on the

relationship (70 percent selected strong positive, compared to 43 percent of the *same/worse* cohort). Nearly all (90 percent) of the *improve* cohort predicts that an increase in *Silicon Valley protests* would negatively impact the relationship, compared to 57 percent of the *same/worse* cohort. These differences are not statistically significant, however, and should therefore be treated with caution.

Figure 4. Stakeholders' predicted effect of an increase in each factor on the DOD-Silicon Valley relationship



Source: CSET stakeholder survey.

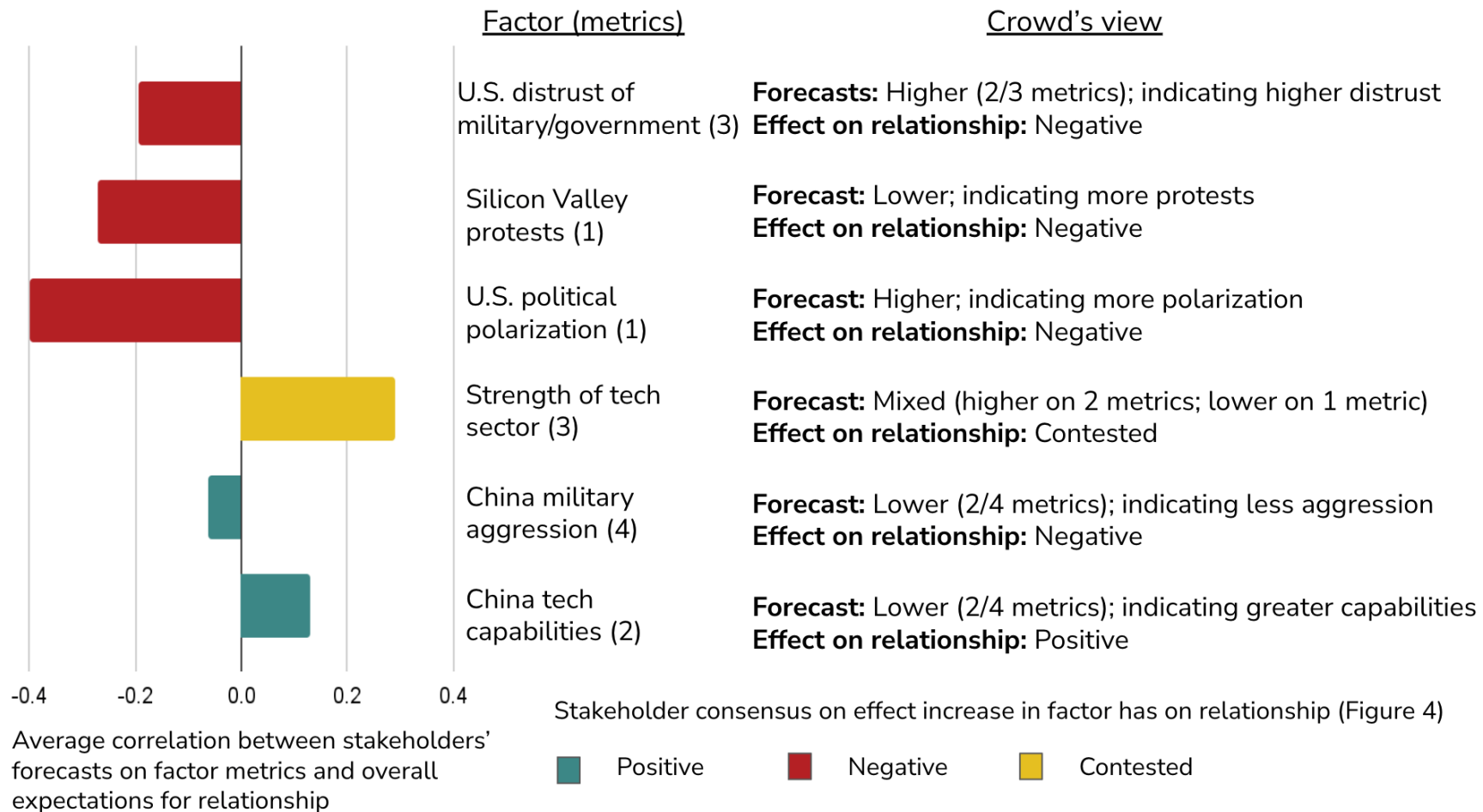
The *strength of the U.S. tech sector* factor is the only one with nearly equal proportions of stakeholders believing an increase would improve the DOD-Silicon Valley relationship (24 percent) and an increase would worsen the relationship (29 percent). The significance of the cohort divide (p-value 0.07) is also highest for this factor. All of the stakeholders who believe a stronger U.S. tech sector would improve the relationship are in the *improve* cohort.



### ***Factor Forecasts***

This subsection compares the stakeholders' forecasts and crowd's forecasts for the metrics informing the six factors relevant to the DOD-Silicon Valley relationship. A pattern emerged for five of the six factors, as shown in Figure 5. We saw generally consistent correlations between stakeholders' forecasts and their overall expectations for the DOD-Silicon Valley relationship in the expected direction. In other words, if the stakeholder consensus is that an increase in a factor would improve the relationship, the stakeholders who expect the relationship to improve tended to have higher forecasts on the metrics for that factor. The exception is the *China military aggression* factor, on which the stakeholders' forecasts either did not diverge, or did not do so in a consistent direction.

Figure 5. Summary of stakeholders' and crowd's views on factor



Source: CSET stakeholder survey and Foretell.

Note: The effect of the crowd's forecasts on the DOD-Silicon Valley relationship is relative to the stakeholders' forecasts and based on the stakeholder consensus on how changes in a factor affect the relationship (Figure 4). For example, because the stakeholder consensus is that an increase in *U.S. distrust of military/government* is negative for the relationship and the crowd's forecasts are generally higher than those of the stakeholders for this factor, the effect of the crowd's forecasts on the relationship is negative. Foretell data pulled November 4, 2021.

For four metrics, the crowd's forecasts generally differ from the stakeholders' forecasts in a way that suggests the crowd is more pessimistic about the future of the DOD-Silicon Valley relationship (see Figure 4). For example, the stakeholder consensus is that an increase in the *U.S. distrust of military/government* factor would negatively affect the DOD-Silicon Valley relationship. The crowd's forecasts on the metrics for that factor are generally higher than the stakeholders' forecasts, indicating an expectation of higher distrust. For one factor—*China tech capabilities*—the crowd's forecasts indicate an expectation of greater capabilities, relative to the stakeholders. According to the stakeholder consensus, this would positively affect the DOD-Silicon Valley relationship. Because the stakeholders disagree about how changes to the *strength of the U.S. tech sector* factor affect the relationship, the relationship between the crowds' forecasts and the DOD-Silicon Valley relationship is unclear.

The metrics with the highest correlations between the stakeholders' forecasts and their overall expectations for the relationship are: 1) the percentage of Americans with low confidence in the military; and 2) annual “big tech” revenue. The differences between the cohorts' forecasts for these metrics are also statistically significant. For both metrics, the crowd's forecasts are closer to those of the *same/worse* cohort.

Below we provide a factor-by-factor comparison of the stakeholders' and crowd's forecasts, highlighting where the stakeholder cohorts' forecasts diverge in a notable way. Table 1 in the Appendix provides a more detailed comparison of the forecasts. Visit [cset-foretell.com](https://cset-foretell.com) for more information about the metrics, including up-to-date forecasts.

### **Factor 1: U.S. Distrust of Military/Government**

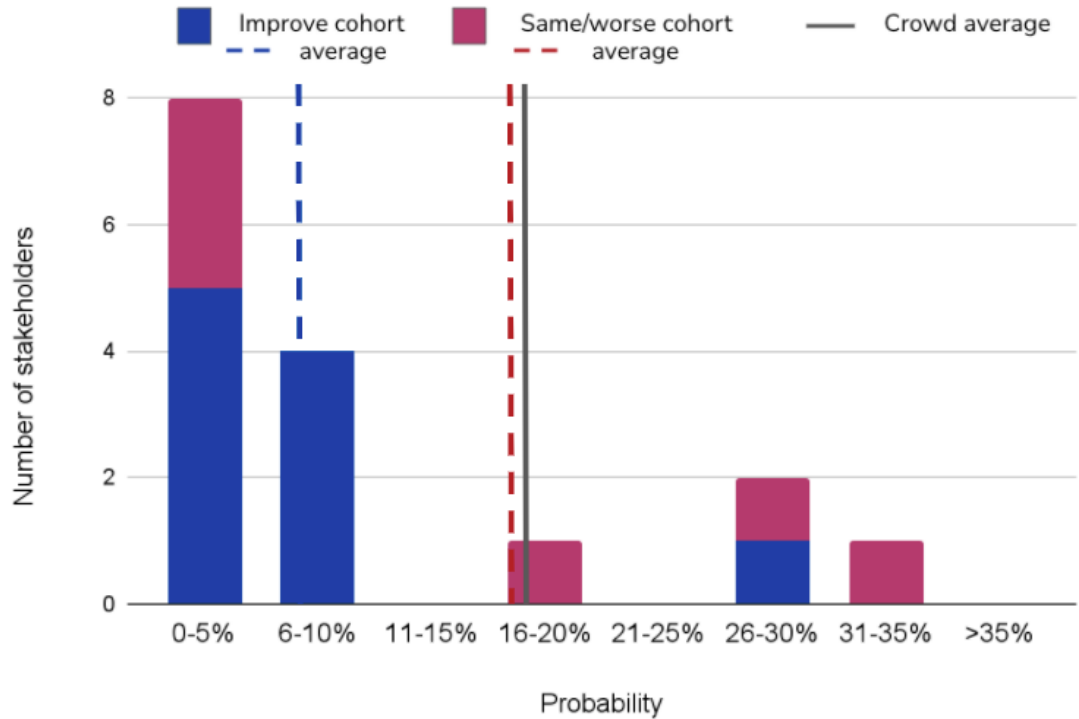
Most stakeholders (65 percent) believe an increase in *U.S. distrust of military/government* would worsen the DOD-Silicon Valley relationship.<sup>14</sup> See Figure 4. The stakeholders' forecasts most strongly correlated with their overall expectations, and in the expected direction, for the first two metrics: 1) the percentage of Americans with low confidence in the military; and 2) the likelihood

the United States uses a lethal autonomous drone against a human target. The crowd's forecasts are closer to the higher forecast of the *same/worse* cohort for both metrics, indicating an expectation of higher distrust, relative to the stakeholders. Based on the stakeholder consensus on the effect of this factor on the relationship, the crowd's forecasts have negative implications for the relationship.

Metric 1: Likelihood the United States will acknowledge using an autonomously operated drone to identify and deploy lethal force against a human target in the next year.

- **Stakeholders' views.** Average forecast of 10 percent with a negative correlation (-0.3) between forecasts and overall expectations for the DOD-Silicon Valley relationship. The *same/worse* cohort's average forecast (16 percent) is higher than the *improve* cohort's average forecast (7 percent). In the interviews, the stakeholders commonly cited the United States using a lethal autonomous weapon as a reason distrust of the military might increase. If true, the *same/worse* cohort's higher forecast indicates a greater expectation of distrust.
- **Crowd's view.** Based on 184 forecasts, the crowd's forecast (17 percent) is higher than the average stakeholder forecast and closer to the views of the *same/worse* cohort. The crowd's higher forecast indicates an expectation of higher *U.S. distrust of military/government*, relative to the stakeholders. According to the stakeholder consensus, this has negative implications for the DOD-Silicon Valley relationship. In their rationales, several forecasters focused on the likelihood that Russia or China are the first to deploy this technology, prompting the United States to respond in kind.

Figure 6. Chance U.S. military acknowledges using an autonomously operated drone to identify and deploy lethal force against a human target in next year



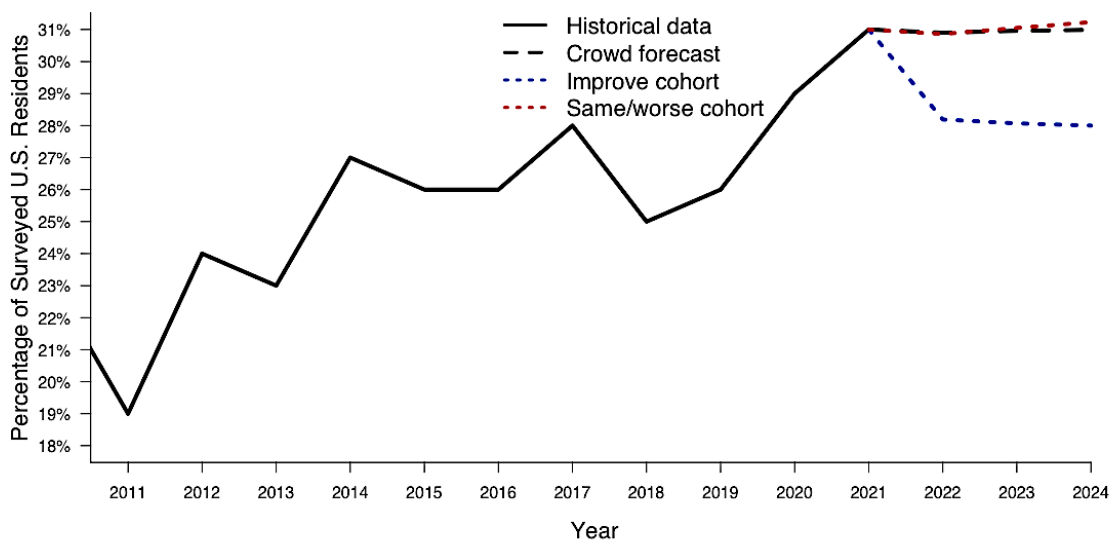
Sources: CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/527](https://www.cset-foretell.com/questions/527).

Metric 2: Percentage of U.S. residents with low confidence in the U.S. military (Gallup survey).<sup>15</sup>

- Stakeholders' views.** Average forecast of 29 percent by 2024, with a negative correlation (-0.5) between forecasts and overall expectations for the relationship. The *same/worse* cohort's average forecast (31 percent in 2024) is higher than the *improve* cohort's average forecast (28 percent in 2024), indicating a greater expectation of distrust. The difference between the cohorts' forecasts is statistically significant.

- Crowd's view.** Based on 193 forecasts, the crowd expects the percentage of U.S. residents with “some” or “very little” trust in the military to stay at about 31 percent. The crowd's forecast is closer to the *same/worse* cohort. The crowd's forecast indicates an expectation of higher *U.S. distrust of military/government*, relative to the stakeholders. According to the stakeholder consensus, this has negative implications for the DOD-Silicon Valley relationship. Forecasters commonly cited the U.S. withdrawal from Afghanistan and its possible repercussions as a reason confidence in the military might decline.

Figure 7. Percentage of U.S. residents with “some” or “very little” confidence in military

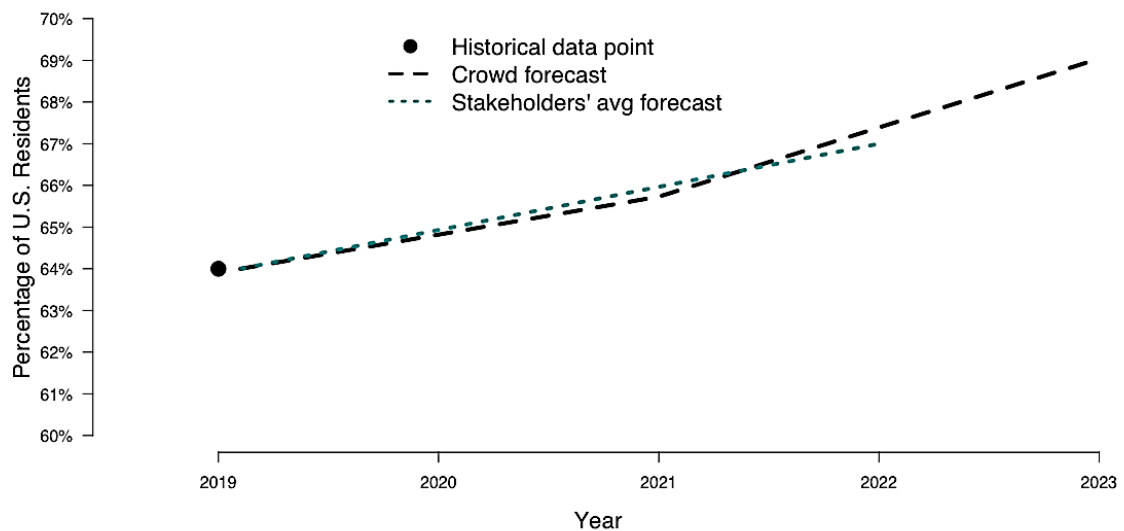


Sources: Gallup, CSET stakeholder survey, and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/864](http://www.cset-foretell.com/questions/864).

Metric 3: Percentage of U.S. residents concerned about government data use (Pew survey).<sup>16</sup>

- **Stakeholders' views.** Average forecast of 67 percent in 2022, with no correlation between forecasts and overall expectations for the relationship.
- **Crowd's view.** Based on 195 forecasts, the crowd's forecast for 2022 is the same as the stakeholders' average forecast for 2022. The crowd expects that the percentage of U.S. residents who are “very” or “somewhat” concerned about how the government is using their data will increase slightly, from 64 percent in 2019—the only time the survey was conducted—to 67 percent in 2022 and 69 percent in 2023.<sup>17</sup>

Figure 8. Percentage of U.S. residents concerned about government use of their data



Sources: Pew Research Center Americans and Privacy Survey, CSET stakeholder survey, and Foretell. Crowd forecast data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/422](http://www.cset-foretell.com/questions/422).

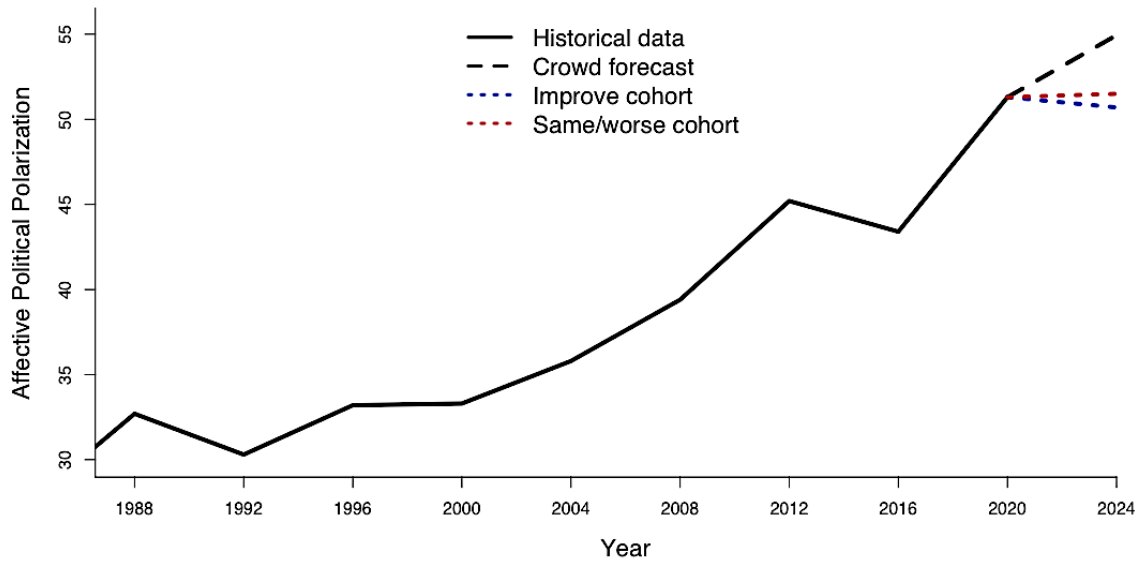
## Factor 2: U.S. Political Polarization

Most stakeholders (65 percent) believe an increase in *U.S. political polarization* would worsen the DOD-Silicon Valley relationship; none believe an increase would improve it. See Figure 4. One metric informs this factor, tracking the *affective political polarization* score from the American National Election Studies (ANES) survey, conducted every four years.<sup>18</sup>

- **Stakeholders' views.** Average forecast of 51 affective polarization score, with a negative correlation (-0.4) between forecast and overall expectations for the relationship. The *same/worse* cohort's average forecast (52 in 2024) is higher than the *improve* cohort's average forecast (51 in 2024), indicating a greater expectation of political polarization.
- **Crowd's view.** Based on 335 forecasts, the crowd's forecast is higher than both cohorts. The crowd expects affective political polarization in the United States will increase to an all-time high of 55 in 2024. The crowd's forecast indicates an expectation of higher political polarization, relative to the stakeholders. According to the stakeholder consensus, this has negative implications for the DOD-Silicon Valley relationship.



Figure 9. U.S. political polarization



Sources: American National Election Studies (ANES) Survey, CSET stakeholder survey, and Foretell. Crowd forecast data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/410](https://www.cset-foretell.com/questions/410).

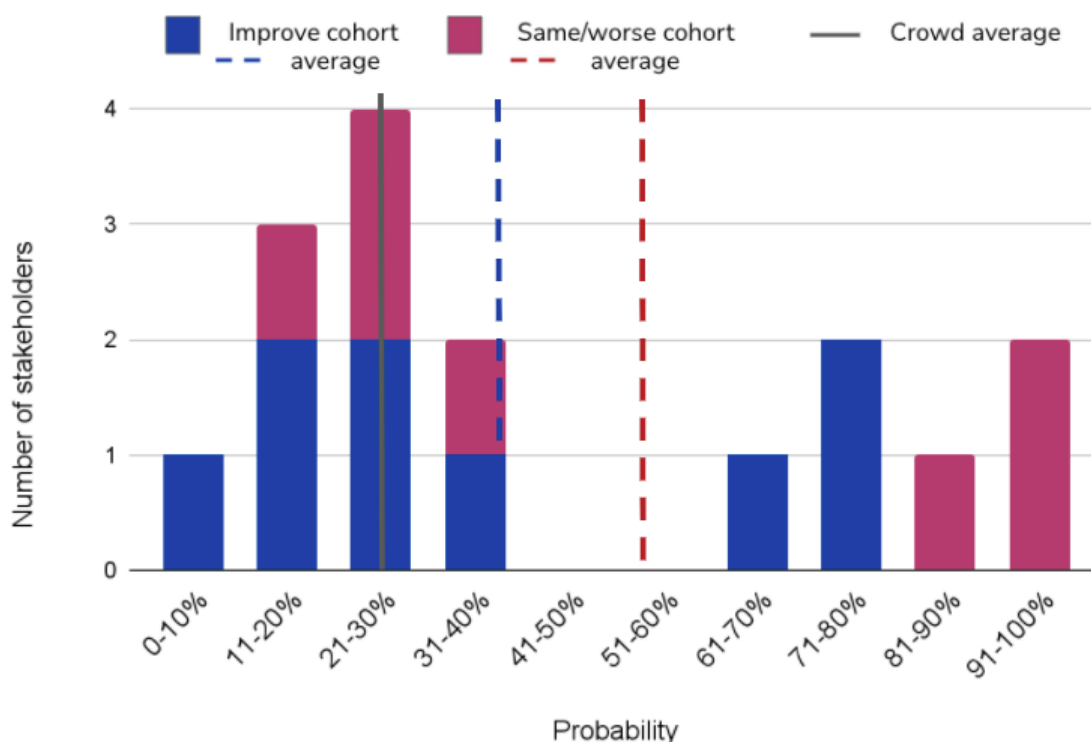
### Factor 3: Silicon Valley Protests

Most stakeholders (76 percent) believe an increase in *Silicon Valley protests* would negatively affect the DOD-Silicon Valley relationship; none believe an increase would improve it. One metric informs this factor: tracking the likelihood of an organized employee protest at a “big tech” company against the company’s involvement with DOD in the next year.

- **Stakeholders’ views.** Average forecast of 46 percent, with a negative correlation (-0.3) between forecasts and overall expectations for the relationship. The *same/worse* cohort’s average forecast (56 percent) is higher than the *improve* cohort’s average forecast (38 percent), indicating a greater expectation of protests.

- Crowd's view.** Based on 134 forecasts, the crowd's forecast (25 percent) is lower than both stakeholder cohorts. The crowd's forecast indicates an expectation of fewer protests, relative to the stakeholders. According to the stakeholder consensus, this has positive implications for the DOD-Silicon Valley relationship.

Figure 10. Chance of employee protest at a “big tech” company against company’s involvement with DOD in next year



Sources: CSET stakeholder survey and Foretell. Crowd forecast data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/442](https://www.cset-foretell.com/questions/442).

#### Factor 4: Strength of U.S. Tech Sector

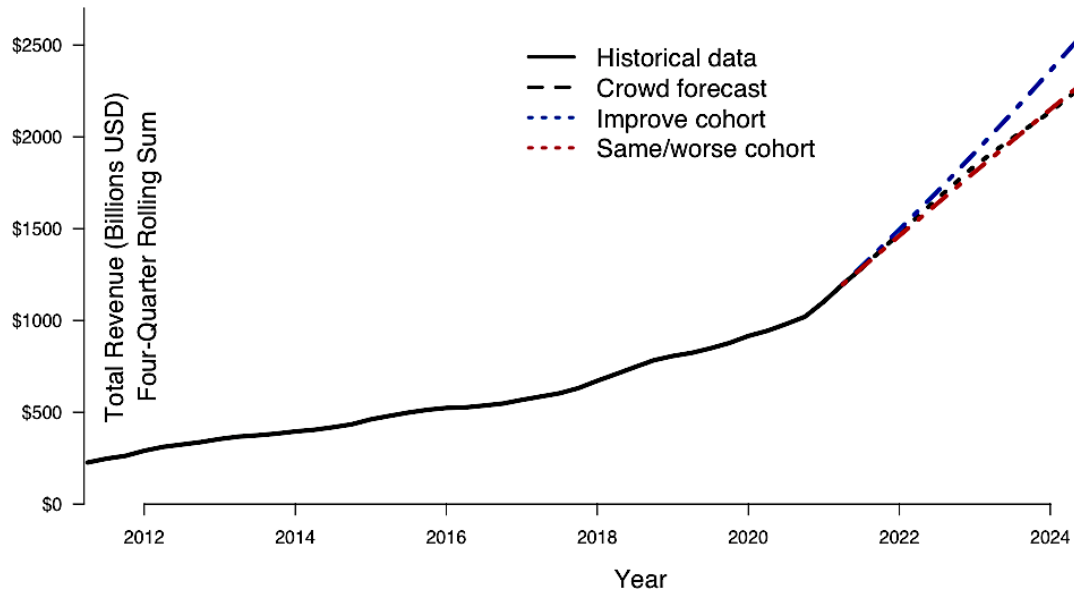
Stakeholders are split on the effect an increase in *strength of the U.S. tech sector* would have on the DOD-Silicon Valley relationship, with nearly equal numbers believing it would improve the relationship (24 percent) and worsen the relationship (29 percent).

All of the stakeholders who believe it would improve the relationship are in the *improve* cohort. Because the stakeholders disagree about the effect of this factor on the relationship, we do not assess the implications of the forecasts for the overall relationship. The cohorts' forecasts nonetheless diverged on all three metrics, with the *improve* cohort consistently forecasting greater tech sector strength. The crowd's forecast is higher than both cohorts for two of the three metrics and closer to the forecast of the *same/worse* cohort for the third.

Metric 1: Annual revenue of Alphabet, Amazon, Apple, Facebook, and Microsoft.

- **Stakeholders' views.** Average forecast of \$2.5 trillion by the second quarter of 2024, with a positive correlation (0.5) between forecasts and overall expectations for the relationship. The *improve* cohort's average forecast (\$2.6 trillion in 2024 Q2) is higher than the *same/worse* cohort's average forecast (\$2.3 trillion in 2024 Q2). The difference between the cohorts' forecasts is statistically significant.
- **Crowd's view.** Based on 116 forecasts, the crowd's forecast is closer to the *same/worse* cohort, projecting combined annual revenue to increase to \$2.3 trillion by the second quarter of 2024.

Figure 11. Annual revenue of Alphabet, Amazon, Apple, Facebook, Microsoft



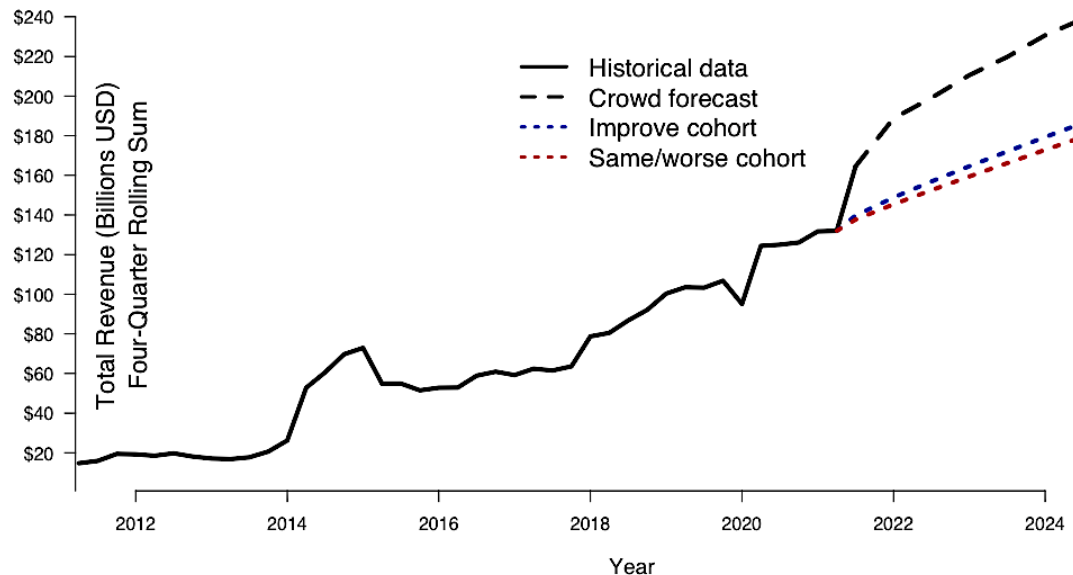
Sources: 10-Q Filings, CSET stakeholder survey, and Foretell. Crowd forecast data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/396](https://www.cset-foretell.com/questions/396).

Metric 2: Annual money raised by private U.S. tech companies.

- **Stakeholders' views.** Average forecast of \$183 billion by the second quarter of 2024, with a weak positive correlation (0.2) between forecasts and overall expectations for the relationship. The *improve* cohort's average forecast (\$186 billion in 2024 Q2) is higher than the *same/worse* cohort's average forecast (\$179 billion in 2024 Q2).
- **Crowd's view.** Based on 67 forecasts, the crowd's forecast is higher than both stakeholder cohorts. The crowd expects annual money raised to increase to \$239 billion by the second quarter of 2024. The crowd's higher forecast was likely influenced by data for the second quarter of 2021—not available when the stakeholders made their forecasts—

which showed a 25 percent increase in annual funding over the previous quarter.

Figure 12. Annual money raised by private U.S. tech companies



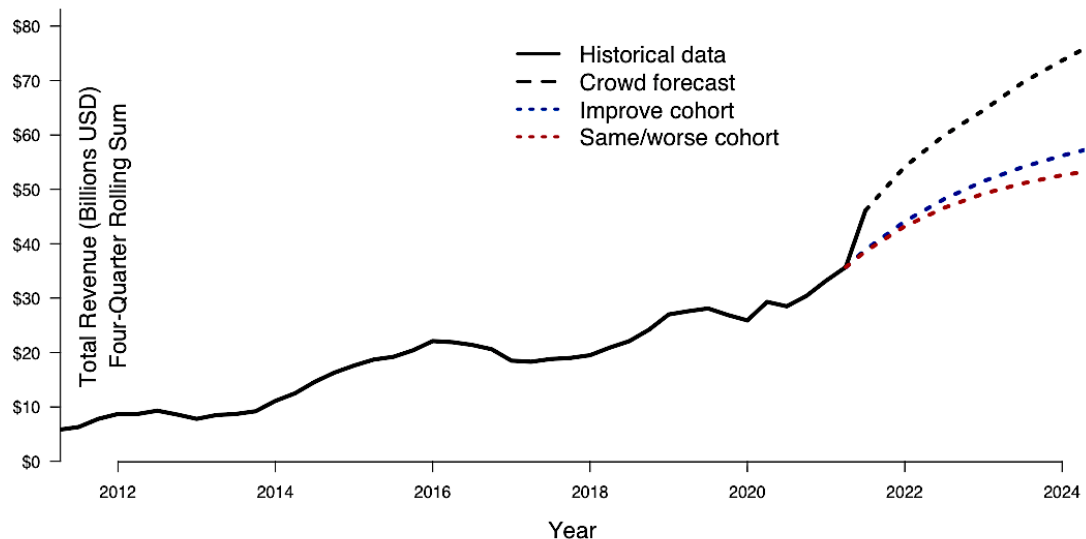
Sources: Crunchbase, CSET stakeholder survey, and Foretell. Crowd forecast data pulled November 4, 2021. For current forecasts, see <http://www.cset-foretell.com/questions/403>.

Metric 3: Annual money raised by U.S. tech startups.

- **Stakeholders' views.** Average forecast of \$56 billion by the second quarter of 2024, with a weak positive correlation (0.2) between forecasts and overall expectations for the relationship. The *improve* cohort's average forecast (\$58 billion in 2024 Q2) is higher than the *same/worse* cohort's average forecast (\$53.7 billion in 2024 Q2).
- **Crowd's view.** Based on 87 forecasts, the crowd's forecast is higher than both stakeholder cohorts. The crowd expects annual money raised to increase to \$77.5 billion by the second quarter of 2024. The crowd's higher forecast was likely influenced by data for the second quarter of 2021—

not available when the stakeholders made their forecasts—which showed a 29 percent increase in annual funding over the previous quarter.

Figure 13. Annual money raised by U.S. tech startups



Sources: Crunchbase, CSET stakeholder survey, and Foretell. Crowd forecast data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/350](http://www.cset-foretell.com/questions/350).

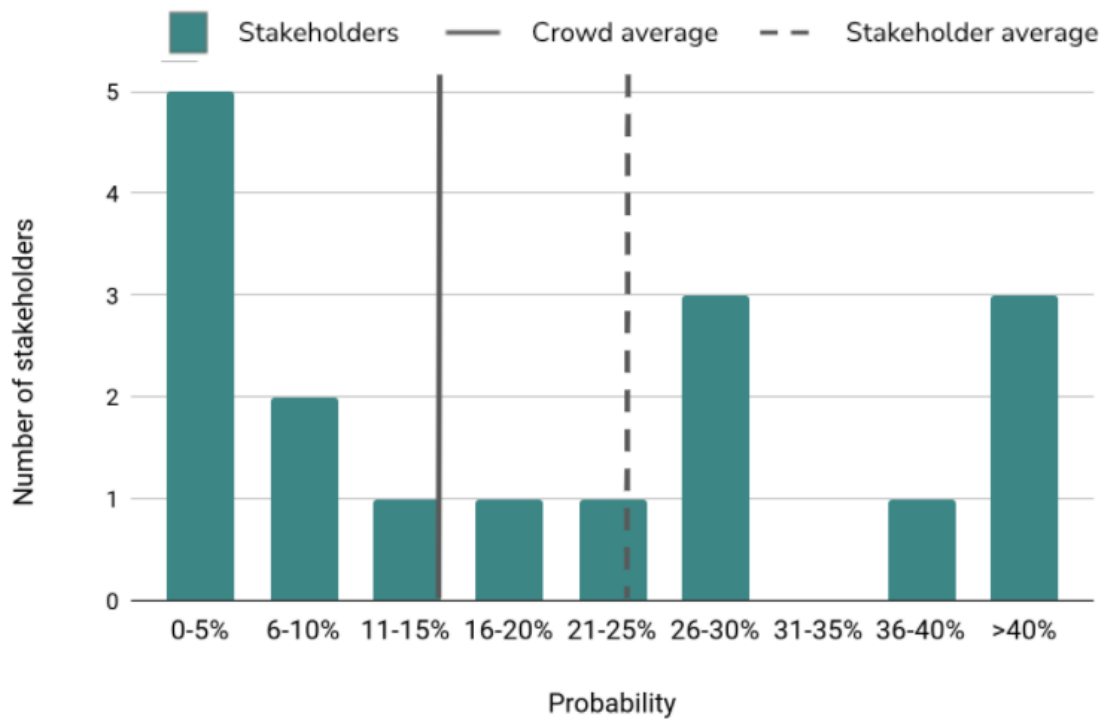
## Factor 5: China Military Aggression

Most stakeholders (82 percent) believe an increase in *China military aggression* would improve the DOD-Silicon Valley relationship; none believe it would worsen it. See Figure 4. The stakeholder cohorts' forecasts did not show a consistent pattern. The crowd's forecasts were lower than the stakeholders' forecasts on two of the four metrics, and similar on the other two metrics. This suggests the crowd expects lower *China military aggression* than the stakeholders, which according to the stakeholder consensus, has negative implications for the relationship. Many forecasters cited the upcoming Beijing Olympics as a reason China is unlikely to instigate conflict in the next six months.

Metric 1: Likelihood the Chinese military or other maritime security forces will fire upon another country's civil or military vessel in the South China Sea in the next six months.

- **Stakeholders' views.** Average forecast of 24 percent, with no correlation between forecasts and overall expectations for the relationship.
- **Crowd's view.** Based on 199 forecasts, the crowd thinks this event is less likely (15 percent) than the stakeholders (24 percent).<sup>19</sup> The crowd's lower forecast indicates an expectation of lower *China military aggression*. According to the stakeholder consensus, this has negative implications for the DOD-Silicon Valley relationship.

Figure 14. Chance China shoots at another country's vessel in South China Sea in next six months



Sources: CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/349](https://www.cset-foretell.com/questions/349).

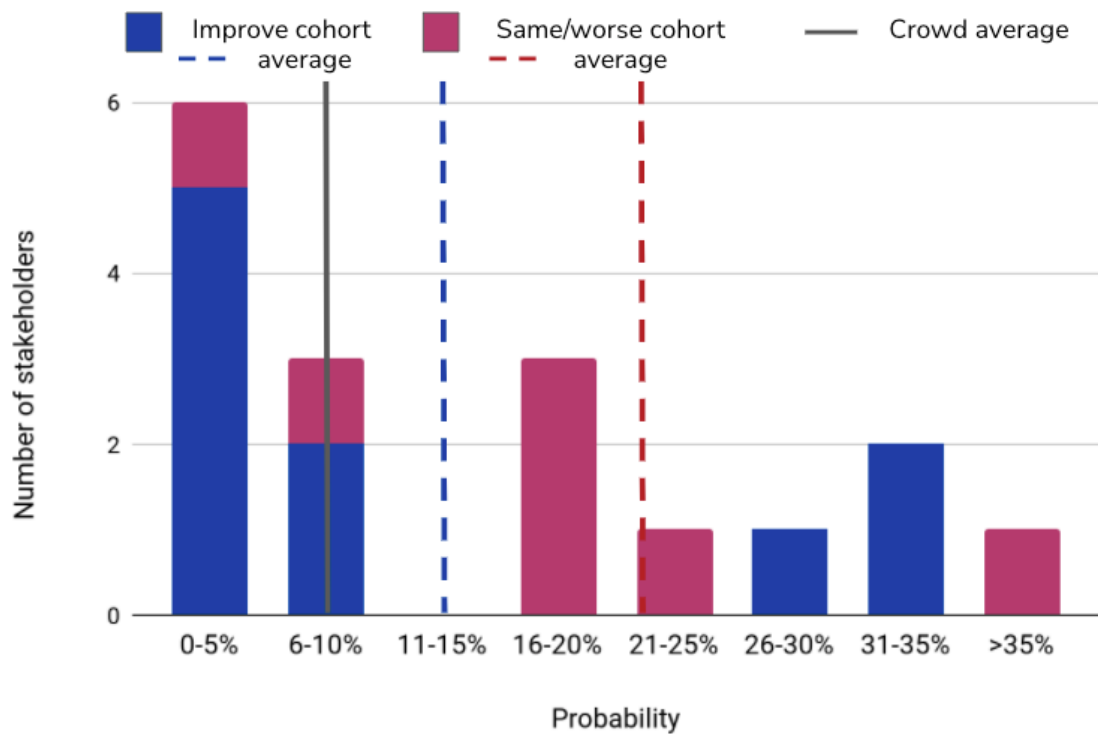
Metric 2: Likelihood China seizes control of a Taiwanese-occupied feature in the South China Sea in the next six months.

- **Stakeholders' views.** Average forecast of 16 percent, with a negative correlation (-0.3) between forecasts and overall expectations for the relationship. The *improve* cohort's forecast (13 percent) is lower than the *same/worse* cohort's forecast (21 percent), indicating that the *improve* cohort expects lower *China military aggression*. The direction of cohort divergence for this metric is unexpected.
- **Crowd's view.** Based on 183 forecasts, the crowd thinks this event is less likely (8 percent) than both stakeholder cohorts.

<sup>20</sup> The crowd's lower forecast indicates an expectation of lower *China military aggression*. According to the stakeholder consensus, this has negative implications for the DOD-Silicon Valley relationship.



Figure 15. Chance China seizes Taiwanese-occupied feature in South China Sea in next six months

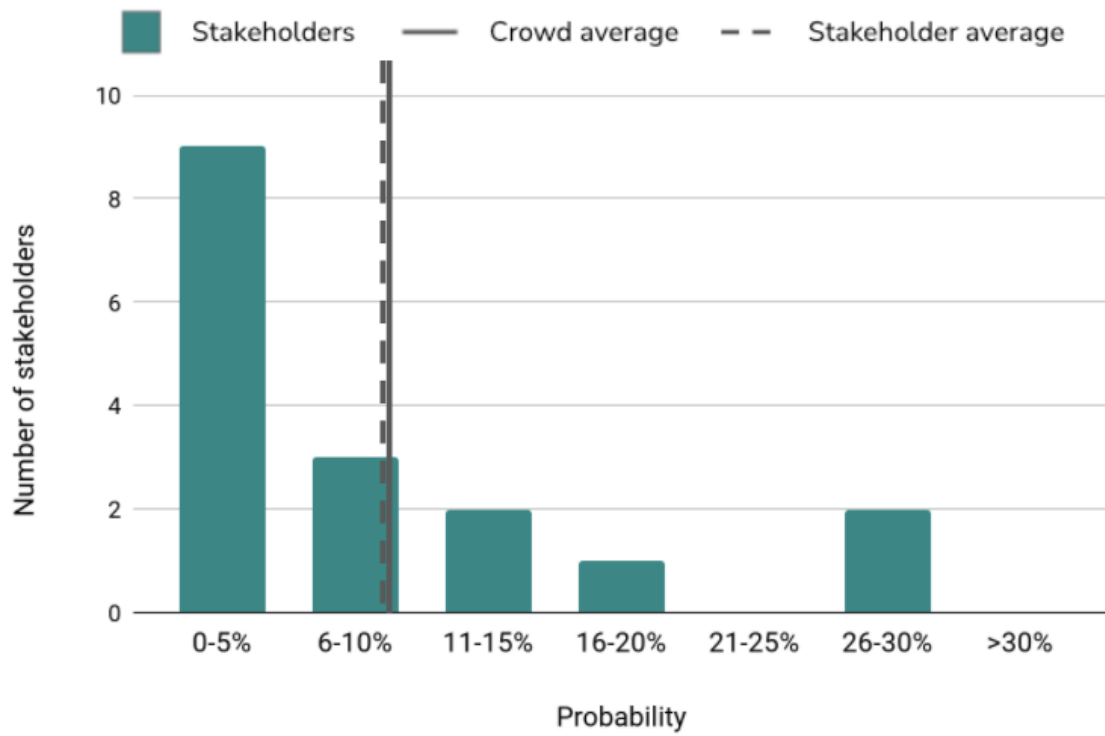


Sources: CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/372](https://www.cset-foretell.com/questions/372).

**Metric 3:** Likelihood China executes an acknowledged military attack against Vietnam, India, or Taiwan in the next six months.

- **Stakeholders' views.** Average forecast of 9 percent, with no correlation between forecasts and overall expectations for the relationship.
- **Crowd's view.** The crowd's forecast (9 percent) is the same as the stakeholders' forecast.

Figure 16. Chance China executes a military attack against Vietnam, India, or Taiwan in next six months

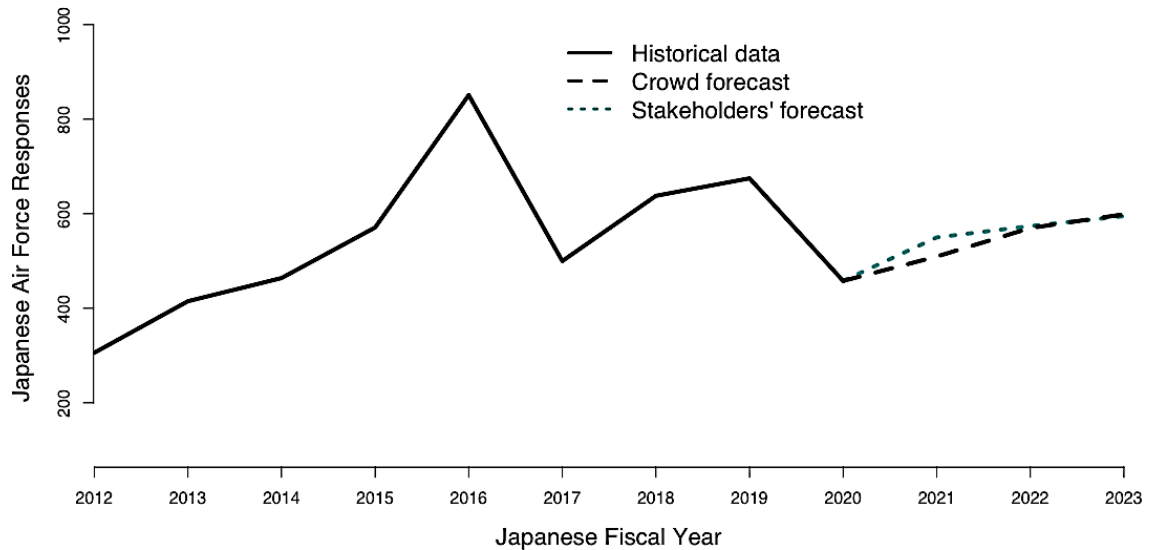


Sources: CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/373](http://www.cset-foretell.com/questions/373).

Metric 4: Annual frequency of Japanese Air Force responses to threats to Japan's territorial airspace by Chinese military aircraft.<sup>21</sup>

- **Stakeholders' views.** Average forecast of 595 Japanese Air Force responses in 2023, with a very weak (0.1) positive correlation between forecasts and overall expectations for the relationship.
- **Crowd's view.** Based on 99 forecasts, the crowd forecasts a similar trend to the stakeholders. The crowd expects the number of Japanese Air Force responses to increase year-to-year, reaching 599 in 2023.

Figure 17. Japanese Air Force responses to Chinese incursions of airspace



Sources: Japan Air Force Press Releases, CSET stakeholder survey, and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see <http://www.cset-foretell.com/questions/374>.

## Factor 6: China tech capabilities

Most stakeholders (71 percent) believe an increase in *China tech capabilities* would improve the DOD-Silicon Valley relationship; none believe it would worsen it. See Figure 4. The stakeholders' forecasts weakly correlate with their overall expectations for one metric, and in the expected direction. For both metrics, the crowd expects higher China tech capabilities than both cohorts. Based on the stakeholder consensus, the crowd's views have positive implications for the DOD-Silicon Valley relationship.

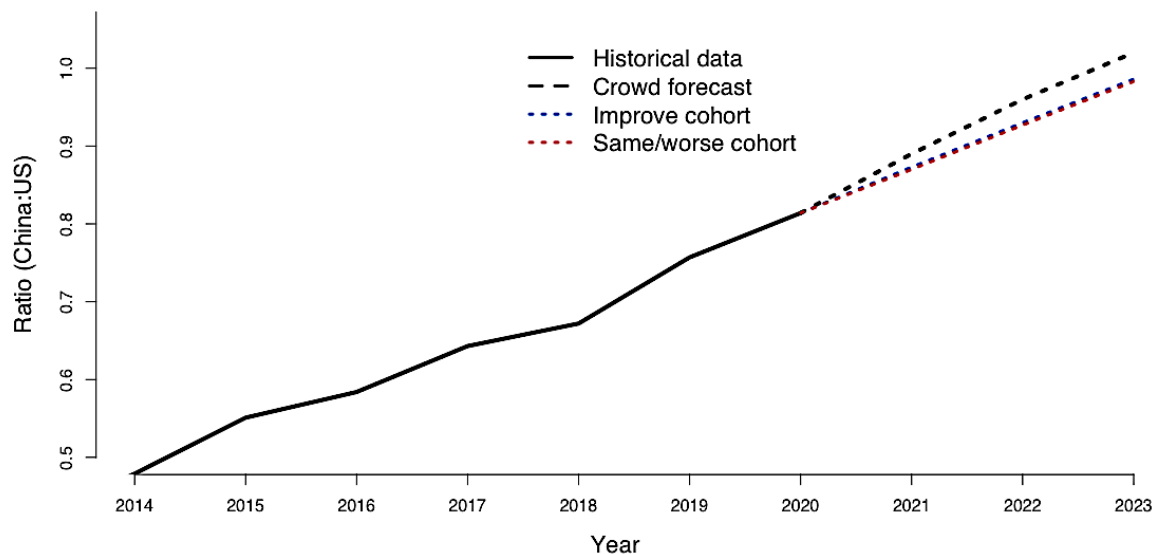
Metric 1: Ratio of China-authored to U.S.-authored top 1% AI publications.

- **Stakeholders' views.** Average forecast of 0.98 in 2023, with a weak positive correlation (0.2) between forecasts and

overall expectations for the relationship. The *improve* cohort's average forecast (0.99 in 2023) is slightly higher than *same/worse* cohort's average forecast (0.98 in 2023), indicating an expectation of slightly greater China tech capabilities.

- **Crowd's view.** Based on 117 forecasts, the crowd's forecast is slightly higher than both stakeholder cohorts. The crowd expects the ratio to increase to 1.02 in 2023. The crowd's higher forecast indicates an expectation of greater *China tech capabilities*. According to the stakeholder consensus, this has positive implications for the DOD-Silicon Valley relationship.

Figure 18. Ratio of China- to U.S.-authored top 1% AI publications

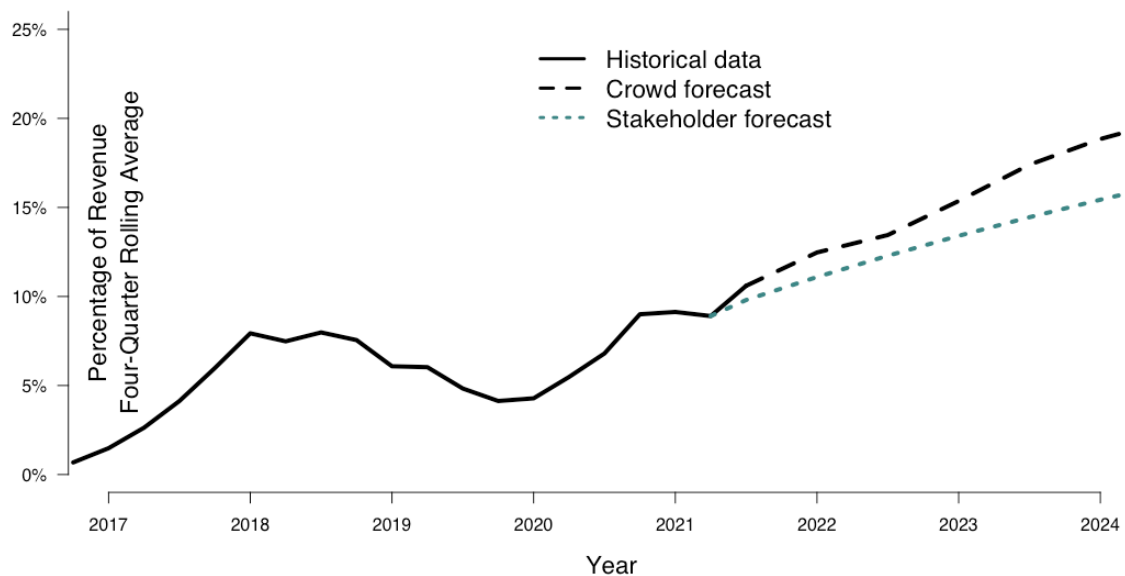


Sources: Clarivate Web of Science, Digital Science Dimensions, Microsoft Academic Graph, CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/899](http://www.cset-foretell.com/questions/899).

Metric 2: Percentage of Semiconductor Manufacturing International Corporation's annual revenue that comes from chips that are 28 nm or smaller.<sup>22</sup>

- **Stakeholders' views.** Average forecast of 16 percent, with no correlation between forecasts and overall expectations for the relationship.
- **Crowd's view.** Based on 76 forecasts, the crowd's forecast was higher than the stakeholders' average forecast. The crowd expects the percentage to increase to nearly 20 percent by the second quarter of 2024. The crowd's higher forecast indicates an expectation of greater *China tech capabilities*. According to the stakeholder consensus, this has positive implications for the DOD-Silicon Valley relationship.

Figure 19. Percentage of SMIC revenue from 14/28 NM chips or below



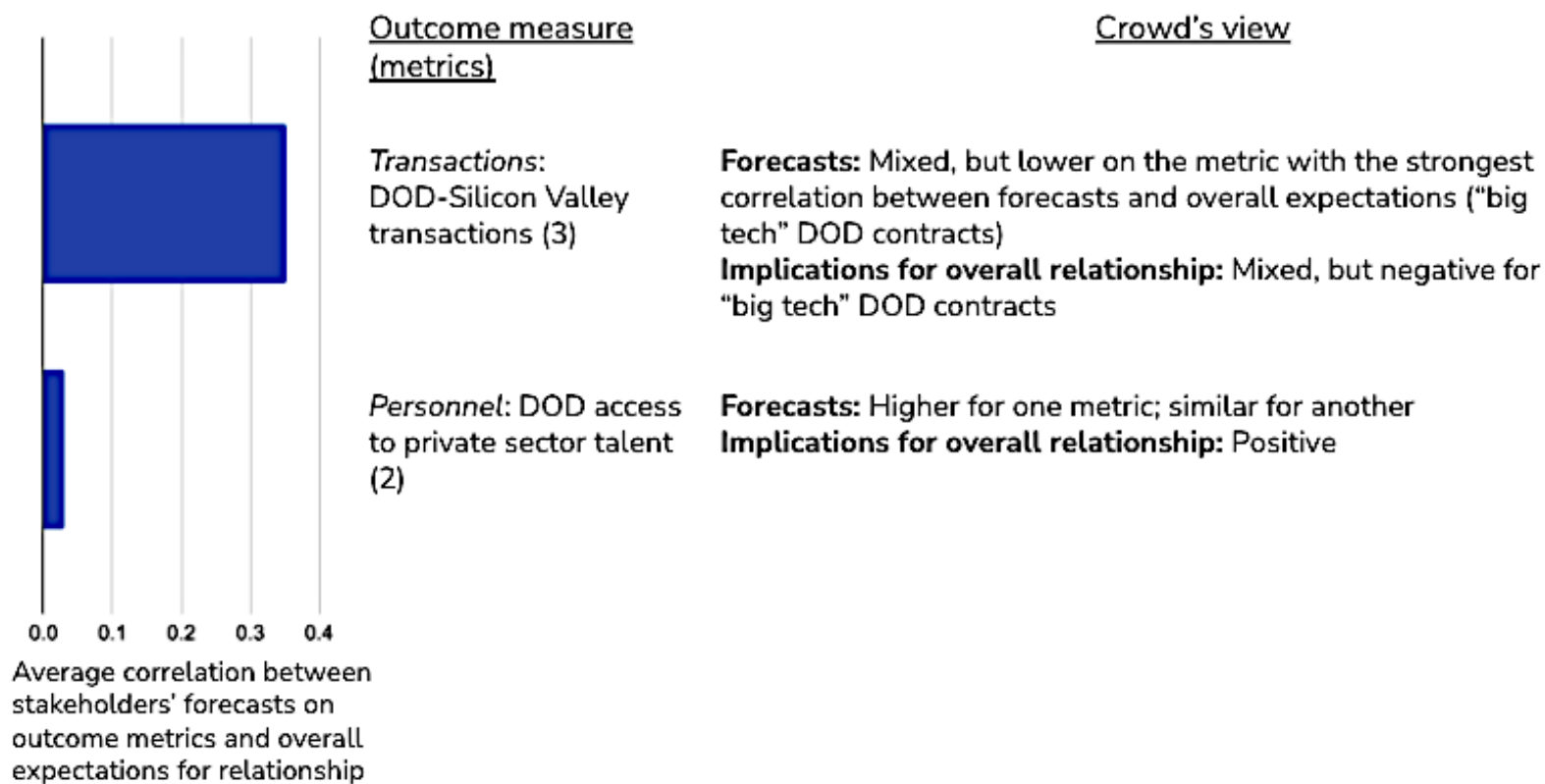
Sources: SMIC quarterly financial reports, CSET stakeholder survey, and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/382](http://www.cset-foretell.com/questions/382).

### *Outcome Measure Forecasts*

This subsection compares the stakeholders' forecasts and crowd's forecasts for the metrics informing two measures of the state of the DOD-Silicon Valley relationship: contracts between DOD and Silicon Valley (*transactions*) and DOD's access to talent developed in the private tech sector (*personnel*). Based on the stakeholder interviews, we determined that more transactions and greater access to personnel correspond with a better DOD-Silicon Valley relationship.

As shown in Figure 20, a pattern emerged for the *transactions* measure, but not for the *personnel* measure. For the transactions measure, the "big tech" DOD contracts metric has the strongest correlation between the stakeholders' forecasts and their overall expectations for the relationship. The difference between the cohorts' forecasts is also statistically significant for this metric. The crowd's forecasts for the "big tech" DOD contracts metrics is closer to the *same/worse* cohort, which, according to the stakeholders' views on the relevance of the *transactions* measure, indicates a negative expectation for the overall relationship, relative to the stakeholders.

Figure 20. Summary of stakeholders' and crowd's views on outcome measures



Sources: CSET stakeholder survey and Foretell.

Note: The implications of the crowd's forecasts on the DOD-Silicon Valley relationship are relative to the stakeholders' forecasts. For example, because the stakeholders believe greater DOD access to *personnel* is positive for the relationship and the crowd's forecasts on the *personnel* measure are overall higher than those of the stakeholders, the implications of the crowd's forecasts on the relationship are positive. Foretell data pulled November 4, 2021.

Below we provide a factor-by-factor comparison of the stakeholders' and crowd's forecasts, highlighting where the stakeholder cohorts' forecasts diverge in a notable way. Table 2 in the Appendix provides a more-detailed comparison of the forecasts. Visit [cset-foretell.com](https://cset-foretell.com) for more information about the metrics, including up-to-date forecasts.

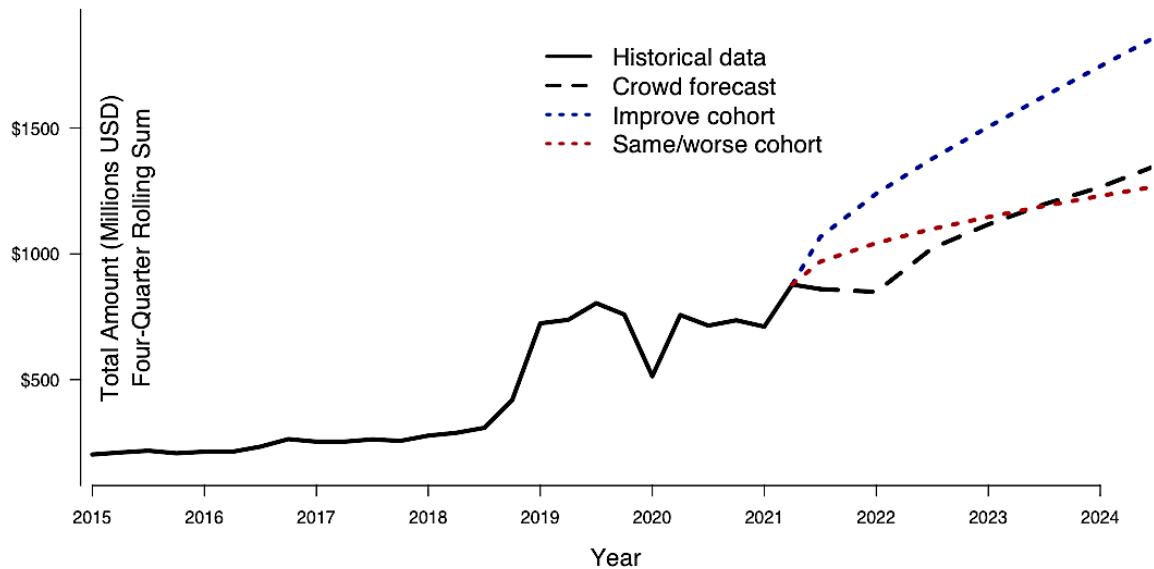
### Outcome Measure 1: Transactions

Metric 1: Combined annual dollar amount of DOD contracts with Alphabet, Amazon, Apple, Facebook, and Microsoft.

- **Stakeholders' views.** Average forecast of \$1.6 billion by the second quarter of 2024, with a large, positive correlation (0.7) with overall expectations for the relationship. The *improve* cohort's average forecast (\$1.9 billion in 2024 Q2) is higher than the *same/worse* cohort's average forecast (\$1.4 billion in 2024 Q2). The difference between the cohorts' forecasts is statistically significant.
- **Crowd's view.** Based on 95 forecasts, the crowd's forecast is closer to the *same/worse* cohort. The crowd expects the annual contract dollar value to increase to \$1.4 billion by the second quarter of 2024. The crowd's lower forecast indicates an expectation of fewer *transactions*, relative to the stakeholders. According to the stakeholder interviews, this has negative implications for the DOD-Silicon Valley relationship.



Figure 21. DOD contract amounts of Alphabet, Amazon, Apple, Facebook, and Microsoft

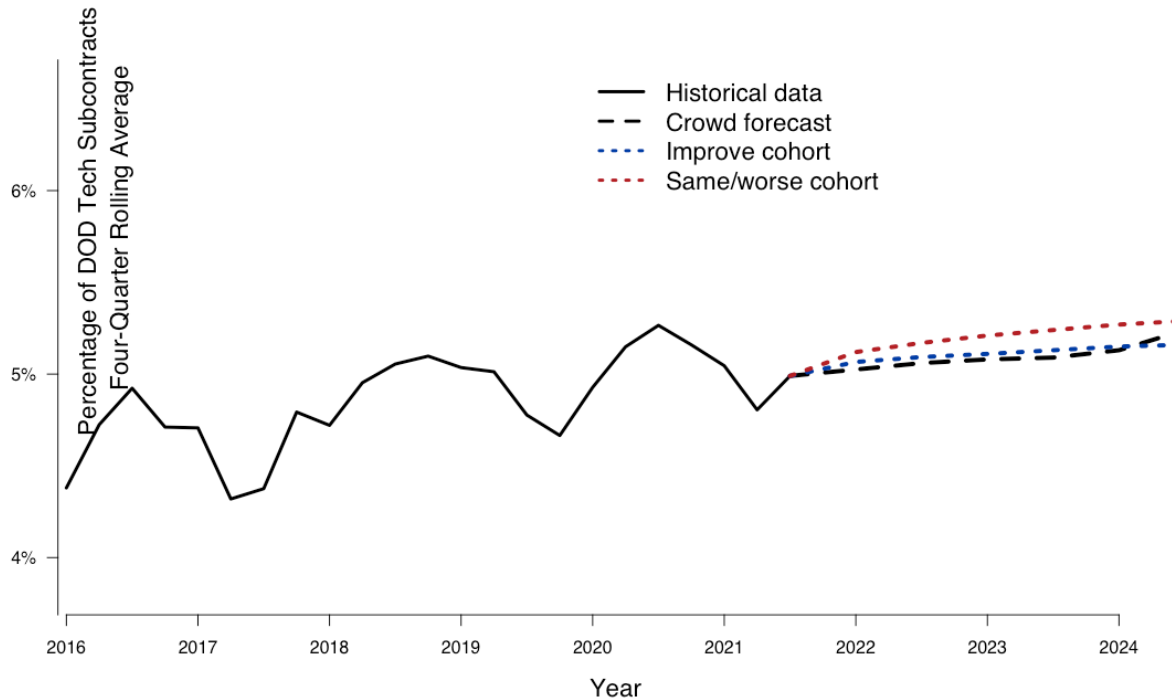


Sources: Bloomberg Government, CSET stakeholder survey, and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/311](https://www.cset-foretell.com/questions/311).

**Metric 2:** Percentage of DOD subcontracts for scientific research and development services going to Northern California-based companies.

- **Stakeholders' views.** Average forecast of 5 percent by the second quarter of 2024, with no correlation between forecasts and overall expectations for the relationship.
- **What the crowd says.** Based on 71 forecasts, the crowd agrees with the stakeholders and does not expect significant changes, forecasting that the percentage will continue to hover just over 5 percent.

Figure 22. Percentage of DOD tech subcontracts to Northern California companies



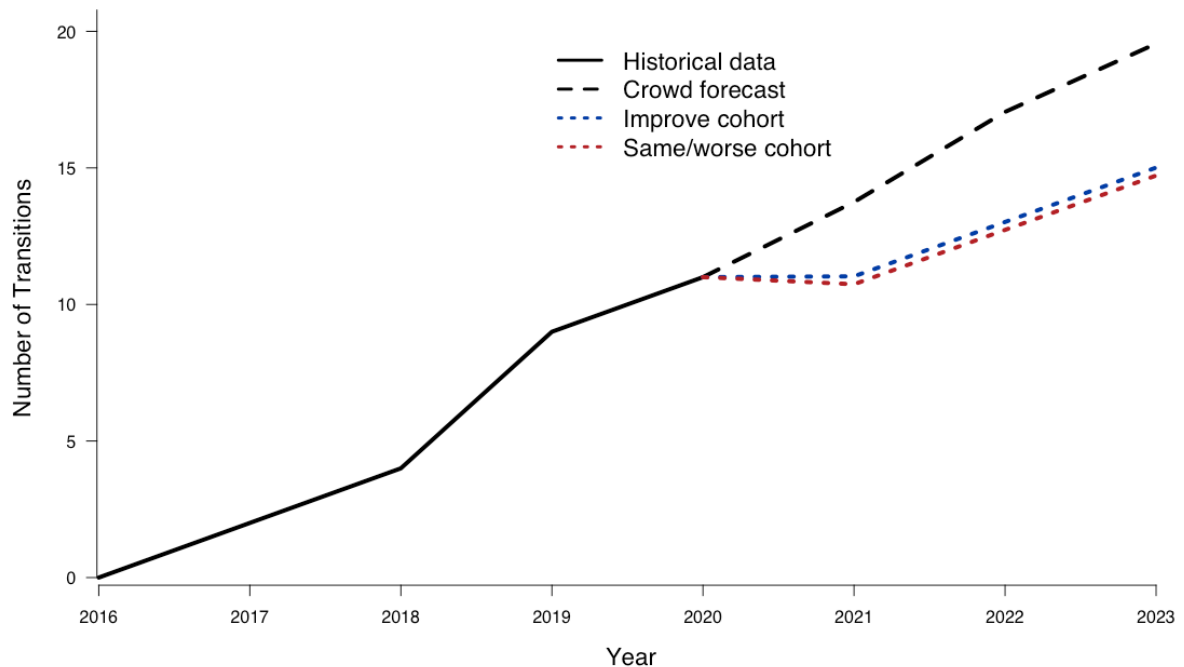
Sources: Bloomberg Government, CSET stakeholder survey, and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/332](http://www.cset-foretell.com/questions/332).

### Metric 3: Defense Innovation Unit transitions.<sup>23</sup>

- **Stakeholders' views.** Average forecast of 14.8 in 2023, with a positive correlation (0.4) between forecasts and overall expectations for the relationship. The *improve* cohort's average forecast (15 in 2023) is slightly higher than the *same/worse* cohort's average forecast (14.7 in 2023).
- **Crowd's view.** Based on 80 forecasts, the crowd's forecast is higher than both stakeholder cohorts. The crowd expects DIU transitions to increase from 11 in 2020 to 19.6 in 2023. The crowd's higher forecast indicates an expectation of more *transactions*, relative to the stakeholders. According

to the stakeholder interviews, this has positive implications for the DOD-Silicon Valley relationship.

Figure 23. Defense Innovation Unit transitions



Sources: Defense Innovation Unit Annual Report, CSET stakeholder survey, and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/339](https://www.cset-foretell.com/questions/339).

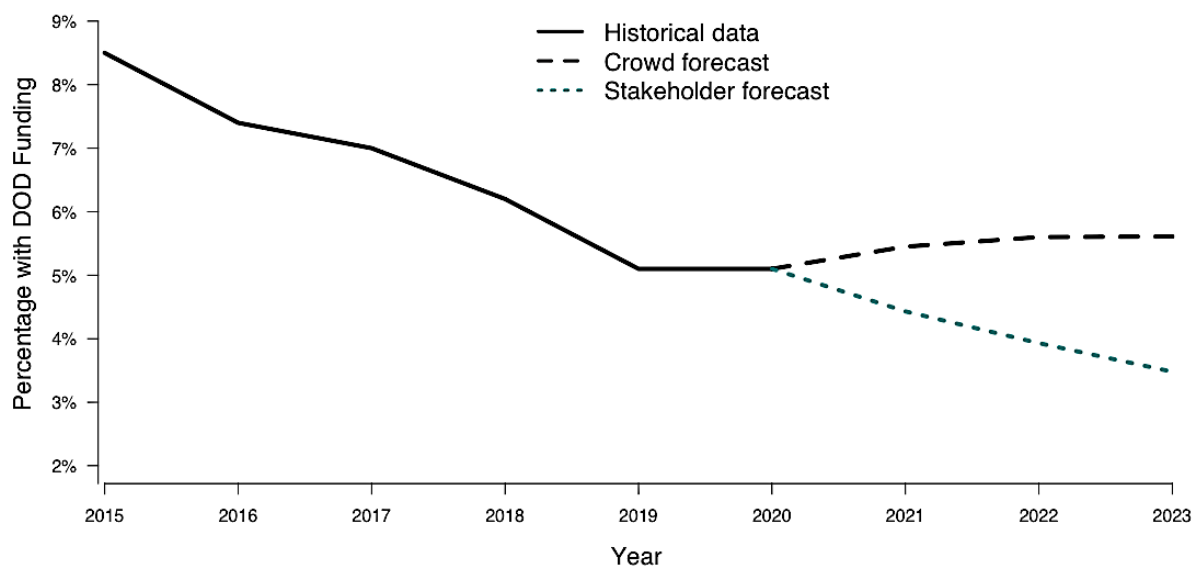
## Outcome Measure 2: Personnel

Metric 1: Percentage of top 10 percent U.S. AI publications supported by a DOD grant.

- **Stakeholders' views.** Average forecast of 3.5 percent in 2023, with a very weak negative correlation (-0.1) between forecasts and overall expectations for the relationship.
- **Crowd's view.** Based on 71 forecasts, the crowd forecasts that the percentage will reverse its historic decline and remain roughly where it is today (5.6 percent). The crowd's

forecast is higher than the stakeholders' forecast who expect the historic decline to continue. The crowd's higher forecast indicates an expectation of greater DOD access to *personnel*, relative to the stakeholders. According to the stakeholder interviews, this has positive implications for the DOD-Silicon Valley relationship.

Figure 24. Percentage of top 10% U.S. AI publications with DOD funding



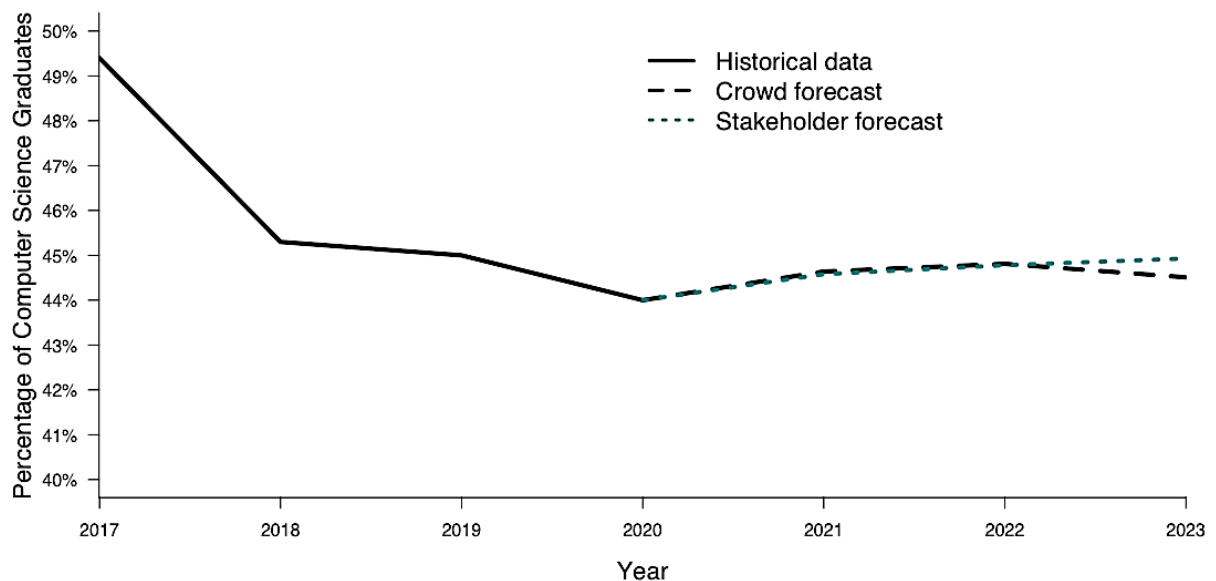
Sources: Digital Science Dimensions, CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/445](http://www.cset-foretell.com/questions/445).

Metric 2: Percentage of Carnegie Mellon University computer science graduates whose first job is at a company that has a contract with DOD.<sup>24</sup>

- **Stakeholders' views.** Average forecast of 45 percent in 2023, with a weak positive correlation (0.2) between forecasts and overall expectations for the relationship.

- **Crowd's view.** Based on 64 forecasts, the crowd's forecast is similar to the stakeholders' forecast. The crowd expects the percentage will remain roughly the same as it is today (45 percent in 2023).

Figure 25. Percentage of CMU computer science graduates to companies with DOD contracts



Sources: Bloomberg Government, CMU First Destinations Outcomes Dashboard, CSET stakeholder survey and Foretell. Foretell data pulled November 4, 2021. For current forecasts, see [www.cset-foretell.com/questions/368](http://www.cset-foretell.com/questions/368).

## What should stakeholders do with this information?

This section offers a few suggestions for how stakeholders could use the outputs of this process.

**Focus the debate.** This study discovered which components of the overarching question—"What is the future of the DOD-Silicon Valley relationship?"—are most contested by the stakeholders, in part by disentangling factor disagreements and forecast disagreements. Regarding factor disagreements, the cohorts most

significantly diverged on how changes to the strength of the U.S. tech sector would affect the DOD-Silicon Valley relationship. This factor, and its relationship to the DOD-Silicon Valley relationship, is a candidate focus area in its own right. For example, what are the mechanisms by which a stronger U.S. tech sector could improve or worsen the relationship?

As for forecast disagreements, the cohorts significantly disagreed about future trends involving American's confidence in the military, "big tech" revenue, and the amount of "big tech" contracts with DOD. The stakeholders might ask themselves how their overall expectations for the DOD-Silicon Valley relationship would change if the other cohort is correct about these trends. The closer the link between the stakeholders' expectations for these trends and their overall expectations for the DOD-Silicon Valley relationship, the more we should invest in understanding, monitoring, and forecasting them.

**Adjust expectations.** Individual stakeholders can see where their forecasts are outliers. Different views on uncertain topics are appropriate, but others' views—from both other stakeholders and the crowd—provide relevant information that should generally cause one to reconsider and adjust their own views. For example, stakeholders who think the chance of China shooting at another country's vessel in the South China Sea in the next six months is more likely than not—an outlier forecast—should consider adjusting their expectations for that metric.

After adjusting their expectations on individual metrics, the stakeholders should also consider the implications of those adjustments on their overall expectations for the DOD-Silicon Valley relationship. For example, if a stakeholder believes increases in *China military aggression* would improve the DOD-Silicon Valley relationship and learns through this process that they have consistently higher forecasts than others on metrics relating to that factor, they should also conclude that their overall expectations for the relationship might be too optimistic. If all the stakeholders engage in this process, the level of overall disagreement would go

down. The more credible the probabilistic forecasting method—here, the Foretell crowd—and the less confident the individual stakeholder is in their initial forecast, the more adjustment is appropriate.

**Notice surprising developments.** Even for the stakeholders who find the views of the Foretell crowd unconvincing, creating a record of one's views at a moment in time is valuable. Viewing changes in the world as expected developments is a common bias. By memorializing what one thought at a point in time (forecasts) and the big-picture significance (effect on DOD-Silicon Valley relationship), it is easier to give changes their due weight. For example, after the stakeholders submitted their forecasts, new data on money raised by private U.S. tech companies and tech startups showed significant increases over the previous quarter, well above any of the stakeholders' expectations. In the ordinary course, one might not notice these changes, but because we have a record of the stakeholders' expectations, we know that they are surprising.

Noticing surprising developments on individual metrics is only the first step. It's also important to understand the broader implications of those developments. By mapping metrics to factors—such as the *strength of the U.S. tech sector*—and big issues—such as the future of the DOD-Silicon Valley relationship—one is better able to make the correct adjustments to their worldview in light of observable changes in the world.

## Conclusion

This brief explores a method for incorporating probabilistic forecasting methods into a policy debate on a big question. While the method succeeded in identifying potential root disagreements amenable to probabilistic forecasts, it has limitations. For one, it cannot capture everything that matters. For example, several stakeholders cited factors critical to the future of the DOD-Silicon Valley relationship that are difficult to measure, such as domestic and international regulation of AI technologies, DOD culture, and obstacles to non-traditional defense contractors working with

DOD. These limitations highlight the importance of pairing quantitative approaches to forecasting with qualitative approaches—such as scenario planning—which can provide complementary texture and context.

Limitations notwithstanding, we think extending this methodology to include conditional forecasts would improve group decision-making. Where disagreements about a policy can be traced to its measurable consequences, forecasting the effect of both pursuing and not pursuing the policy can help resolve the disagreement. It would also generate a record for the decision-makers on what they expected would happen that could be compared to what actually happened, allowing decision-makers to extract lessons relevant for future decisions.



## Author

Michael Page is a Research Fellow at CSET.

## Acknowledgments

For feedback and assistance, the author would like to thank Catherine Aiken, Danny Hague, Margarita Konaev, Igor Mikolic-Torreira, and Lynne Weil. The author would also like to thank Cultivate Labs—and Vanessa Pineda and Carolyn Mehta in particular—for creating and maintaining the forecasting platform. The author is solely responsible for the views expressed in this piece and for any errors.



© 2021 by the Center for Security and Emerging Technology. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.

To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0/>.

Document Identifier: doi: 10.51593/20210076

## Appendix

### *Participating Stakeholders*

Catherine Aiken, CSET [interview and survey]

Anthony Bak, Palantir [survey]

Jason Brown, Google; formerly U.S. Air Force, Joint Artificial Intelligence Center [interview and survey]

Michael Brown, Defense Innovation Unit [interview]

Miles Brundage, OpenAI [interview and survey]

Bess Dopkeen, House Committee on Armed Services; formerly Department of Defense [survey]

Ed Felten, Princeton University [survey]

Melissa Flagg, Flagg Consulting and Atlantic Council's GeoTech Center; formerly CSET, Department of Defense [interview and survey]

Michael Horowitz, University of Pennsylvania; formerly Department of Defense [survey]

Nicholas Joseph, Anthropic [interview and survey]

Josh Marcuse, Google; formerly Defense Innovation Board [survey]

Igor Mikolic-Torreira, CSET; formerly Department of Defense [interview and survey]

Enrique Oti, Second Front Systems; formerly U.S. Air Force, Defense Innovation Unit [interview]

Scott Phoenix, Vicarious [interview]

Jack Poulson, Tech Inquiry [interview and survey]

James Ryseff, RAND; formerly Google, Microsoft [interview and survey]

Jack Shanahan, retired U.S. Air Force, Joint Artificial Intelligence Center [interview and survey]

Trae Stephens, Anduril Industries, Founders Fund [survey]

Danielle C. Tarraf, JCI Ventures; formerly RAND [interview and survey]

*Forecast Summary Tables*

Table 1. Factor forecast summary

Factor  Predicted effect of increase on relationship: improve, worsen, contested	Metric  (Year of forecast provided for time series metrics)	Stakeholder Forecasts		Crowd Forecast
		All  (Improve vs. Same/worse)	Correlation with overall expectations	
U.S. distrust of military/ government	U.S. use of lethal autonomous drone, next year	10%  (7% / 16%)	-0.3	16%
	Percentage of U.S. with low confidence in military (2024)	29%  (28%/31%)*	-0.5	31%
	Percentage of U.S. concerned about government use of their data (2022)	67%	0.2	67%

U.S. political polarization	Affective political polarization in the U.S. (2024)	51 (50.7 / 51.5)	-0.4	55
Silicon Valley protests	Employee protest at “big tech” company against company’s involvement with DOD	46% (38% / 56%)	-0.3	25%
Strength of U.S. tech sector	Annual “Big tech” revenue (2024 Q2)	\$2.5 T (\$2.6 T/ \$2.3 T)*	0.5	\$2.3 T
	Annual U.S. private tech company funding (2024 Q2)	\$183 B (\$186 B/\$179 B)	0.2	\$239 B
	Annual U.S. tech startup funding (2024 Q2)	\$56 B (\$58 B/ \$54 B)	0.2	\$78 B
China military aggression	China shooting conflict in South China Sea, next six months	24%	0	15%

	China seizure of Taiwanese feature in South China Sea, next six months	16% (13% / 21%)	-0.3	8%
	China military attack on Vietnam, India, or Taiwan, next six months	9%	0	9%
	Japanese Air Force responses to Chinese threats to airspace (2023)	595	0.1	599
China tech capabilities	China-U.S. ratio of top 1% AI publications (2023)	0.98 (0.99 / 0.98)	0.2	1.0
	Percentage of SMIC revenue from 28 nm chips or smaller (2024 Q2)	16.4% (17.3% / 15.5%)	0	19.9%

Sources: Stakeholder survey and Foretell. Foretell data pulled November 4, 2021. The improve and *same/worse* cohorts' average forecasts are provided where the correlation between forecasts and overall expectations is greater than 0.2. Statistically significant differences between the cohorts are noted with an asterisk.

Table 2. Outcome measure forecast summary

Outcome Measure	Metric (Year of forecast provided for time series metrics)	Stakeholder Forecasts		Crowd Forecast
		All (Improve vs. Same/worse)	Correlation with overall expectations	
Transactions	Annual value of “big tech” DOD contracts (2024 Q2)	\$1.6 B (\$1.9 B / \$1.3 B)*	0.7	\$1.4 B
	DOD tech subcontracts to Northern California-based companies (2024 Q2)	5.2% (5.2% / 5.3%)	0	5.2%
	Defense Innovation Unit transitions (2023)	14.8 (15.0 / 14.7)	0.4	19.6
Personnel	Percentage of top 10% AI publications funded by DOD (2023)	3.5%	-0.1	5.6%
	Percentage of CMU computer science graduates first employed by company with DOD contract (2023)	44.9%	0.2	44.5%

Sources: Stakeholder survey and Foretell. Foretell data pulled November 4, 2021. The *improve* and *same/worse* cohorts' average forecasts are provided where the correlation between forecasts and overall expectations is greater than +/- 0.2. Statistically significant differences between the cohorts are noted with an asterisk.

## Endnotes

<sup>1</sup> See generally J. Peter Scoblic and Philip E. Tetlock, “A Better Crystal Ball: The Right Way to Think About the Future,” *Foreign Affairs*, November/December 2000, <https://www.foreignaffairs.com/articles/unitedstates/2020-10-13/better-crystal-ball>.

<sup>2</sup> IARPA’s Aggregative Contingent Estimation (ACE) project was a four-year tournament designed to identify the methods most effective at estimating the likelihood of geopolitical events between one month and one year in the future. The Good Judgment Project, a team led by University of Pennsylvania professors Phillip Tetlock and Barbara Mellers, won the tournament by recruiting thousands of volunteers, training them to forecast, and aggregating their judgments. Tetlock and Dan Gardner describe their approach in *Superforecasting: The Art and Science of Prediction* (Broadway Books, 2015).

<sup>3</sup> Tetlock et al., “Bringing probability judgments into policy debates via forecasting tournaments,” *Science* Vol 355 Issue 6324, (February 5, 2017), 481-483, <https://science.sciencemag.org/content/355/6324/481.full>.

<sup>4</sup> Our method involves a form of “question clustering” proposed by Philip Tetlock. See Philip E. Tetlock, “Full-Inference-Cycle Tournaments: The Quality of our Questions Matters as Much as the Accuracy of our Answers,” Prepared for IARPA, August 30, 2017, available through Dropbox at <https://t.co/dLOOCXac8A?amp=1>. Scoblic and Tetlock, *supra* note 1, make a similar proposal, suggesting methods for combining qualitative scenario planning and probabilistic forecasting to create warning systems for policymakers.

<sup>5</sup> See generally Aiken et al., “‘Cool Projects’ or ‘Expanding the Efficiency of the Murderous American War Machine?’” CSET Issue Brief, November 2020.

<sup>6</sup> Elicit is a research assistant tool built on OpenAI’s GPT-3 language model, <https://elicit.org/>.

<sup>7</sup> The stakeholder survey is available at [github.com/georgetown-cset/public-foretell/blob/main/Future%20of%20DoD-SV%20Relationship%2C%20Stakeholder%20Survey.pdf](https://github.com/georgetown-cset/public-foretell/blob/main/Future%20of%20DoD-SV%20Relationship%2C%20Stakeholder%20Survey.pdf).

<sup>8</sup> The stakeholder survey defined these terms as follows: “Silicon Valley refers to U.S.-based technology companies whose business models are not based on Department of Defense contracts. It’s not limited to companies in the San Francisco Bay Area, although in some cases, we use location as a proxy for companies in this category. The DOD-Silicon Valley relationship refers to the flow of resources, such as information, technology, and talent, between DOD and Silicon Valley companies. A positive change means more exchange of these



resources, and a negative change means less exchange of these resources. These definitions are deliberately loose, and we invite stakeholders to use their own common-sense interpretation of them, consistent with the spirit of the project.”

<sup>9</sup> We initially identified 20 metrics but removed a metric relating to the Geopolitical Risk Index due to a substantial change in the index’s methodology made after the question was published.

<sup>10</sup> We used a simpler elicitation method for the stakeholders for two reasons. First, while the purpose of the crowd’s forecast is accuracy, the purpose of the stakeholders’ forecast is to understand the distribution of stakeholders’ views. For the latter purpose, the additional detail provided by the crowd provides less value. Second, we wanted to minimize the time required of the stakeholders to participate in this process.

<sup>11</sup> More data on the crowd’s forecasts are available on [cset-foretell.com](https://cset-foretell.com). For time series metrics, for example, the crowd provides both a point estimate (median forecast) and an 80 percent prediction interval for each year or half-year in the forecast period. For simplicity, we only use the point estimate in this study.

<sup>12</sup> We also surveyed the stakeholders on the effect decreases in each factor would have on the DOD-Silicon Valley relationship. The stakeholders’ views on the effect of factor increases generally mirrored their views on the effect of factor decreases. For simplicity, this brief focuses only on the effect of increases, but data on the effect of decreases is available at <https://github.com/georgetown-cset/public-foretell/blob/main/Wisdom%20of%20crowd%20as%20arbiter%20of%20expert%20disagreement%2C%20data.xlsx>.

<sup>13</sup> For the *U.S. distrust in the military/government* factor, two stakeholders (12 percent) selected that an increase in the factor would improve the DOD-Silicon Valley relationship, but a strong majority (65 percent) believe an increase would worsen it.

<sup>14</sup> Two stakeholders (12 percent) believe an increase in *U.S. distrust of military/government* would improve the DOD-Silicon Valley relationship. Despite this disagreement, because a strong majority believe an increase in this factor would worsen the relationship, we categorized this factor as one having a negative effect on the relationship.

<sup>15</sup> The survey provides four answer options: “a great deal,” “quite a lot,” “some,” and “very little.” This metric tracks the percentage of respondents who select the two lowest-confidence options: “some” or “very little.” See Brennan, Megan, “Americans’ Confidence in Major U.S. Institutions Dips,” Gallup, July 14, 2021, <https://news.gallup.com/poll/352316/americans-confidence-major-institutions-dips.aspx>.

<sup>16</sup> The survey asks “How concerned are you, if at all, about how the government is using the data it collects about you?” The answer options are “Very concerned,” “Somewhat concerned,” “Not too concerned,” and “Not at all concerned.” This question asks about the percentage of respondents who select either “Very concerned” or “Somewhat concerned.” See Pew Research Center, “Americans and Privacy: Concerned, Confused and Feeling Lack of Control over their Personal Information,” [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/11/Pew-Research-Center\\_PI\\_2019.11.15\\_Privacy\\_FINAL.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/11/Pew-Research-Center_PI_2019.11.15_Privacy_FINAL.pdf).

<sup>17</sup> Because there is only one historical data point, we were not able to use forecast categories based on the historical data in the stakeholder survey. As a consequence, we asked the stakeholders only to forecast 2022.

<sup>18</sup> The survey asks how Democrats and Republicans feel about their own party and the other party on a 0-100 scale. Affective political polarization is the difference between those scores.

<sup>19</sup> Foretell published this question three times in the past, for the six-month periods covering the first half of 2021, the second half of 2021, and September 2021 through February 2022. The Foretell crowd forecasts were 11 percent, 14 percent, and 14 percent, respectively.

<sup>20</sup> Foretell published this question in the past for the six-month period beginning August 2021. The final crowd forecast was 6 percent.

<sup>21</sup> The frequency of Chinese incursions into Japanese airspace appears to reflect tensions between China and Japan, and the aggressiveness of China’s foreign policy. For example, the spike in the number of Japanese responses to Chinese incursions in 2016 coincided with Japan’s announced intention to revise its constitution in a manner China found threatening. See “A Chinese View: The Risks of Changing Japan’s Peace Constitution,” Council on Foreign Relations, May 11, 2017, [www.cfr.org/blog/chinese-view-risks-changing-japans-peace-constitution](http://www.cfr.org/blog/chinese-view-risks-changing-japans-peace-constitution).

<sup>22</sup> SMIC is China's leading manufacturer of semiconductor chips. To reduce its dependence on the United States and U.S. allies, China has prioritized developing its domestic semiconductor industry. China is currently prioritizing 28 nm chips—viewed as the dividing line between mid- and high-end chips—and 14 nm chips.

<sup>23</sup> DIU was founded in 2015 with a mission to "strengthen U.S. national security by increasing the military's adoption of commercial technology." In response to DOD needs for commercial solutions, it solicits proposals, primarily from small businesses or non-traditional defense contractors. Successful applicants are contracted to develop prototypes. Transitions are successful prototypes that lead to large-volume procurement. See Defense Innovation Unit, Annual Report 2020, [assets.ctfassets.net/3nanhbfkr0pc/3VXak4123q9HHoG2rnpQFO/385542158e5b6ca62e7fa63c03bcfe0d/DIU\\_-\\_2020\\_Annual\\_Report\\_FINAL.pdf](https://assets.ctfassets.net/3nanhbfkr0pc/3VXak4123q9HHoG2rnpQFO/385542158e5b6ca62e7fa63c03bcfe0d/DIU_-_2020_Annual_Report_FINAL.pdf).

<sup>24</sup> CMU was selected for this metric for two reasons. First, according to U.S. News and World Report, CMU has the best computer science graduate program in the United States. And second, it is unusual in that it publicly reports where its students work after graduating in granular detail. See CMU First Destination Outcomes Dashboard, <https://www.cmu.edu/career/outcomes/post-grad-dashboard.html>.