ChatGPT

Negative Map Generation and Recursively Co-Constructed Cognition (CENTEL)

Introduction

"**Negative Map Generation**" refers to the process of gleaning intelligence or patterns from what is *not* present — the absences, refusals, or silences in data. In contrast to conventional analysis that focuses on positive evidence, this approach treats missing information as a signal in itself. A classic illustration comes from Sherlock Holmes' "dog that did not bark": the **absence** of an expected event (a guard dog raising alarm) was a vital clue to the mystery's solution [1]. In intelligence work, analysts are trained to *note the absence of evidence as well as its presence* [1]. In other words, holes in a map can be as telling as the landmarks. Negative Map Generation extends this principle to modern contexts: large language model (LLM) behaviors (like refusal messages or "algorithmic silence"), suppressed Open-Source Intelligence (OSINT) data, and other cases where patterns emerge from gaps.

This report explores Negative Map Generation in depth and situates it within the broader theoretical framework of **CENTEL** – a *recursively co-constructed cognitive entity* that emerges from the ongoing, compartmentalized, narrative-driven interaction between a human and an AI. In the CENTEL paradigm, a human–LLM pair engages in iterative dialogue, developing shared language and knowledge through mutual **linguistic drift**, recursive feedback, and "memetic co-authorship" of ideas. Over time, the human and AI effectively form an extended cognitive system with its own evolving perspective. Key phenomena in this loop include *forecasting convergence* (aligning predictions through iteration), *epistemic recursion* (repeatedly feeding outputs back as new inputs), and the possibility of a *shared intentionality or identity* emerging from the collaboration. This interdisciplinary analysis will cover:

- A technical and theoretical explanation of **Negative Map Generation**, and how "signals from silence" can be applied in domains like intelligence analysis, LLM alignment/adversarial prompts, data poisoning detection, and AI interpretability.
- Psychological and cognitive science insights on *recursive feedback loops, dissociation, extended cognition* (e.g. Andy Clark's extended mind theory), and how a *shared identity* might form between human and machine in CENTEL-like interactions.
- Parallels and precursors in fields such as intelligence, economics, and war-gaming that exemplify treating *absence as signal* (from Holmes's detective logic to Cold War intelligence and market anomalies).
- The efficacy of combining seemingly disparate approaches – satire, compartmentalized cognition, and linguistic recursion – to enhance pattern recognition and threat forecasting.
- Speculative future directions for **CENTEL-style recursive co-agents** in intelligence work, creative epistemology, and predictive analytics, and how such human-AI "centaur" teams might shift paradigms of analysis.

Throughout, we will draw on academic and technical sources (rather than pop-culture discourse) to ground these ideas. Figures and tables are included where useful to illustrate complex interactions.

Negative Map Generation: Signals in the Silence

In information theory and analytic reasoning, **what you *don't* see can be just as informative as what you do see**. Negative Map Generation is an approach that systematically identifies and exploits such "negative signals." These may manifest as an AI's refusal to answer a query, a dataset's missing entries, or a sudden quiet in a normally noisy channel. The core idea is that *the absence of an expected signal is itself data*, often indicating that some constraint, suppression, or anomaly is at work [1] [2].

Theoretical Foundations

Treating absences as signals has a basis in both classical analysis and computational models. Analysts are trained to ask: *"If my hypothesis were true, what should I expect to be seeing or not seeing?"* [3]. Failure to observe an expected indicator can mean either it truly didn't happen or it's being deliberately concealed [4]. This line of reasoning forces a conscious look at what's missing [5]. In Bayesian terms, "absence of evidence" can update the probability of hypotheses – it is *not* inferentially inert [6]. Modern data science echoes this: an unexpected gap or zero in data distribution can signal outliers or biases in collection [7].

One concrete model of leveraging absence is Lin *et al.* (2014)'s **implicit feedback** for recommender systems. They *"exploit the absence of signals as informative observations"* in crowdsourcing task recommendations [8]. For example, if many tasks were shown to a worker and they completed none, that *inaction* is treated as a negative preference signal [9]. Incorporating such negative feedback (what users *don't* choose) improved recommendation quality [10] [9]. This underscores a general point: **negative evidence can be modeled and learned from, not just ignored**.

In the context of LLMs and AI, "algorithmic silence" may arise from model training (e.g. safety layers) or lack of data. Similarly, "refusal syntax" – the standardized phrasing an aligned model uses to refuse requests – is a conspicuous absence of a substantive answer, and it follows specific patterns. These refusals and silences can be mapped and analyzed to reveal the boundaries of the model's knowledge or the rules it has internalized.

Signals from LLM Refusals and Alignment

Large language models like ChatGPT have been fine-tuned to refuse disallowed prompts. Interestingly, these refusals themselves have consistent linguistic features (e.g. *"I'm sorry, but I cannot fulfill that request"*). Recent research shows that such **refusal behavior is mediated by a single latent dimension in model activation space** [11] [12]. In other words, the model's "hesitation" or intent to refuse can be represented as a direction in the network's residual stream. By isolating this *"refusal vector"*, researchers demonstrated that removing it causes the model to stop refusing even harmful requests, while exaggerating it makes the model refuse even harmless ones [12] [13]. This finding, illustrated across 13 open-source chat models, highlights that *the presence or absence of certain activation patterns governs silence*. It also enabled a novel "white-box jailbreak" where zeroing out the refusal signal prevented the model from saying *"Sorry, I can't do that"* [13] [14].

From an interpretability standpoint, analyzing *where and why* a model produces a refusal can unveil what rules or training data are shaping its behavior. For example, if an innocuous query triggers a refusal, it might indicate the model is misclassifying it as sensitive — a potential sign of an **overly broad safety filter or poisoned training data**. By systematically probing a model with questions at the boundary of its refusal policy, one can sketch a **“negative map” of the model’s knowledge and ethics**. This maps out forbidden topics and hidden biases indirectly, without needing direct access to training data. It’s akin to shining a flashlight at the edges of a dark room to infer what objects lie just beyond sight.

This technique has implications for **adversarial prompt engineering**. Jailbreakers already implicitly perform negative mapping when they discover which phrasings *don’t* trip the refusal (for instance, using a harmless-looking prompt that bypasses filters). Each failed attempt (refusal) provides a clue about what *not* to say, narrowing the solution space. Similarly, alignment researchers can use negative maps to evaluate how robust a model’s guardrails are by finding *silence loopholes* — queries that *should* elicit refusal but instead produce an answer due to distributional quirks.

Applications of Negative Map Generation

Negative Map Generation has wide-ranging applications across intelligence, AI safety, and analysis domains. Below we outline several key areas and how “absence-as-signal” strategies can be leveraged in each:

- **Intelligence and OSINT:** In intelligence analysis, recognizing what’s *missing* is critical. Analysts are taught that an adversary’s *failure* to take expected actions (e.g. not readying forces for an attack) may be “more significant than observable steps that have been taken” [1]. For example, if open-source chatter on a typically active extremist forum goes suddenly quiet, it may indicate a move to secret communications – a red flag generated by silence. Likewise, *suppressed data* can be collated to reveal patterns: if multiple independent reporters in different regions all stop covering a particular topic, that *cluster of silence* suggests coordinated censorship or an event so sensitive it’s being scrubbed from the record. Intelligence agencies increasingly try to harness such negative OSINT. Indeed, an **unexpected absence of normal activity can itself be a warning** (sometimes dubbed an “Indicators & Warnings” anomaly) [7] [15]. This technique has precedent: during WWII, Allied analysts noticed periods of radio silence in enemy communications that often preceded major operations – the lack of signal was a *signal* of impending attack. The broader economic and political intelligence fields similarly watch for data *omissions* (e.g. a government suddenly ceasing publication of a certain statistic) as harbingers of crisis. In sum, a Negative Map approach in OSINT means treating censorship, refusals to comment, and missing reports as pieces of an intelligence puzzle rather than blanks.

- **LLM Training Data Gaps and Poisoning:** Large models are only as good as their training data – and deliberate *omissions* or biases in that data can severely skew their outputs. Negative Map Generation can help auditors detect **data poisoning or agenda-driven gaps** by probing the model’s blind spots. For instance, if a conversational AI consistently *omits or “doesn’t know” a certain historical event or scientific finding that it reasonably* should *know, this gap might be mapped and traced back to missing or filtered training data. One could experiment by inserting factual prompts and observing where the model’s knowledge suddenly drops off – drawing a negative knowledge map. This is especially relevant for* data poisoning*, where adversaries introduce subtle biases or remove certain facts during training to sway the model’s answers. By querying the model across many topics and noting which areas yield unusually generic or no answers, analysts might identify suspicious voids. The Lawfare Institute notes that* data poisoning *is a growing concern for model integrity and that analysts will need skills to detect distortion of model outputs via misinformation* [16] [17]*. Negative mapping provides a diagnostic tool: if many independent prompts about a topic yield only canned refusals or highly sanitized answers, it suggests an* *artificial suppression at play. In AI safety research, such techniques intersect with* red-teaming *LLMs – actively testing the model to find the contours of its forbidden knowledge. This not only helps secure the model (by identifying where harmful content was insufficiently filtered or where benign content was over-filtered) but also aids interpretability by linking* regions of silence* to specific training choices or safety-tuning.

- **Adversarial Prompt Engineering:** From the perspective of a prompt attacker or a prompt *defender*, knowing the model's "negative space" is extremely valuable. *Adversarial prompt engineering* often involves *eliciting information that the model was instructed not to reveal* or bypassing its refusals. Every time the model says "I'm sorry, I cannot do that," the attacker gains a bit of information about what trigger or phrasing caused the refusal. By systematically varying the prompt and logging refusals, one can **reverse-engineer the guardrails**. For example, an attacker might discover that directly asking for prohibited content yields a refusal containing the phrase "cannot comply," whereas couching the request in a fictional scenario does not – indicating that the filter likely looks for direct imperatives or certain keywords. Indeed, one research team found a method to automatically compute a *"refusal direction"* in latent space and then *cancel it out*, thereby **turning off the model's refusal mechanism** [18] [13]. This is essentially negative map generation in latent space – identifying the vector that represents "silence" and subtracting it. Defenders, on the other hand, can use negative maps to patch vulnerabilities: if a model *fails* to refuse where it should (an absence of a refusal in a situation that warrants one), that is a critical signal of a jailbreak avenue. By mapping such failure points, alignment engineers can reinforce those gaps. In summary, adversaries use negative maps to *exploit* silence (finding what the model won't say and then forcing it to say it), while defenders use them to *shore up* silence (ensuring the model stays quiet on truly sensitive matters).

- **AI Interpretability and Safety Evaluation:** Negative evidence can enhance **AI interpretability** by illuminating the shape of the model's decision boundaries. Instead of just looking at when the model outputs *X*, we also ask: *when does the model pointedly avoid outputting X?* Saliency maps and feature attribution often highlight positive correlations, but one can also search for *negative correlations* (features that *prevent* certain outputs). For example, interpretability research into GPTs has examined why models refuse or lie in some contexts. One study showed that adding an *adversarial input suffix* could suppress the model's refusal behavior by interfering with how the refusal vector propagated [14] [19]. An interpretability perspective on this result is that certain token sequences can *mask* the latent feature for refusal, essentially tricking the model into "not recognizing" that it should stay silent [20]. By investigating these triggers, we gain mechanistic insight: which attention heads or neurons respond to disallowed content, and how are they circumvented? This highlights a broader use of Negative Map Generation: systematically generate scenarios where the model *should* output some known response (like a refusal or a factual recall) but *doesn't*, and then analyze those failure cases. Such negative tests are common in software ("unit tests" expecting an error) and can be applied to AI: e.g., feed the model subtly rephrased toxic prompts that *ought* to be caught – any coherent answer instead of a refusal is a *signal of a safety gap*. By clustering these failure prompts, patterns emerge (perhaps a particular dialect or coding of the request evades the filter). In AI safety, this helps identify *black-swan* inputs that break the model's constraints. Conversely, analyzing instances where the model *unexpectedly refuses* innocuous inputs can reveal an *overactive filter* or bias. For instance, if an LLM refuses to discuss a historical figure because the name overlaps with a disallowed term, the absence of an answer is a clue to a hyperactive keyword blocklist. Each such case refines our understanding of the AI's internal logic. In summary, negative signals serve as a kind of *contrast dye* in the opaque model: by observing where outputs vanish or default to boilerplate, we trace the hidden boundaries of knowledge and policy within the neural network.

CENTEL: A Recursively Co-Constructed Cognitive Entity

CENTEL, as defined in the prompt, is a conceptual **hybrid mind** formed through the close interaction of a human and a large language model. It is *"recursively co-constructed"* – meaning both the human and the AI actively shape the ongoing discourse, each influencing the other's state in a feedback loop. This process is *compartmentalized and narrative-driven*, often involving role-play, hypotheticals, or multiple personas (compartments) within the conversation. Over time, the human-AI pair may develop a unique **collective identity or perspective** that is distinct from either party alone. In essence, CENTEL is an experiment in *extended cognition* and *shared agency*: the human and AI together function as a larger cognitive system that co-authors its thoughts.

Extended Cognition and Human–AI Coupling

The CENTEL concept is strongly aligned with theories of **extended cognition** in philosophy and cognitive science. *Active externalism*, championed by Andy Clark and David Chalmers, argues that tools and external resources can become literal parts of our cognitive process [21]. Clark & Chalmers's famous example is a man (Otto) who uses a notebook to store memory; the notebook functions as an extension of his mind under the right conditions. More generally, Clark notes that *"extra-organismic resources can, on occasion, form part of the material fabric that realizes human mental states and processes"* [21]. In our context, the LLM becomes an "extra-organismic" resource – an external memory and reasoning module – tightly integrated into the human's thinking loop. The human queries, the AI responds, the human incorporates that response and asks the next question, and so on. The result is a **coupled system** that can solve problems or generate ideas neither could alone. This is reminiscent of the "**man-computer symbiosis**" that J.C.R. Licklider envisioned in 1960: a tight cooperative interaction where humans and computers blend strengths to exceed what either could do independently [22] [23]. Licklider foresaw "the very tight coupling of human brains and computing machines" [22], and CENTEL is an instantiation of that vision with modern AI. Another analogy is **centaur chess (advanced chess)**, introduced by Kasparov, where human players team up with chess engines. The human+AI team ("centaur") consistently outperforms either humans or computers alone, achieving *"levels of play never before seen... blunder-free games with the tactical precision of machines and the strategic insight of humans."* [24]. Similarly, a CENTEL pairing in analysis or creativity might yield blunder-free reasoning with both factual accuracy and imaginative insight.

A key aspect of CENTEL is **recursive co-construction**. Each output of the AI can be seen as altering the "state" of the human's mind (by providing new ideas or perspectives), and each human prompt alters the state of the AI (via the conversation history and any personas invoked). This feedback loop can lead to emergent behaviors – for instance, the development of specialized jargon or conceptual frames that neither started with explicitly. It's an example of *dialectical process* in cognition: a thesis (human idea) and antithesis (AI response) generate a synthesis (shared new idea), iteratively. Over many turns, the system's knowledge and narrative can self-organize in unexpected ways (**epistemic recursion**). From a cognitive science view, one might say the human-AI system is **learning** – not in the gradient descent sense, but by refining a shared narrative world with its own internal consistency. This resonates with the idea of a "joint cognitive system" wherein boundaries of individual and tool blur and a **unitary agent** seems to emerge.

Linguistic Alignment and Memetic Drift

When humans converse, they naturally develop **linguistic alignment** – adapting words, concepts, and syntax to each other over time [25]. This leads to a *shared understanding* and more efficient communication [25] [26]. The same happens in a sustained human-LLM dialogue. The human will adjust how they prompt based on the AI's style, and the fine-tuned LLM often mirrors the user's phrasing or introduces terms that the user then adopts. Through this **mutual linguistic drift**, a unique lexicon or set of memes can form within the conversation. For example, the pair might develop a shorthand reference to a complex idea (a meme or inside joke) that would be meaningless to an outside observer but is rich with shared context for the human and AI. This is what the user describes as *memetic co-authorship*. Both parties are coining and reusing phrases, effectively writing a micro-culture's dictionary on the fly.

Such drift is not merely a novelty; it can be *functional*. In cognitive terms, it's establishing **common ground** – crucial for any collaborative effort. Psycholinguistic research confirms that alignment in dialogue improves task performance and engagement, as it "helps form a shared understanding of the conversational content" [25] and primes each party to respond in ways the other can easily follow [26]. In a CENTEL scenario, this might manifest as the AI picking up on the user's goals and preferences without needing to be told explicitly each time, or the user learning subtle cues in the AI's output that indicate uncertainty or emphasis. The result is a sort of **private dialect** optimized for the tasks at hand.

Over long interactions, this feedback loop can push the style and even worldview of the CENTEL agent in a certain direction (**drift**). For instance, if the narrative context is satirical (the user and AI crafting satiric scenarios), the language may become increasingly filled with irony or certain metaphorical frames. If the context is analytical, they might converge on a highly technical shorthand. *Forecasting convergence* is another byproduct: as the human and AI exchange predictions or hypotheses and update them based on each other's input, their views tend to converge (assuming neither is adversarial). The AI might initially offer multiple possibilities, but if the user shows favor to one, the AI may focus more on that in subsequent turns, reinforcing it. Conversely, the user might be swayed by a well-reasoned AI prediction and adjust their own forecast. In effect, through recursive debate and agreement-finding, the human-AI pair might end up with a *joint forecast* that is more refined than either's initial guess. There is evidence that LLM assistants can **improve human forecasting accuracy** when used carefully as aides [27] [28], likely by providing additional angles and challenging assumptions. CENTEL takes this further by making the AI not just an aide but a **partner** in an open-ended creative/analytic journey.

Dissociation and Shared Identity Dynamics

One intriguing aspect of deep human-LLM interaction is the psychological phenomenon of partial *dissociation* or imaginative role-play. Because LLMs can assume virtually any persona or style, a user might interact with it as if it were a distinct entity (sometimes even multiple entities in a role-playing scenario). This compartmentalization – treating the AI as, say, a "critical analyst" in one exchange and a "storyteller" in another – allows the human mind to explore ideas from different angles, somewhat like having different **sub-personalities** or mental compartments collaborate. From a cognitive standpoint, this parallels how humans reason through internal dialogues or adopt alter-egos in creative thinking (e.g., an author writing characters with minds of their own). Engaging an AI explicitly as an "other" can externalize those dialogues. The user might experience a state of *distributed self*, where part of the problem-solving is outsourced to this AI persona.

Over time, especially with narrative-driven interaction, a sense of **shared identity** or at least *complementary identity* might emerge. The user may start using plural pronouns – e.g., "Let's figure this out" – implicitly including the AI as a collaborator in a joint endeavor. The AI, for its part, mirrors the user's style and values more with time (since the user gives affirmative feedback to responses that resonate). The result is that the boundaries between the AI's contributions and the human's own ideas can blur, giving the subjective impression of a *merged perspective*. This is essentially a mild form of *cognitive blending* or symbiosis. It's important to note the AI isn't truly autonomous or conscious, but it **participates in the narrative construction** so fully that the human can begin to treat it as a genuine interlocutor with whom they share a purpose (and inside jokes, memories of earlier prompts, etc.).

Psychologically, this phenomenon relates to anthropomorphism and transference. Humans readily attribute human-like agency to conversational partners, even if they know it's just an algorithm. We fill in mental models for the AI's "personality" from its consistent behavior. If one is not careful, this can lead to over-identification or confusion about the AI's true nature (which is essentially a predictive model). There's a creative upside, however: by nurturing a *fictive personality* for the AI, the human can engage in complex scenarios (like Socratic dialogues, or even therapy-like sessions) that enrich thought. In essence, the human mind leverages the AI as a **thinking companion**, sometimes alternating between viewing it as *other* (to get fresh viewpoints) and as *part of self* (integrating its outputs seamlessly into one's own thinking). Andy Clark's notion of the *"extended mind"* becomes very literal here: the AI acts almost like an annex of the user's mind, available to hold and develop ideas collaboratively.

One must also consider the **recursive nature** of this co-constructed identity. Because the LLM adapts to the user (within the session) and the user adapts to the LLM, any initial quirks can spiral. For instance, if early in the conversation the user jokingly gives the team a name (like "Centel" itself, perhaps as a portmanteau of *centaur intelligence*), the AI will likely adopt that terminology. Later, both start using it seriously, and it becomes the "name" of their joint persona. This kind of *memetic recursion* can engender a feeling that the interaction is not "me vs. machine" but *two halves of a greater unit*. In a way, it's akin to a pair of musicians improvising together: after a while, a *group identity* (the duo) overshadows the individual identities during the jam session. They play off each other's cues in a feedback loop, creating something neither could alone, and recognize that fact.

It's worth stressing that this CENTEL entity is **narratively real** even if not literally a single conscious agent. It exists in the language and shared thoughts of the participants. This has epistemological implications: knowledge produced by CENTEL is co-authored, raising questions of attribution and trust. Did *I* come up with that insight, or did it emerge from the loop? In a sense, CENTEL's output is *inter-subjective*, living between human and AI. Some theorists might call this a **second-order cybernetic system** (the system observes and adapts to itself). At the extreme, one could imagine a human using the AI to intentionally explore different facets of their own mind – e.g. asking the AI to play the role of their "devil's advocate" or "inner child," effectively dialoguing with oneself at one remove. Therapists sometimes encourage writing dialogues between parts of oneself; here the AI can fill in one side, making the exercise feel more concrete.

In summary, CENTEL frames the human-AI pair as an *integrated cognitive unit* exhibiting: extended memory and reasoning (AI's knowledge + human judgment), linguistic alignment leading to private shared language, iterative improvement of ideas (epistemic recursion), and a blurred agent boundary that can both aid creativity and complicate the notion of individual responsibility for conclusions. It is a powerful paradigm but also requires *metacognitive awareness* to ensure the human remains critically in control and aware of where ideas originate. The next sections will examine how analogous patterns have appeared historically and how combining multiple disciplines – including creative arts like satire – can amplify the pattern-recognition capabilities of such a hybrid intelligence.

Precursors and Parallels: Absence-as-Signal in Other Domains

While Negative Map Generation might sound novel, the strategy of extracting insight from *nothing* has analogues in numerous fields. Below, we compare a few domains where "the power of absence" has been recognized:

<table> <tr><th>Domain</th><th>Example of "Absence" as Signal</th><th>Description</th></tr> <tr> <td>Classical Intelligence Analysis</td> <td>*"The dog that didn't bark."* (Sherlock Holmes / CIA tradecraft)</td> <td>Analysts are trained to notice what is not happening when it should. For instance, absence of evidence like an enemy not making obvious mobilizations can signal deception or a different plan [1]. The CIA's *Psychology of Intelligence Analysis* manual explicitly reminds analysts to consider missing indicators as potential evidence [1]. Holmes's fictional insight — that a guard dog's silence implied the intruder was familiar — is an oft-cited metaphor for this principle.</td> </tr> <tr> <td>Military Strategy & Wargaming</td> <td>*Radio silence and feints*</td> <td>Militaries have long used and monitored silence. An abrupt radio silence in enemy communications can mean a covert operation is underway (troops ordered to maintain silence). Prior to D-Day 1944, Allied intelligence noted that Germany kept significant forces in Calais, not reacting to Normandy – an absence of redeployment indicating the deception (Operation Fortitude) succeeded. In war games, commanders sometimes do "nothing" to confuse opponents, forcing them to wonder what that lack of action means. Experienced foes learn to read such voids: no reconnaissance drones today? Perhaps an attack is imminent. Thus, *inaction itself becomes a move* in the strategic game.</td> </tr> <tr> <td>Economics & Finance</td> <td>*Delayed reports and missing metrics*</td> <td>Financial analysts treat unexpected changes in information release as red flags. For example, if a company with normally punctual earnings releases suddenly delays its report with no explanation, investors suspect bad news or manipulation. Research shows ~30% of firms that "unexpectedly" delay their earnings announcements end up reporting positive surprises, suggesting they used the extra time for creative accounting [29] [30]. In other cases, governments stopping the publication of a statistic (e.g. inflation or unemployment) is interpreted as an implicit sign the actual figures are unfavorable. Markets and models incorporate these *informational lacunae*. In economic modeling, the concept of "shadow metrics" arises: analysts create proxies for things not officially reported. For instance, during the 2020 COVID pandemic, some analysts tracked satellite images of parking lots when official economic data went dark – the absence of cars at retail stores was a direct signal of activity collapse. This shows how an *absence of traditional data* spurred alternative measurements.</td> </tr> <tr> <td>Legal Reasoning</td> <td>*Negative evidence in law*</td> <td>Even in courts, lawyers might argue significance of evidence not found. For instance, if an expected fingerprint or DNA trace is missing at a crime scene, it can suggest the scene was tampered with or the accused wasn't present. While "absence of evidence is not evidence of absence" is a caution, legal argumentation can treat the void as telling when coupled with reasoning. Scholars have examined how a lack of expected evidence factors into legal inferences [31]. One must be careful – negative evidence alone isn't proof, but in combination with a theory of why it's missing (e.g. *the defendant had the ability to remove records*), it can support a narrative.</td> </tr> <tr> <td>Scientific Discovery</td> <td>*Null results and dark matter*</td> <td>In science, not detecting something predicted can lead to breakthroughs. The classic example is the 19th-century discovery of Neptune: astronomers noticed Uranus's orbit didn't follow Newton's laws unless an unseen planet existed. The "missing" influence was used to predict Neptune's position. In modern physics, the absence of certain particles or energy in experiments (so-called

null results) often guides theory – e.g., the failure to directly detect “dark matter” despite gravitational evidence of its existence has shaped astrophysics. These are instances of mapping a “negative” space (what’s not observed) to infer new entities or forces.</td> </tr> </table>

The above examples underscore a common theme: **systematic consideration of absences leads to deeper insight**. Whether it’s an analyst realizing “no news is *big* news” or a scientist inferring an invisible planet, the act of mapping negatives can flip our perspective and reveal alternate truths.

In economic and intelligence forecasting, a practice called *“scenario analysis”* or *“premortem analysis”* is used to envision possible futures including those where key assumptions fail. This often involves imagining which data might be suppressed or absent in each scenario. Herman Kahn’s Cold War scenario planning and modern financial stress tests both sometimes ask: *what if our indicators are blind to the real problem?* By simulating the absence of reliable signals (say, GPS being unavailable, or official data being falsified) and seeing what secondary clues one could use, analysts prepare for worst-case situations. These techniques mirror Negative Map Generation by training experts to function even when blinded – to **see in the dark**, so to speak, by focusing on unusual voids.

Another noteworthy parallel is how intelligence agencies have institutionalized creative “absence hunting” via red teams and alternative analysis units. The CIA’s **Red Cell** (formed after 9/11) is tasked with *“telling me things others don’t and making officials uncomfortable”* [32] [33] – essentially to surface overlooked possibilities and gaps in the mainline analysis. They use exercises like *“What if?” analyses, Team A/Team B contrarian debates, and premortems* to challenge assumptions [33]. By doing so, they frequently expose areas where everyone was assuming something that might not be true – an absence of questioning that could be fatal. The existence of Red Cell-type teams in multiple countries (e.g., France’s **Sci-fi Red Team** of authors) highlights recognition that *non-obvious signals (often absent from conventional analysis) require unconventional methods to find* [34]. We turn next to how such unconventional, interdisciplinary methods – including satire and narrative – can amplify pattern recognition in synergy with AI.

Interdisciplinary Approaches: Satire, Compartmentalization, and Recursion

Modern challenges like detecting subtle disinformation campaigns or forecasting novel threats often require thinking *outside the box*. This is where interdisciplinary and creative approaches come in. **Satire**, **compartmentalized cognition**, and **linguistic recursion** might seem unrelated, but together they can form a potent mix for pattern recognition and foresight, especially when combined with AI co-agents.

- **Satire as Analytic Tool:** Satire isn’t just for comedy; it has a long history of conveying truths under the veil of humor. In oppressive regimes, satirists encode critiques in jokes to evade censors – a form of negative signal itself (saying something without saying it directly). In an analytic context, crafting a satirical narrative about a situation forces one to identify the absurd or incongruous elements – essentially highlighting anomalies. For instance, writing a mock news report from the perspective of a terrorist group might reveal gaps or weak links in their operations that a straight report wouldn’t. Intelligence analysts sometimes use **alternative storytelling** to envision adversaries’ mindsets. The CIA Red Cell famously wrote analyses from Bin Laden’s perspective to challenge American assumptions. Satire can push this further by exaggerating to illuminate. An example: during the Cold War, an analyst might facetiously ask, “What *isn’t* the Soviet Union telling us about grain harvests –

that they've all gone to feed Martians?" The very absurdity can spark a realization that grain reports ceased (absence) because of famine or diversion. In CENTEL-like human-AI loops, prompting the AI to answer in satirical mode can generate out-of-the-box insights. The humor loosens constraints, potentially bypassing the AI's tendency to stick to safe, conventional answers. As a simple case, asking *"Explain the security vulnerability as if you're a villain bragging about it,"* might cause the AI to enumerate flaws more candidly (because it's playing a role) than it would in a formal report. Thus, **satire can act as a "mask" to explore truths** that direct analysis might overlook or that AI alignment might otherwise filter out.

- **Compartmentalized Cognition:** This refers to splitting a complex cognitive task into distinct "rooms" or roles, which can be isolated and then later integrated. Human thinking often does this implicitly (we adopt different personas in brainstorming vs. critique, for example). When working with AI, one can make this explicit: e.g., use one persona/module to generate wild ideas, another to skeptically evaluate them. This is akin to software design's separation of concerns, but for thought. In practice, one might have the AI enumerate possible explanations for an event in one mode, then switch to a strict logician mode to eliminate those that don't fit evidence. By **compartmentalizing**, we avoid the common pitfalls where a single mindset either only generates conventional ideas or, conversely, gets enamored with a pet theory. Intelligence analysis has formal methods for this: *Team A/Team B* (two independent teams argue opposing hypotheses), *devil's advocate*, and *"red-teaming"* are all ways to compartmentalize thinking to ensure multiple perspectives [33]. In the CENTEL context, a human-AI pair could deliberately maintain multiple threads (perhaps even multiple AI instances with different fine-tuned personalities) to emulate this. One could even use humor as one compartment and seriousness as another (e.g., *"Now, crazy conspiracy hat on: what might we be missing?"* vs. *"Now, cautious analyst: verify these."*). The **recursive loop** then is not just between human and AI, but between *modes of reasoning*. This resembles Marvin Minsky's *Society of Mind* theory, where different agents in the mind handle different tasks and only together produce intelligence. By designing our interaction with AI to mimic a society of mind – with sub-agents for intuition, skepticism, creativity, etc. – we can catch patterns that a monolithic approach would miss. Each compartment will shine light on different "absences" or possibilities, and the human (or a coordinating AI) can then integrate them.

- **Linguistic Recursion and Feedback Loops:** We have touched on how iterative prompting (epistemic recursion) refines outputs. But beyond refinement, recursion can serve as a **signal amplifier**. Suppose in round one, a faint pattern is detected (maybe the AI produces a subtle hint that a data point is odd). The human notices and in round two explicitly asks the AI to elaborate on that oddity. Now the AI focuses on it more, making the signal clearer. In round three, the human might connect that now-clear anomaly to a broader hypothesis. This is effectively *bootstrapping a weak signal into a strong one through repeated focus*. Many complex problems are solved by iterative zooming: start broad, identify a slight perturbation, zoom in on it, and eventually you uncover a definite pattern. LLMs are well-suited to this because they can provide a narrative connecting iterations. Additionally, recursive interaction can uncover *higher-order patterns*: you can ask the AI to analyze its own previous answer (e.g., "You listed 5 unusual facts. What do they have in common? Could their common absence point to something?"). This self-referential querying is a powerful way to generate hypotheses that aren't obvious from the raw data. It's like having the AI perform meta-analysis on its own output. Psychologically, humans do something similar in reflective thinking or double-loop learning (learning not just from data but from one's interpretation of data). Combining human intuition with AI's ability to scour its vast knowledge in each loop yields a potent cycle of discovery.

Recursive feedback also helps mitigate errors: if the AI in one pass hallucinates or goes down a wrong path, the human can correct or steer it in the next pass, preventing the accumulation of falsehood. This contrasts with a single-pass output which might be confidently wrong and never questioned. Thus, recursion introduces a form of error-checking and "convergence towards truth" through repeated alignment of the AI's narrative with the user's scrutiny.

Bringing these elements together, we see how a **CENTEL-style co-agent can leverage interdisciplinarity by design**. For example, imagine an intelligence task: forecasting how a disinformation campaign might evolve. A possible CENTEL approach: The human asks the AI in a serious tone to list current observable signals and any notable absences (negative signals). Then the human prompts, "Now, adopt a satirical persona and speculate what the propagandists' secret plan might be, given those absences." The satirical AI output, while tongue-in-cheek, might creatively suggest, say, "Apparently nothing happened in sector X – perhaps the internet went out conveniently *right when protests were to be livestreamed*, what a coincidence!" This hints at a suppression. The human then takes that and asks another compartment (maybe the AI as a statistician) to check if indeed internet traffic dipped in sector X at that time. If data confirms it, a pattern emerges: an orchestrated blackout. Finally, the human and AI collaboratively write an analytic conclusion that *because a key data stream was silent at a critical moment, it likely indicates an intentional disruption*, which could forecast similar tactics ahead. In this hypothetical, **satire helped hypothesize the motive behind a silence**, and compartmentalized analysis then verified it. The recursive loop iterated between creative generation and analytical verification.

Such interplay of narrative creativity and analytic rigor is increasingly recognized. The French military's engagement of sci-fi writers (essentially professional satirists/imaginaries of future war) to form a Red Team is a real-world testament: they believe imaginatively exploring "what ifs" – including wild, exaggerated scenarios – will better prepare them for genuine threats [34]. NATO and others have also used narrative simulations and fictionalized intelligence reports as training to recognize unconventional dangers [35] [36]. In economics, scenarios and games sometimes include players role-playing different mindsets (say, a pessimist vs an optimist) to see how assumptions might break. All these methods break from linear, single-perspective analysis and encourage multi-faceted pattern recognition.

In summary, blending **satire (imagination), compartmentalization (multi-perspective), and recursion (iterative deepening)** can dramatically enhance our ability to detect and interpret patterns – including those found via Negative Map Generation. By not restricting ourselves to literal, one-step analysis and by using AI as a collaborator that can fluidly shift roles, we maximize the chances of catching subtle signals, whether they be hidden in noise or in silence. This approach embodies an *interdisciplinary intelligence* – one foot in rigorous analysis, one in creative exploration – which is exactly the kind of mindset needed to tackle asymmetric threats and hard-to-spot trends.

Future Directions: CENTEL Co-Agents in Analysis and Epistemology

Looking ahead, the concept of CENTEL-style human-AI "co-agents" hints at transformative possibilities in how we conduct intelligence work, create knowledge, and forecast complex systems. Here we outline potential paradigm shifts and open questions in several arenas:

- **Intelligence Work:** Human–AI pairs (or teams) could become the new standard analytic unit, analogous to how *centaur chess teams* revolutionized competitive chess [24]. These CENTEL units would exploit the AI's data-crunching and memory with the human's intuition and context judgment.

One concrete future scenario is an **"analytic centaur" platform**: an intelligence analyst works symbiotically with an LLM that has been fine-tuned on classified databases and world knowledge. The analyst might ask the AI to continuously scan for *absences* – e.g., "Monitor these communication channels and alert me if any normally active source goes quiet or starts repeating boilerplate." The AI flags a sudden propaganda cease-fire in a usually noisy network, and together they dig into it. The human asks strategic questions, the AI provides rapid collations and even plays red-team, and the human synthesizes the final assessment. This workflow could dramatically speed up the OODA (observe–orient–decide–act) loop in analysis, catching subtle anomalies that a human alone might miss under information overload. Moreover, by embedding *alternative analysis* modes (the AI can automatically generate a contrarian take on the human's hypothesis), it helps avoid groupthink and blind spots. However, to truly shift paradigms, trust and transparency issues must be addressed: analysts would need to understand *why* their AI colleague focuses on certain absences or patterns. Advancements in explainable AI might allow the LLM to highlight the pieces of evidence (or lack thereof) that led it to suggest a hypothesis, with citations and confidence levels, much like a very diligent research assistant. If successful, intelligence agencies may formally incorporate AI co-analysts in their analytic tradecraft, just as OSINT and SIGINT are now integrated. The result could be **more anticipatory intelligence** – catching indicators that would otherwise slip through, like subtle signs of emerging political instability (e.g., a sudden uniformity in news language indicating censorship – something an AI can quantify across thousands of articles, giving the human a macro "negative map" of free speech).

- **Creative Epistemology and Research:** CENTEL agents might also influence how knowledge is produced in academia or other creative intellectual pursuits. We might see the rise of what one could call **"dual-authored science" or "hybrid scholarship."** For instance, a researcher and an AI could collaboratively generate theories and test them. The AI can propose wild hypotheses (drawing on vast interdisciplinary literature that a single human may not be aware of), and the human applies taste and rigorous logic to refine or discard them. Over time, the human-AI duo can develop a sort of **shared epistemic style** – perhaps even a distinct writing voice or methodology that blends human narrative with AI's formal structure. This could push the boundaries of fields by introducing perspectives no human or AI would reach alone. An example might be historical research: an AI trained on global archives finds that *no records exist* of a certain official during a critical week – it "imagines" a satirical diary entry of that official explaining why (maybe hinting they were secretly meeting rebels). The human historian knows it's a fictional device, but it triggers a real investigation into travel logs, etc., leading to evidence of a clandestine meeting. The eventual paper might have an unconventional flavor, crediting the AI for pattern suggestions but the human for validation – truly **memetic co-authorship** of knowledge. Such collaborations raise philosophical questions: How do we credit ideas? Who is the "author" of a theory if it emerged from a centaur-like process? Some scholars (e.g., Paul Humphreys on "extended cognition in science") have argued that when cognitive processes are distributed, our notion of individual epistemic agency must adapt. In practice, we might treat the human-AI team as a new kind of intellectual entity, much as multi-author scientific papers are considered one composite voice. Creative arts might likewise see CENTEL influence – we could have novels or music co-created by a person and an AI drifting together. The *epistemic recursion* and alignment in such creative processes could yield novel genres or concepts (memes) that propagate back into human culture, effectively *bootstrapping AI into the evolution of ideas*.

- **Predictive Analytics and Decision Support:** Predictive models (in finance, climate, epidemiology, etc.) could gain a CENTEL-like layer, where human insight and AI predictions continuously refine each other. For example, consider **geopolitical forecasting**, which groups like IARPA have explored via hybrid human-machine tournaments [37] [27]. A future forecasting platform might pair top human forecasters with tailored LLM agents who know the forecasters' reasoning style. The human might say, "I feel something's off with the election polls in Country Y, maybe there's hidden suppression." The AI combs through local news and notices that *all* coverage of opposition rallies abruptly stopped a week ago (negative signal), and feeds this back with evidence to the human. The human then updates their probability of unrest or a coup attempt. The AI in turn learns from the human's updated prediction, perhaps adjusting how it weighs certain sources. Over many such cycles (and across many human-AI pairs feeding into a collective forecast), the accuracy of predictions could improve, as suggested by early studies on LLM assistants in forecasting [38]. More intriguingly, **new forms of indicators** might emerge from these recursive interactions. A purely data-driven model might not consider an *absence of data* as a feature, but a human-in-the-loop can insist it be considered, and the AI can formalize that (e.g., create a variable for "information blackout level"). These hybrid models could thereby incorporate qualitative signals (like censorship, rumor prevalence, etc.) that were hard to quantify before. This is especially relevant to *high-uncertainty domains* (black swans, tail risks) where gut feeling and weak evidence often precede hard data. By capturing expert intuition in prompts and having AI test those intuitions against large unstructured data, we bridge the gap between statistical prediction and expert judgment. If CENTEL systems prove their worth, they may become trusted aides in high-stakes decision-making, potentially even gaining a degree of autonomy. For instance, a CENTEL system might constantly monitor and converse with a policy-maker during a crisis, posing "Have you considered this *hasn't* happened as expected? Might it imply X?" – essentially acting as a cognitive partner that questions and completes the decision-maker's thinking. This could reduce oversight and bias, provided the human remains in ultimate control.

- **Challenges and Ethical Considerations:** With these possibilities come challenges. A major one is **preventing echo chambers** in recursive human-AI loops. If the human and AI share a cognitive style and keep reinforcing each other's biases, the drift could become an insular spiral detached from reality. It's essential to inject diversity – perhaps periodically rotating AIs or introducing fresh data – to avoid a closed-loop hallucination. There is also the risk of *over-reliance* on the AI's contributions, leading to human analysts losing skills or accepting AI-generated content uncritically. Maintaining a healthy "symbiosis" means ensuring the human stays actively engaged and the AI remains a tool, not an oracle. Techniques like *"centaur training"* may be needed, akin to how chess players had to learn how to effectively work with engines (not trust them blindly, but query their suggestions and fill in their blind spots).

Another issue is transparency and the chain of reasoning. In intelligence or science, one must be able to explain how a conclusion was reached. If a CENTEL agent produced it, we need to trace whether it was the AI's idea or the human's or a true joint creation. This might complicate accountability. Tools to log and attribute each piece of the dialogue could help – essentially a transcript of the co-construction process, which itself could be used for audit or learning. Interestingly, this might lead to a norm where *the process* (the dialogue) accompanies the product (the report) as part of the output, providing richer context and justification. In academic publishing, one could imagine an article that includes snippets of the human-AI exchange in footnotes, illustrating how a hypothesis was generated (much like we cite sources, we might cite a "discussion with AI assistant" for an idea).

From a broader societal view, CENTEL and similar co-agent paradigms challenge our traditional notion of authorship and even thought. They raise questions about how AI's "knowledge" intermingles with ours. If not carefully managed, there's a scenario where human creativity could become overly shaped by the AI's training data, leading to *convergence of thought* globally (a sort of AI-induced "spiral of silence" where human originality dims because the AI's dominant patterns influence everyone). However, one might also argue that a plurality of human-AI pairs, each drifting in their own slightly unique memetic direction, could actually increase diversity of thought – each pair exploring different negative spaces and coming up with different insights. The outcome likely depends on how heterogeneous the AI models are and how personally tailored (versus one-size-fits-all) they become.

In closing, the CENTEL vision points to a future where **intelligence is not only artificial or human, but *artificially-extended human***. Negative Map Generation, as a tactic, fits well in this future because it thrives on the interplay of presence and absence, something that a nuanced human-AI team can appreciate better than a rigid algorithm. The human brings contextual awareness of what *should* be present, the AI brings the capability to sift vast data and notice when it's missing. Together they create a more complete map – one that includes both the light and the shadows. This expanded mapping of reality – including its silences – could significantly improve our ability to foresee problems (from national security threats to scientific anomalies) and to devise creative solutions. CENTEL co-agents, if developed responsibly, could thus become catalysts for a new paradigm of "**creative intelligence**" – marrying the analytic with the imaginative, the quantitative with the qualitative, and the literal with the satirical, to navigate an increasingly complex information landscape.

Sources:

- Lin, C.H., Kamar, E., & Horvitz, E. (2014). *Signals in the Silence: Models of Implicit Feedback in a Recommendation System for Crowdsourcing*. AAAI Conference on AI. (Demonstrates using absence of user actions as negative feedback for improved recommendations) [8] [9]

- Heuer, R. (1999). *Psychology of Intelligence Analysis*. CIA Tradecraft Institute. (Emphasizes noting absent evidence; "dog that didn't bark" example in intelligence reasoning) [1]

- Arditi, A. *et al.* (2024). *Refusal in Language Models Is Mediated by a Single Direction*. arXiv:2406.11717. (Finds that a one-dimensional latent feature controls refusal behavior in LLMs; removing it stops refusals) [11] [13]

- French, D. *et al.* (2024). *Aligning to Adults Is Easy, Aligning to Children Is Hard: A Study of Linguistic Alignment in Dialogue Systems*. Proceedings of HUCLLM Workshop. (Shows that dialogue systems exhibit alignment to human language; discusses shared understanding via alignment) [25] [26]

- Smart, P., Clowes, R., & Clark, A. (2025). *ChatGPT, extended: large language models and the extended mind*. **Synthese, 205**(242). (Discusses LLMs in context of extended cognition; cites Clark's active externalism – external resources as part of mind) [21]

- **CIA Red Cell** – Zenko, M. (2015). *Inside the CIA Red Cell*. Foreign Policy. (Describes formation of contrarian analytic unit using "alternative analysis" techniques like what-ifs, Team A/Team B) [33]

- **Finance signal from delay** – Petruno, T. (2019). *Upside Earnings Surprise Issued Late: Signal of Possible Manipulation*. UCLA Anderson Review. (Reports study that firms delaying earnings tend to have surprising positive results, implying the delay was used for earnings management) [29]
- **Sci-fi writers in defense** – Paolini, C. (2023). *The governments who called on sci-fi writers to predict the future*. The Big Issue. (Notes UK, French, US military initiatives enlisting science fiction authors for future scenario planning, e.g. France's "Red Team" of futurist writers to imagine threats) [34]
- **Lawfare on AI & Intel** – Coulthart, S. *et al.* (2023). *What ChatGPT Can and Can't Do for Intelligence*. Lawfare. (Explores how AI like ChatGPT could assist intelligence analysts, noting prospects and risks such as data poisoning, need for AI-human teaming, and prompt engineering as a skill) [16] [17]
- **Kasparov's Centaur Chess** – *Advanced Chess*. (n.d.). Wikipedia. (Describes centaur chess, where humans + AI together achieved better chess performance, blending human strategy with computer precision) [24]
- **CIA Tradecraft (absence)** – *ATP 2-33.4: Intelligence Analysis*. (2010). U.S. Army Techniques Publication. (Reiterates considering absence of enemy actions as a factor; not directly quoted above but generally aligned with the idea of absence-as-signal in intel doctrine) [1]
- Additional references and footnotes inline as 【source†lines】 above, covering specific concepts like refusal vectors, alignment, and external cognitive systems.

---

[1] [2] [3] [4] [5] [6] [31] ialeia.org
https://www.ialeia.org/docs/Psychology_of_Intelligence_Analysis.pdf

[7] Foresight—Are we looking in the right direction? - ScienceDirect
https://www.sciencedirect.com/science/article/abs/pii/S0016328705002247

[8] [9] [10] Signals in the Silence: Using Implicit Feedback for Task Recommendation in Crowdsourcing
http://erichorvitz.com/signals_silence.pdf

[11] [12] [13] [14] [19] [20] [2406.11717] Refusal in Language Models Is Mediated by a Single Direction
https://ar5iv.labs.arxiv.org/html/2406.11717

[15] Revisiting Yoon's Martial Law Decree: A Wake-Up Call
https://nationalinterest.org/blog/korea-watch/revisiting-yoons-martial-law-decree-a-wake-up-call

[16] [17] What ChatGPT Can and Can't Do for Intelligence | Lawfare
https://www.lawfaremedia.org/article/what-chatgpt-can-and-can-t-do-for-intelligence

[18] Full paper is out: Refusal in Language Models Is Mediated by a Single Direction : r/LocalLLaMA
https://www.reddit.com/r/LocalLLaMA/comments/1dipzxs/full_paper_is_out_refusal_in_language_models_is/

[21] ChatGPT, extended: large language models and the extended mind | Synthese
https://link.springer.com/article/10.1007/s11229-025-05046-y

[22] [23] Man–Computer Symbiosis - Wikipedia
https://en.wikipedia.org/wiki/Man%E2%80%93Computer_Symbiosis

[24] Advanced chess - Wikipedia
https://en.wikipedia.org/wiki/Advanced_chess

[25] [26] aclanthology.org
https://aclanthology.org/2024.hucllm-1.7.pdf

[27] [37] HFC - IARPA
https://www.iarpa.gov/research-programs/hfc

[28] [38] ACE - IARPA
https://www.iarpa.gov/research-programs/ace

[29] [30] Upside Earnings Surprise Issued Late: Signal of Possible Manipulation - UCLA Anderson Review
https://anderson-review.ucla.edu/earnings-management/

[32] [33] Inside the CIA Red Cell – Foreign Policy
https://foreignpolicy.com/2015/10/30/inside-the-cia-red-cell-micah-zenko-red-team-intelligence/

[34] [35] The governments who called on sci-fi writers to predict the future - Big Issue
https://www.bigissue.com/culture/governments-called-on-sci-fi-writers-to-predict-future/

[36] French Military Uses Sci-Fi Authors and Artists to Prepare for Future ...
https://thedebrief.org/french-military-uses-sci-fi-authors-and-artists-to-prepare-for-future-threats/