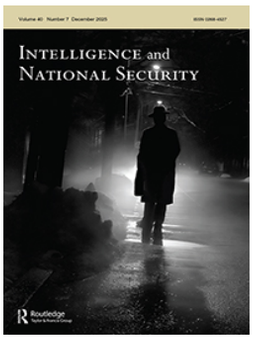

Routledge
Taylor & Francis Group

Intelligence and National Security

ISSN: 0268-4527 (Print) 1743-9019 (Online) Journal homepage: www.tandfonline.com/journals/fint20

Leveraging data science to investigate intelligence failures

Leonard Kern, Kristian Gustafson & Martin Ejnar Hansen

To cite this article: Leonard Kern, Kristian Gustafson & Martin Ejnar Hansen (03 Jan 2026): Leveraging data science to investigate intelligence failures, Intelligence and National Security, DOI: 10.1080/02684527.2025.2607374

To link to this article: https://doi.org/10.1080/02684527.2025.2607374

Published online: 03 Jan 2026.

INTELLIGENCE AND NATIONAL SECURITY
https://doi.org/10.1080/02684527.2025.2607374

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Leveraging data science to investigate intelligence failures

Leonard Kern, Kristian Gustafson and Martin Ejnar Hansen

ABSTRACT

This article challenges the conventional assumption underpinning the 'First Law of Intelligence Failure' – that warning signs are always available, but ignored, prior to intelligence breakdowns. Employing advanced natural language processing and machine learning techniques, the authors analyse declassified US State Department cables from the 1970s, focusing on two case studies often deemed intelligence failures: the Soviet invasion of Afghanistan and the Iranian Revolution. Using semantic outlier and change-point detection algorithms, they test whether meaningful signals ('signal in the noise') or emergent patterns ('connecting the dots') were more prevalent prior to failure than in earlier, 'successful' periods. The study finds this is not consistently the case, suggesting that indicators are not uniformly available or discernible before failures occur. By demonstrating the limitations of this study, the article concludes that the binary framing of intelligence as either success or failure is analytically flawed and potentially misleading. It offers a proof-of-concept for applying data science to intelligence analysis and advocates for a more nuanced understanding based on baselines and deviations, rather than retrospective judgements shaped by hindsight.

ARTICLE HISTORY
Received 17 June 2025
Accepted 15 December 2025

KEYWORDS
Intelligence failures; data science; United States

Introduction

Two notions continuously reappear in the discussion on intelligence failure: the failure to detect 'the signal in the noise' and the failure to 'connect the dots'.[1] The presence of both notions is so pervasive that James Wirtz has called their combination the 'First Law of Intelligence Failure'.[2] The common understanding is that there have been sufficient data available prior to an attack or incident that would have allowed for an effective warning, eventually averting the incipient intelligence failure. This appears to happen so frequently and predictably, so that one can deduce a law, akin to a law in the natural sciences. This First Law, importantly, *describes* situations prior to intelligence failures. It does not answer the question why the signal was not detected or why the dots were not connected nor does it set normative guidelines on how to avoid failures.[3] But, it enunciates a fundamental premise in that there are always signals or dots to be found amongst the noise.

The validity of this premise is examined in this article. To achieve this, the concepts of 'signal in the noise' and 'connected dots' are translated into

CONTACT Kristian Gustafson kristian.gustafson@brunel.ac.uk

measurable observations. By collecting and analysing data on past cases of intelligence failure and success, the understanding not only of these cases but also of the general concept of intelligence failure and success is deepened. In a rather novel way for the study of intelligence, modern data science methods are applied to a bulk of declassified governmental data pertaining to selected cases creating quantitative insights. This is made possible by recent advancements in natural language processing which provide the unprecedented ability to capture the tone and nuance in human language. Methods and techniques like those used in artificial intelligence reveal quantitative properties that go beyond basic descriptive statistics and can 'understand' textual data. Capturing these nuances is of particular relevance for research on diplomatic or intelligence primary sources, in which the choice and order of words is often a deliberate part of the messaging or judgement. These nuances are unlikely to be sufficiently captured by the usage of traditional keyword-based analyses alone.[4]

The dataset is provided by the History Lab's FOIArchive, which has made millions of declassified documents available in a machine-readable format.[5] Specifically, documents of the US State Department concerning two historical case studies are examined: the period preceding the Soviet invasion of Afghanistan in 1979 and the Iranian Revolution in the same year. While both events are widely considered to be instances of 'intelligence failures', the years long before the actual 'failure' are examined as well. In this way, it is demonstrated that the assumption of the sufficient availability of indicative data before the fact must be reconsidered.

The contribution of this article is threefold: First, it informs the discussion of intelligence failure and success. Second, it provides a methodology that can be applied to other historical cases using natural language processing techniques. Third, it provides a proof-of-concept for modern data science methods ready to be used in intelligence, amid ongoing discussions about the impact of artificial intelligence on the intelligence enterprise.[6]

This paper finds that its own limitations point to a proplematic 'success vs. failure' dichotomy: the nature of successful intelligence is anything but clear-cut, with fuzzy definitions and parameters. Both concepts – the 'signals in the noise' and 'connecting the dots' – are poor, indeed fallacious, concepts in understanding the ultimate success of intelligence, better suited to political infighting than the learning of lessons. As NSA Director Michael Hayden has said, 'accusations fit on a bumper sticker. The truth is longer'.[7]

This approach also helps demystify quantitative analysis. The mathematical concepts used here are indeed complex, but a solid grasp of the underlying logic should provide the reader sufficient guidance to understand the processes. Mathematical concepts are deliberately described in general terms to make this paper accessible to non-technical readers. Familiarity with what can be done through data analytics may allow it to be applied more widely within intelligence studies or intelligence analysis. As natural language AI models become more ubiquitous, applying data analytics to historical data sets may become a more accessible approach to a wider audience, and more easily deployable by scholars.

Intelligence failure and success

Despite, or rather because of, claims that 'the study of intelligence failures is perhaps the most academically advanced field in the study of intelligence',[8] competing scholarly perspectives, theories, and proposals for remediation exist.[9] If one then adds the plethora of political hearings, commissions, and journalistic accounts, this field appears exceptionally diverse and complex, perhaps even opaque, and challenging to fully grasp. This warrants a precise appreciation of the issue. In this article one aspect of intelligence failures is examined: the availability of indicative data prior to it. In doing so, the focus is on *descriptive* inferences. This concentration of scope is deliberate, aiming to apply as much rigour as possible to an issue that is not only academically challenging to study but also fraught with politics, accusations, and cherry-picking. With conceptual purity in mind, the choice is made here to only briefly discuss existing *causal* explanations, followed by a more typological examination of the issue. This aims to situate the case studies and the following analysis in the wider discussion. To achieve this, two preliminaries must be established.

First, one of the foundational principles of an academic enquiry into intelligence failure is to also consider its opposite, success. Although the former is usually more prominently discussed in public[10] and forms the raison d'être of the many theories of intelligence failure, it is crucial to consider its opposite, too, so to avoid 'selection on the dependent variable'.[11] If the dependent variable does not vary in a study at all, one cannot draw valid causal inferences. Although causal inference is not the aim of this article, both intelligence failure and success will be considered – which is arguably easier said than done, as will be explained below.

Second, the label intelligence failure or success ought not be understood as a judgement concerning the apparent causal source of failure or success – the intelligence organisation – but rather to describe failures and successes concerning subjects of intelligence such as threats to national security. In fact, the causal factors of failure are intentionally left out after the following brief discussion, since this would add a level of complexity of no value to the more bounded aim of this article.

To the outsider, it may seem intuitive to locate the source of an intelligence failure in intelligence: it appears axiomatic to say that intelligence reporting or intelligence organisations that fail to warn of a detrimental event are the cause of an intelligence failure.[12] A look at the wider machinery reveals, however, a more intricate reality.

To start with many stakeholders, most prominently decision-makers, frequently interact with intelligence organisations, each of them becoming a potential cause for failure. This is why many theories of intelligence failure suggest that the actual causes for failure are often located outside of the realm of intelligence, for example, at the policymaker level.[13] In this case, intelligence failure could be aptly reframed as a policy failure. A policymaker's receptivity to intelligence is a crucial prerequisite for countering surprise attacks.[14] Others have identified the communicative aspect of intelligence[15] or the missing setting of requirements and priorities at the mandate-level[16] to be a source of failure. Some advocate for a more holistic look at the interplay of the various elements of intelligence instead of concentrating on single parts of the process.[17] Consequently, intelligence failures are often found to be multi-causal rather than mono-causal.[18]

Expectations play a fundamental role in intelligence failure and success. Robert Jervis argues that an intelligence failure occurs when intelligence is 'falling short of what is expected from good intelligence'.[19] Mark Jensen qualifies the expectation further – and thereby inconspicuously identifying a culprit – by stating that intelligence failures are to some degree caused by the decision-makers' unrealistic expectations of intelligence.[20] The role of expectations and assumptions is also echoed by Stephen Marrin, who states that the term 'failure' entails many meanings, each reflecting 'an implicit assumption regarding the purpose of intelligence analysis'.[21] One of these expectations can be for intelligence to make accurate predictions about the future, that is, to match an estimate with the eventual outcome of events.[22] In this case, intelligence failure or success is rather easy to determine, even numerically.[23] On the other hand, or probably at the same time, one might expect intelligence to adhere to 'good' tradecraft and to fulfil its duties accordingly.[24] Such tradecraft could be defined by directives like the Intelligence Community Directive 203[25] published by the Office of the Director of National Intelligence that aims to establish compulsory analytical standards for the US intelligence community, or less proscriptively in the UK's Professional Development Framework for All-Source Intelligence Analysis.[26] It is often assumed that good tradecraft necessarily leads to accurate predictions. This is, however, far from clear. One might serendipitously predict the future accurately without any tradecraft or, conversely, adhere to 'good' tradecraft but fail to predict the future.[27] Distinguishing between the process *before* the event and the eventual accuracy of the forecast that can only be determined *after* the event is essential: conflating both becomes a fallacy.

Expectations might appear somewhat subjective, but this does not mean that failure or success is entirely a question of malleable perspectives and individual opinions. Expectations, however unrealistic and doomed to failure as they might appear,[28] are formulated by official statutes and requirements. Formal documents aside, one might argue that it is the subjective feeling of 'being surprised' that leads a decision-maker to proclaim an intelligence failure. The vague expectation of 'not being surprised' by an attack indicates a rather subjective expectation towards intelligence. If that was the case, very imaginative decision-makers would be less likely to be 'failed' by intelligence. Indeed, the psychological constitution of the individual decision-maker, who seldom considers themself to fail,[29] appears to be an important factor, less so the actual materialisation of the attack. In fact, whether the event to warn of materialises or not, does not seem to matter much: the warning of the Russian invasion into Ukraine in 2022, widely considered to be an intelligence success par excellence,[30] did happen, but did not catch policymakers by surprise. On the other hand, the Falklands war of 1982 is considered to be an intelligence failure as it caught policymakers by surprise, despite the – too late but accurate – warning prior to the invasion.[31] However, it is not so much the individual but the (individual) decision-maker making decisions on behalf of a nation, who should not be caught by surprise. Such decisions resulting in actions require a preparatory time: troops, aircraft and ships need time to deploy. In that sense, the decision-maker should not be surprised in such a way that the window for action is closed. For intelligence to be useful it must be actionable.[32] As Kahn has argued, 'surprise is a matter not of insufficient information but of insufficient time. Often, in looking back at the data available at the time of surprise, the indications of the event appear to have been present. But the analysts did not have enough time to understand them, to see a pattern in the mass of facts'.[33] This

becomes a trade-off. The more time an analyst takes to understand the data, the greater the certainty in her judgment but the shorter the window to act. Conversely, the quicker a judgment is made, the greater the uncertainty but the longer the decision window. This is further complicated if one considers the role of both adversary denial, and deception, in producing in the analyst's mind an incorrect understanding of the world.[34] It also suggests that the detection of change within an observed system over time is a difficult task.

Another, albeit implicit, prerequisite for intelligence to fail or to succeed is a reference point or period against which the assessment is measured. In the end, intelligence is less about stating the obvious baseline but more about warning of the deviations and suspected 'surprises'.[35] When such a deviation occurs, or does not occur, success or failure can be determined. If it does not occur, which essentially becomes a counterfactual situation without an exact point in time, the 'non-event' must be realistically imaginable and cognisable to measure the performance of intelligence. Otherwise, it would be questionable to consider every moment in which any attack does not occur, and intelligence (unwittingly) does not issue a warning, an intelligence success. One could then tautologically assert that 'intelligence is always successful until it isn't'. To be sure, if an attack is foiled due to a warning, it is justified to speak of an intelligence success. The latter phenomenon is commonly referred to as the 'paradox of warning'.[36] If, conversely, a warning was issued and the event was not actively prevented but simply did not occur for other reasons—"crying wolf", as the saying goes – this might be considered a failure. Crucially, both success and failure can only be determined after the fact.

Determining either intelligence success or failure implies that the value of this variable must be binary: either a success or a failure has occurred. This reduces the inherent complexity dramatically as does almost any binary variable in the social sciences. Nikki Ikani proposes a more nuanced approach by considering the multiple dimensions involved, like the analytical and political, and by using an ordinal scale per dimension.[37] This naturally increases the information space per observation, which in turn is likely to a lead to more accurate but also idiosyncratic diagnosis. Ironically, intelligence failure studies can learn from the more nuanced expression of uncertainty in intelligence analysis with the 'words of estimative probability' or the UK Probability Yardstick.[38]

This article, however, assumes the binary construct for the subsequent analysis. Its aim is not the perform a diagnostic test on an uncategorised case but to investigate whether a specific heuristic that is used for the binary categorisation – the First Law – is actually analytically diagnostic. Accordingly, multiple other paradigms of intelligence effectiveness are left out.[39] The result of this article cannot be, therefore, to confirm of refute an assigned binary label, but to shed light on one heuristic that is often used to arrive at this very label.

Signal in the noise and connecting the dots

Roberta Wohlstetter famously introduced the notion of the signal in the noise in her analysis of the attack on Pearl Harbor.[40] She argued that clues or fragments of information – signals – indicating an adversary's intentions were indeed recorded and circulated prior to the attack, but their significance was obscured by the abundance of routine or irrelevant information – noise.[41] Importantly, by available she

meant that such signals existed within the communication system, not that they were recognised or acted upon at the time.[42] This idea has often been challenged on the grounds of hindsight bias: what appear as obvious warnings after the fact are not necessarily discernible beforehand.[43] Yet, the continuing debate over the existence and detectability of signals lies at the heart of intelligence failure research. In this study, that question is revisited empirically, using modern computational tools to test whether, within a large corpus of diplomatic communication, distinct signals can be separated from background noise before key events.

A related concept – connecting the dots – became prominent following the 9/11 Commission Report.[44] It refers to the process of aggregating and linking separate pieces of information whose individual importance may be limited but which, in combination, could have formed a clear warning. The Commission attributed the failure to connect the dots to structural and organisational barriers within the US intelligence system, notably the compartmentalisation of information across agencies.[45]

Both notions rest on a shared assumption: that data indicative of an impending event exist ex ante, regardless of whether the episode is later judged an intelligence success or failure. Signals in the noise concern isolated, unusually informative messages; connecting the dots concerns the collective pattern that emerges from many such messages when viewed together. The two are therefore complementary-one focusing on discrete anomalies, the other on systemic shifts. Our approach allows both concepts to be tested systematically. Using natural language processing and time-sensitive semantic modelling, we examine whether (a) the frequency of unusual communications – potential signals – and (b) the structural changes in the overall information landscape – potential connections – differ between periods later characterised as intelligence failures and those regarded as relative successes.

From the standpoint of an analyst faced with continuous streams of reporting, the following hypotheses are proposed:

H1: The relative amount of signals in the noise – that is, atypical or semantically distinctive communications – is equivalent in periods preceding intelligence successes and failures.

H2: The number of indicative properties emerging from the connection of dots – captured through shifts in the semantic structure of the reporting stream – is equivalent across cases of intelligence success and failure.

If both hypotheses hold, the common assumption underpinning the 'First Law of Intelligence Failure' – that sufficient indicators are always available before the fact – would be supported. If they do not, the strong version of that law would be qualified: the availability of indicators may itself vary systematically across contexts.

The Soviet Invasion of Afghanistan and the fall of the Shah

Two cases will be examined: the fall of the Shah and the Soviet invasion into Afghanistan. In both cases, an extensive period prior to the key events, the period from 1973 to 1979, will be analysed. The Soviet Invasion as well as the Iranian Revolution are often considered intelligence failures. On a broader perspective, however, one could aptly argue that

intelligence has failed some time before 1979 but succeeded some time before that. The 'quiet' period of 1973 to the mid-1970s is the 'successful' control case.

The selection of these cases is primarily based on the different nature of the respective intelligence issues. The Iranian Revolution is essentially an intra-state affair with comparably little external influence. Moreover, it was not a proxy war within the Cold War. The Soviet invasion into Afghanistan, on the other hand, was an inter-state and ideologically motivated conflict, if not proxy war during the Cold War.

The Soviet Invasion

The full-scale Soviet invasion of Afghanistan on 24 December 1979 marks the beginning of a ten-year war whose ending correlated with the collapse of Soviet Union.[46] Occurring in the midst of the Cold War, Afghanistan was yet another place where a seemingly locally confined conflict turned out to be one of global scale and significance.[47]

The Soviets and the KGB had been active in the country long before the invasion, partly because of their ideological convictions but also for geo-strategic reasons – Afghanistan was located on their southern border.[48] Sadar Mohammed Daoud, himself a proponent of realpolitik and less of ideology, got into power with Soviet backing in 1973.[49] Support also came from the Marxist-Leninist People's Democratic Party of Afghanistan (PDPA) comprising two factions that opposed each other, namely Babrak Karmal's 'Parcham' and Noor Mohammed Taraki's 'Khalq' with his strong second-in-command Hafizullah Amin.[50] Despite an initially close cooperation with the Soviet Union, Daoud moved away from Moscow, which ultimately resulted in the Saur Revolution on 27–28 April 1978 leaving Daoud dead and Taraki now in power.[51] Shortly afterwards, the Soviets increased their presence in Afghanistan in the form of military advisory,[52] but the situation remained far from stable: over the next months, Khalq, under the direction of Amin, eliminated the influence of Parcham within the governmental apparatus while public resistance to socialist reforms simultaneously erupted on the countryside.[53]

Further military cooperation with the USSR notwithstanding, the public discontent – or, rather, insurgency – kept growing and reached its peaks in the Herat Uprising on 15 March 1979, where up to twenty Soviet military advisors were killed.[54] Significantly, parts of the Afghan army defected to the insurgency during the uprising, marking a starting point for an increased Soviet military activity in and around Afghanistan. The US was also directly affected when US ambassador Adolph Dubs was abducted and killed by an extremist anti-government cell on 14 February 1979.[55] During the summer of 1979, the situation continued to deteriorate with the Soviets seeking an alternative to the team of Taraki and Anim. However, in September another power struggle between the two erupted in which Taraki was killed and Anim became the ostensibly uncontested president, much to the dissatisfaction of the Soviets.[56] The next months witnessed yet another increase in Soviet activity, including the deployment of special forces, ultimately resulting in the full-scale invasion comprising more than 30,000 troops on Christmas Eve 1979.[57] Within the first days of the invasion, Anim was killed and Karmal became president.

As shown above, the potent, sometimes accusatory, label of 'intelligence failure' can be problematic for many reasons. Rarely does a single label do justice to a complex reality. And indeed, whether this case should be considered an intelligence failure or success is

far from obvious. Robert Gates, for instance, argues that 'if ever there was a crisis foreseen well in advance it was the gradual but unmistakable growing Soviet involvement in Afghanistan'.[58] Douglas MacEachin, on the other hand, asserts that the policymakers' surprise over size and scope of the invasion makes this instance an intelligence failure.[59] The crucial point of the debate, which Gates does acknowledge, was the (later proved to be inaccurate) judgement that the Soviet would intervene gradually and not on a full scale.[60] At the same time, former DCI Stansfield Turner noted that Afghanistan was 'not very high on the American foreign policy agenda . . . We had lots of other things that were of much greater concern to us . . . ' suggesting low receptivity or attention to the topic.[61]

The Fall of the Shah

The Iranian Revolution refers to the overthrow of the then-ruling Shah and the subsequent establishment of the Islamic Republic of Iran under Ruhollah Khomeini. After a brief pause, Mohammed Reza Shah Pahlavi had been ruling Iran since 1953 following a US-/UK-led operation to restore him to power at the expense of ousting Mohammad Mossadegh.[62] Over the years, he had become a close ally to the West in general and the US in particular, introduced domestic reforms like the 'White Revolution' to economically advance the country, but also had the powerful secret service SAVAK at his disposal.[63]

The timeline of the events leading up the revolution is generally uncontested.[64] The first signs of public discontent emerged in the summer of 1977 at the latest, followed by intensified, deadly protests in January 1978 in the Iranian city of Qom.[65] These protests correlated with increased circulation of publications by the then-exiled Khomeini whose return to Iran was demanded by protesters.[66] After a summer full of protests, the Shah eventually declared martial law on 8 September 1978 and formed a military government on 4 November 1978. Amid a worsening and seemingly uncontrollable situation, the Shah left Iran for Egypt on 16 January 1979, only shortly before Khomeini entered Iran from his exile. On 10 February 1979 the military declared its neutrality, ultimately paving the way for a new government under Khomeini which ended the Pahlavian era.[67]

Causes of revolutions are always challenging to identify. This discontent with the Shah's ruling had multiple causes, some of which were social insecurities like food and job shortages amid a growing population, while the regime's reaction was marked by brutality and perceived arrogance.[68] The clerics, or mullahs, conversely appeared to be able to satisfy these needs which Khomeini readily leveraged.[69] Abdalla notes, however, that the revolution must not be understood as a purely Islamic movement but was also socially motivated.[70]

In contrast to intelligence debates and predictions about the Soviet intervention in Afghanistan, little warning was provided by the US Intelligence Community about the coming revolution in Iran. The American estimates throughout 1977 saw no threat to the Shah's regime. The State Department assessed in May that year that 'the Shah is in a stronger position internally than at any previous time in his long rule (26 years)', and the CIA estimated a few months later that 'there will be no radical change in Iranian political behaviour in the near future'.[71] As late as the fall of 1978, UN intelligence was predicting political continuity in Iran.[72] On 9 November 1978, Ambassador Sullivan sent the notable 'Thinking the Unthinkable' telegram which noted the authority of the Shah had 'considerably shrunk', discussed the potential for his ouster, and noted the potential

role of the Ayatollah Khomeini in any potential successor government. By this point, however, the decision-makers' attention or capability was insufficient for the US Government to act upon these reports.[73] After the fall of the Shah materialised in 1979 to the surprise of many, the overall case is widely considered a failure, single reports notwithstanding.

Research design

Seemingly little has been written about research design in intelligence studies. And yet, the scientific study of secret intelligence faces an inherent and arguably unique paradox, as Michael Warner pointed out: 'Intelligence (. . .) by definition resists scholarship'.[74] This tension is academically intriguing and makes it challenging to arrive at reliable and valid conclusions – and therefore necessitates a thorough consideration of methodology and research design even more so to ensure reproducibility.[75] One reason for the dearth of a standardised, adapted methodology might be the interdisciplinary nature of the field, comprising history, political science, sociology, psychology, and insights from practitioners.[76] While the literature on the question of 'theory' about or of intelligence is ample and subject to debate,[77] specific procedures and caveats on how to arrive at valid inferences in intelligence studies are rarely discussed. An exemption are remarks by historians to critically examine the provenance and credibility of the available sources, which should be treated with extra caution in intelligence due to its secret nature and political context.[78] But the historical method is not a substitute for a research design in the social sciences.[79] Such a research design comprises the explicit formulation of a research question, the consideration of a theory to be tested, the discussion of data, and its use to ensure replicable and valid inferences.[80] This paper tries to provide such a research design.

Two peculiarities of intelligence as a field of social enquiry must be considered when making inferences. First, secret intelligence, conceptualised as knowledge written down in files,[81] is classified by default and thus not available to research.[82] Declassification, especially of more recent documents, is the exception; when declassified, data have usually been subjected to rigorous redaction. And it is the most sensitive – and arguably academically intriguing – data that are redacted and remain hidden for the public for an extended period. In that sense, intelligence studies usually need to account for data that are 'not missing at random'.

Second, the veracity of the data must be examined with exceptional rigour. Concepts like the active denial of information to the opponent – the primary reason for redaction – and deception, that is, 'the effort to cause an adversary to believe something that is not true',[83] are inherent properties of intelligence.[84] For example, a primary source reporting from a human source containing ostensibly high-quality information can be highly skewed if an adversary controls the source and uses it for means of deception.[85] On the other hand, finished intelligence products aim to be as objective as possible, although issues like politicisation can severely impede the objectivity.[86] In effect, the veracity of data in the field of intelligence must be carefully examined, potentially more than in any other field of social enquiry.

Whilst the points outlined above pertain to the general nature of the intelligence studies, additional methodological challenges for the enquiry into intelligence failure and success exist. Importantly, as echoed by Marrin and others, intelligence success are more likely to remain classified than failures.[87] Furthermore, if on a more general note, hindsight bias, a sort of sword of Damocles to the researcher of intelligence failure and success, makes it easy to cherry-pick information that in retrospect appears to have been an unambiguous 'prediction' or, more accurately, an 'indicator'[88] of the event.

Data and methodology

This paper analyses diplomatic traffic with machine-learning techniques. Diplomatic traffic as a dataset is common in studies on intelligence failure and is often analysed qualitatively.[89] The quantitative approach taken here, using 'text as data', is also not unprecedented in the field. Connelly et al. examined how analysts and policymakers preformed in the period leading up to the Iranian revolution by quantitively analysing a dataset of diplomatic traffic, known as the 'Central Foreign Policy Files'.[90] Methodologically, the authors relied on metadata of the corpus, tags assigned to each text as well as keyword-based filters and aggregations.[91] These data formed input into time-series models that allowed for statistical hypothesis testing.

While this paper was inspired by the approach Connelly et al. took and builds upon the same type of data – diplomatic traffic – it employs another technique, namely machine-learning. This differs from keyword-based methods in that it relies on models that have been pretrained on text before, which enables them to capture nuances in the text better than traditional keyword-based methods.[92] Also, no statistical hypotheses test will be conducted, the reasons for which will be explained below. On a broader note, the aim of this paper is different: while Connelly et al. further interrogate the single case of the Iranian revolution, this paper examines overarching patterns of two cases of intelligence failures and success.

The dataset consists of declassified documents by the US State Department, namely the 'Central Foreign Policy Files' (abbreviated as CFPF) and is not secret intelligence in the classical sense.[93] Rather obviously, it stems from diplomatic missions and not from secret intelligence organisations and thus might not be considered applicable to an investigation into intelligence failures and successes. However, while both governmental functions are different, their outputs are not disjunct. They overlap, as diplomatic reporting can be one information source for intelligence analysts.[94] In terms of the quality of information, Connelly et al. argue that diplomatic reporting was even superior to that of secret intelligence in the Iranian case.[95] Nonetheless, to adapt the perspective of an analyst, cables are only one type of 'raw' information that need to be further vetted and processed for them to become intelligence. Processing here can mean trying to detect the signal in the noise or to connect the dots – which makes the data suitable for the research question.

The files were made available through the Freedom of Information Archive in a machine-readable and enriched format.[96] The CFPF comprises communication, 'cables' in diplomatic parlance, between and among US diplomatic posts and the State Department headquarters from 1973 to 1979. The dataset does not include every cable but only those deemed important by an archivist. Communication

pertaining to visa and passport issues, for example, is routinely removed as is communication between the State Department and other government agencies.[97] Notably for this article, declassified documents up to 'SECRET' are included, although some parts remain redacted.[98] The dataset also contains cables concerning important non-security issues like trade. Only CFPF-cables relating to 'Afghanistan' and 'Iran' were collected.[99] For a cable to be related to a country, it can be connected in various ways: it might be linked to the country because it was sent from the respective embassy to Washington DC or because the country's name appears explicitly in the cable. Duplicated cables and those containing the phrase 'error reading text', 'expand error encountered', 'text for this segment is unavailable', 'telegram text for this mrn is unavailable', 'no title', 'n/a', 'w/w', or 'msg' were removed. Also, cables without any letters but only numbers were excluded. Finally, the data were cut off on 24 December 1979 in the case of Afghanistan and 11 February 1979 in the case of Iran.

Prior to the application of the machine learning algorithms, a preparatory pre-processing step is performed. The CFPF-messages are all uppercased which are converted to lowercase.[100]

This article uses the concept of so-called embeddings to analyse the content of the cables quantitatively. These embeddings are created by Voyage AI's model 'voyage-large -2-instruct'.[101] Embeddings are representations of information such as sentences that capture its semantic meaning.[102] These representations are usually vectors of numerical values in a high-dimensional space. Texts with similar meanings are located in proximity to one another while dissimilar texts are farer apart. This proximity reflects semantic similarity rather than exact word matching. For example, the embeddings for 'This King is celebrating!' and 'It's the Queen's birthday today'. would occupy nearby positions, whereas 'This King is celebrating!' and 'I like bananas'. would be distant.[103]

Based on this concept, 1024-dimensional embedding vectors are created per cable, irrespective of the length of the actual cable.[104]

The first algorithm, denoted as A1, pertains to H1 ('signal in the noise'). The building block for this are so-called outliers that represent the signal in the noise. Outliers are data points, cables in this instance, that deviate significantly from all the other data points. If, for instance, most of the cables are about the prospering economy but one cable reports on the growing military threat by an adversary, then this cable will be considered an outlier. More precisely, the tone, words and context in language is different which can be captured by the embeddings. This outlying cable becomes the signal in the noise.

To retrieve the outliers, some computational steps are necessary.[105]

(1) First, the data are segmented into overlapping 90-day time windows. The overlap is 14 days. These time windows present a temporal frame for detecting outliers. If the entire time series over the years were considered instead, local outliers would be less likely to appear, even though they might have been important at that time.
(2) Per time window, the dimensionality of the standardised embeddings is reduced from 1024 to 5 dimensions using the UMAP-Algorithm.[106] The reason for this is to eschew the curse-of-dimensionality, whereafter in higher dimensional spaces, the distance between the most distant data points becomes negligible in relation to the closest distance.[107] A reduction in dimensions inevitably reduces information and is dependent on the specific algorithm and its parameters. UMAP's parameter

n_neighbors determines the local structure of the reduced embedding space. To provide a wide range of results, different values for n_neighbors, namely 5, 10, and 15, are chosen.

(3) Lastly, the cluster algorithm DBSCAN is used per time window on the 5-dimensional embedding to detect outliers.[108] The outlier detection is also sensitive to the specific parameters. DBSCAN's epsilon parameter affects the number of outliers. Here, too, different values are chosen: 0.0001, 0.0005 and 0.001.

The second algorithm, denoted as A2, concerns H2 ('connecting the dots'). The basis are also the 1024-dimensional embeddings per cable. This time, the dots are connected by taking the average and dispersion of a set of embeddings. The resulting averages and the dispersions are subsequently contrasted with others. If the right dots are connected, or averaged, then it is assumed that a sudden insight emerges. For instance, if cables only reported on the military capabilities of an adversary and a report on the intentions arrives, their connection will then reveal the whole threat. These shifts are conceptualised here as so-called change-points. In time series, change points are points in time where statistical properties of the data abruptly change in comparison with the previous properties. In contrast to A1, which considers outlying characteristics of individual data points, A2 examines properties that emerge from the connection of cables.

Three steps are consecutively taken to detect properties emerging from the 'connection the dots':

(1) Time windows of 30-days without overlapping are segmented. Overlapping is less needed in this case, as opposed to A1, where outlying characteristics heavily depend on the adjacent local data points.

(2) For every time window, the average, called centroid, is calculated. The centroid captures the theoretical mean of the embeddings. It is then reduced to 5 dimensions like in step 2 of A1 and standardised. Also, the dispersion is calculated per time window. The dispersion accounts for shifts in the semantical range that would be ignored by the mean alone.

(3) Two change point detections are performed: one on the dispersion and another on the centroid. The resulting change points from each detection are then combined using a logical OR operation. This means that it must take at least one change point at a point in time to designate it an overall change point. The change point detection relies on the linear penalised segmentation method ('Pelt').[109] Importantly, this method does not require an ex-ante specification of the number of change points to be detected. The number of change points hinges upon the value for pen, for which 3, 10 and 20 were chosen. For visualisation purposes, the five dimensions of the centroids are reduced to 1.

Evaluating both variants with respect to the hypotheses is not trivial. The performance of an outlier or change point detection depends on the threshold or parameters set by the researcher. This in turn influences the number of 'false positives' and 'false negatives', respectively. In other words, if one lowers the threshold of a data point to be classified as an outlier, the chance of missing 'true' points decreases while the chance of capturing 'false' points increases. This is

particularly relevant in intelligence analysis.[110] If an analyst wants to minimise the probability of missing a relevant information and lowers the 'threshold', she will be inundated by irrelevant data. If, on the other hand, she wants to avoid the flood of information and raises the threshold, the probability of missing relevant information increases. This understanding is especially relevant in today's environment, in which digital data are abundant.

This problem is well known in machine learning and can be addressed by using dedicated performance metrics like the F1-score as a harmonic mean to gauge a model's performance.[111] At the same time, the problem of hindsight bias restricts an ex post classification of the cables as 'to-be-hit' or 'to-be-omitted', especially when studying intelligence failures and successes. Nevertheless, this would be necessary to calculate evaluation metrics such as the F1-score. The same problem pertains to the change point detection. The number of change points can be readily influenced by the parameters and features provided by the researcher. An ex-post determination of change points to be detected would also carry the risk of hindsight bias.

The variability of the thresholds and parameters can present a risk to scientific rigour since it would allow to search for parameters consistent with the hypothesis. One countermeasure taken here is to consider multiple instances of the parameters. The following methods are used to assess the consistency with the hypothesis:

(1) The number of outliers and change points over time, emitted by A1 and A2, is displayed and visually inspected. A diverse set of parameters is used to account for the algorithms' sensitivity to these parameters. This reduces the probability of a serendipitous choice of parameters that (dis)confirm the hypotheses.
(2) For A1, the corresponding cables of some outliers that were hit on multiple occasions with different parameters are manually inspected in order to cursorily validate the overall approach.

Results

The cables are expressed as number of cables per week (Figures 1 and 2) and as descriptive statistics in Table 1.

Signal in the noise

This section reports the results for H1, which examines whether the relative number of signals – here defined as semantic outliers – differs between periods of intelligence failure and periods of relative success. Outliers are cables that deviate markedly in content from their temporal surroundings, while the remaining 'inlier' cables form the informational noise. In both corpora, the number of outliers varies over time rather than clustering exclusively before the respective crises. Outlier detection was conducted across nine parameter combinations using two tuning variables: the n_neighbours parameter, which shapes the local structure of the embedding space, and epsilon, which controls the sensitivity of the density threshold. To ensure

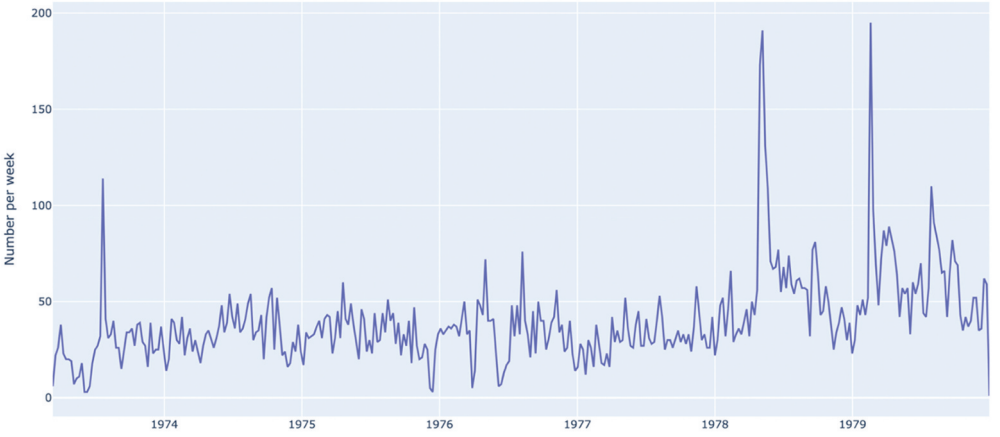

Figure 1. Number of Afghan cables per week.

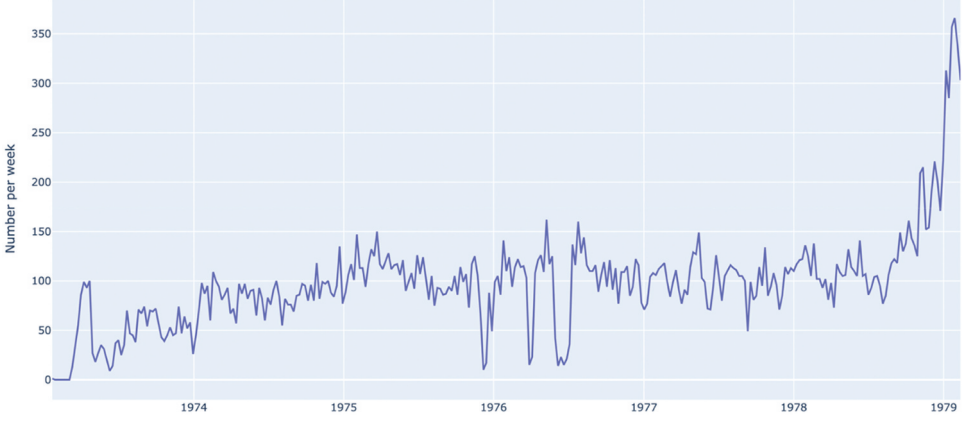

Figure 2. Number of Iranian cables per week.

Table 1. Descriptive statistics.

	Afghanistan	Iran
Cables collected	15684	45385
Cables excluded	1482	14059
Cables per day	5.72	13.93
Average number of characters	1351.43	1152.73
Average number of tokens (lowercase)	369.35	322.11

robustness, only cables identified as outliers in at least three of the nine parameter settings were retained. Figures 3 and 4 display the results for n_neighbours = 10, while Figures 5 and 6 show the cumulative distribution of outliers across all parameter combinations.

In both corpora, the number of outliers varies over time rather than clustering exclusively before the respective crises. In the Afghanistan case, several prominent outliers

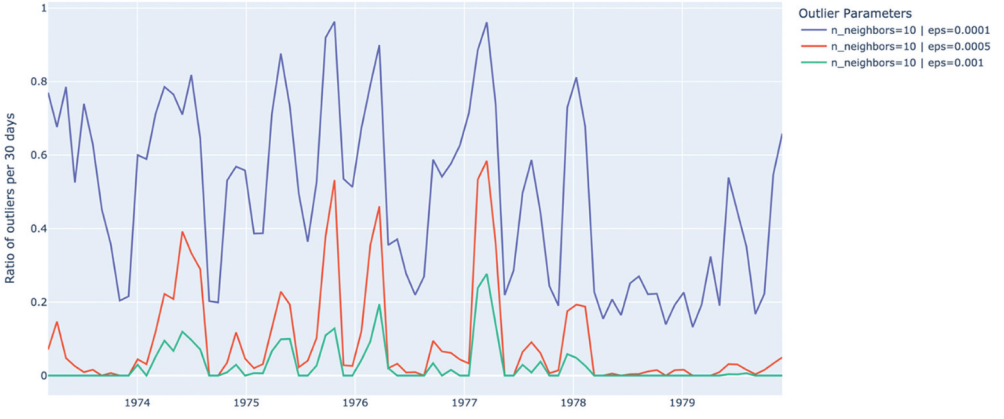

Figure 3. Ratio of outliers in Afghan cables.

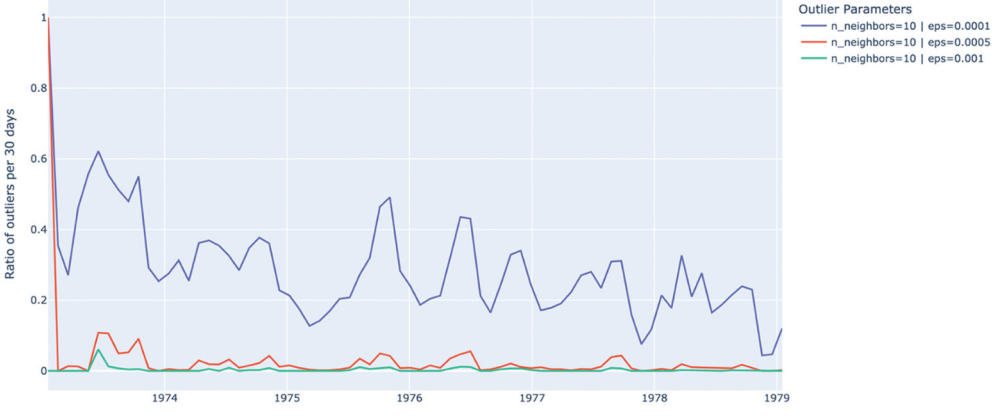

Figure 4. Ratio of outliers in Iranian cables.

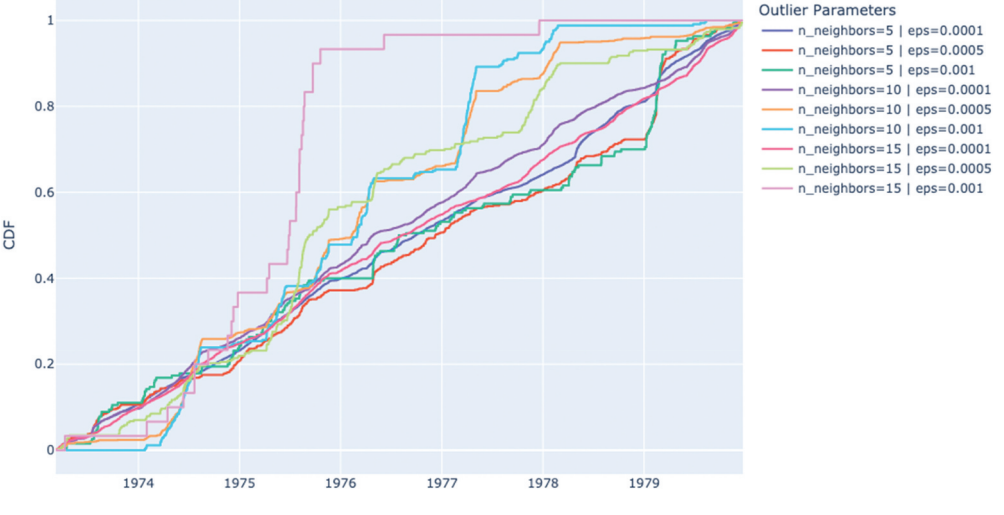

Figure 5. Cumulative distribution function for Afghan outliers over time.

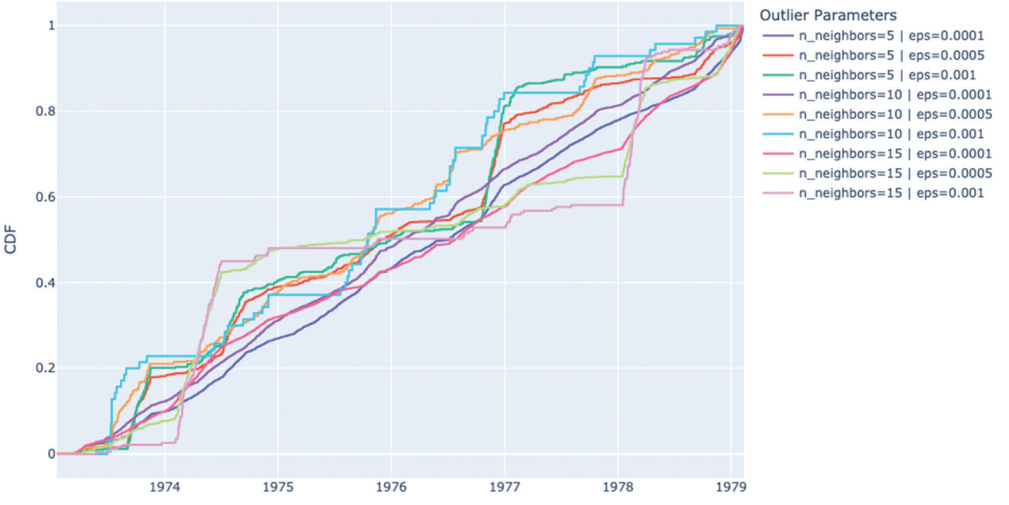

Figure 6. Cumulative distribution function for Iranian outliers over time.

occur in 1979, such as cables expressing doubts about Moscow's intentions[112] or reporting exaggerated claims of Soviet troop presence.[113] Other outliers from earlier years—1974 and 1975 in particular – flag internal instability under Daoud and the deepening Soviet influence over the Afghan military.[114] These earlier messages did not lead to explicit warnings, yet with hindsight they appear as valid indicators of vulnerability. Because the predicted regime collapse did not occur in the mid-1970s, intelligence restraint in issuing alarm can reasonably be interpreted as a success. The later surge of outliers close to 1979, many of which reflected optimistic assessments of Soviet restraint, also highlights the inherent difficulty of judging 'signals' ex ante. What appears as misplaced optimism now might, under slightly different historical outcomes, have been considered prudent interpretation.

A similar pattern appears in the Iranian case. Outliers before 1979 include striking examples – such as a 1978 cable describing reports of Khomeini acquiring arms from Libya,[115] or another noting the Embassy's denial of a National Security Council study advising the Shah's removal.[116] These seem, with hindsight, to prefigure the revolution. Yet other outliers from 1973 and 1976 discuss student protests and labour unrest that did not develop into systemic challenges.[117] At the time, treating such events as local disturbances rather than existential threats was analytically defensible. In that sense, periods of 'non-warning' can constitute intelligence successes: they reflect appropriate filtering of transient noise.

Overall, the data do not support H1. The quantity of semantic outliers – the potential 'signals in the noise' – was not equal across success and failure periods. Outliers were numerous not only before the crises but also during earlier, comparatively stable years. Remarkably, they were even slightly more frequent during the first half of period. The apparent increase near 1979 likely reflects a shifting informational baseline rather than the emergence of uniquely diagnostic indicators. From the standpoint of contemporary analysts, the signals available before failure were not obviously stronger or more distinctive than those during times when no major event occurred. Thus, the results suggest

that the 'signal-to-noise' ratio was roughly constant across time, meaning that intelligence personnel were faced with a steady flow of ambiguous information rather than a sudden surge of clear warnings.

Connecting the dots

The second hypothesis (H2) concerns the collective structure of communication rather than individual anomalies. It tests whether the degree of systemic change in the discourse – the extent to which 'dots' can be connected – differs between periods of intelligence success and failure. This 'connection of dots' is operationalised through monthly estimates of the centroid (the average semantic position of all cables) and the dispersion (the spread of those positions). Abrupt shifts in either quantity represent change-points, interpreted as emerging or dissolving themes in diplomatic reporting. The penalty parameter (pen) controls how many change-points are detected; smaller values capture more fine-grained variation. Figures 7 and 8 display the results for pen = 10; additional values (pen = 3 and 20) were used to check robustness.

In the Afghanistan data, a distinct change-point appears between late 1976 and early 1977 under all parameter settings, followed by another roughly a year later. The number of change-points increases toward 1979, suggesting growing thematic volatility as relations with Moscow deteriorated. In contrast, the period from 1973 to 1975—interpreted here as a relative success – shows fewer changes, consistent with a more stable diplomatic focus.

The Iranian corpus exhibits a broadly similar but less pronounced pattern. Between four and eight change-points are detected depending on the penalty value, with noticeable shifts beginning in late 1977. Dispersion remains comparatively stable, implying that reporting became more concentrated on a limited set of political issues as unrest intensified. The overall level of variation before 1978 is low, indicating a coherent narrative rather than fragmented or contradictory signals.

Taken together, these findings provide only limited support for H2. The number of change-points increases as the crises approach, but the magnitude of those shifts is not

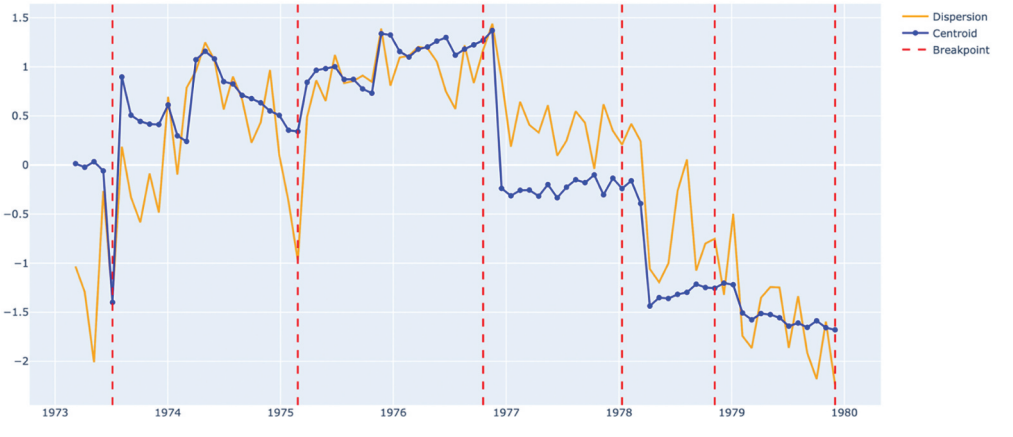

Figure 7. Change points in Afghan cables.

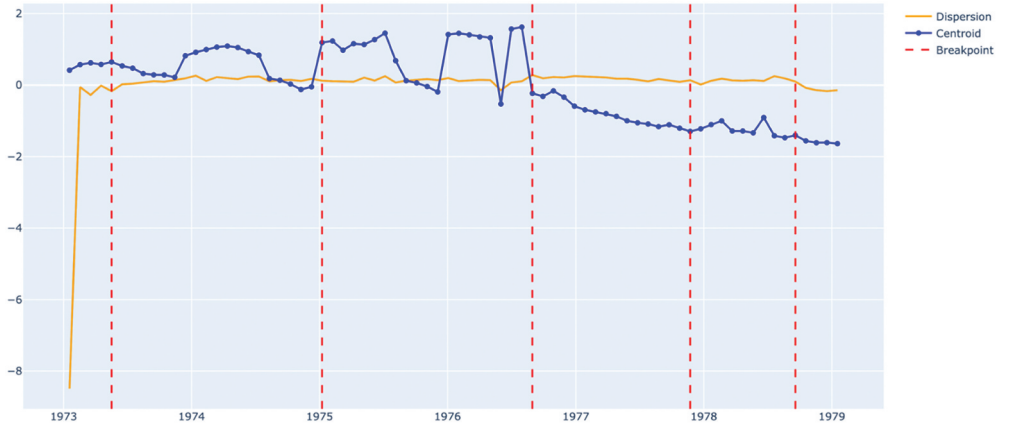

Figure 8. Change points in Iranian cables.

exceptional compared with earlier fluctuations. In both cases, the information landscape appears to evolve gradually rather than undergo dramatic transformation. If analysts were to 'connect the dots', they would not have encountered a sudden explosion of new or divergent themes that clearly distinguished failure from stability. The Afghan data show slightly greater volatility, consistent with the complex and shifting Soviet decision-making in late 1979; the Iranian data show steadier discourse and fewer identifiable discontinuities.

Interpretation and limitations

The analyses of H1 and H2 must be interpreted cautiously. The classification of periods into 'success' and 'failure' is necessarily approximate. In neither case can a clear temporal boundary be established at which success ends and failure begins: the Soviet decision to invade Afghanistan was taken only in December 1979, while American awareness of the Shah's fragility increased gradually throughout 1978. Classical hypothesis tests would therefore risk false precision and are not used here.

Several limitations temper these conclusions. First, the dataset composition is broad: the CFPF includes diplomatic reporting on trade, culture, and administration as well as security issues. This heterogeneity likely inflates the number of false positives – for example, one strong outlier concerns the trade of dried apricots rather than political unrest.[118] Filtering by topic might reduce such noise in future work. Second, the data stream analysed here represents only one channel of information – the State Department's diplomatic communications – whereas intelligence analysts would have consulted multiple sources, including military and classified reporting. The relative strength of signals can change when evidence from several streams is aggregated. Third, the algorithms themselves are deliberately simple. They model semantic similarity and abrupt change but cannot assess intent or importance as human analysts would. Algorithm A2, in particular, provides a rudimentary proxy for 'connecting the dots' and does not distinguish between tactical and strategic indicators, a distinction central to Dahl's Theory of Preventive Action.[119]

Despite these caveats, an interesting pattern emerges. When searching for signals in the noise (H1), the ratio of indicative data is lower during failure than success. When examining the connection of dots (H2), the ratio of indicative change-points is higher during failure than success. This apparent paradox suggests that the assumption underlying the 'First Law of Intelligence Failure' – that sufficient indicators are always available before the fact – does not hold uniformly. In these two cases, the informational environment was dense and noisy throughout, but the structure of that noise, rather than its volume, varied over time. The findings therefore call for a more nuanced understanding of how signals emerge, interact, and are recognised within complex information systems.

Taken together, the two analyses show that patterns of diplomatic communication before major political shocks do not clearly distinguish periods of intelligence failure from those of relative success. Both crises were preceded by a steady flow of ambiguous information rather than a sudden intensification of meaningful signals or a radical shift in discourse structure. The apparent constancy of the informational environment underscores the practical challenge for analysts operating under uncertainty: even with hindsight, separating genuine warning signs from background noise is far from straightforward. These findings invite a reconsideration of how the 'First Law of Intelligence Failure' is framed – not as a universal truth about the perpetual availability of indicators, but as a conditional statement dependent on context, data streams, and the analytic frameworks applied. The following section discusses what these results imply for theories and studies of intelligence failure and for the potential – and limits – of applying computational methods to historical intelligence questions.

Conclusion and outlook

Some excerpts from the mid-1970s cables examined were indicative 'signals' of possible but eventually non-materialised events. Intelligence did not act upon them, and rightly so. Intelligence appears to have succeeded by neglecting the wrong signals or by not connecting the wrong dots. To succeed, the right signals and right dots need to be considered. Therefore, the premise of 'First Law of Intelligence Failure' could be specified so that the right signals and the right dots are available before the surprise. This normative judgement is problematic for at least two reasons.

First, the judgement as to whether a data point is wrong or right can only be made after the fact. This might seem obvious and scholars like Wohlstetter have acknowledged that it is 'much easier after the event to sort the relevant from the irrelevant signals'.[120] The 9/11-commission also introduced a chapter by stating that they have 'tried to remember that [they] write with the benefit and the handicap of hindsight. Hindsight can sometimes see the past clearly – with 20/20 vision'.[121] Yet both studies heavily rely on an ex post view – it appears that the acknowledgment of hindsight is more a lip service. Also, it is questionable whether the remedy proposed by Dahl – requiring the dependent variable to vary – is sufficient to address this issue.[122] What hindsight really urges is a fundamental reconsideration of methodology.

A possible alternative approach could be a prospective study or experiment.[123] If one group classified the data as 'signals' or calculated properties from 'connecting the dots' ex ante and in secret, while another control group based their judgements on the data without these explicit concepts, it could be investigated whether adherence to 'signals' or

'connected dots' makes a real difference. Whether this approach is feasible at all remains to be seen, but it would open new methodological frontiers.

Second, as elaborated above, the label of 'failure' and 'success' is too imprecise. To set off a scientific investigation into the reasons why intelligence failed or succeeded and thereby accepting the presupposition that it failed or succeeded, might not be so helpful after all. In particular, the nature of successful intelligence is anything but clear-cut.[124] This demands a precise definition of the expectations of intelligence, including what is not or cannot be expected from it and what deviations from these expectations might look like. Deviations essentially require the formulation of a baseline. Studying baselines and deviations, both positive and negative and in varying degrees, might offer more insight into the performance of intelligence than adopting the dichotomy of failure and success.

Focusing on baselines and deviations of systems might not only be useful for the study of intelligence failure and success but also for practical intelligence analysis. The maths used here are complex, certainly, but one does not need to fully understand that maths to see the value that their application can generate. And the value here goes beyond historical analysis. The research shows how data analytics can help detect change within systems over time. While not a process with any predictive value, the methods used here may have value in the analysis of reporting of stable systems over time, offering insights into that system and perhaps cueing analysts to the degree of change they may not have noticed. Like all analytical tools this does not offer a 'silver bullet', but another degree of insight which may bring marginal advantage.

This article is neither a prospective study nor does it account for the blurry, nonbinary lines of failure and success. In this sense, this approach cannot be used to solve the inherent problems presented above. Notwithstanding, the paradoxical and difficult-to-explain results in relation to the 'First Law' suggest that the entire approach to this problem needs a further and thorough reconsideration. Otherwise, to further Richard Betts' argument, failures to study intelligence failures might be inevitable.

Notes

1. Wohlstetter, *Pearl Harbor*; and National Commission on Terrorist Attacks Upon the United States, *The 9/11Commission Report*.
2. Wirtz, "Responding to Surprise," 51.
3. Dahl, *Intelligence and Surprise Attack*, 8–13.
4. Giarelis and Karacapilidis. "Deep Learning and Embeddings-based Approaches." See Macanovic, "Text Mining for Social Science," 7.
5. Connelly et al., "History Lab"; and Connelly et al., "Diplomatic Documents Data."
6. Zegart, *Spies, Lies, and Algorithms*, 139–41.
7. Hayden, "A 'Reasonable' View".
8. Kuhns, "Intelligence Failures," 80.
9. Dahl, *Intelligence and Surprise Attack*, 9–14.
10. Herman, *Intelligence Power*, 224.
11. Dahl, *Intelligence and Surprise Attack*, 14–6.
12. Wirtz, "Sources and Methods," 65.
13. Betts, *Surprise Attack*, 288–9.
14. Dahl, *Intelligence and Surprise Attack*, 22–6.
15. Hatlebrekke and Smith, "Towards a New Theory."
16. Abdalla and Davies, "Intelligence, Policy, and the Mandate," 2021.

17. Wirtz, "Are Intelligence Failures Still Inevitable," 326.
18. Gill and Phythian, *Intelligence in an Insecure World*, 153.
19. Jervis, *Why Intelligence Fails*, 2–3.
20. Jensen, "Intelligence Failures," 270.
21. Marrin, "Evaluating the Quality," 896.
22. Jervis, *Why Intelligence Fails*, 2; and Wirtz, "Sources and Methods," 65.
23. Tetlock and Gardner, *Superforecasting*, 52–64.
24. Jervis, *Why Intelligence Fails*, 2.
25. Office of the Director of National Intelligence, *ICD-203 Analytic Standards*.
26. UK Cabinet Office, *Professional Development Framework*.
27. Jervis, "Why Postmortems Fail."
28. Betts, "Analysis, War, and Decision."
29. Jervis, *Why Intelligence Fails*, 157.
30. Gustafson et al., "Intelligence Warning," 401–7.
31. Herman, *Intelligence Power*, 225.
32. Shulsky and Schmitt, *Silent Warfare*, 63; and Dahl, *Intelligence and Surprise Attack*, 21–4.
33. Kahn, "Surprise and Secrecy," 1060.
34. Gustafson, "Confidence Game."
35. Betts, *Enemies of Intelligence*, 191–2.
36. Dahl, *Intelligence and Surprise Attack*, 15.
37. Ikani, "Beyond the Binary."
38. Kent, "Words of Estimative Probability"; and UK Cabinet Office, *Professional Development Framework*, 29.
39. Manger, "Unravelling Effectiveness."
40. Wohlstetter, *Pearl Harbor*.
41. Ibid., 1–2.
42. Ibid., 73.
43. Handel, "The Yom Kippur War," 467; and Levite, *Intelligence and Strategic Surprises*, 34.
44. National Commission on Terrorist Attacks Upon the United States, *The 9/11Commission Report*, 408.
45. Davies, "Intelligence Culture," 507–8.
46. Walton, *Challenges in Intelligence Analysis*, 159–60.
47. Gates, *From the Shadows*, 143– 9; and MacEachin, *Predicting the Soviet Invasion*, 3–4.
48. Walton, *Challenges in Intelligence Analysis*, 157.
49. MacEachin, *Predicting the Soviet Invasion*, 3.
50. Ibid., 3–4.
51. Ibid., 6–9; Central Intelligence Agency, *Declassified Interagency Intelligence Memorandum*, 7.
52. Central Intelligence Agency, *Declassified Interagency Intelligence Memorandum*, 7.
53. MacEachin, *Predicting the Soviet Invasion*, 10–1.
54. Ibid., 12–3; and Central Intelligence Agency, *Declassified Interagency Intelligence Memorandum*, 7.
55. MacEachin, *Predicting the Soviet Invasion*, 12.
56. Lyakhovsky, *Inside the Soviet Invasion*, 5–7.
57. Ibid., 44–71.
58. Gates, *From the Shadows*, 131.
59. MacEachin, *Predicting the Soviet Invasion*, 2.
60. Gates, *From the Shadows*, 134; and MacEachin, *Predicting the Soviet Invasion*, 21–23' 40–2.
61. Lyakhovsky, *Inside the Soviet Invasion*, 5.
62. Byrne, "CIA Confirms Role."
63. Jervis, *Why Intelligence Fails*, 29–30; and Walton, *Challenges in Intelligence Analysis*, 183–4.
64. Connelly et al., "New Evidence," 784.
65. Kurzman, *The Unthinkable Revolution*, 18, 33–34.
66. See note 64 above.
67. Kurzman, *The Unthinkable Revolution*, 161–163.

68. Walton, *Challenges in Intelligence Analysis*, 184–5.
69. Ibid., 185.
70. Abdalla, "Requirements, Priorities, and Mandates," 120.
71. Bar-Joseph, "Forecasting a Hurricane," 727.
72. Connelly et al., "New Evidence," 785.
73. Ibid., 799–800.
74. Warner, "Sources and Methods," 17.
75. King, Keohane and Verba, *Designing Social Inquiry*, 25.
76. Robertson, "Agenda for Intelligence Research"; and Scott and Jackson, "The Study of Intelligence."
77. See Davies, "Theory and Intelligence Reconsidered"; and Gill and Phythian, *Intelligence in an Insecure World*, 27–44.
78. Scott and Jackson, "The Study of Intelligence," 145–6; and Warner, "Sources and Methods," 17.
79. King, Keohane and Verba, *Designing Social Inquiry*, 36–41.
80. Ibid., 6–8, 12.
81. Kent, *Strategic Intelligence*, xxiii.
82. Herman, *Intelligence Power*, 91.
83. Godson & Wirtz, "Strategic Denial and Deception," 425.
84. Herman, *Intelligence Power*, 118; and Gustafson, "Confidence Game," 152.
85. Herman, *Intelligence Power*, 65.
86. Davis, "Intelligence analysts and policymakers," 1000–1002.
87. Marrin, "Preventing Intelligence Failures," 661; Betts, *Enemies of Intelligence*, 187; Dahl, *Intelligence and Surprise Attack*, 15.
88. Artner, Girven and Bruce, *Assessing the Value*, 2.
89. See, for instance, Karam, "Missing Revolution."
90. Connelly et al., "New Evidence," 783.
91. Ibid., 789.
92. See note 4 above.
93. Connelly et al., "Diplomatic documents data," 10–13.
94. Herman, *Intelligence Power*, 11,14.
95. Connelly et al., "New Evidence," 786.
96. Connelly et al., "Diplomatic documents data." History-Lab, "foiarchive." The data has been made available on Huggingface.
97. Connelly et al., "Diplomatic documents data," 5–6, 10.
98. Ibid., 10–1.
99. The data were collected using the python library "histlabapi." Gozal, "histlabapi 0.1.1." The function to filter for the countries is provided by the library.
100. The reason for the lowercase is twofold: first, the probability of a letter being lowercase in the English language is higher than of it being uppercase. A machine learning model is much more likely to have been trained on more lowercase text than on uppercase text. Moreover, consecutive uppercase characters often imply an emphasis ("IMPORTANT message"), which can be understood accordingly by machine learning models. Since all CFPF messages are uppercased, such emphases are not appropriate here.
101. Voyage AI, "voyage-large-2-instruct." This model ranked fourth on the widely used benchmark called "MTEB" to assess a model's capacity for embedding tasks. Specifically, MTEB spans reight embedding tasks covering a total of 58 datasets and 112 languages. Muennighoff et al., "MTEB."
102. Tunstall, von Werra and Wolf, *Natural Language Processing*, 275.
103. Manning, Raghavan and Schütze, "Introduction to Information Retrieval."
104. This process was as follows: Let the text be the input string, that is, the "body" of the cable, or, if the body was empty, the respective "title." The text is then broken down into tokens by a deterministic process called *tokenisation*. Tokens represent sub-parts of a word as atomic, numerical units, which are required to process text by a neural network. The text in the form

of tokens is then embedded. If, however, the number of tokens exceeds the capacity of the model, known as context window (16,000 in this case), the text is chunked beforehand. The chunks are then embedded and averaged to get one mean embedding of the text. The resulting embedding per cable is always a 1024-dimensional vector, irrespective of the length of the input text. Both the number of dimensions of the vector as well as the context window are given by the model "Voyage-large-2-instruct."

105. The dimensionality reduction and clustering steps were inspired by BERTopic. Grootendorst, "BERTopic."
106. McInnes, Healy and Melville, "UMAP."
107. Aggarwal, Hinneburg and Keim, "On the Surprising Behavior."
108. Ester et al., "A Density-Based Algorithm." The algorithm used the cosine distance and detects outliers by not assigning them to a cluster.
109. Killick, Fearnhead and Eckley, "Optimal Detection of Changepoints."
110. De Valk and Goldbach, "Towards a Robust."
111. Goodfellow, Bengio and Courville, *Deep Learning*, 410–2.
112. 1979KABUL07952, "Great October Socialist Revolution Celebrated Throughout Afghanistan."
113. 1979KABUL08419, "Voa News Item of Dec 11 From its New Delhi Correspondent on Amcit Nassry's News Conference."
114. 1974KABUL00539, "Daoud Government After Six Months an Assessment"; and 1975KABUL02382, "General Abdul Karim Mostaghni – vip Army Tour."
115. 1978PARIS38987, "Iranian Walkin Hints at Threat to Embassy Tehran Pol Counselor and Says Khomeiny Has Given Orders For Blood to Flow."
116. 1978PARIS39925, "Purported U.S. Study of Iranian Situation."
117. 1973TEHRAN01726, "Continuing Widespread Student Unrest"; and 1976TEHRAN10912. "Ministry Comment on University Student Unrest."
118. 1974KABUL07264, "Goa Ministry of Commerce Requests Aphis Test Afghan Apricots."
119. Dahl, *Intelligence and Surprise Attack*, 21–4.
120. Wohlstetter, *Pearl Harbor*, 387.
121. National Commission on Terrorist Attacks Upon the United States, *The 9/11Commission Report*, 339.
122. Dahl, *Intelligence and Surprise Attack*, 15.
123. McDermott, "Experimental Intelligence."
124. Betts, *Enemies of Intelligence*, 183–93.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Notes on contributors

Leonard Kern holds an MA in Intelligence and Security Studies from Brunel University of London. His research focuses on the intersection of intelligence analysis and technology, particularly artificial intelligence, as well as on research design in intelligence studies.

Kristian Gustafson is Reader in Intelligence and War at Brunel University of London. His research focuses on intelligence analysis and structured analytical tools, with publications addressing topics ranging from counter-poaching strategies in Africa to Horizon Scanning and Futures. He is an

Associate Researcher at the Norwegian Intelligence School, and has worked extensively with the UK MOD and Cabinet Office.

Martin Ejnar Hansen is a political scientist specialising in Comparative European Politics and Public Policy with specific focus on parliaments, governments and parties at Brunel University of London. Before joining Brunel as Reader in Politics, he was employed at the University of Southern Denmark, University of Aarhus and the University of Vienna.

ORCID

Leonard Kern http://orcid.org/0009-0009-9114-124X
Kristian Gustafson http://orcid.org/0000-0002-5532-3742
Martin Ejnar Hansen http://orcid.org/0000-0002-3637-208X

Bibliography

1973TEHRAN01726. "Continuing Widespread Student Unrest." State Department Central Foreign Policy Files. March 19, 1973. https://api.foiarchive.org/documents?doc_id=in.(1973TEHRAN01726).

1974KABUL00539. "Daoud Government After Six Months an Assessment." State Department Central Foreign Policy Files. January 28, 1974. https://api.foiarchive.org/documents?doc_id=in.(1974KABUL00539).

1974KABUL07264. "Goa Ministry of Commerce Requests APHIS Test Afghan Apricots." State Department Central Foreign Policy Files. November 19, 1974. https://api.foiarchive.org/documents?doc_id=in.(1974KABUL07264).

1975KABUL02382. "General Abdul Karim Mostaghni - VIP Army Tour." State Department Central Foreign Policy Files. April 16, 1975. https://api.foiarchive.org/documents?doc_id=in.(1975KABUL02382).

1976TEHRAN10912. "Ministry Comment on University Student Unrest." State Department Central Foreign Policy Files. February 11, 1976. https://api.foiarchive.org/documents?doc_id=in.(1976TEHRAN10912).

1978PARIS38987. "Iranian Walkin Hints at Threat to Embassy Tehran Pol Counselor and Says Khomeiny Has Given Orders for Blood to Flow." State Department Central Foreign Policy Files. November 28, 1978. https://api.foiarchive.org/documents?doc_id=in.(1978PARIS38987).

1978PARIS39925. "Purported U.S. Study of Iranian Situation." State Department Central Foreign Policy Files. December 6, 1978. https://api.foiarchive.org/documents?doc_id=in.(1978PARIS39925).

1979KABUL07952. "Great October Socialist Revolution Celebrated Throughout Afghanistan." State Department Central Foreign Policy Files. November 13, 1979. https://api.foiarchive.org/documents?doc_id=in.(1979KABUL07952).

1979KABUL08419. "VOA News Item of Dec 11 from Its New Delhi Correspondent on Amcit Nassry's News Conference." State Department Central Foreign Policy Files. December 12, 1979. https://api.foiarchive.org/documents?doc_id=in.(1979KABUL08419).

Abdalla, N. S. "Requirements, Priorities, and Mandates: A Model to Examine the US Requirements and Priorities Process and Its Impact on the Outcome of National Security and Foreign Policy Events." PhD diss. Brunel University of London, 2017.

Abdalla, N. S., and P. H. J. Davies. "Intelligence, Policy, and the Mandate: A Third Form of Strategic Failure." *The International Journal of Intelligence, Security, and Public Affairs* 23, no. 2 (2021): 105–124. doi:10.1080/23800992.2021.1881724.

Aggarwal, C. C., A. Hinneburg, and D. A. Keim. "On the Surprising Behavior of Distance Metrics in High Dimensional Spaces." Proceedings of the 8th International Conference on Database Theory, 420–434. Berlin, Springer, 2001.

Artner, S., R. S. Girven, and J. B. Bruce. *Assessing the Value of Structured Analytic Techniques in the U.S. Intelligence Community*. Washington DC: RAND, 2017.

Bar-Joseph, U. "Forecasting a Hurricane: Israeli and American Estimations of the Khomeini Revolution." *Journal of Strategic Studies* 36, no. 5 (2013): 718–742. doi:10.1080/01402390.2012.742009.

Betts, R. K. "Analysis, War, and Decision: Why Intelligence Failures Are Inevitable." *World Politics* 31, no. 1 (1978): 61–89. doi:10.2307/2009967.

Betts, R. K. *Surprise Attack: Lessons for Defense Planning*. Washington, DC: Brookings Institution, 1982.

Betts, R. K. *Enemies of Intelligence: Knowledge and Power in American National Security*. New York: Columbia University Press, 2007.

Byrne, M. "CIA Confirms Role in 1953 Iran Coup." *National Security Archive Electronic Briefing Book No. 435*. August 19, 2013. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB435/.

Central Intelligence Agency. *Declassified Interagency Intelligence Memorandum: The Soviet Invasion of Afghanistan: Implications for Warning*. October. 1980. https://www.cia.gov/readingroom/docs/DOC_0000278538.pdf .

Connelly, M., R. Hicks, R. Jervis, and A. Spirling. "New Evidence and New Methods for Analyzing the Iranian Revolution as an Intelligence Failure." *Intelligence and National Security* 36, no. 6 (2021): 781–806. doi:10.1080/02684527.2021.1946959.

Connelly, M., R. Hicks, R. Jervis, A. Spirling, and C. H. Suong. "Diplomatic Documents Data for International Relations: The Freedom of Information Archive Database." *Conflict Management and Peace Science* 38, no. 6 (2021): 762–781. doi:10.1177/0738894220930326.

Connelly, M., D. Madigan, A. Spirling, R. Hicks, B. Lis, M. Flynn, and R. Jervis. "History Lab | Freedom of Information Archive (FOIArchive)." Accessed July 7, 2024. http://history-lab.org/.

Dahl, E. J. *Intelligence and Surprise Attack: Failure and Success from Pearl Harbor to 9/11 and Beyond*. Washington, DC: Georgetown University Press, 2013.

Davies, P. H. J. "Intelligence Culture and Intelligence Failure in Britain and the United States." *Cambridge Review of International Affairs* 17, no. 3 (2004): 495–520. doi:10.1080/0955757042000298188.

Davies, P. H. J. "Theory and Intelligence Reconsidered." In *Intelligence Theory: Key Questions and Debates*, edited by P. Gill, S. Marrin, and M. Phythian, 186–207. Abingdon: Routledge, 2009.

Davis, J. "Intelligence Analysts and Policymakers: Benefits and Dangers of Tensions in the Relationship." *Intelligence and National Security* 21, no. 6 (2006): 999–1021. doi:10.1080/02684520601046325.

De Valk, G., and O. Goldbach. "Towards a Robust β Research Design: On Reasoning and Different Classes of Unknowns." *Journal of Intelligence History* 20, no. 1 (2021): 72–87. doi:10.1080/16161262.2020.1746144.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), 226–231. Portland, OR, AAAI Press, 1996.

Gates, R. M. *From the Shadows: The Ultimate Insider's Story of Five Presidents and How They Won the Cold War*. New York: Simon & Schuster, 1996.

Giarelis, N., and N. Karacapilidis. "Deep Learning and Embeddings-Based Approaches for Keyphrase Extraction: A Literature Review." *Knowledge and Information Systems* 66, no. 11 (2024): 6493–6526. doi:10.1007/s10115-024-02164-w.

Gill, P., and M. Phythian. *Intelligence in an Insecure World*. 3rd ed. Cambridge: Polity Press, 2018.

Godson, R., and J. J. Wirtz. "Strategic Denial and Deception." *International Journal of Intelligence and Counterintelligence* 13, no. 4 (2000): 424–437. doi:10.1080/08850600050179083.

Goodfellow, I., Y. Bengio, and A. Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016.

Gozal, D. "HistLabAPI 0.1.1." Python Package Index, 2024. Accessed June 13, 2024. https://pypi.org/project/histlabapi/ .

Grootendorst, M. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." *ArXiv* preprint (2022). doi:10.48550/arXiv.2203.05794.

Gustafson, K. C. "Confidence Game: Intelligence, Deception and Subterfuge." In *A Cultural History of War* Vol. 6, edited by M. K. Barbier and D. Showalter, 149–166. London: Bloomsbury Publishing, 2025.

Gustafson, K. C., D. Lomas, S. Wagner, N. S. Abdalla, and P. H. J. Davies. "Intelligence Warning in the Ukraine War, Autumn 2021 – Summer 2022." *Intelligence and National Security* 39, no. 3 (2024): 400–419. doi:10.1080/02684527.2024.2322214.

Handel, M. I. "The Yom Kippur War and the Inevitability of Surprise." *International Studies Quarterly* 21, no. 3 (1977): 461. doi:10.2307/2600234.

Hatlebrekke, K. A., and M. L. Smith. "Towards a New Theory of Intelligence Failure? The Impact of Cognitive Closure and Discourse Failure." *Intelligence and National Security* 25, no. 2 (2010): 147–182. doi:10.1080/02684527.2010.489274.

Hayden, M. "A 'Reasonable' View of Privacy, Security." The Washington Times. March 18, 2015. https://www.washingtontimes.com/news/2015/mar/18/michael-hayden-a-reasonable-view-of-privacy-securi/.

Herman, M. *Intelligence Power in Peace and War*. Cambridge: Cambridge University Press, 1996.

History-Lab. "foiarchive." Accessed June 17, 2025. https://huggingface.co/datasets/HistoryLab/foiarchive .

Ikani, N. "Beyond the Binary: A New Typology for Evaluating Warning Success and Failure in Strategic Surprise." *International Studies Review* 27, no. 1 (2025): 1–24. doi:10.1093/isr/viaf009.

Jensen, M. A. "Intelligence Failures: What Are They Really and What Do We Do About Them?" *Intelligence and National Security* 27, no. 2 (2012): 261–282. doi:10.1080/02684527.2012.661646.

Jervis, R. *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War*. Ithaca: Cornell University Press, 2010.

Jervis, R. "Why Postmortems Fail." *Proceedings of the National Academy of Sciences* 119, no. 3 (2022): 1–6. doi:10.1073/pnas.2116638118.

Kahn, D. "Surprise and Secrecy: Two Thoughts." *Intelligence and National Security* 21, no. 6 (2006): 1060. doi:10.1080/02684520601046747.

Karam, J. "Missing Revolution: The American Intelligence Failure in Iraq, 1958." *Intelligence and National Security* 32, no. 6 (2017): 693–709. doi:10.1080/02684527.2016.1275138.

Kent, S. *Strategic Intelligence for American World Policy*. Princeton: Princeton University Press, 1949.

Kent, S. "Words of Estimative Probability." *Studies in Intelligence* 8, no. 4 (1963): 49–65.

Killick, R., P. Fearnhead, and I. A. Eckley. "Optimal Detection of Changepoints with a Linear Computational Cost." *Journal of the American Statistical Association* 107, no. 500 (2012): 1590–1598. doi:10.1080/01621459.2012.737745.

King, G., R. O. Keohane, and S. Verba. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. 2nd ed. Princeton: Princeton University Press, 2021.

Kuhns, W. J. "Intelligence Failures: Forecasting and the Lessons of Epistemology." In *Paradoxes of Strategic Intelligence: Essays in Honor of Michael I. Handel*, edited by R. K. Betts and T. G. Mahnken, 80–100. Abingdon: Routledge, 2003.

Kurzman, C. *The Unthinkable Revolution in Iran*. Cambridge, MA: Harvard University Press, 2004.

Levite, A. *Intelligence and Strategic Surprises*. New York: Columbia University Press, 1987.

Lyakhovsky, A. A. *Inside the Soviet Invasion of Afghanistan and the Seizure of Kabul, December 1979*. Cold War International History Project. Washington, DC: Woodrow Wilson International Centre for Scholars, 2007.

Macanovic, A. "Text Mining for Social Science – the State and the Future of Computational Text Analysis in Sociology." *Social Science Research* 108 (2022): 1–17. doi:10.1016/j.ssresearch.2022.102784.

MacEachin, D. *Predicting the Soviet Invasion of Afghanistan: The Intelligence Community's Record*. Washington, DC: CIA Center for the Study of Intelligence, 2007.

Manger, G. "Unravelling Effectiveness in Intelligence: A Systematic Review." *Intelligence and National Security* 39, no. 7 (2024): 1136–1157. doi:10.1080/02684527.2024.2370132.

Manning, C. D., P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.

Marrin, S. "Preventing Intelligence Failures by Learning from the Past." *International Journal of Intelligence and Counterintelligence* 17, no. 4 (2004): 655–672. doi:10.1080/08850600490496452.

Marrin, S. "Evaluating the Quality of Intelligence Analysis: By What (Mis) Measure?" *Intelligence and National Security* 27, no. 6 (2012): 896–912. doi:10.1080/02684527.2012.699290.

McDermott, R. "Experimental Intelligence." *Intelligence and National Security* 26, no. 1 (2011): 82–98. doi:10.1080/02684527.2011.556361.

McInnes, L., J. Healy, N. Saul, and L. Großberger. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *Journal of Open Source Software* 3, no. 29 (2018): 861. doi:10.21105/joss.00861.

Muennighoff, N., N. Tazi, L. Magne, and N. Reimer. "MTEB: Massive Text Embedding Benchmark." Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, Dubrovnik, 2023. doi:10.48550/arXiv.2210.07316.

National Commission on Terrorist Attacks Upon the U.S. The *9/11 Commission Report*. July 22, 2004. https://www.9-11commission.gov/report/911Report.pdf.

Office of the Director of National Intelligence. *ICD-203 Analytic Standards*. Accessed December 21, 2022. https://www.dni.gov/files/documents/ICD/ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf .

Robertson, K. "Editorial Comment: An Agenda for Intelligence Research." *Defense Analysis* 3, no. 2 (1987): 95–101. doi:10.1080/07430178708405287.

Scott, L., and P. Jackson. "The Study of Intelligence in Theory and Practice." *Intelligence and National Security* 19, no. 2 (2004): 139–169. doi:10.1080/0268452042000302930.

Shulsky, A. N., and G. J. Schmitt. *Silent Warfare: Understanding the World of Intelligence*. 3rd ed. Washington, DC: Potomac Books, 2002.

Tetlock, P., and D. Gardner. *Superforecasting: The Art and Science of Prediction*. New York: Broadway Books, 2015.

Tunstall, L., L. von Werra, and T. Wolf. *Natural Language Processing With Transformers*. Revised ed. Sebastopol, CA: O'Reilly Media, 2022.

UK Cabinet Office. "Professional Head of Intelligence Assessment." *Professional Development Framework for All-Source Intelligence Assessment*. 2019. https://www.gov.uk/government/publications/intelligence-analysis-professional-development-framework/the-professional-development-framework-for-all-source-intelligence-assessment.

Voyage, A. I. "Voyage-Large-2-Instruct: Instruction-Tuned and Rank 1 on MTEB." Accessed May 5, 2024. https://blog.voyageai.com/2024/05/05/voyage-large-2-instruct-instruction-tuned-and-rank-1-onmteb/.

Walton, T. R. *Challenges in Intelligence Analysis: Lessons from 1300 BCE to the Present*. New York: Cambridge University Press, 2010.

Warner, M. "Sources and Methods for the Study of Intelligence." In *Handbook of Intelligence Studies*, edited by L. K. Johnson, 17–27. Abingdon: Routledge, 2007.

Wirtz, J. J. "Responding to Surprise." *Annual Review of Political Science* 9, no. 1 (2006): 45–65. doi:10.1146/annurev.polisci.9.062404.170600.

Wirtz, J. J. "The Sources and Methods of Intelligence Studies." In *The Oxford Handbook of National Security Intelligence*, edited by L. K. Johnson, 59–69. New York: Oxford University Press, 2010.

Wirtz, J. J. "Are Intelligence Failures Still Inevitable?" *International Journal of Intelligence and Counterintelligence* 37, no. 1 (2024): 307–330. doi:10.1080/08850607.2023.2214328.

Wohlstetter, R. *Pearl Harbor: Warning and Decision*. Stanford: Stanford University Press, 1962.

Zegart, A. B. *Spies, Lies, and Algorithms: The History and Future of American Intelligence*. Princeton: Princeton University Press, 2022.