HSPI The Open Access Publisher

JAIRI ISSN 3069-762X JOURNAL OF ARTIFICIAL INTELLIGENCE RESEARCH AND INNOVATION

Position Paper

Understanding AI-enabled Outputs: A Conceptual and Analytical Primer for All-source Intelligence Analysts

George Ellison*

Centre for Intelligence Studies, University of Lancashire, PR1 9HE, UK

Abstract

Intelligence analysts are increasingly able, and required, to consume and interpret outputs produced by artificial intelligence (AI) enabled tools – yet most receive little training in what these outputs actually represent, or how these might be robustly evaluated. This primer addresses that gap. It starts from the premise that analysts do not need to know how to develop or operate AI-enabled tools in order to use these tools' outputs critically and judiciously. But analysts do need sufficient conceptual and analytical understanding – and a modicum of technical knowledge – to evaluate these outputs competently. Three principles provide the framework for this understanding: First, the consequential distinction between AI-*facilitated* outputs – where automation improves the pace, scale and fidelity of data collection, processing and analytical procedures that analysts could otherwise perform; and AI-*generated* outputs – where many of the novel insights these outputs support could not have been produced by analysts working independently; Second, the conceptually challenging yet fundamental difference between mechanistic prediction and interpolative or extrapolative estimation; and Third, the critical dependencies and substantive limitations that govern the reproducibility and practical utility of all AI-*facilitated* and AI-*generated* outputs. Together these principles constitute the conceptual and analytical foundations of the AI literacy training that all-source intelligence analysts should receive – the case for which is presented in a companion piece to this position paper.

More Information

***Corresponding author:** George Ellison, Centre for Intelligence Studies, University of Lancashire, PR1 9HE, UK; gthellison@lancashire.ac.uk

Submitted: June 18, 2026
Accepted: June 24, 2026
Published: June 26, 2026

Citation: Ellison G. Understanding AI-enabled Outputs: A Conceptual and Analytical Primer for All-source Intelligence Analysts. J Artif Intell Res Innov. 2026; 2(1): 57-65. Available from: https://dx.doi.org/10.29328/journal.jairi.1001019

Keywords: Artificial intelligence; Intelligence analysis; Decision support; AI literacy; Human-AI teaming; Predictive estimation; Mechanistic prediction

The epistemic and cognitive impact of 'artificial intelligence'

'Artificial intelligence' or 'AI' has become a troublesome, catch-all term [1] encompassing a growing array of computational techniques that use rules-based software. These include those that:

- simply automate existing, or novel, analyst-defined procedures (known as "[un-intelligent] rules engines"); or
- can independently respond and adapt to the results of user-defined procedures (known as "intelligent rules engines"); or
- include components of, and contributions from, each of these [2].

All of these techniques can appear 'intelligent' in their ability to perform tasks systematically and automatically, and thereby *facilitate* the production of what Donald Rumsfeld (2002) [3] famously called 'known unknowns' – i.e. "things that we know we don't know" – and to do so at a scale and pace, and with a level of fidelity, that would otherwise require substantial time and effort on the part of skilled and experienced analysts. But the second (and, therefore, the third) of these techniques can also *generate* estimates of what Rumsfeld (2002) [3] called 'unknown unknowns' – i.e. "things that we don't know we don't know". These become apparent from hitherto hidden patterns within the data analyzed, which reflect and reveal dataset features that would otherwise have been impossible for even the most skilled human analysts to detect (and even if unconstrained by the time and effort they would require to do so).

Notwithstanding the unprecedented advance in analytical productivity, innovation and discovery that these computational techniques provide, the broader contemporary use of the term 'artificial intelligence' (or simply, 'AI') risks mistaking the outputs of AI-enabled tools as equivalent to human intelligence, intuition and sense-making by conflating:

- AI as a discipline with AI as an entity;
- AI as an aspiration with AI as a reality; and
- AI as a tool with AI as an independent agent.

As Sanguinetti (2025) [1] points out, this conflation has led to "misunderstandings [that]… have spread throughout the lexicon of the discipline", and the widespread misrepresentation and misunderstanding of what AI-enabled tools can *actually* achieve. Indeed, the adoption of such terminology seems designed to imply that some AI techniques are *fully* autonomous, or even sentient. In reality, the most immediate advances that AI-enabled tools have achieved are simply a consequence of the automation of processes that humans themselves are already able to perform – albeit at a far slower pace, a far lower scale, and with far less fidelity than AI-enabled tools commonly achieve.

Terminology that encourages this misleading representation includes: 'machine *learning*'; 'deep *learning*'; '*predictive* analytics'; and even 'AI *hallucinations*' – each of which can imply capabilities and forms of agency that exceed those such tools actually possess [1]. Those who develop and operate the tools that produce AI-enabled outputs – as well as those who commission these tools and consume their outputs – should all be aware of the insidious effects these semantic sleights of hand can have on more prosaic yet more accurate and literal conceptualizations of the:

- tasks that AI-enabled techniques can (and cannot) perform; and
- questions and problems for which AI-enabled techniques can (and cannot) provide meaningful or actionable answers, solutions and explanations [4].

Yet this can nonetheless be challenging, since the outputs produced by AI-enabled tools have *both* epistemic *and* cognitive effects:

- 'Epistemic' as in the empirical and experiential evidence on which human knowledge of the world is based; and
- 'Cognitive' as in the variable and often imperfect information processing mechanisms, reasoning skills and heuristics that humans employ to evaluate and integrate such evidence into the corpus of knowledge and understanding they hold to be true.

Without sufficient conceptual understanding of how AI-enabled tools produce these outputs, those who commission or consume them risk misunderstanding both:

- what these outputs mean; and
- how they can (and should) be used.

In the absence of this understanding, the insights that commissioners and customers derive from the outputs of AI-enabled tools are vulnerable to a range of *inferential* errors and biases. This is particularly the case within 'disciplines of uncertainty' – such as intelligence analysis [5] – where the allure of AI-enabled tools lies not only in the productivity gains these tools can offer, but also in their ability to reveal patterns hidden so deep within the datasets available, that they would otherwise be challenging, impracticable or even impossible for human analysts to identify *without* such tools.

Yet the attraction of AI-enabled tools as a novel source of information wherever uncertainty is pervasive, also heightens the need for analysts to clearly understand:

- what AI-enabled tools are actually doing;
- what their outputs actually represent; and
- how any inherent dependencies and limitations might be evaluated, attenuated or accommodated within the insights that analysts then derive from the outputs these tools produce.

These vulnerabilities are not merely semantic abstractions. They arise precisely because the outputs of AI-enabled tools are increasingly being treated as meaningful sources of information in contexts where analysts must *still* decide:

- what these outputs mean;
- whether they can be trusted; and
- what inferences and insights these might support.

The twin aims of this position paper are therefore to:

- provide an accessible conceptual and analytical primer to accompany a companion piece in which the case for providing foundational 'AI literacy' training [6] to all-source intelligence analysts is made [7]; and
- situate this against the backdrop of AI-related policies and practices being developed to support the integration of AI-enabled tools within intelligence workflows.

The three principles that follow are intended to address the first of these aims; and although none of these principles requires specialist technical training to understand or apply, each can help analysts navigate the analytical and inferential pitfalls that AI-enabled techniques entail. These principles mirror the more technical considerations that analysts using computer-enabled statistical techniques have needed to bear in mind when applying these to analyze and interrogate large and complex datasets. Indeed, these techniques include early versions of 'intelligent rules engines' – such as: multivariable linear and logistic regression; and analysis of covariance

(ANCOVA) – and much might be learned from the challenges analysts face when attempting to use these within intelligence settings [8].

As for the second of these aims, this will be addressed by critically examining Deloitte's early "task-level" review exploring the potential impact of AI on intelligence analysis – a review that presented a compelling, if somewhat partial, case for the integration of AI-enabled tools within intelligence workflows [2]; and one that has to some extent been overtaken by events. In this sense, the Deloitte review is used here as a prominent, policy-relevant and analyst-accessible point of departure, rather than as a comprehensive authority on the broader conceptual and analytical issues examined in this position paper – issues that have been usefully and more comprehensively explored in a number of subsequent reports [9-15].

Meanwhile, to make the terminology used throughout this position paper more accessible to its intended audience, Table 1 provides working definitions of the core concepts on which the arguments that follow rely. Although many of these arguments emerged in relation to earlier rules-based and machine-learning systems, they also apply to more recent AI-enabled tools, including foundation models, large language models, retrieval-augmented systems, and related forms of human-AI teaming.

The future of intelligence analysis in the age of AI

Although the Deloitte review [2] paid little heed to the conceptual and analytical understanding that intelligence analysts require to critically and competently consume AI-enabled outputs, it enthusiastically endorsed a growing role for AI within intelligence analysis. In particular, it summarized the key incentives for, and many benefits of, integrating AI-enabled tools into intelligence workflows – and emphasized the efficiencies that AI might bring by:

- automating "lower-value" tasks;
- supporting "higher-value" analytical and inferential tasks; and
- providing more timely, comprehensive and customer-focused decision support [2].

However, any such efficiencies are likely to be substantively attenuated by the time, resource and training costs required to implement, maintain and sustain this shift in analytical practice. Indeed, Mitchell, et al. [2] acknowledged a number of limitations and uncertainties that undermined the findings of their review, and cited three specific caveats that are particularly relevant to the training needs of all-source intelligence analysts:

- First, that "intelligence work never ends" – meaning that prioritization will need to remain a central and essential component of professional intelligence practice irrespective of any improvements in information processing and analytical capacity that AI-enabled tools might deliver;
- Second, that "AI is not the solution to every problem" – meaning that the utility of AI-enabled tools will depend in no small part on the nature of the task, the datasets available, the net improvements that can be achieved, and the costs of implementation; and
- Third, that the "impact of AI on analysts' cognitive biases remains uncertain" – meaning that intelligence agencies will need to carefully monitor the cognitive capabilities and wellbeing of analysts using or teaming with AI-enabled tools. This is not a trivial concern given the tangible risk that familiarity with AI-enabled outputs might make analysts – like other consumers of AI-enabled tools – more trusting and less critical of these [2,16-18].

Beyond these caveats, there are at least four additional considerations that Mitchell, et al. [2] overlooked that are also relevant to the realizable utility of AI-enabled outputs in all-source intelligence analysis. Each of these is worth examining in some detail:

1. The unprecedented (and essentially unique) nature of many intelligence requirements

A variable (but potentially substantial) proportion of intelligence requirements concern phenomena, entities, processes, or their characteristics, that are historically unprecedented, or so spatio-temporally contingent as to be essentially (and quite literally) unique. As a result, the data available for these may lack the volume or variety required to justify or support the use of AI-enabled information processing and analytical techniques. Under such circumstances – and where substantial uncertainty resides in the paucity of definitive evidence – analysts will need to retain the specialist cognitive and analytical skills required to generate more theorized and speculative, yet resolutely robust, assessments through which to provide decision support [5,19].

2. Data classification, storage and fusion constraints

Combining datasets and fusing information from different systems, often held at different classifications, poses a long-standing challenge to automated processing in intelligence analysis [20]. This becomes more difficult still when the information streams and datasets involved are large, require substantial and secure storage space, and must be hosted within systems that can support the necessary processing, automation and analytical software – and from which any resulting algorithms and outputs can be extracted or released.

3. AI is evolving, and so are the analysts who are expected to use it

Much of Mitchell, et al.'s [2] enthusiasm for integrating

AI-enabled automation and analysis within intelligence workflows reflected the state of AI technology in 2019 (when their review was published), and the contemporaneous capabilities of intelligence analysts whose exposure to analog vs. digital technologies now include those born in the early 1960s through to the early 2000s. Current and future cohorts of intelligence analysts – including yesteryear's 'analog natives', today's 'digital natives', tomorrow's 'AI natives', and (perhaps) next year's 'quantum natives' [18] – may require very different forms of training and support to strengthen and sustain their analytical judgment, assessment skills and capacity for human-AI teaming. And whatever the future brings, it is unlikely to entail exactly the same discrete set of challenges that intelligence analysts face today. This means not only that the Deloitte review (published in 2019) is already substantively out-of-date, but that the doctrinal, training-related and technical infrastructure required to integrate AI into intelligence workflows will need continual revision as both the technology, and the socio-technical environments that shape analysts' cognitive capabilities, continue to evolve.

4. The people who can do this well are scarce and in demand

Analysts trained to consume, commission or produce AI-enabled outputs – and to integrate such outputs judiciously within intelligence analysis and assessment for decision support across complex, unpredictable and uncertain multi-domain contexts – remain in short supply and in high demand. For the foreseeable future, intelligence agencies are therefore likely to face stiff competition when recruiting (and retaining) personnel with the competencies, motivation and aptitude required to develop and deploy these skills.

Taken together, these additional considerations suggest that the integration of AI-enabled tools into intelligence workflows cannot be framed as simply a matter of providing and integrating these more powerful tools. It will substantively depend upon an analyst workforce that can understand what these tools are doing, what sorts of outputs they produce, and what claims can be confidently made on the basis of the outputs they provide. With this in mind, the three principles that follow address the most pertinent conceptual and analytical considerations analysts need to understand if their practice is to keep pace with the system-wide transformation that the integration of AI-enabled tools will require.

AI-enabled outputs: A conceptual and analytical primer for intelligence analysts

Principle 1: The important distinction between AI-facilitated and AI-generated outputs

The distinction between AI-*facilitated* and AI-*generated* outputs (as summarized in Table 1) is central to understanding what intelligence analysts most need to know about AI-enabled tools and their outputs. Much of the excitement surrounding AI focuses on the unique quasi-'predictive' capabilities of a subset of AI techniques that involve 'intelligent rules engines' – particularly machine learning and deep learning [2]. Yet it is important not to ignore how tasks performed by much simpler, and far less inscrutable '[un-intelligent] rules engines' can greatly facilitate the outputs that analysts themselves can already generate, and the insights these can support. This is because whenever the automated tasks AI-enabled tools perform enable the human analysts involved to undertake analyses, produce outputs and derive insights – by, for example: observing and measuring; collating and combining; classifying and categorizing; or summarizing and visualizing information – then the uplift in (human-derived, yet AI-*facilitated*) insight can be just as profound, and potentially more relevant, meaningful and useful, as that based on outputs *generated* by AI-enabled tools.

In this sense, the key strengths of AI-*facilitated* outputs are their (relative) simplicity, accessibility and interrogability, and the benefits that accrue when human analysts remain central to the analytical process [12]. These benefits include the broader contextual understanding that only humans can provide, such as the attribution of meaningfulness and utility, and the filtering of discernible and emergent patterns according to their actionability. This can help ensure that the outputs identified, and any insights derived therefrom, are those that not only help reduce a decision-maker's uncertainty, but also offer more appropriate (that is, more relevant, meaningful and actionable) decision support. Added to this is the substantive benefit of being able to replicate whatever outputs AI-*facilitated* techniques have produced, since these predominantly involve the automation of information collection, processing and analytical tasks that analysts themselves *could* perform [7].

In contrast, AI-*generated* outputs comprise those produced by a sub-set of semi-autonomous AI techniques. These techniques exploit the statistical-associational properties of relationships amongst variables within a dataset to identify optimal algorithmic (i.e. arithmetic and algebraic) solutions to user-specified computational tasks. This is an approach that was (and is still) widely practiced using computer-based statistical packages decades before high-powered computational capacity enhanced the scale, scope and pace with which such analyses could be performed semi-autonomously [21]. These include:

- *'supervised machine learning'* – which involves comparing multiple permutations of alternative algorithms to identify (or 'train') the optimal algorithm that: most consistently or accurately generates similar values to those that are available for a pre-specified 'target variable' or 'target dataset property or parameter' within the 'training' dataset concerned. These algorithms can then be used to estimate, impute or 'predict' (through interpolation or extrapolation) values for such 'targets' within subsequent, comparable

Table 1: Working definitions of the core concepts used in this position paper, and their relevance to intelligence analysis practitioners

Term	Definition	Relevance to intelligence practitioners
AI-enabled tools	Computational tools that automate, augment, or otherwise support analytical tasks using rules-based, statistical, or learning-based procedures	Analysts need to know what kind of tool produced an output, because that affects: how readily its procedures can be understood; whether its results can be evaluated; and the degree of inferential caution required
Rules engines	Tools that apply predefined (and in some cases condition-responsive) computational rules to data in order to sort, filter, classify, or process information	This helps analysts distinguish comparatively transparent forms of automation from less transparent model-based procedures; and therefore identify outputs that are usually easier to replicate, evaluate, and explain
Machine learning	A subset of AI-enabled procedures in which models are trained on data to identify patterns, classify cases, estimate specified outcomes, or detect relationships within and across variables	This often produces outputs that analysts alone could not readily derive, and whose underlying procedures and assumptions may be less transparent, making the results harder to interrogate, evaluate, and explain
Deep learning	A more complex subset of machine learning that uses layered model architectures to identify and learn patterns from large or complex datasets	This can produce highly useful outputs, but is often less transparent than simpler statistical or rule-based approaches, making it harder for analysts to confidently evaluate, understand or explain to others how its results were produced
Generative AI	Tools – including large language models and other foundation-model-based systems – that generate synthetic text, images, data, or other content from patterns learned from large training corpora or other underlying datasets	These tools raise issues of provenance, attribution, fabrication, deception, and interpretability – requiring analysts to be especially cautious as to whether apparently coherent outputs are evidentially grounded, verifiable, and fit for analytical use
AI-facilitated outputs	Outputs produced where automation improves the speed, scale, consistency, or fidelity of analytical tasks analysts themselves could ordinarily perform	These outputs are often easier to understand and interrogate because the underlying task remains recognizably analytical, even if automation allows it to be performed faster, at greater scale, and with greater fidelity
AI-generated outputs	Outputs involving model-based estimation, inference, clustering, or generation that go beyond what analysts alone could ordinarily produce	These outputs require greater analytical scrutiny and inferential caution, because their apparent novelty or usefulness may exceed the analyst's ability to evaluate: how the results were produced; on what assumptions these depend; or how robustly they can be validated
Predictive estimation	Statistical or algorithmic estimation of unknown, unmeasured, or future features, usually through interpolation or extrapolation from multivariable associational patterns present within available datasets, rather than through mechanistic prediction in any causal sense	This reminds analysts that such outputs should be: treated as context-bound *estimates* rather than self-validating *predictions*; and evaluated in light of data scope, representativeness, uncertainty, and the stability of the conditions on which both the estimation process and the resulting estimate depend
Mechanistic prediction	The use of definitive evidence regarding the mechanistic structure of systems and processes on which precise predictions of past, current and future outcomes can be determined – including evidence from experimentation, causal inference modelling of observational data, or knowledge of system/process design	Analysts will have a working mechanistic understanding of how the design of systems and procedures can make their outcomes predictably dependable. However, a lack of definitive evidence regarding these designs, together with prevailing contextual uncertainty, will limit analysts' ability to derive dependable 'mechanistic predictions' from them
Human-AI teaming	The use of AI-enabled tools – including interactive and retrieval-augmented systems – to support, extend, or inform human analysis without displacing human responsibility for interpretation and judgment	This clarifies that AI-enabled tools may assist analysis without replacing the analyst's responsibility to: critically interpret these tools' outputs; supply context; recognize limitations; and remain accountable for any judgments and assessments these outputs inform

datasets where these values are unattainable or have yet to have occurred, been measured, observed, or recorded; and

- '*unsupervised machine-learning*' – which' operates in a similar fashion, but in lieu of a 'target variable' or 'target dataset property or parameter' against which the optimal algorithm can be calibrated (or 'trained'), the 'target' involved is one of a number of unmeasured (i.e. 'latent') and (perhaps only loosely) user-defined dataset properties, or parameters. These include: 'latent variables' (i.e. unmeasured and therefore hidden variables, that can nonetheless be elucidated from user-specified statistical-associational properties evident within the dataset); or a discrete number of 'latent classes' of cases (i.e. unmeasured and therefore hidden groups or clusters of cases that share more similar statistical-associational features than cases more closely aligned with other groups or clusters).

These techniques make a powerful contribution to the perceived utility of AI-enabled tools, as a result of their ability to estimate the value of missing values, or hidden properties, parameters and features of datasets – and to do so (in a good many datasets) with apparent facility and acuity. This offers what appear to be hitherto unattainable 'predictive estimates' of two principal sources of uncertainty: 'known unknowns'; and 'unknown unknowns' [3,19,22,23].

Indeed, the ability of this subset of AI techniques to offer predictive estimates of hitherto 'unknown unknown' phenomena, entities, processes, or characteristics thereof, means these offer an unprecedented advance in measurement – and one that seems likely to eclipse even the most dramatic step-changes in 'knowing' that have previously occurred, such as the invention of: optical instruments (microscopes and telescopes) [24,25] in the 16th and 17th centuries; and sensors for measuring the full extent of the electromagnetic spectrum (in the 19th and early 20th century) [26]. Moreover, 'intelligent rules engines' look likely to dramatically expand the capabilities of all preceding technologies, while simultaneously capitalizing on the data these technologies generate so as to reveal entities, structures, phenomena, processes and mechanisms that might well go some way beyond our wildest expectations. Moreover, this distinction remains pertinent even where contemporary systems combine multiple capabilities – as in foundation-model-based, retrieval-augmented, or interactive AI tools – since the critical question for analysts remains whether the output primarily *facilitates* recognizable analytical tasks or *generates* novel outputs or content beyond what analysts on their own could ordinarily produce.

For the analyst, the practical importance of this distinction is straightforward. AI-*facilitated* outputs can often be interrogated, replicated or checked against processes and judgments that trained analysts understand well, and are able to produce themselves. AI-*generated* outputs, by contrast, are more likely to support insights that analysts cannot independently derive, and therefore demand a different and more conceptually advanced form of scrutiny. Distinguishing between these two broad categories of output is therefore a necessary first step in knowing what questions to ask of AI-enabled tools, and what kinds of inferences and claims might safely be made on the basis of their outputs.

Box 1: An example of AI-*facilitated* outputs within the context of intelligence analysis

A tool provides automated information-extraction and cross-document triage that supports analysts to extract relevant information, and collate this within and across reports, so as to: identify any recurring entities, phenomena, processes or characteristics thereof; and facilitate the visualisation of relationships amongst and between these, both faster and more consistently than would be possible manually without substantial additional analytical time and resource.

Relevance: The outputs *facilitated* by this AI-enabled tool have substantial operational utility, and remain intelligible to the analysts concerned, and are amenable to independent replication and validation by the analysts.

Box 2: An example of AI-*generated* outputs within the context of intelligence analysis

A clustering tool trained to perform anomaly-detection identifies a pattern in logistics activity hidden within SIGINT and GEOINT data, and flags the possibility of covert preparation for an imminent change in an adversary's operational posture.

Relevance: The pattern identified may prove invaluable and hitherto inaccessible, but analysts must nonetheless ensure that: the tool's outputs have been robustly evaluated; the training and operational datasets are similarly representative, operationally equivalent, contextually consistent, and thereby transferable; and the pattern identified is therefore not merely the result of chance, artifact, bias or deception.

Principle 2: 'Predictive analytics' generates descriptive estimates, not bona fide 'predictions'

In common parlance, a 'prediction' reflects the ability to accurately determine something that has yet to occur, or an observation or measurement that has yet to be made of something in the past, present or future. Yet, as Table 1 indicates, many of the so-called 'predictions' that AI-enabled tools can provide can be more accurately understood as forms of interpolative or extrapolative *estimation* than as mechanistic predictions in any stronger causal sense (i.e. as dependable predictions based on definitive understanding of the underlying mechanisms and processes involved). Nonetheless, as a concept, 'prediction' has particular resonance in contexts where there is little if any knowledge or understanding of the mechanisms involved in producing whatever it is to which the prediction relates. Indeed, so-called 'intelligence failures' are often erroneously attributed to those predictive assessments and judgments that, despite a high absolute or relative 'estimative probability' [27], do not subsequently occur [28].

Under such circumstances, a prediction that turns out to be true is commonly taken as proof of specialist predictive expertise, or even of mystical or prophetic powers. This is not without justification if the basis on which the successful prediction was made is unclear or unknown to anyone else (including, perhaps, the person making the prediction…). Moreover, even predictions that would have later turned out to be wildly inaccurate can inspire sufficient confidence to take on a life of their own as 'self-fulfilling' or 'self-defeating' prophecies. In such cases, the prediction acts as an intervention because it leads those who believe it to precipitate or prevent, respectively, whatever it was that had been predicted to occur.

Unsurprisingly, then, the use of the term 'prediction' to describe the statistical estimates that AI techniques known as 'intelligent rules engines' can generate imbues such techniques, and the very notion of AI, with capabilities that far exceed those of any human forecaster. At the same time, the complex and opaque algorithmic solutions (i.e. arithmetic and algebraic solutions) – that these AI-enabled techniques construct, only serve to enhance their allure as something we know or believe to be true but cannot interrogate (and might never fully understand or explain).

In reality, the facts of the matter are somewhat more mundane. Yet even those who (should) fully grasp the basis on which these AI techniques generate their so-called 'predictions' can struggle to resist the temptation to view AI as anything other than uncannily prescient. This is evident in the case of the Google engineer [29] and the celebrated evolutionary biologist [30] who came to believe or infer that the chatbots they had helped develop, or had 'conversed' with, were sentient and 'alive'. Both cases serve as vivid examples of the substantive cognitive impacts [4] that AI-enabled tools, their outputs, and the insights these support, can have.

For intelligence analysts, the practical implication is straightforward. AI-enabled 'predictions' should be approached not as privileged glimpses of what is likely or going to happen, but as context-dependent statistical estimates whose meaning, utility and trustworthiness depend on: the datasets from which they were derived; the tools used in their estimation; the conditions under which these tools were applied; and the extent to which these datasets and conditions are comparable to those that currently prevail.

Principle 3: The three dependencies and five limitations of AI-enabled 'predictive estimation'

All currently available semi-autonomous AI-enabled analytical tools depend on three principal conditions if they are to generate quasi-'predictive' estimates of unmeasured, unknown or hitherto unknowable variables, dataset properties or parameters. These are that:

- the dataset(s) concerned contain the 'informational perspectives' and 'statistical power' necessary to algorithmically characterize the patterns on which to generate interpolative or extrapolative estimates of unmeasured, unknown or hitherto unknowable variables or dataset properties or parameters;
- the analysts concerned have sufficient prior knowledge and understanding (i.e. empirical and theoretical evidence, subject matter expertise and domain-specific awareness) [12] to interpret, evaluate and validate the meaning and utility of any patterns identified, and the 'predictions' these support; and
- the algorithmic identification of these patterns can subsequently be faithfully and usefully replicated in broadly comparable datasets that have been derived using similar 'data- and dataset-generating mechanisms and procedures'.

Furthermore, in contemporary generative or retrieval-augmented systems, these dependencies may extend not only to the model's training data, but also to the provenance, completeness, timeliness, and retrieval logic of the external information on which any particular outputs depend. Nonetheless, even when the dependencies listed above have been satisfied, there remain five principal limitations of all 'predictions' generated by currently available, semi-autonomous, AI-enabled analytical tools. Namely that these 'predictions':

- will be more accurate estimates of the sample average value of the 'predicted' feature, than of their specific value for any individual case within the sample analyzed;
- will not necessarily be an accurate estimate of the population average value of the 'predicted' feature, unless the sample of cases analyzed is truly representative of the wider population of cases from which the sample was drawn;
- will not necessarily provide interpretable evidence of the direction, strength or precision of any (direct or indirect) causal relationships amongst the variables retained in the optimal predictive algorithm (since such inferences require carefully designed models) [31];
- will rarely offer accurate estimates of future phenomena, entities, processes, or their characteristics, and only then if the data and dataset generating mechanisms and procedures involved are not sensitive to, or dependent on, known, unknown, unpredicted, unpredictable or unprecedented spatio-temporal changes in external factors; and
- will not be definitively validated (and cannot be considered 'proven') by the subsequent occurrence of whatever it was that had been predicted, since this may still have occurred due to error or bias; or as a result of prediction-induced interventions causing 'self-fulfilling' or 'self-defeating' prophecies.

These dependencies and limitations are not abstract concerns; they point directly to the questions analysts should ask when deciding how much weight to place on AI-enabled outputs. Indeed, evaluating the validity and practical utility of any meaningful and actionable inferences that can reliably be drawn from AI-*generated* 'predictions', and from insights based thereon, requires substantial knowledge and understanding relevant to:

- the datasets *on* which, and the spatio-temporal contexts *in* which, the 'predictive' algorithms were trained and subsequently applied; and
- the subject matter expertise (i.e. the empirical, experiential and theoretical evidence, and plausible speculation) and domain-specific awareness necessary to offer robust assessments of any credible alternative explanations for the 'predictions' and insights concerned.

Some of the knowledge, understanding and skills relevant to evaluate the utility, meaning and validity of AI-*generated* outputs is already covered within the post-basic, specialist training – and subsequent in-service supervision and support – that most all-source intelligence analysts will receive. However, augmented AI literacy training [6] will be required to cover the more technical dependencies and limitations of AI-*generated* outputs if analysts are to be sufficiently competent and confident when applying these techniques – or when using insights produced by others who applied them – so as to strengthen rather than compromise their analyses and assessments.

For the analyst, the practical implications are clear: AI-*generated* outputs should never be treated as self-evident, self-explanatory or self-validating, but always as context-bound estimates whose trustworthiness depends on whether: their dependencies have been satisfied; and their limitations properly understood.

Taken together, these three principles (and the dependencies and limitations relevant to the last of these) can inform a discrete set of questions that intelligence analysts should ask when evaluating and interpreting AI-enabled outputs – questions that have been summarized in Table 2.

Table 2: Questions intelligence analysts should ask of AI-enabled outputs.

Question	Why it matters
What kind of output is this: AI-facilitated or AI-generated?	This distinction determines the level and type of scrutiny required
On what dataset (or datasets) was the output based?	Output quality depends on data quality – including the scope, coverage, and relevance of the datasets on which the AI-enabled tool was trained, and subsequently applied
Were the datasets representative of the problem and context at hand?	Unrepresentative datasets can (for some applications) weaken the accuracy, and inferential validity (and therefore utility) of outputs produced by AI-enabled tools
What exactly is being estimated, classified, clustered, or generated by this AI-enabled tool?	Determining precisely what the output represents helps limit over-interpretation, and helps analysts avoid overstating what the tool can reveal
Was the output validated, and if so in what way(s) and under what conditions?	Validation procedures are necessary to ensure the outputs of AI-enabled tools are robust; but validation in one context may not guarantee that the tool will necessarily provide comparably valid outputs in all other contexts
Were the conditions under which the tool was trained or applied comparable?	Changes – over time or place – in the conditions under which AI-enabled tools are trained and applied can substantively reduce the reliability and/or validity of the outputs these tools provide
What information is available regarding residual uncertainty?	Analysts need to be aware of any residual (and potentially irreducible) uncertainty so as to be able to calibrate their analytical confidence in the inferential insights they infer from AI-enabled outputs
Could the output reflect error, bias, artefact, adversarial manipulation, or classification constraints?	Analysts should remain alert to the possibility that alternative mechanisms might be responsible for the data-, capability-, and context-dependent outputs provided by AI-enabled tools – including the risk of chance, poor practice or deliberate manipulation and misrepresentation
What human judgment remains necessary before the output can inform assessment?	Analysts should retain responsibility and accountability for the inferential interpretations and insights they derive from the outputs of AI-enabled tools, and for the subsequent inclusion of these in intelligence assessments

Conclusion

This position paper has argued that intelligence analysts do not need to become AI developers or operators in order to use AI-enabled outputs well. They do, however, need sufficient conceptual and analytical understanding, and a modicum of technical knowledge, to: distinguish between AI-*facilitated* and AI-*generated* outputs; interpret so-called 'predictions' as context-bound *estimates*; and evaluate the dependencies and limitations that condition the utility of these *estimates*.

The practical implication is straightforward: AI literacy should be treated as a professional baseline for all intelligence analysts who are expected to consume AI-enabled outputs critically, competently, confidently, and accountably. In this sense, the effective integration of AI into intelligence workflows depends not only on the sophistication of the tools available, but also on the quality of human judgment brought to bear on the outputs these tools produce.

Acknowledgement

This position paper benefitted substantively from: insights shared by Thea de Wet from the Centre for Academic Technologies at the University of Johannesburg; technical explanations provided by former colleagues at Leeds Institute for Data Analytics (LIDA) – including Mark Gilthorpe, Marc de Kamps and Peter Tennant; and opportunities to strengthen both: the contemporary relevance of these arguments with proficient operators of AI-enabled tools – including Jim G, Dan M, Leon C, Arjun J and James J; and the accessibility of these arguments to intelligence practitioners – including Matthew B, Matt L, Dan C, Ben D and Dan D. My sincere thanks to you all.

References

1. Sanguinetti P. Why the term artificial intelligence is misleading. IE Insights. 2025 Jan 22. Available from: https://www.ie.edu/insights/articles/why-the-term-artificial-intelligence-is-misleading/

2. Mitchell K, Mariani J, Routh A, Keyal A, Mirkow A. The future of intelligence analysis: a task-level view of the impact of artificial intelligence on intel analysis. Deloitte Center for Government Insights. 2019 Dec 11. Available from: https://www.deloitte.com/us/en/insights/industry/government-public-sector-services/artificial-intelligence-impact-on-future-intelligence-analysis.html

3. Donald Rumsfeld. DoD News Briefing. United States Department of Defense. 2002 Feb 12. Available from: https://web.archive.org/web/20160406235718/

4. Thierry G. We need to stop pretending AI is intelligent - here's how. The Conversation. 2025 Apr 14. Available from: https://theconversation.com/we-need-to-stop-pretending-ai-is-intelligent-heres-how-254090

5. Pili G. A new theory of surprise: unraveling the logic of uncertainty and knowledge. Intelligence and National Security. 2023;39:695-708. Available from: https://doi.org/10.1080/02684527.2023.2278845

6. Konishi Y. What is needed for AI literacy? Priorities for the Japanese economy in 2016. Research Institute of Economy, Trade and Industry. 2015 Oct 10. Available from: https://www.rieti.go.jp/en/columns/s16_0014.html

7. Ellison G. All-source intelligence analysts as confident consumers of AI-enabled outputs: the pressing case for AI literacy training. Preprints. 2026c Jun 18:2026061383. Available from: https://doi.org/10.20944/preprints202606.1383.v1

8. Ellison G. Statistics in intelligence analysis: 'A little learning is a dang'rous thing'. The RUSI Journal. 2026a; in press. Available from: https://doi.org/10.1080/03071847.2026.2683229

9. Katz B. The analytic edge: leveraging emerging technologies to transform intelligence analysis. Centre for Strategic and International Studies. 2020 Oct 9. Available from: https://www.csis.org/analysis/analytic-edge-leveraging-emerging-technologies-transform-intelligence-analysis

10. Ish D, Ettinger J, Ferris C. Evaluating the effectiveness of artificial intelligence systems in intelligence analysis. Deloitte National Defense Research Institute. 2021 Aug 26. Available from: https://www.rand.org/pubs/research_reports/RRA464-1.html

11. Knack A, Carter R, Babuta A. Human-machine teaming in intelligence analysis: requirements for developing trust in machine learning systems. Alan Turing Institute: Centre for Emerging Technology and Security. 2022 Dec. Available from https://cetas.turing.ac.uk/sites/default/files/2022-12/cetas_research_report_-_hmt_and_intelligence_analysis_vfinal.pdf

12. NASEM (National Academies of Sciences, Engineering, and Medicine. Human-AI teaming: state-of-the-art and research needs). Washington

(DC): The National Academies Press; 2022. Available from: https://doi.org/10.17226/26355

13. Blanchard A. Taddeo M. The ethics of artificial intelligence for intelligence analysis: a review of the key challenges with recommendations. Digital Society. 2023; 2:12. Available from: https://doi.org/10.1007/s44206-023-00036-4

14. Dstl (Defence Science and Technology Laboratory). Human-centred ways of working with AI in intelligence analysis. UK Ministry of Defence. 2023 Jul 26. Available from: https://www.gov.uk/government/publications/human-centred-ways-of-working-with-ai-in-intelligence-analysis/human-centred-ways-of-working-with-ai-in-intelligence-analysis

15. ASPI (Australian Strategic Policy Institute). The future of intelligence analysis: U.S.-Australia project on AI and human machine teaming. Australian Strategic Policy Institute. 2024 Sep. Available from: https://www.scsp.ai/wp-content/uploads/2024/09/AI-The-Future-of-Intelligence-Analysis-SCSP-ASPI-Report.pdf

16. Heuer RJ. Psychology of intelligence analysis. Center for the Study of Intelligence, Central Intelligence Agency; 1999. Available from: https://www.ialeia.org/docs/Psychology_of_Intelligence_Analysis.pdf

17. Gillespie N, Lockey S, Curtis C, Pool J, Akbari A. Trust in artificial intelligence: a global study. The University of Queensland and KPMG Australia. 2023. Available from: https://doi.org/10.14264/00d3c94

18. Dolman J. Naive or native? Do digital natives really understand their world? The AI English Teacher. 2024 Sep 11. Available from: https://theaienglishteacher.wordpress.com/2024/09/11/naive-or-native-do-digital-natives-really-understand-their-world/

19. Ellison G. Might “uncertain military leadership” minimize the risks and optimize the opportunities of strategic uncertainty. Contemporary Military Challenges. In press.

20. Hanna M, Granzow D, Bolte B, Alvarado A. NATO intelligence and information sharing: improving NATO strategy for stabilization and reconstruction operations. Connections Quarterly Journal. 2017;16:5-33. Available from: https://doi.org/10.11610/Connections.16.4.01

21. Ellison G. What “typical insights” might “core statistical analytical techniques” provide to intelligence analysts? International Journal of Applied Mathematics, Sciences, and Technology for National Defense. In Press.

22. Luft J, Ingham H. The Johari window: a graphic model of interpersonal awareness. Proceedings of the Western Training Laboratory in Group Development. 1955 Aug:1-246. Cited in: Luft J. The Johari window: a graphic model of awareness in interpersonal relations. NTL reading book for human relations training. NTL Institute. 1982. Available from: https://web.archive.org/web/20260205064708/https:/ombuds.columbia.edu/si tes/ombuds.columbia.edu/files/content/pics/30 Anniv/The Johari window_A graphic model of awareness in interpersonal relations.pdf

23. Davies A, Thomson M. Known unknowns: uncertainty about the future of the Asia-Pacific. Australian Strategic Policy Institute Special Report. 2010;35:1-16. Available from: https://web.archive.org/web/20140213200346/https:/www.aspi.org.au/publica tions/special-report-issue-35-known-unknowns-uncertainty-about-the-futureof-the-asia-pacific/SR35_known_unknowns.pdf

24. Wall WA. History of optical telescopes in astronomy. Switzerland: Springer Nature; 2018. Available from: https://doi.org/10.1007/978-3-319-99088-0

25. Lopez SG. A brief history of optical microscopy. John Innes Centre Blog. 2021 May 12. Available from: https://www.jic.ac.uk/blog/a-brief-history-of-optical-microscopy/

26. Ball DW. The electromagnetic spectrum: a history. Spectroscopy. 2007;22:14-20. Available from: https://www.spectroscopyonline.com/view/electromagnetic-spectrum-history

27. Dhami MK, Mandel DR. Words or numbers? Communicating probability in intelligence analysis. American Psychologist. 2021;76:549-560. Available from: https://doi.org/10.1037/amp0000637

28. Jensen MA. Intelligence failures: what are they really and what do we do about them? Intelligence and National Security. 2012;27:261-282. Available from: https://doi.org/10.1080/02684527.2012.661646

29. de Cosmo L. Google engineer claims AI chatbot is sentient: why that matters. Scientific American. 2022 Jul 12. Available from: https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/

30. Moss MF, Koplin J. Is Richard Dawkins right about Claude? No. But it’s not surprising AI chatbots feel conscious to us. The Conversation. 2026 May 7. Available from: https://theconversation.com/is-richard-dawkins-right-about-claude-no-but-its-not-surprising-ai-chatbots-feel-conscious-to-us-282151

31. Ellison G, Rhoma H. Directed acyclic graphs as conceptual and analytical tools in applied and theoretical epidemiology: advances, setbacks and future possibilities. Mathematical Biosciences and Engineering. 2025;22:1280-1306. Available from: https://doi.org/10.3934/mbe.2025048